DISSERTATION

MODELING BURN PROBABILITY PATTERNS FOR LARGE FIRES

Submitted by

Pamela Sue Ziesler

Department of Forest and Rangeland Stewardship

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2013

Doctoral Committee:

Advisor: Douglas B. Rideout

Robin Reich
Yu Wei
Robert Kling

ABSTRACT

MODELING BURN PROBABILITY PATTERNS FOR LARGE FIRES

I present a set of techniques for modeling burn probability patterns for large wildland fires. The resulting models address an important goal of a large fire risk analysis by estimating large fire burn probabilities. The intent was to develop models of burn probability using data that are widely available or easily calculated and that achieve acceptable predictive performance. Two models were successfully estimated using variables that may be extracted directly or easily calculated from standard GIS layers and other sources and they had 'good' predictive ability with AUCs of 0.81 and 0.83.

The ultimate intended use for the models is strategic program planning when information about future fire weather and event durations is unavailable and estimates of the average probabilistic shape and extent of large fires on a landscape are needed. Four primary objectives were to: 1) estimate models from historical fire data that are appropriate for strategic program planning, 2) incorporate the effect of barriers to the spread of fires across a landscape, 3) account for the average effect of weather streams and management actions on large fires without using detailed information on weather, fire duration or management tactics, and 4) investigate methods for addressing the spreading, connected nature of large fires on a landscape within a regression model. Models like these can provide finer detail than most landscape-wide models of burn probability and they have advantages over simulation methods because they do not require multiple runs of spread simulation models or information on fire duration or hourly weather events.

To model burn probability patterns, I organized historical fire data from Yellowstone National Park, U.S.A. into a set of grids; one grid per fire. I incorporated explanatory variables

such as fuel type, topography data and fire season indicators and I captured some spatial relationships through the use of distance, direction and other geometric variables. The data set observations are highly correlated and I investigated two approaches to account for and incorporate this correlation: one employed an autoregressive covariance structure and the other utilized a variable to account for the effects that neighboring cells may have on average burn probability. The two approaches yielded models with estimated coefficients that are consistent with fire behavior theory and that reflect how fires usually behave on the study site landscape. Both models compared well with the predictive ability of other fire probability models in the literature. Based on their predictive performance, this was a successful first attempt at addressing the research objectives and for estimating regression models to predict burn probability patterns for large fires.

ACKNOWLEDGMENTS

This work was funded by the McIntire-Stennis research program at Colorado State University. I thank professors Doug Rideout, Robin Reich, Yu Wei and Robert Kling for their input, support and patience throughout my work. I also thank Nicole Kernohan of the Department of Forest and Rangeland Stewardship for her GIS and programming assistance. I will always smile when I think of Pam and Niki's Adventures in Flatland. Finally, I could not have done any of this without support from Clay, my family and my friends – thank you all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Risk analysis is a core task in wildland fire program management. Predicting the location and spatial extent of landscape fires, quantifying the potential damage from these fires, and estimating the effects of management actions on mitigating fire risk are topics of great interest among researchers worldwide. While it has long been acknowledged that the systems affecting wildland fire are inherently uncertain, few tools and techniques have been available to quantify that uncertainty until recently. Recent advances in spatial statistics, geographic information systems (GIS) and fire spread modeling permit researchers to quantify wildland fire risk in scientifically sound ways. One of the most important components of any risk analysis is estimating the probabilities associated with the hazardous event (Brillinger 2003; Finney 2005). The techniques and models presented in this paper address part of this critical component of fire risk analysis by demonstrating a tractable approach to model burn probability patterns for large wildland fires given an ignition.

One type of fire probability modeling concerns only ignitions. I discuss it here briefly because it is a significant portion of the literature concerning wildland fire probability modeling. Several statistical methods have been used to accurately and reliably predict ignition hotspots and spatial patterns of ignitions. Logistic regression is the most popular approach and it allows modelers to interpret the effects of individual covariates on the probability of ignition. Ignition probability modeling is often split into different models for human-caused and lightning-caused fires (Amatulli et al. 2007) because many ignition probability researchers are solely concerned with fire prevention (de Vasconcelos et al. 2001; Vilar del Hoyo et al. 2011). Vega Garcia et al. (1995) used logistic regression to estimate daily probability of occurrence of human caused

forest fires in Alberta, Canada. They were able to predict 79% of fire-days and 81.5% of no fire-days within their data set and had an accuracy of 74% when the model was tested on an outside data set. Badia et al. (2011) estimated ignition probabilities in Wildland-Urban Interface (WUI) areas of Catalonia, Spain. They discovered that high ignition probabilities in this region are correlated with human activity variables and land use. For example, human activity variables such as distance to urban areas and road networks are correlated with ignition probability in dispersed housing areas while land use and seasonal temperatures are correlated with ignitions in agro-forestry areas. Romero-Calcerrada et al. (2008) used weights-of-evidence modeling and discovered that human access to the landscape (proximity to urban areas and roads) was strongly associated with ignition patterns. Another study investigated the effect of spatial accuracy of fire occurrence data on human-caused fire risk prediction (Vilar del Hoyo et al. 2011) in the Mediterranean region. They estimated two models using the same predictors (natural factors and socio-economic factors) but different response variables: one used the x,y coordinate of ignition points as the response variable and the other used a Kernel model with density of ignition as the response variable. They discovered that differences between the models were slight and models such as these may be incorporated into regional fire risk management systems to improve risk management activities.

Another variety of fire probability modeling attempts to estimate models of landscape-wide burn probabilities using historical fire information. The most common method is to estimate models using logistic regression, although the size of landscape patches, the significance of explanatory variables and the use of spatial variables vary by the management question addressed. For example, Preisler et al. (2004), Mermoz et al. (2005), Hernandez-Leal et al. (2006) and Lozano et al. (2007) used logistic regression to model landscape burn probability

and evaluate fire risk for broad areas, such as a state or national park. Gonzalez et al. (2006) and Chou et al. (1990; 1993) focused on smaller areas to fit logistic models for the probability of burning for individual forested areas. Chou et al. (1990) also investigated the role of spatial autocorrelation in the response variable (burn) and discovered that including terms that capture this autocorrelation significantly improved the explanatory power of the model. Statistical techniques other than logistic regression can also be used to estimate burn probabilities such as classification and regression trees (CART) (Lozano et al. 2008), weights-of-evidence modeling (Dickson et al. 2006) and ANOVA methods (Cardille and Ventura 2001). Statistical models based on historical information are usually estimated for broad areas and may not contain enough detail to be used for estimating conditional burn probability patterns from a particular ignition. This broad scale can also make it difficult to investigate the effects of barriers to spread on the probability of burning or how changes in local fuel patterns affect the probability of burning.

In contrast to using direct statistical modeling to estimate probabilities of burning, simulation techniques can be used to model burn probabilities by combining the results of multiple simulated ignitions with varying inputs or outputs. Most simulation-based techniques for estimating the probability of spread/burn use known distributions of fire modeling inputs such as ignition location, fire duration, fire size, wind speed and wind direction. The probability of fire spread is usually modeled by varying input conditions for a single ignition location while landscape-level probability of burning is typically estimated by varying input conditions for many ignition locations. Both use multiple draws from known distributions of fire modeling inputs in combination with a large number of model runs (500-10,000). Li and Apps (1996) used mathematical models for stand growth dynamics to evaluate how the probability of susceptibility to fire varied with distance from the fire origin. Hargrove et al. (2000) evaluated

the probability of fire spread from an ignition to neighboring areas. These studies emphasized that probability of spread depends on distance from ignition and on conditions in the neighborhood of the targeted point. In Canada, Anderson et al. (2007) have applied random perturbations to a simulated weather forecast to produce probabilistic fire perimeter outputs with a fire growth model and Anderson (2010) developed a model to predict the probable extent of a wildland fire if it was allowed to grow over the course of a fire season. Parisien et al. (2010) also used fire growth algorithms to produce spatial estimates of the relative likelihood of burning and, unlike many of the studies designed to directly inform local management decisions, they examined the relative contribution of factors that affect the mean and spatial variability in burn probability. A slightly different approach using simulation methods involves generating stochastic output rather than drawing from distributions of variable inputs (Boychuk et al. 2009). This technique uses a Markov chain and information on burn states and transition rates for each point in a lattice. The transition rates incorporate information about weather and depend on the state of neighboring cells.

Some probability models work in concert with established landscape fire modeling systems. The modeling system LANDIS has been used to examine fuel and fire risk dynamics (He et al. 2004) and to estimate the probability of fire occurrence and the probability of fire spread in the Missouri Ozarks (Yang et al. 2008). Wei et al. (2008) developed a fire risk distribution map for the southern Sierra planning unit in California by accumulating fire risk across a landscape using probability theory and outputs from the FlamMap (Finney 2006) system. Carmel et al. (2009) used the related modeling system FARSITE (Finney 1998) to estimate high-resolution spatial patterns of fire and fire risk in northwestern Israel. A collection of related simulation systems are frequently used for wildland fire decision support in the U.S.:

the Rare Event Risk Assessment Process RERAP (USDA Forest Service 2000) based on work by Wiitala and Carlton (1994) assumes that fire spreads in one direction while FSPro (USDA Forest Service 2012) maps the probability that fire will visit each point in a landscape during a specified time.

Statistical models of burn probability based on historical information are usually estimated for broad areas and do not contain enough detail to be used for estimating probability of burn patterns from an ignition. This broad scale may also make it difficult to investigate the effect that changes in fuel characteristics or barriers to spread may have on the probability of burning. Fire simulation methods provide information about likely fire sizes and shapes and can incorporate barriers to spread, but usually require specific assumptions about weather and fire duration. In this research, I explored the middle-ground between broad-scale statistical modeling of burn probabilities based on historical fire information and simulation methods that spread a fire from an ignition. The goal was to develop statistical models of burn probability for large fires using data that are widely available or easily calculated and that achieve reasonable predictive ability. Some key objectives were to: 1) estimate models from historical fire data that are appropriate for strategic program planning, 2) incorporate the effect of barriers to the spread of fires across a landscape, 3) account for the average effect of weather streams and management actions on large fires without using detailed information on weather, fire duration or management tactics, and 4) to investigate methods for addressing the spreading, connected nature of large fires on landscape within a regression model.

The eventual intended use for the models is strategic program planning when information about future weather and fire duration is unavailable and the average probabilistic shape and extent of large fires on a landscape are desired. It is important to note that I am not attempting to

model burn probabilities for a specific fire event with a given set of weather information; instead, I am averaging across the historical weather, fire management actions and fire-ending events for the study site's recent large fires. The presented models provide finer detail than most landscape models of burn probability and they have advantages over simulation methods for strategic planning because they do not require multiple runs of spread simulation models or the selection of wind speed(s) and direction(s) or fire duration. These models are not intended as a replacement for simulation or other methods for estimating landscape burn or spread probabilities, particularly for areas where historical fire information is poor or for the management of individual large fire events. Instead, they complement these methods by estimating average probabilistic fire shapes and extents as influenced by local variations in vegetation and topography.

Chapter 2 discusses the first of two models that I developed to address the goal and objectives described above. The model uses generalized logistic regression estimation and incorporates an autoregressive covariance structure as a first attempt to account for the spatial nature of the data set. A second model is presented in chapter 3 where I introduce a neighbor state variable that is intended to approximately capture the spread process for a landscape-scale wildland fire. Both models generated acceptable probability of burn patterns for large fires and had good predictive ability. Chapter 4 discusses some possible applications for the models and chapter 5 provides a brief summary and conclusions with suggestions for future work.

CHAPTER 2

CONDITIONAL PROBABILITY MODEL WITH AUTOREGRESSIVE COVARIANCE

This chapter discusses a model for conditional burn probability that uses an autoregressive covariance structure. The model was published in the International Journal of Wildland Fire in 2013, volume 22, issue 5 (Ziesler et al. 2013) and this chapter follows the general format of the published article with the addition of a few figures and descriptions to highlight how some of the variables were developed. Introductory material and related research from the literature was covered in chapter 1 and is therefore omitted from this chapter. A brief summary of the article is presented next followed by details of model development, results and performance.

To model average conditional burn probability patterns, I organized historical fire data from Yellowstone National Park into a set of grids; one grid per fire. I captured various spatial relationships inherent in the gridded data through use of geometric variables in the main model and by incorporating an autoregressive covariance structure. The final model had 'good' predictive ability with an AUC of 0.81 (1.0 is perfect prediction) and the estimated coefficients are consistent with theory and reflect how fires usually behave on the study site landscape. This technique produced a predictive model with finer detail than most landscape-wide models of burn probability and it has advantages over simulation methods for strategic planning because it does not require multiple runs of spread simulation models or information on fire duration.

Study Area

The study area is Yellowstone National Park located in the western interior of the U.S.A. The park covers over 2.2 million acres, primarily in the state of Wyoming with small sections in

Montana and Idaho. It is the core of the Greater Yellowstone Ecosystem and is one of the largest intact temperate-zone ecosystems on earth (Yellowstone National Park 2010). Approximately 64% of the park is covered by lodgepole pine (*Pinus contorta*) forest with other forest types occupying an additional 16% of the park. Figure 1 shows a typical area of lodgepole forest regenerated after the 1988 fires behind the famous Old Faithful geyser.

Figure 1: Lodgepole pine forest near Old Faithful geyser

The remaining 20% of the land base is divided between grassland areas (15%) and water (5%). Elevations range from 1600 m to over 3400 m above sea level and there is a wide variation in average precipitation from 25 cm per year along the north boundary to about 200 cm per year in the southwest corner. The park has a well-documented fire history and an extensive database with the data necessary for modeling burn probabilities, such as vegetation, topography, historical fire perimeters and ignition locations.

Data Source and Format

The U.S. National Park Service (NPS) provided the base GIS coverages with ignition locations, fire perimeters, fire behavior fuel model, aspect, elevation and slope. All coverages were converted from vector to raster format with 480 m x 480 m (23 ha) cells. There is a tradeoff between capturing landscape detail with small raster cells and having too many cells (Brillinger 2003; Mermoz et al. 2005; Preisler et al. 2004). The 480 m x 480 m cell size permits covering large areas, such as a national park, while maintaining a reasonable number of cells for statistical estimation and it is similar to the cell size used by Lozano et al. (2008) for modeling severe fire events. One of my objectives is to discover if a large cell size provides enough detail to estimate conditional burn probabilities.

The data encompass all of Yellowstone National Park within the park boundary with a buffer of cells outside the park to account for fires spreading beyond the park boundary. Of 38,559 cells within the park boundary there are 35,090 burnable cells and 3,469 non-burnable cells. The non-burnable cells consist of rock, water, permanent snow, and developed areas with little vegetation. While these non-burnable areas do not have estimated burn probabilities, their locations relative to burnable areas are incorporated into the models to account for changes in probability near barriers to spread.

The data cover an 18-year period after the 1988 fire season (1990-2007). It is recent enough to assume a consistent and modern approach to the management of fires and long enough to capture a wide range of fire season conditions. The selected time period has an average maximum fire season (June-September) temperature of 27.6°C which is similar to the mean for the preceding 40 years (1950-1989) and within the 95% confidence bound (mean = 27.3°C, standard deviation = 1.4°C) (Western Regional Climate Center 2006a). Average fire season and

yearly precipitation in Yellowstone National Park for the selected period and the preceding 40 years are also similar. The mean fire season precipitation for the selected period is 14.5 cm versus 15.2 cm (standard deviation 5.1 cm) for the preceding period and the mean total yearly precipitation for the two periods is 35.6 cm versus 37.6 cm (standard deviation 7.9 cm) (Western Regional Climate Center 2006b). The large fires of the selected period occurred under short-term drought conditions that varied from no drought to extreme drought (Enloe 2011) and the fires occurred under every category of the ENSO (El Niño – Southern Oscillation) phenomenon from strong La Niña to strong El Niño. While it is difficult to predict future conditions under which fires may occur, I believe the selected time period is acceptable to represent a wide range of fire conditions for Yellowstone National Park and for developing and evaluating the models.

Fire Perimeter Data

Fire spread from an ignition (conditional burning of cells near an ignition) is a spatial phenomenon requiring a data structure that accounts for the two dimensional nature of fire spread. Instead of observing the park as a single two-dimensional landscape, each fire is treated as a separate instance of a "fire event landscape" and I estimated probabilities within this smaller "landscape". There were 504 ignitions inside the park boundary from 1990-2007 with 40 of those fires becoming 'large' in that they spread beyond the ignition cell at the selected cell size of 480 m x 480m.

Each of the 40 large fires was mapped on to a 21 x 21 grid of cells centered on the ignition (see Appendix B for large fire grids). The conditional probability models calculated the probability that each cell in the grid will burn when a fire starts in the ignition cell. I refer to the cell for which I am calculating a probability as the 'target' cell. There is no single 'target' as there often is for simulation; instead, the final models are applied individually to every burnable

cell in the grid (every cell is a target in turn) and each cell receives an estimated probability that the cell will burn given a start in the ignition cell. The large fire models are therefore interpreted as 'the probability of burning a target cell given an ignition *and* spread to at least one other cell'. The 21 x 21 grid is large enough to capture the basic shape of the fire perimeters and small enough to yield a manageable number of data points. Fire sizes vary from two cells burned (ignition plus one cell) to 292 cells burned. The mean fire size is 44.1 cells (standard deviation 68.8 cells). Six of the 40 large fires burned only two cells (ignition plus one) and half of the large fires burned 21 cells or less. The raster point nearest the ignition for each fire was measured from the ignition coverage and identified as the ignition cell. Perimeter data from the NPS was used to establish the "burn footprint". Fire footprints use a binary raster format with "0" denoting a cell outside the perimeter (cell unburned) and "1" denoting a cell inside the perimeter (cell burned).

All burnable cells in each fire grid represent a separate observation in the dataset. There were 15,619 target cells in the 40 grids (15,619 = 441 cells/grid * 40 grids – 40 ignitions – 1,981 non-burnable cells) with 1,272 cells burned and 14,347 cells unburned. Despite the imbalance between burned and unburned cells, I elected to use all 15,619 cells as observations for estimating the models to maintain and to exploit the spatial relationships among the cells in a grid.

Variables Identifying the Relationship of the Target Cell to the Ignition Cell

Fires are affected by complex processes such as varying wind speeds and directions, precipitation events, variations in temperature and humidity, and fire management tactics. Unfortunately, accurate information for these dynamic processes is unavailable for strategic planning. By design and by the spreading nature of wildland fires, all cells in a fire grid have a

spatial relationship with the ignition cell. I capitalized on this relationship to capture elements of the spread process that are otherwise difficult to measure by using the final fire size and shape as proxies for the effects of changes in the weather and fire event management activities. I used two characteristics to describe fire size and shape in the grid: the distance of a target cell from the ignition and the direction of a target cell from the ignition.

Two distance variables were tested: one quantitative and one categorical. I evaluated several quantitative measures of distance and selected a square root function of distance from the ignition. I further tested the distance relationship by estimating coefficients for a different distance variable based on how burn probability changes in concentric rings about the ignition. A plot of the ring variable coefficients supported the square root function of distance as an appropriate functional form for this study site. While I was developing and testing the models it became apparent that I needed to correct for over-predictions near the ignition. To address this, I developed a categorical distance variable that divided the grid into three concentric sectors surrounding the ignition to account for the wide range of fire sizes in the data set and partially correct for over-prediction. Sector 1 contains the 80 cells nearest the ignition, sector 2 contains the next 144 cells and sector 3 is the outermost 216 cells (Figure 2). Both the square root form of the quantitative distance variable and the number and size of the sectors for the categorical variables were selected using the quasi-likelihood information criterion statistic, QIC (Pan 2001). QIC is a useful extension of the AIC, Akaike Information Criterion (Akaike 1973), that properly adjusts for quasi-likelihood methods and their asymptotic properties. Given a set of candidate models, the one with the lowest QIC is preferred.

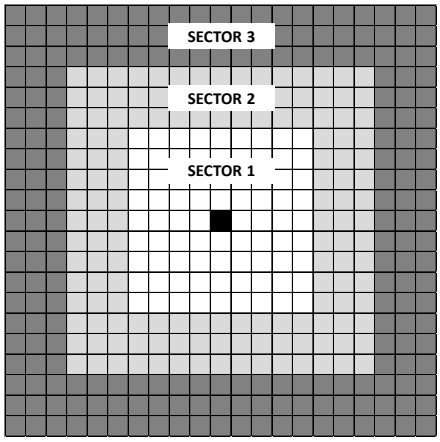


Figure 2: Sector variable

Because there is no consistent reference direction from the ignition, I used the eight cardinal and inter-cardinal direction classes to describe direction from the ignition: North, Northeast, East, Southeast, South, Southwest, West and Northwest. Direction from ignition is illustrated in Figure 3.

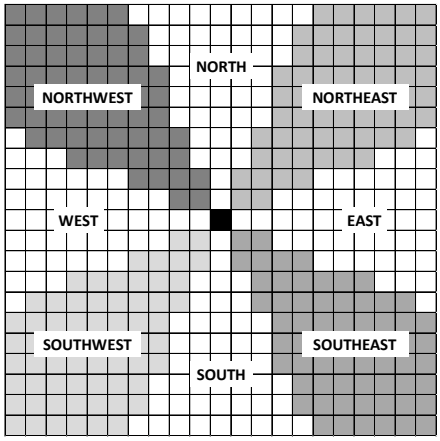


Figure 3: Direction from ignition variable

Fuels Data

Fuel characteristics appear to strongly influence fire occurrence (Amatulli 2007; Krawchuk 2006; and others) and are expected to significantly contribute to estimating conditional burn probability. I examined fuel model effects by incorporating the standard fire behavior fuel model codes from the set described by Scott and Burgan (2005). Categories present in the database for Yellowstone National Park are: Grass1, Grass2, GrassShrub1, GrassShrub2, Shrub2, TimberLitter (a combination of TimberLitter3 and a very small amount of TimberLitter1), TimberUnderstory4 and TimberUnderstory5. See Appendix D.2 for a raster graphic of fuel models in the park. Generally, I expected a grass model to increase the probability of a cell burning and a shrub model to decrease the probability of burning when compared to timber models.

Topography Data

Based on conventional wildland fire behavior calculations (Rothermel 1983) I identified slope, elevation, and aspect as likely topographic variables to test for including in each model. The attributes of slope, elevation and aspect were derived from Digital Elevation Model (DEM) data supplied by the NPS. Slope is measured in degrees from horizontal and elevation is measured in meters above sea level (see graphics D.4 and D.5, respectively). Because there is no reference direction for aspect, I converted it from degrees from north to the eight cardinal and inter-cardinal directions: north, northeast, east, southeast, south, southwest, west and northwest (see graphic D.6) plus flat. I evaluated slope, elevation and aspect independently and in various combinations. I also developed variables to represent change in elevation and change in slope between neighboring cells and the target cell to measure the effect of upslope or downslope

movement into a target cell. Figure 4 illustrates one method for calculating change in elevation or slope and figure 5 illustrates a different method.

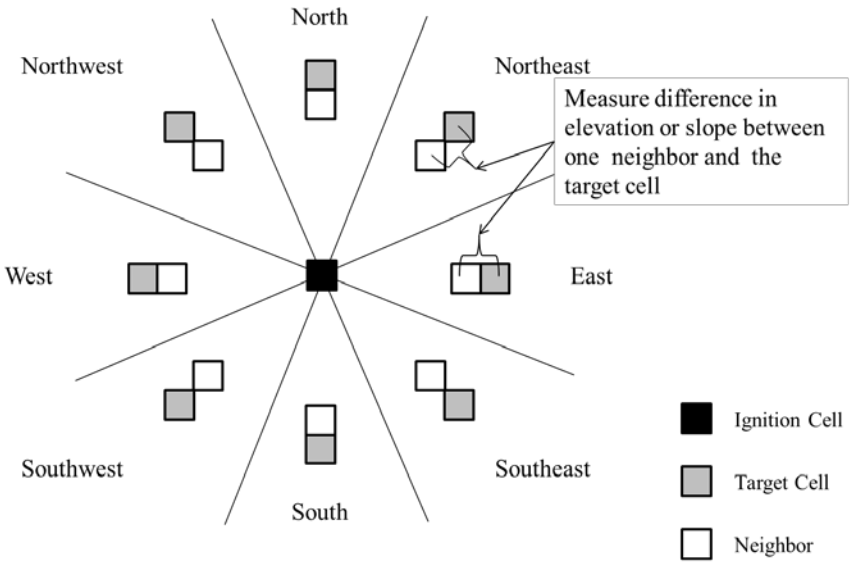


Figure 4: Method 1 for calculating change in elevation or slope.

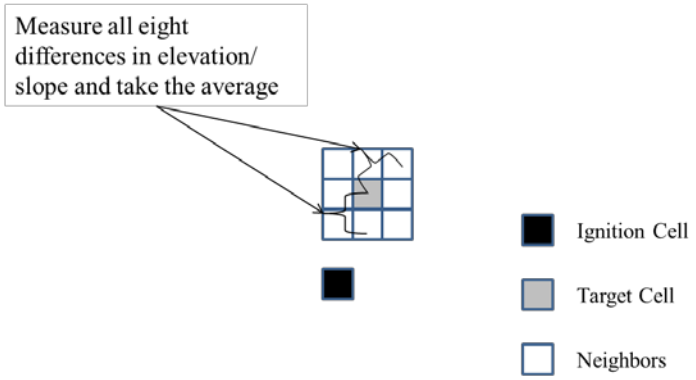
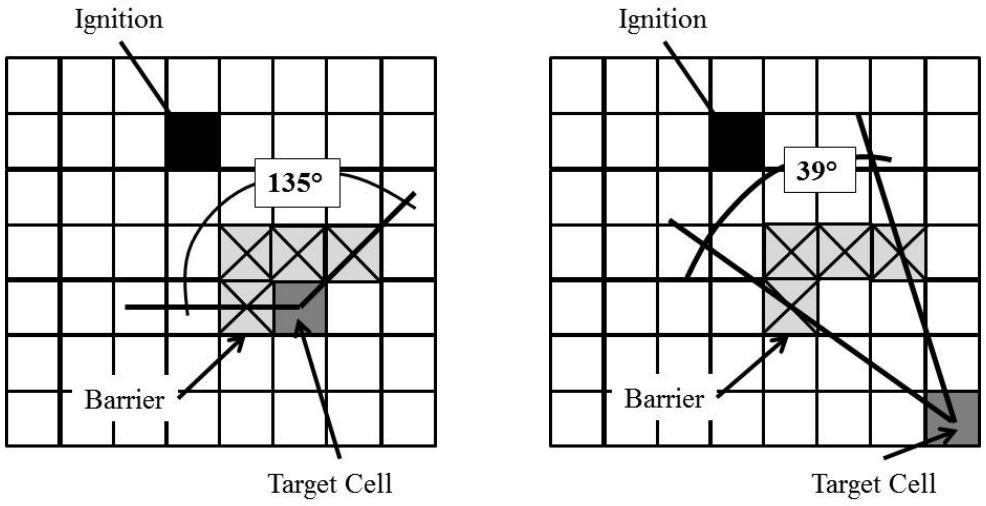


Figure 5: Method 2 for calculating change in elevation or slope.

Method 1 considers only one nearest neighbor and assumes the fire is moving radially outward from the ignition and spreads from the single neighbor into the target cell; method 2 takes the average change in elevation or slope for all eight neighbors.

To account for non-burnable barriers to spread, I developed a variable to capture the change in the conditional probability of burning behind a two-dimensional barrier between the target cell and the ignition cell (linear barriers such as rivers and streams require a different approach and are not addressed). Because fires burn around or jump over some barriers, I expected barriers to reduce but not eliminate the probability of a fire burning cells behind the barrier. I used the 'view angle' generated by the target cell's geometric relationship with the barrier to capture both the size of the barrier and the distance of the target cell from the barrier (Figure 6).



(a) (b)

Figure 6: View angles for cells behind a barrier to spread. (a) Target cell close to a barrier generating a view angle of 135º; (b) Target cell further from a barrier generating a view angle of 39°.

If the view angle is large the target cell is close to the barrier or the barrier is large, or both; if the view angle is small the target cell is far from the barrier or the barrier is small, or both. A view angle of zero indicates no barrier between the target cell and the ignition. I hypothesized the view angle effect to be negative because as view angle increases I expect the cells behind the barrier to have a lower conditional probability of burning.

I conducted early tests of slope, elevation and aspect independently and in various combinations. I also evaluated change in elevation and change in slope (both measure the effect of upslope/downslope into a target cell). Only the slope variable was statistically significant and retained in the model; the others were dropped due to very high p-values (> 0.5). I suspect that the effects of elevation, aspect and changes in elevation/slope are either averaged out over the large grid spacing or are partially captured by other variables at the selected scale, such as fuel model capturing some effects of aspect.

Variables to Capture Effects of Different Fire Seasons

Climate variables are known to have a strong influence on fire incidence (Roman-Cuesta et al. 2003; Shoennagel et al. 2005). To account for the effect of average fire season conditions on conditional burn probability I incorporated two climate variables, one based on a monthly drought indicator to capture local seasonal effects and the other based on the El Niño / Southern Oscillation (ENSO) phenomenon to capture wider, regional seasonal effects and the movement of storm systems. The drought variable is based on the Palmer Z drought index (Enloe 2011) and it attempts to capture the impact of monthly drought on burn probability. I evaluated several combinations of Palmer Z indices and selected a two-category variable as the best fitting (by QIC): one category represents no drought to moderate drought (-1.99 and above) and the other category represents severe to extreme drought (-2.00 and below). The ENSO phenomenon can

be measured quantitatively by the Southern Oscillation Index (SOI) and qualitatively by the El Niño / La Niña classifications. The effects of ENSO phase on regional weather patterns and wildland fire activity are complex (Shoennagel et al. 2005) so I considered several approaches to including this information in the models. I tested the fit of models for both the numeric SOI and the categorical ENSO phase and tested each variable type using the index for the year prior to a fire season (June-November average in the prior year) and the index over the fire season (June-November of the same year). Based on QIC, the categorical ENSO variable measured over the fire season performed best. Categories are: strong El Niño, moderate El Niño, neither Niño, moderate La Niña and strong La Niña using the breakpoints described by Redmond (2007). Another drought variable, the long-term Palmer Drought Severity Index (PDSI), was also considered but did not exhibit enough variability across the period to be of use in statistical modeling. Table 1 summarizes the final list of variables used in the model.

A Mathematical Formulation of Conditional Burn Probability

The mathematical formulation of the problem in this section follows from a structurally similar study on spatially correlated neuroimaging data for stroke patients (Albert and McShane 1995). The burn state for burnable cells is modeled as a binary response where "Y" takes the form of 'cell burns' (1) or 'cell does not burn' (0). The mean of Y provides an estimate of the probability of a cell burning. Let Y_{ij} denote the binary response for a target cell at grid location j for fire i, where i = 1 to I fires in the data set (I = 40) and j = 1 to J burnable cells in the 21x21-cell fire grid. The number of burnable cells, J, may be different for each fire i depending on the

Table 1: Variable list for estimating conditional burn probability autoregressive model

Variable Name	Classes	Description	Units
Dir	North, Northeast, East, Southeast, South, Southwest, West, Northwest	Direction from ignition	none
Dsqroot		Square root of (distance from ignition/1000)	$(\text{meters}/1000)^{1/2}$
Sector	1, 2, 3	Membership in the ring of cells nearest the ignition (1), middle ring (2) or outer ring (3)	none
Slope		Slope at the center of a raster cell	degrees
Viewangle		View angle of a spread barrier	degrees
FuelModel	Grass1, Grass2 GrassShrub1 GrassShrub2 Shrub2, TimberLitter TimberUnderstory4 TimberUnderstory5	Fire behavior fuel model from Scott and Burgan (2005)	none
PalmerZ	Severe and Extreme Drought, Moderate to No Drought	Aggregated Palmer Z index – monthly drought indicator	none
ENSO	StrongElNiño, ElNiño, NeitherNiño LaNiña StrongLaNiña	Category based on June – November average of Southern Oscillation Index: calculated over a given fire season	none

presence of non-burnable cells in the fire grid (water, rock, etc.); the smallest number of burnable cells in a grid for the data set is 32, the largest number of burnable cells is the maximum possible, 440 (21x21=441 minus the ignition).

Burn response Y_{ij} is asserted to be a function of cell attributes $\mathbf{X}_{ij}$ (see Table 1 for attribute definitions). Let $\mathbf{X}_{ij}$ denote the vector containing target cell attributes at location j for fire i:

$$\mathbf{X}_{ij} = (\text{Dir}_{ij}, \text{Dsqroot}_{ij}, \text{Sector}_{ij}, \text{Slope}_{ij}, \text{ViewAngle}_{ij}, \text{FuelModel}_{ij}, \text{PalmerZ}_{ij}, \text{ENSO}_{ij}) \quad (1)$$

My interest is to relate target cell-specific information to the burn frequency and estimate the average probability of a fire burning a target cell, P_{ij}, conditioned on the observed cell attributes, $\mathbf{X}_{ij}$:

$$P_{ij} = \text{Prob}(Y_{ij}= 1 \mid \mathbf{X}_{ij}) \quad (2)$$

The mean probability model to be estimated takes the form:

$$P_{ij} = [\exp(\mathbf{X}_{ij}\,\boldsymbol{\beta}) / (1+ \exp(\mathbf{X}_{ij}\,\boldsymbol{\beta}))] + \boldsymbol{\varepsilon}_{ij} \quad (3)$$

where a logit link function is selected to accommodate the binary response and $\boldsymbol{\beta}$ denotes a vector of unknown parameters, constant for all fires, describing the effect of target cell attributes on the mean probability of a fire burning a target cell, P_{ij}. The error term, $\boldsymbol{\varepsilon}_{ij}$, is assumed to be identically distributed binomial. The data set structure and underlying fire spread process yielded a collection of correlated observations and require a Generalized Estimating Equations (GEE) approach as described next.

Estimation by classical logistic regression is inappropriate because the assumption of spatial independence among observations is violated due to correlations among grid cells.

Failure to account for lack of independence can lead to improper model selection and incorrect interpretation of covariate effects. One form of correlation present in the input data is that of 'fire event' correlation. This correlation is a natural part of the data because all cells in a fire grid belong to the same fire event. The correlated data case can be described by the same mean and variance functions as the basic mathematical formulation (3), but a lack of independence among observations requires that the covariance be modeled as well. The Generalized Estimating Equations (GEE) procedure proposed by Liang and Zeger (1986) (see also (Zeger and Liang 1986)) takes advantage of information across fire events to estimate a working correlation matrix that is incorporated into the covariance structure. Because I was interested in averaging the conditional probability of burning across fire events, fire event dependence can be modeled as a 'repeated measures' case for individual fires (see Albert and McShane (1995) and others). Each fire is a "subject" and data points are generated across space for each 'subject' rather than across time, as in the more familiar case of repeated measures data for a subject through time (Figure 7).

Familiar selection criteria such as the Akaike information criterion, AIC, are not available for the GEE quasi-likelihood method so a different criterion must be used. Variables were assessed and selected using the quasi-likelihood information criterion statistic, QIC (Pan 2001), as described above. QIC is a useful extension of AIC that balances model fit against model simplicity (provides a penalty as variables are added) and properly adjusts for quasi-likelihood methods and their asymptotic properties.

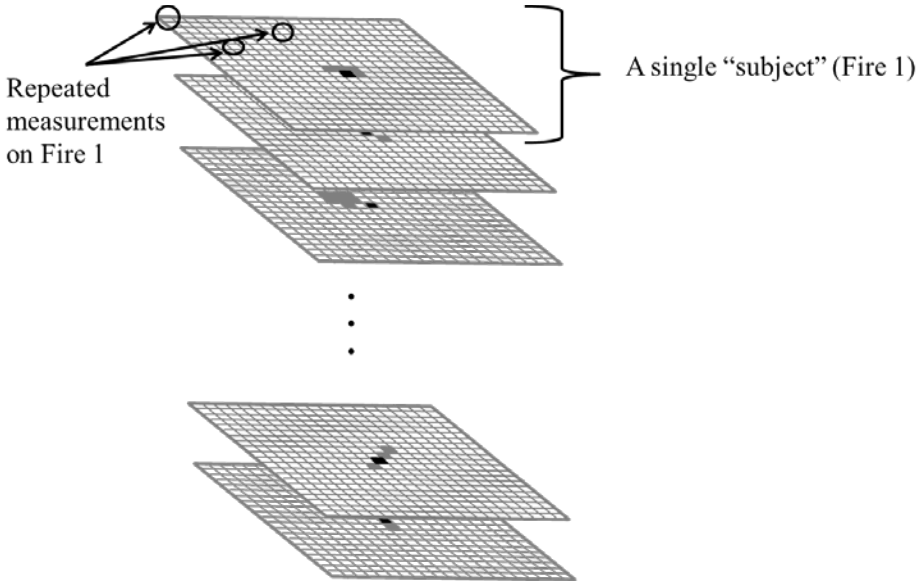


Figure 7: Each fire is a separate subject; each cell in a fire grid is a repeated measurement on that fire. Black is the ignition cell; grey indicates a burned cell inside the fire perimeter (response = 1); white indicates an unburned cell outside the fire perimeter (response = 0).

Spatially Autoregressive Covariance Structure

Fire event correlation is only one type of correlation present in the data set. The grid structure imposes a spatial relationship on all cells in a fire grid. A simple approach is to incorporate this relationship into the covariance structure. The model has a mean described by the mathematical formulation given in (3) and a covariance described by (4) to impose a distance-based exponential covariance function on the model:

$$\boldsymbol{\Omega} = e^{-0.473*(\text{distance between cells})} \qquad (4)$$

The exponent (-0.473 * distance between cells) was selected as the best fitting exponential decay function based on the QIC statistic. I estimated the model using the generalized linear models (GENMOD) procedure within SAS software, Version 9.3 of the SAS System for Windows,

copyright © 2002-2010 SAS Institute Inc. I invoked the repeated measures option to account for fire event dependence and specified a user-defined working correlation matrix of the form given in (4).

Results

Table 2 displays estimated coefficients, standard errors and p-values for the model.

Table 2: Estimated model coefficients, standard errors and p-values for autoregressive model

Coefficients in **bold** are significant at 0.05-level

Parameter	Coefficient	Std. Error	p-value
Intercept	**-2.1019**	0.3892	< 0.0001
Dir – East	**0.8161**	0.3234	0.0116
Dir – North	**1.1723**	0.3050	0.0001
Dir – Northeast	**1.6205**	0.3368	< 0.0001
Dir – Northwest	**0.4127**	0.1811	0.0226
Dir – South	-0.1066	0.2305	0.6438
Dir – Southeast	0.2603	0.2826	0.3570
Dir – Southwest	0.2727	0.1690	0.1066
Dir – West	Reference		
Dsqroot	**-2.3027**	0.2368	< 0.0001
Sector – 1	**-0.4748**	0.1209	< 0.0001
Sector – 2	**-0.3562**	0.0875	< 0.0001
Sector – 3	Reference		
Slope	**0.0254**	0.0103	0.0138
Viewangle	**-0.0138**	0.0051	0.0067
Fuel – Grass1	**2.0609**	0.2594	< 0.0001
Fuel – Grass2	0.1792	0.0975	0.0661
Fuel – GrassShrub1	**0.9834**	0.1877	< 0.0001
Fuel – GrassShrub2	-0.0661	0.1589	0.6772
Fuel – Shrub2	0.1431	0.1991	0.4725
Fuel – TimberLitter	**1.6928**	0.3379	< 0.0001
Fuel – TimberUnderstory4	**0.6809**	0.1262	< 0.0001
Fuel – TimberUnderstory5	Reference		
PalmerZ – Drought	**1.5901**	0.6292	0.0115
PalmerZ – No Drought	Reference		
ENSO – StongElNiño	**2.6847**	0.9391	0.0043
ENSO – ElNiño	**1.6843**	0.5165	0.0011
ENSO – NeitherNiño	1.0466	0.6552	0.1102
ENSO – LaNiña	**1.5373**	0.6304	0.0147
ENSO – StrongLaNiña	Reference		

The direction classes of East, North, Northeast and Northwest were statistically significant at the 0.05-level and positive indicating a higher likelihood of burning in those directions than the reference direction, West. The square root of distance coefficient was negative and significant indicating that the conditional probability of burning decreases with distance from the ignition. The sector variable was significant and had coefficients that adjusted probabilities for the cells nearer the ignition downward, with a greater downward adjustment in sector 1 than in sector 2. The slope coefficient was positive and significant indicating an increase in burn probability with increasing slope. The other topographic variable, barriers to spread as measured by view angle, was statistically significant and the estimated coefficient was negative implying that the probability of burning decreases as the view angle increases.

Estimated coefficients for fuel model categories Grass1, GrassShrub1, TimberLitter and TimberUnderstory4 were positive and highly statistically significant and Grass2 was nearly significant ($p=0.0661$), indicating a higher probability of burning than the reference category TimberUnderstory5. Estimated coefficients for fuels represented by GrassShrub2 and Shrub2 were not statistically significantly different than the reference category.

The aggregate Palmer Z variable and ENSO variable were both statistically significant. The coefficient for the Palmer Z variable category representing severe and extreme monthly drought was positive and corresponds to a higher probability of burning than the reference condition representing moderate to no drought. All ENSO categories except neither Niño were statistically significant and coefficients were positive with respect to the reference category strong La Niña. The ranking of categories (highest to lowest) were strong El Niño – El Niño – La Niña – neither Niño – strong La Niña. Note that the strong El Niño category has the highest conditional probability of burning, the strong La Niña has the lowest probability of burning and

that the rankings do not follow the SOI magnitude rankings of strong El Niño – El Niño – neither Niño – La Niña– strong La Niña, justifying the choice of the categorical variable over the quantitative SOI variable.

Discussion

The primary intent of this research was to discover if the proposed model of conditional burn probability is estimable using standard statistical procedures and if so, to see if it produces acceptable results. The coarseness of the lattice and vegetation information did not produce any problems with estimation and the model coefficients are as expected and correspond well with theory. Figure 8 illustrates the original burn pattern and the modeled conditional probability of burn pattern for a representative ignition location in the data set – the August 2002 Phlox fire. Panel (a) shows the ignition cell (black), the actual burn footprint (grey) and all non-burnable regions (hashed cells). Panel (b) shows the estimated conditional probability of burning for each target cell using the coefficients in Table 2 along with the proper input data for each target cell. Modeled burn probabilities are generally higher near the ignition and lower near the edges of the grid, as expected. Areas of higher probability that are surrounded by areas of lower probability, such as those near the right edge of the fifth and sixth row from the top in Fig. 8(b), have prediction intervals that overlap. For example, the prediction interval for the lighter colored cell in sixth row from the top, third column from the right is 0.05-0.30 and the interval for the darker cell in the same row and second column from the right is 0.13-0.54.

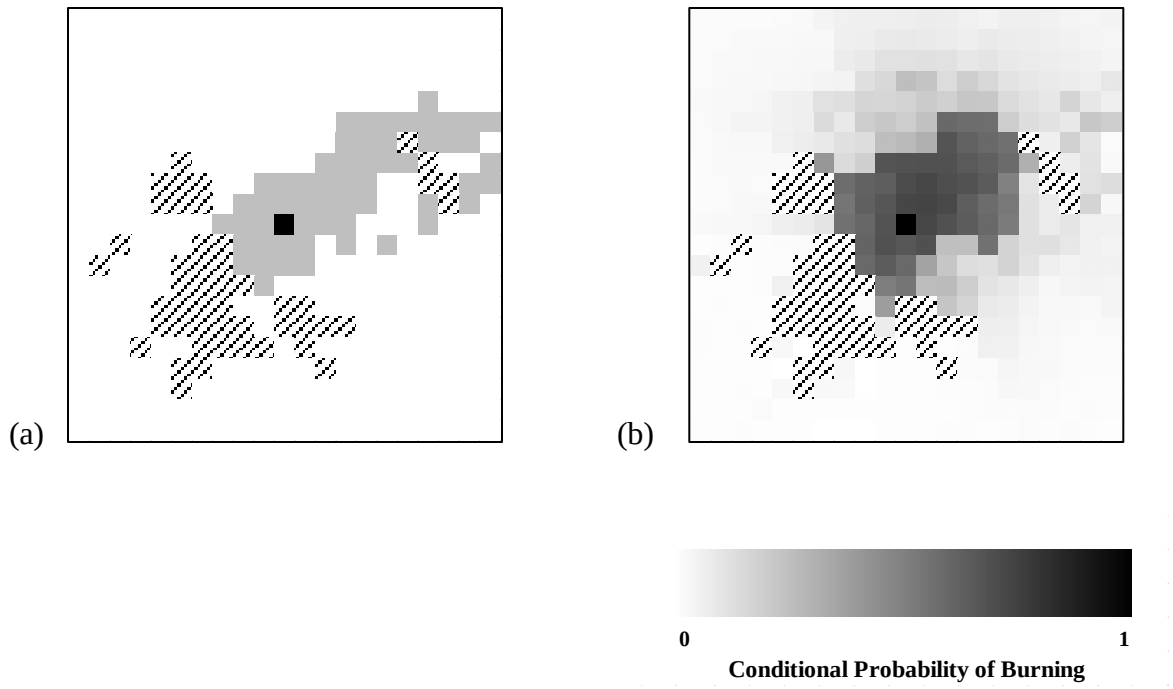


Figure 8: Actual burn pattern and estimated probability of burn pattern for the Phlox fire (2002). (a) Actual burn pattern: ignition cell is black (center); grey cells indicate the burned area; hashed cells are barriers to spread. (b) Modeled probabilities of burning for the Phlox fire.

The ranking by magnitude of the direction classes (Northeast was the largest, followed by North, East and Northwest) supports the common contention that Yellowstone's large fires are shaped by prevailing south to southwest winds. The sector variable improved the predictive ability by reducing over-prediction (false alarms) near the ignition that would occur if corrections for variations in large fire sizes were not included in the model. The negative square root of distance coefficient reflects the expected decrease in burn probability with increasing distance from the ignition. This effect implies that average large fire spread is affected by complex time-dependent factors, such as weather and management actions, and care should be exercised when modeling large fires. The universality of the square root form of distance could be investigated further by examining other data sets. The negative coefficient for the view angle

barrier variable supports the hypothesis that target cells behind or close to large barriers are less likely to burn than cells behind small barriers or that are far away from barriers (cet. par.).

The ranking of fuel model coefficients by magnitude was as expected when interpreted for the particular condition of 'large fires' in Yellowstone National Park. An important point is that many fires in Yellowstone are allowed to burn unhindered under most conditions if they do not threaten developed areas in or near the park. The GrassShrub2 fuels and the Shrub2 fuels are not statistically different than the reference TimberUnderstory5 fuels indicating that the contribution to burn probability for these three fuels is similar. These fuel complexes are usually the least likely to burn under large fire conditions in the park because it is less likely for heavier sagebrush types and moister, denser timber types to burn compared to other fuel types. The ranking of the other types was somewhat surprising, at first. Typically, models Grass2 and TimberUnderstory4 are used to represent fuel types with more aggressive fire behavior (higher spread rates, longer flame lengths) than those represented by Grass1 and TimberLitter models. However, the modeled coefficients indicate that Grass1 and TimberLitter contribute to higher conditional probabilities of burning when compared to Grass2 and TimberUnderstory4. When I inspected a fuel model map with a large fire perimeters overlay I observed two factors that explain this result: Grass1 and TimberLitter occur in relatively smaller patches than the other fuel types and when a fire with enough energy to become "large" encounters one of the patches, the patch is almost entirely contained within the fire perimeter. Also, it is reasonable to expect that the fire management philosophy in the park would not hinder fire spreading through patches of less aggressive fuels. These observations support the model results that when a large fire encounters Grass1 or TimberLitter they are more likely to burn than other fuel types. Note that

the model data and results do not indicate the severity or completeness of burning, merely that the area is considered 'burned' because it falls within the recorded fire perimeter.

Climate variables that are correlated with fire season severity and weather patterns also contributed to the models in accordance with expectations. The sign of the Palmer Z variable indicates that more severe drought increases the conditional probability of burning, as expected. The effects of the ENSO variable are particularly interesting. Strong El Niño promotes the highest conditional probability of burning in Yellowstone National Park, as expected (Shoennagel et al. 2005). However, the ENSO phase categories are not ranked from low SOI to high SOI (or vice versa) suggesting that, while SOI certainly affects probability of burning, numeric SOI is not directly proportional to fire activity in the study area. Interestingly, the more moderate categories of El Niño and La Niña have coefficients that are similar in magnitude which suggests that conditional probabilities of burning will be similar under those conditions (cet. par.).

Predictive Performance

A general measure of predictive performance for a binary response model is the area under the ROC (Receiver Operating Characteristics) curve, known as AUC. For this model, the AUC is equivalent to the probability that the model will rank a randomly chosen burned cell higher than a randomly chosen unburned cell (Fawcett 2006). Perfect predictive performance yields an AUC of 1.0 and predictive performance equivalent to random guessing yields an AUC of 0.5, so all realistic predictive models should have an AUC between 0.5 and 1.0. Table 3 lists the AUC, the probability threshold used to establish if a cell is considered 'burned', the percent correct burn predictions and the false alarm rate. The expected predictive performance of the

model is considered ‘good’ with an AUC of 0.81, percent correct burn predictions of 0.84 and a false alarm rate of 0.22.

Table 3: Model predictive performance for autoregressive model

AUC	Optimal Probability Threshold for Burned Cell	Percent Correct Burn Predictions	False Alarm Rate (predicted to burn, but did not)
0.81	0.13	0.84	0.22

I further investigated the predictive ability of the model by removing a single fire from the data set, recalculating coefficients and predicting the conditional burn probabilities of the removed fire for each of the 40 fires. The model performed very well by generating an AUC of 0.85 or better for 17 of the 40 fires and an AUC of 0.90 or better for eight of the fires. The AUCs for seven of the predicted fires were considered poor (<0.60). The hit rates (correctly predicted to burn) for the best performing fires were perfect (1.0) combined with false alarm rates (incorrectly predicted to burn) of less than 0.16. This model has acceptable predictive performance when compared to other regression models of wildland fire based on historical data; such models have AUCs in the range 0.64 to 0.71 (Krawchuk et al. 2006; Vilar del Hoyo et al. 2011). They also are comparable to some gridded simulation models; these models have hit rates of 0.79-0.93 and false alarm rates of 0.24-0.39 (Anderson et al. 2007).

Summary

The technique presented in this chapter for predicting the conditional probability of burning for large fires addresses an important goal of a large fire risk analysis – estimation of large fire burn probabilities. The model incorporates variables that are known to affect fire behavior and that may be extracted directly or easily calculated from standard GIS layers and

other sources. It is capable of modeling the decline in probability outward from an ignition that one expects to see in an averaged probability of burning for real fires on a landscape. With appropriate care in selecting variables, this approach could be used to support strategic wildland fire program planning when information is needed on probabilistic averages for shape and extent of large wildland fires.

CHAPTER 3

CONDITIONAL PROBABILITY OF BURNING MODEL WITH NEIGHBOR VARIABLE

In this chapter I discuss a statistical model of burn probability that incorporates a new variable to account for the effects that burned neighboring cells may have on average burn probability. This is a first attempt to incorporate such a variable to represent the spread process for large fires in a gridded regression model of burn probability. After incorporating this new variable and one other representing location within the park to account for potential differences in large fire management activities across the park, I follow the same general procedure as in the previous chapter to account for barriers to the spread of fires across a landscape and average effect of weather events and management actions on large fires without using detailed information on weather, fire duration or management tactics.

Study Area and Variables

The study area is Yellowstone National Park located in the western interior of the USA. The base data set is the same as the one in the previous chapter and covers the 18-year period from 1990 to 2007. The data encompass all of Yellowstone National Park within the park boundary with a buffer of cells outside the park to account for large fires that occurred near the boundary and may have spread beyond it. The large fires of the selected period occurred under short-term drought conditions that varied from no drought to extreme drought (Enloe 2011) and the fires occurred under every category of the ENSO (El Niño – Southern Oscillation) phenomenon from strong La Niña to strong El Niño. While it is difficult to predict future conditions under which fires may occur, the selected time period represents a wide range of fire

conditions for Yellowstone National Park and is appropriate for developing and evaluating the model.

I used six GIS coverages from the US National Park Service (fire perimeter, ignition location, fire behavior fuel model, aspect, elevation, slope) and the same cell size (480 m x 480 m cells) as for the previous model. Fire perimeters are in binary raster format with "0" denoting a cell outside the perimeter (cell unburned) and "1" denoting a cell inside the perimeter (cell burned). The raster point nearest the ignition for each fire was labeled as the ignition cell. Fuel models are identified using the standard fire behavior fuel model codes from the set described by Scott and Burgan (2005). Categories present in the database are: Grass1, Grass2, GrassShrub1, GrassShrub2, Shrub2, TimberLitter (a combination of TimberLitter1 and TimberLitter3), TimberUnderstory4 and TimberUnderstory5. I tested slope, elevation and aspect independently and in various combinations, including changes in elevation and slope from neighboring cells into the target cell. All three topography variables (elevation, aspect, slope) and their combinations were eventually dropped from the model because all were highly insignificant (p-values > 0.6). The final variable list for this model is presented in Table 4. All but two of the variables (square root of distance from the ignition and view angle) are categorical.

As I discovered in the previous study, variables representing location and spatial relationships are important to predicting conditional burn probability. I retained the direction and distance variables that capture a target cell's relationship with the ignition and that, when combined with burn pattern data, capture final fire size and shape. Direction from the ignition is described by one of the eight direction classes: North, Northeast, East, Southeast, South, Southwest, West or Northwest.

Table 4: Variable list for estimating conditional burn probability for neighbor model

Variable Name	Classes	Description	Units
Dir	North, Northeast, East, Southeast, South, Southwest, West, Northwest	Direction from ignition	none
Dsqroot		Square root of (distance from ignition/1000)	$(\text{meters}/1000)^{1/2}$
Sector	1, 2, 3	Membership in the ring of cells nearest the ignition (1), middle ring (2) or outer ring (3)	none
Viewangle		View angle of a spread barrier	degrees
FuelModel	Grass1, Grass2 GrassShrub1 GrassShrub2 Shrub2, TimberLitter TimberUnderstory4 TimberUnderstory5	Fire behavior fuel model from Scott and Burgan (2005)	none
PalmerZ	Severe and Extreme Drought, Moderate to No Drought	Aggregated Palmer Z index – monthly drought indicator	none
ENSO	StrongElNiño, ElNiño, NeitherNiño, LaNiña, StrongLaNiña	Category based on June – November average of Southern Oscillation Index: calculated over a given fire season	none
Zone	Northwest, NorthCentral, Mirror Plateau, Central Plateau, Southwest, Southeast, Suppression, Outside	Membership in a management zone	none
NeighborState	Non-burnable, Not burned, Burned	Burn status of a target cell's neighbor	none

There are two distance variables, one quantitative and one categorical: the quantitative measure is a square root function of distance from the ignition and the categorical variable divides the grid into three regions surrounding the ignition called 'sectors'. Sector 1 contains the 80 cells nearest the ignition, sector 2 contains the next 144 cells and sector 3 is the outermost 216 cells. A third type of spatial relationship is captured by a variable representing the target cell's geometric relationship with barriers to spread between the target cell and the ignition cell. This 'view angle' is calculated by treating the target cell as a vertex for the angle and extending rays from the vertex to each end of the barrier. The 'view angle' captures both the size of the barrier and the distance of the target cell from the barrier. I expected larger view angles to decrease the conditional probability of burning because fires can be deflected from cells downstream of a large barrier or that cells are less likely to burn when they are close to a barrier.

I evaluated a new location variable to see if I could improve the predictive capability of the model. As a first attempt to capture the general location of each cell within the park, I identified a 'zone' derived from management divisions found in the park's fire management documents (Yellowstone National Park 1992, 2004). Zones are separated by the park's road network and I superimposed an older division that separates out the highly developed areas of the park (referred to as the 'suppression zone'). I also treated the area outside the park boundary as a zone. The zone variable has eight categories: central plateau zone, mirror plateau zone, north central zone (a combination of the northern range and Wilson plateau fire management zones), northwest zone, areas outside the park boundary, southeast zone, highly developed areas requiring protection (suppression zone), and southwest zone. Each target cell is assigned the zone it occupies on the map even if the ignition occurred in an adjacent zone – this is to account for potential differences in management strategy as a fire moves away from its ignition.

I incorporated two climate variables to account for the effect of average fire season conditions on conditional burn probability. One of the climate variables has categories derived from the Palmer Z drought index values given by Enloe (2011). I used a two category form with one category representing little or no drought (-1.99 and higher) and one category representing severe to extreme drought (-2.00 and below). I also used a categorical variable representing the El Niño / Southern Oscillation (ENSO) phenomenon measured over the fire season with categories based on the breakpoints described by Redmond (2007): strong El Niño, moderate El Niño, neither Niño, moderate La Niña and strong La Niña.

A Variable to Capture the Effects of Neighboring Cells on the Conditional Probability of Burning

I expected there to be positive spatial autocorrelation in the data set due to the connected, spreading nature of a wildland fire. While the model in the previous chapter attempted to capture some of this autocorrelation by introducing a spatially autoregressive covariance structure, here I introduced a variable that described the burn state of a target cell's neighbor. For simplicity I assumed that fires spread generally outward from the ignition. A relevant neighbor could be a single cell between the ignition and the target, a collection of cells between the ignition and the target or all eight cells surrounding a target. I examined the case of a single neighbor and used the direction variable to help identify the neighbor (Figure 9). For example, if a cell is classified as a "north" cell I chose the cell immediately to the south as the neighbor of concern. I also tested three immediately adjacent neighbors (based on most likely spread directions) and all eight immediate neighbors but have elected to continue with only a single neighbor because the multi-neighbor structures did not perform well in predicting spatial burn patterns. To incorporate the neighbor information I created a new variable indicating if the

neighbor cell had burned (1), not burned (0) or is not burnable (-1). The next section presents a brief summary of the mathematical development of the conditional probability of burn model developed in the previous chapter.

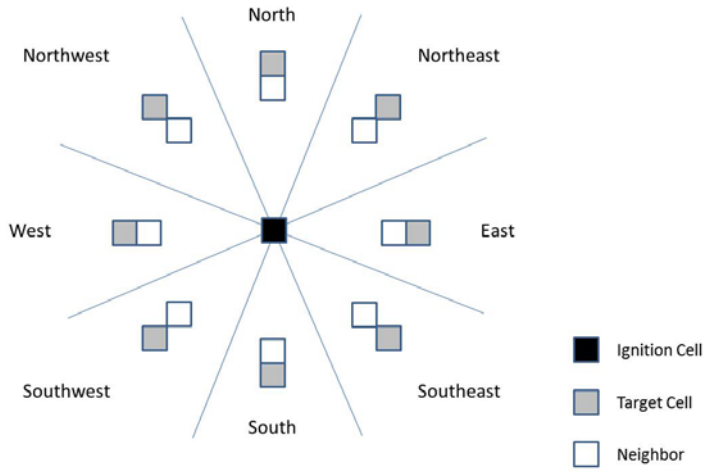


Figure 9: Location of a neighbor in relation to the target cell. Target cells are identified by their direction from ignition.

Conditional Burn Probability Model

The base model to be estimated has the form:

$$P_{ij} = [\exp(\mathbf{X}_{ij}\,\boldsymbol{\beta}) / (1+ \exp(\mathbf{X}_{ij}\,\boldsymbol{\beta}))] + \boldsymbol{\varepsilon}_{ij} \quad (5)$$

The probability of a fire burning a target cell is P_{ij} and $\boldsymbol{\beta}$ denotes a vector of unknown parameters that is constant for all fires. The error term, $\boldsymbol{\varepsilon}_{ij}$, is assumed to be identically distributed binomial and the vector $\mathbf{X}_{ij}$ is a function of the attributes of target cell j for fire i:

$$\mathbf{X}_{ij} = (Dir_{ij}, Dsqroot_{ij}, Sector_{ij}, ViewAngle_{ij}, FuelModel_{ij}, PalmerZ_{ij}, ENSO_{ij}, Zone_{ij}, NeighbourState_{ij}) \quad (6)$$

Subscripts are defined such that i = 1 to I fires in the data set (I = 40) and j = 1 to J burnable cells in the 21x21-cell fire grid. The number of burnable cells, J, may be different for each fire i depending on the presence of non-burnable cells in the fire grid (water, rock, etc.); the smallest number of burnable cells in a grid for the data set is 32, the largest number of burnable cells is the maximum possible, 440 (21x21=441 minus the ignition).

Because I was interested in averaging the conditional probability of burning across fire events, fire event dependence can be modeled as a 'repeated measures' case for individual fires (see Albert and McShane (1995) and others). Each fire is a "subject" and data points are generated across space for each 'subject' rather than across time, as in the more familiar case of repeated measures data for a subject through time. I invoked GEE methods to estimate the model in (5) with the generalized linear models (GENMOD) procedure, SAS software version 9.3 of the SAS System for Windows, copyright © 2002-2010 SAS Institute Inc. I used the repeated measures option to account for fire event dependence and selected a spatially independent covariance structure for the model. GEE estimates are consistent even when the covariance is misspecified, as long as the model of the mean is correctly specified. By using GEE, I estimated 'population-averaged' effects rather than predicting the effects of changing covariates for a given individual. This generates results more appropriate for strategic planning than for individual fire events.

Results and Discussion

Table 5 lists estimated coefficients, standard errors and p-values for the model. The ranking by magnitude of the significant direction classes is as expected and the top three directions ranked by magnitude are Northeast-North-East. The signs and ranking of these three directions is the same as in the previous study and the signs on the coefficients support the

common contention that Yellowstone's fires are shaped by prevailing south to southwest winds. Different than the previous study is that the negative square root of distance coefficient and the sector variable coefficients were insignificant in this model. The new variable indicating the burn state of the single neighbor seems to capture part of the distance from ignition effect and appears to render the distance and sector variables less important. This is not surprising considering that the burn state of a neighbor is expected to have a high influence on the probability of a target cell burning and may make other variables less important. The coefficient for the barrier to spread variable, view angle, is significant and negative supporting the hypothesis that a target cell behind or close to a large barrier is less likely to burn than a cell behind a small barrier or that is far away from a barrier.

The model presented here and the model in the previous chapter had identical rankings by magnitude for the first four significant fuel models shown in Tables 2 and 5. This suggests that Grass1 areas are the most likely to burn (given an ignition) followed by TimberLitter, GrassShrub1 and TimberUnderstory4. The Grass2, GrassShrub2 and Shrub2 fuels are similar to the reference, TimberUnderstory5, indicating that the contribution to burn probability for these fuel types is expected to be similar. These last four fuel complexes may be the least likely to burn in Yellowstone National Park because it is usually less likely for heavier sagebrush types and moister, denser timber types to burn compared to other fuel types.

Table 5: Estimated coefficients, standard errors and p-values for neighbor model

Coefficients in **bold** are significant at 0.05-level

Parameter	Coefficient	Std. Error	p-value
Intercept	**-3.2314**	0.7237	< 0.0001
Dir – East	**1.0063**	0.2661	0.0002
Dir – North	**1.1484**	0.3993	0.0040
Dir – Northeast	**1.3583**	0.3911	0.0005
Dir – Northwest	0.0712	0.2863	0.8036
Dir – South	-0.3144	0.2460	0.2014
Dir – Southeast	-0.3176	0.3166	0.3158
Dir – Southwest	0.0197	0.1639	0.9045
Dir – West	Reference		
Dsqroot	-0.6153	0.4313	0.1537
Sector – 1	0.0119	0.4023	0.9764
Sector – 2	0.2710	0.2090	0.1948
Sector – 3	Reference		
Viewangle	**-0.0102**	0.0033	0.0021
Fuel – Grass1	**3.2223**	0.3164	< 0.0001
Fuel – Grass2	0.2739	0.2637	0.2989
Fuel – GrassShrub1	**1.8280**	0.2570	< 0.0001
Fuel – GrassShrub2	0.1589	0.2795	0.5698
Fuel – Shrub2	0.6719	0.3769	0.0747
Fuel – TimberLitter	**3.0283**	0.4496	< 0.0001
Fuel – TimberUnderstory4	**1.2470**	0.1953	< 0.0001
Fuel – TimberUnderstory5	Reference		
PalmerZ – Drought	0.3818	0.3618	0.2912
PalmerZ – No Drought	Reference		
ENSO – StongElNiño	**3.3355**	0.5257	< 0.0001
ENSO – ElNiño	**2.1574**	0.3862	< 0.0001
ENSO – NeitherNiño	**2.1425**	0.5258	< 0.0001
ENSO – LaNiña	**2.2165**	0.4782	< 0.0001
ENSO – StrongLaNiña	Reference		
Zone – Central Plateau	-0.1072	0.5032	0.8313
Zone – Mirror Plateau	0.1505	0.3626	0.6782
Zone – North Central	-0.6774	0.7253	0.3504
Zone – Northwest	0.0111	0.2634	0.9665
Zone – Outside	0.2125	0.3660	0.5616
Zone – Southeast	0.3793	0.2841	0.1818
Zone – Suppression	**-1.6937**	0.4122	< 0.0001
Zone – Southwest	Reference		
NeighborState – Not Burnable	**-2.5688**	0.4795	< 0.0001
NeighborState – Not Burned	**-4.4685**	0.3136	< 0.0001
NeighborState – Burned	Reference		

The ranking of the first four fuels was somewhat surprising, as in the previous chapter. Typically, Grass2 and TimberUnderstory4 are used to represent fuel types with more aggressive fire behavior (higher spread rates, longer flame lengths) than those represented by Grass1 and TimberLitter models so I expected that Grass 2 and TimberUnderstory4 would have higher conditional burn probabilities compared to Grass1 and TimberLitter. However, the coefficients indicate that Grass1 and TimberLitter contribute to higher conditional probabilities of burning than Grass2 and TimberUnderstory4. Recall that when I inspected a fuel model map with a fire perimeter overlay I observed two factors that explain this result: Grass1 and TimberLitter occur in relatively smaller patches than the other fuel types and when a fire encounters one of the patches, the patch is almost entirely contained within the fire perimeter. Also, it is reasonable to expect that the fire management philosophy in the park would not hinder fire spreading through patches of less aggressive fuels. These observations support the model results that when a fire encounters Grass1 or TimberLitter they are more likely to burn than other fuel types. Note that the model data and results do not indicate the severity or completeness of burning, merely that the area is considered 'burned' because it falls within the recorded fire perimeter.

One of the climate variables I selected to correspond with fire season severity and seasonal weather patterns was significant in this model (ENSO) while the other was not (Palmer Z). Unlike the previous study, all categories of the ENSO variable were significant in this model. As before, strong El Niño conditions promote the highest conditional probability of burning in Yellowstone National Park while strong La Niña (the reference category) promotes the lowest conditional probability of burning, as expected (Shoennagel et al. 2005). Again I observed that numeric SOI is not directly proportional to fire activity in the study area because the three moderate ENSO phase categories are not ranked from low SOI to high SOI (or vice

versa). Having similar coefficients on the three moderate categories of the ENSO variable suggests that conditional probabilities of burning would be similar under moderate ENSO conditions (cet. par.).

The zone variable was not part of previous model but was included here to potentially improve the predictive ability of the model. Location within the park through the zone variable is intended to represent the effects of management decisions regarding large fires that come from strategic land and fire management planning policies for the park (Yellowstone National Park 1992, 2004). Interestingly, only one zone was significantly different than the reference zone of Southwest. The coefficient for the suppression zone was significant and negative implying that there is a lower conditional probability of burning under large fire conditions in developed areas than in other areas, which is to be expected because of the need to prevent large fire activity in developed areas of the park. Lack of significance of the other zone variable coefficients might indicate that, while many management practices vary by location, large fire management policies may be similar across the park outside of developed areas.

The neighbor variable categories were highly significant and the magnitudes are as expected. When a target cell had a neighbor that could not burn (non-burnable) or a neighbor that did not burn, the conditional probability of burning was lower than when the target cell had a neighbor that does burn. Having a neighbor that could have burned but did not reduces the probability of burning more than if the neighbor is not burnable. This variable appears to have qualitatively captured the character of large fire movement on a landscape by correctly identifying that an area is most likely to burn when a neighboring area burns, is less likely to burn when a fire is behind a barrier to spread and that it is least likely to burn if the fire has died down to the point where adjacent burnable fuels have not burned.

Predictive Performance of the Model

A general measure of predictive performance for a binary response model is the area under the ROC (Receiver Operating Characteristics) curve, known as AUC (Fawcett 2006). For the models presented here the AUC is equivalent to the probability that a model will rank a randomly chosen burnt cell higher than a randomly chosen unburnt cell. Perfect predictive performance yields an AUC of 1.0 and predictive performance equivalent to random guessing yields an AUC of 0.5. I applied the model to the original data set and calculated an AUC for the neighbor model. Because I needed to know the burn state of each target cell's neighbor, I used an iterative procedure to calculate the predicted probabilities and tested a set of probability thresholds for identifying a burned cell. For a given probability threshold I started with just the ignition cell in a burned state of 1. If any cells adjacent to the ignition had a probability at the threshold or higher I assigned them a neighbor value of 1 and re-estimated the probabilities. This process continued until no additional cells had a probability above the threshold or I reached the edge of the fire window. I selected the probability threshold that gave the highest AUC.

Table 6 lists the AUC, the probability threshold used to establish if a cell is considered 'burned', the percent correct burn predictions (hit rate) and the percent of cells incorrectly predicted to burn (false alarm rate). This model is considered 'good' with an AUC of 0.83, a hit rate of 0.85 and a relatively low false alarm rate of 0.19. This model had predictive performance that was comparable to or better than other regression models of wildland fire developed from historical data. Models in the literature have AUCs in the range 0.64 to 0.71 (Krawchuk et al. 2006; Vilar del Hoyo et al. 2011) and gridded simulation models of burn probability have hit rates of 0.79-0.93 and false alarm rates of 0.24-0.39 (Anderson et al. 2007). The average hit rate

for this model falls in the same range as those for gridded simulation models and the false alarm rate is lower.

Table 6: Model predictive performance for neighbor model

AUC	Optimal Probability Threshold for Burned Cell	Percent Correct Burn Predictions	False Alarm Rate (predicted to burn, but did not)
0.83	0.38	0.85	0.19

I further investigated the predictive ability of the model by removing a single fire from the data set, recalculating coefficients and predicting the conditional burn probabilities of the cells in the removed fire for each of the 40 fires. The model performed well by generating an AUC of 0.80 or better for 17 of the 40 fires and an AUC of 0.85 or better for six of the fires. The AUCs for eight of the predicted fires were considered poor (<0.60). The hit rates (correctly predicted to burn) for the best performing fires were perfect (1.0) combined with false alarm rates (incorrectly predicted to burn) of less than 0.20. These results compared well with models in the literature and are similar to those for the model in the previous study.

Predicted Probability of Burn Pattern for an Ignition

Although the estimated coefficients reflect population-averaged values and are not strictly appropriate for estimating effects for individual fire events, examining predictive plots for individual ignitions can be instructive and can highlight how the models might work in practice. Figure 10 illustrates the original burn pattern and the modeled average probability pattern for a representative large fire ignition in the data set – the August 2002 Phlox fire. Panel (a) shows the ignition cell (black), the actual burn footprint (grey) and all non-burnable regions

(hashed cells). Panel (b) shows the expected probability of burning for each target cell in the fire grid using the coefficients in Table 5 along with the proper input data for each target cell.

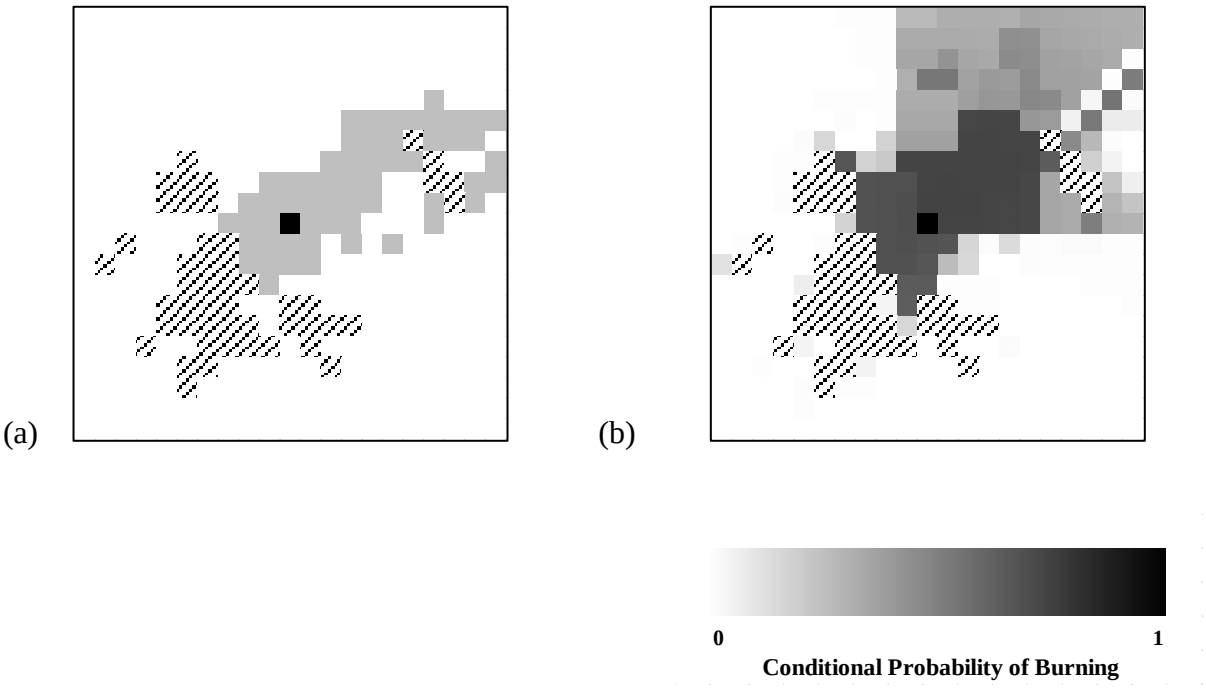


Figure 10: Actual burn pattern and estimated probability of burn pattern for the Phlox fire (2002). (a) Actual burn pattern: ignition cell is black (center); grey cells indicate the burned area; hashed cells are barriers to spread. (b) Estimated probabilities of burning for the Phlox fire.

There are some trends to note about the behavior of the predictive model. In general, modeled burn probabilities are higher near the ignition and lower near the edges of the grid, as expected. The model generally predicts higher probabilities of burning to the north and northeast of the ignition, as expected from knowledge about prevailing winds and large fire shapes in Yellowstone National Park. Barriers to spread behave as expected by reducing the probability of burning behind a barrier and large barriers effectively prevent a fire from moving through the barrier as demonstrated by the barrier to the west and southwest of the ignition. Smaller barriers have the effect of deflecting fire movement and creating a pocket of lower probability beyond

the barrier, as for the small barrier to the northeast of the ignition. The actual footprint of the Phlox fire (Figure 10(a)) shows that the fire burned beyond its small barrier but not immediately behind it. The model approximately captures this in the checkerboard pattern downstream of the barrier. Capturing the effect of such barriers through the view angle variable in a regression model was a key objective in this research and it appears to have been successful by approximating the expected burn pattern of a fire behind a spread barrier.

Another expectation of a population-averaged model useful for strategic planning is that the probabilities will, more or less, smoothly decrease outward from the ignition given the arrangement of fuels, topography and barriers to spread. Although the model from the previous chapter had a smoother decrease of probability outward from the ignition, the present model produces patterns similar to what one would expect a population-averaged probability pattern to look like: the probability of burning is higher near the ignition than in other areas of the grid, the fire is most likely to spread to the northeast, north and east, and barriers to spread behave as expected.

Conclusions

The technique presented in this chapter for modeling conditional probability of burning for large fires addressed some of the goals of a large fire risk analysis by predicting the probabilistic spatial extent of large fires. This model is a first attempt to incorporate a variable representing the spread process for large fires in a gridded regression model of burn probability. I found that the new variable indicating the burn state of a target cell's neighbor improved the predictive ability of the model and model performance was comparable to or better than similar models in the literature. Model results corresponded well with fire behavior theory and management philosophies in Yellowstone National Park. Specifically: the view angle proxy for

barriers to spread made a significant contribution to predicting conditional burn probability and the sign of the coefficient supports the hypothesis that increasing view angles corresponds to a decrease in probability of burning; fuel model coefficients have signs that are consistent with fire behavior theory and likely management practices in the park; probability of burning for developed areas were significantly different than other areas and the sign on the coefficient is in agreement with management practices for areas requiring suppression of fires; and the neighbor variable was highly significant and ranking of the neighbor coefficients by magnitude corresponded well with how large, spreading fires are expected to behave.

CHAPTER 4

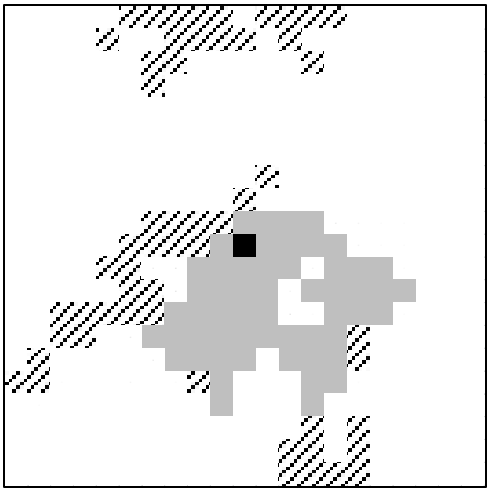
MODEL APPLICATIONS

There are three applications in this chapter: a comparison of the two models for a single ignition, a discussion of how population-averaged large fire probabilities change with monthly drought and ENSO conditions using the autoregressive covariance model, and a test of concept for fire progression modeling using the neighbor model.

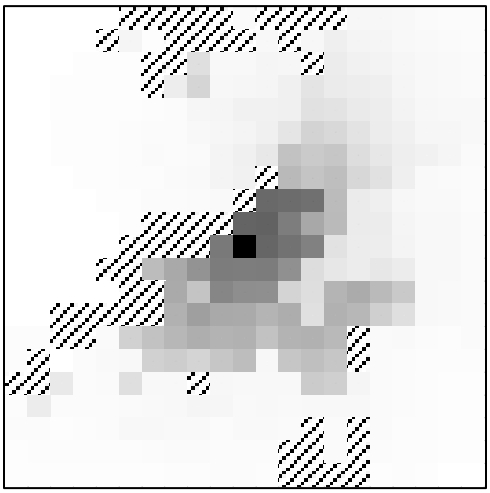
Population-Averaged Large Fire Probability Pattern for an Ignition – Comparing the Two Models

Figures 11 and 12 illustrate the original burn pattern and the modeled conditional probability of burn pattern for two representative large fire ignitions in the data set – the August 2001 Falcon fire (Figure 11) and the August 2002 Phlox fire (Figure 12). Panel (a) of each figure shows the ignition cell (black), the actual burn footprint (grey) and all non-burnable regions (hashed cells). Panels (b) and (c) show the estimated probability of burning for each target cell in the fire grid using the coefficients in Tables 2 and 5 along with the proper input data for each target cell.

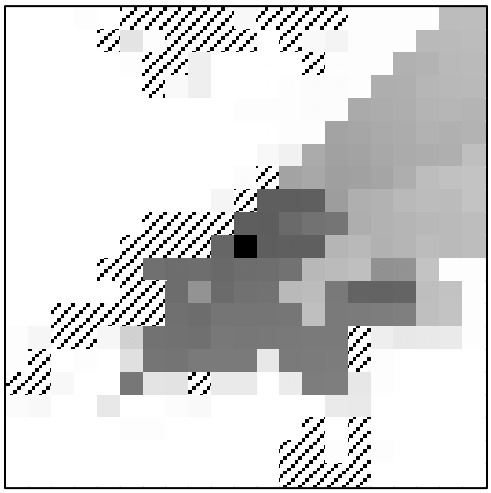
There are some trends to note about the behavior of the predictive models. In general, modeled burn probabilities are higher near the ignition and lower near the edges of the grid for each model, as expected. Both models generally predict higher probabilities of burning to the north and northeast of the ignition, as expected from knowledge about prevailing winds and large fire shapes in Yellowstone National Park.



(a) Actual Burn Pattern

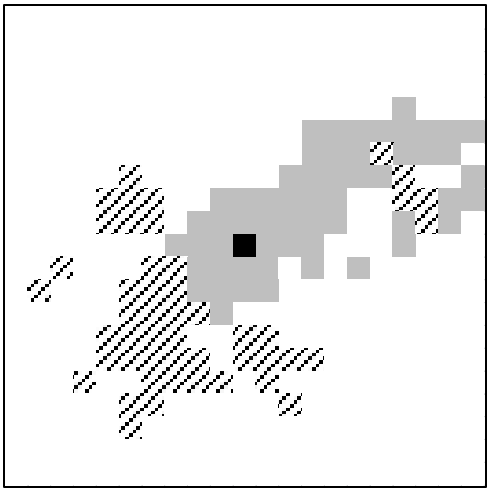


(b) Autoregressive Model Prediction

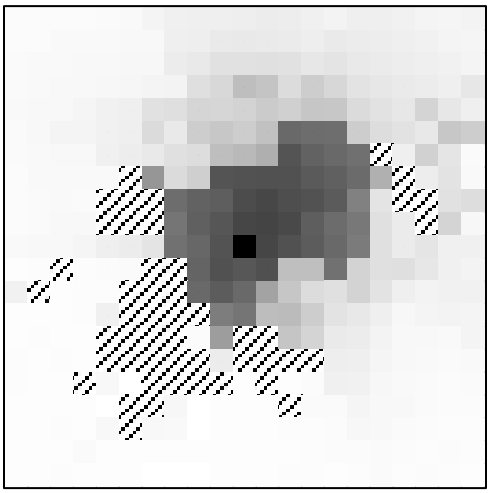


(c) Neighbor Model Prediction

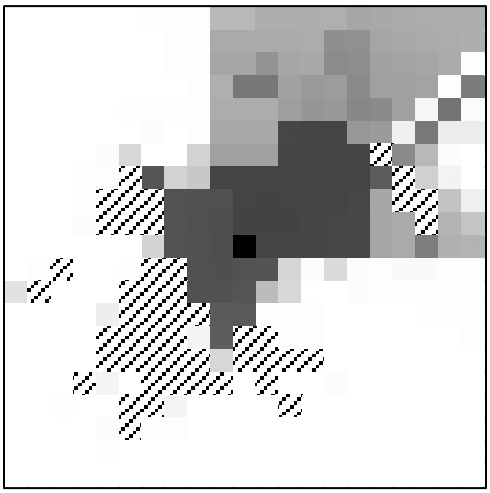
Figure 11: Falcon fire (2001): Ignition cell (black), actual burn footprint is grey in (a); non-burnable regions are hashed cells; probabilities in (b) and (c) are shaded:



(a) Actual Burn Pattern



(b) Autoregressive Model Prediction



(c) Neighbor Model Prediction

Figure 12: Phlox fire (2002): Ignition cell (black), actual burn footprint is grey in (a); non-burnable regions are hashed cells; probabilities in (b) and (c) are shaded:

Barriers to spread behave as expected in both models by reducing the probability of burning behind a barrier. Large barriers effectively prevent a fire from moving across or through the barrier as demonstrated by the barrier to the west and northwest of the ignition in Figure 11 and the barrier to the west and southwest of the ignition in Figure 12. Smaller barriers have the effect of deflecting fire movement and creating a pocket of lower probability beyond the barrier, as in the small barrier to the southeast of the ignition in Figure 11 or the small barrier to the northeast of the ignition in Figure 12. The actual Falcon fire spread out a little to the sides of, but not behind, the small barrier to the southeast while the Phlox fire burned beyond its small barrier but not immediately behind it. Capturing the effect of such barriers through the view angle variable in a regression model was a key objective in this research and it appears to have been successful in approximating the expected burn pattern of a fire behind a spread barrier.

Another expectation of these models is that the population-averaged probabilities will, more or less, smoothly decrease outward from the ignition given the arrangement of fuels, topography and barriers to spread. The autoregressive model has a smoother decrease of probability outward from the ignition than the neighbor variable model. Both models, however, are occasionally prone to slight over-predictions in areas far away from the ignition, such as that along the southern edge of the barrier to the north of the ignition in Figure 11. Over-prediction results from the presence of fuel models that have high probabilities of burning. The quantitative distance to ignition variable and the categorical sector variable were attempts to correct for this type of over-prediction and, while incorporating these variables into the models reduced the likelihood of distant areas of over-prediction, it did not entirely eliminate them.

Population-Averaged Large Fire Probability Pattern Changes under Various Drought and ENSO Conditions

The primary purpose for the models developed here is for strategic wildland fire planning exercises. One possible strategic application is to examine how average large fire probabilities might change under different conditions of monthly drought or for different fire season conditions such as those represented by the ENSO phenomenon. To illustrate how population-averaged large fire probability patterns change under various drought and ENSO conditions, I chose an ignition location outside the original data set near the Grand Canyon of the Yellowstone. I selected an ignition cell to the southeast of the canyon in a broad patch of TimberUnderstory 4 fuels. Figure 13 is a picture taken along the trail system south of the canyon and is representative of the TimberUnderstory 4 fuels present in the canyon area of the park.

Figure 13: Representative TimberUnderstory 4 fuels south of the Grand Canyon of the Yellowstone.

Figure 14 illustrates the canyon wall that sets up a barrier to spread to the north and northwest of the ignition. Figure 15 shows the raster graphic for the fuel model inputs and figure 16 shows the raster graphic for the slope inputs.

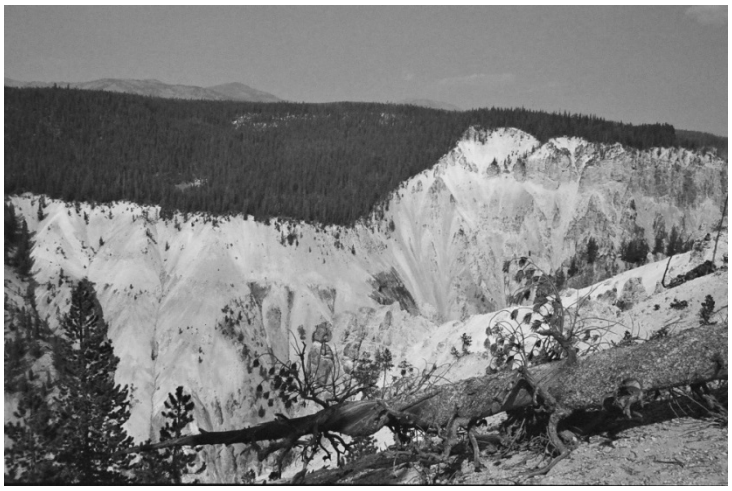


Figure 14: Canyon walls form a barrier to spread.

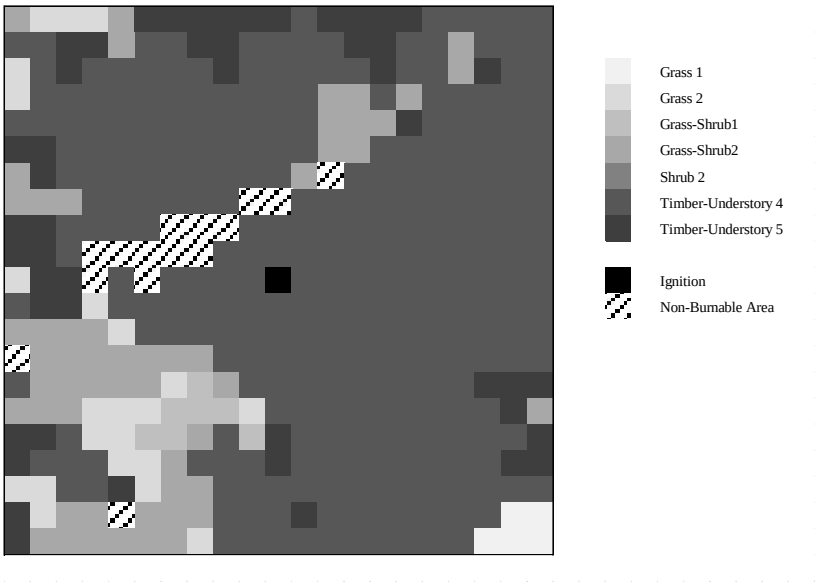


Figure 15: Raster graphic of fuel model for the ignition near the Grand Canyon of the Yellowstone. Hashed areas are non-burnable barriers to spread; the largest barrier is the canyon itself to the north and northwest of the ignition cell (ignition cell is black).

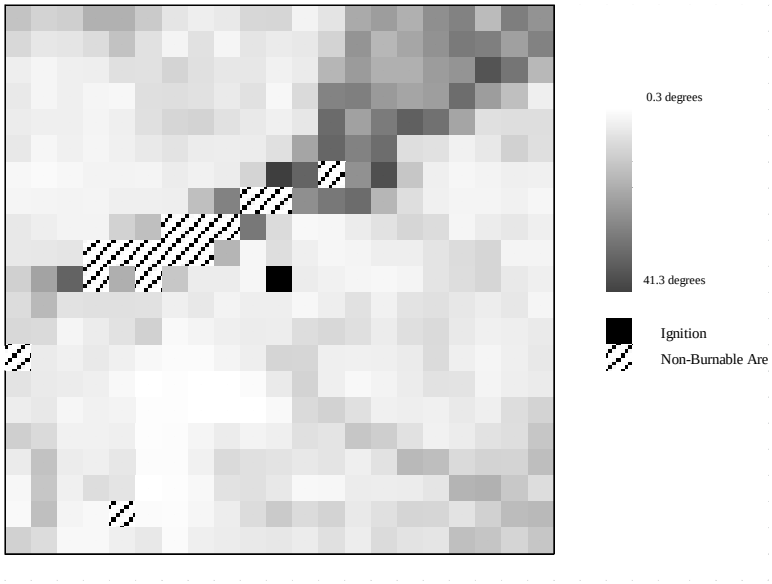
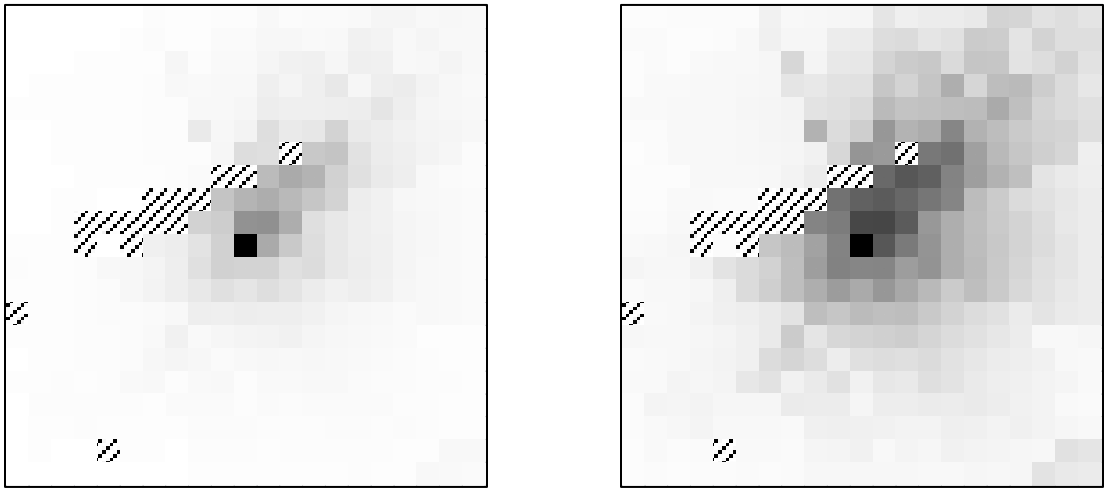


Figure 16: Raster graphic of slope for the ignition near the Grand Canyon of the Yellowstone. Hashed areas are non-burnable barriers to spread; the largest barrier is the canyon itself to the north and northwest of the ignition cell (ignition cell is black).

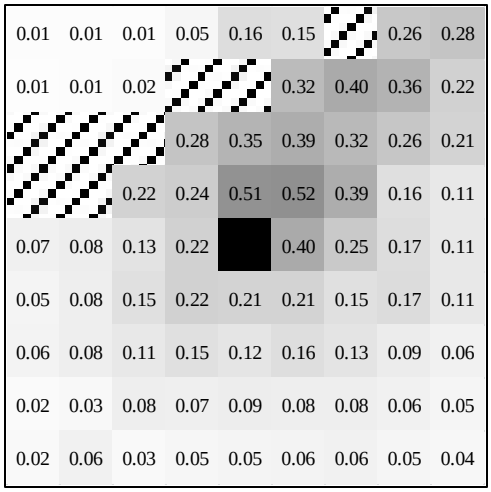
The first comparison is large fire probability patterns under conditions of monthly drought (severe or extreme) and no drought. To facilitate the comparison between drought and no drought I hold the ENSO variable constant at moderate El Niño. Figure 17(a) shows the probability pattern for the canyon area ignition under conditions of no drought while (b) shows the same ignition for severe to extreme drought. In general, drought is predicted to increase the conditional probability of burning given an ignition, as expected. The canyon near the upper and lower falls (denoted by a hashed area in the figures) is expected to provide a barrier to the spread of a large fire to the northwest of the ignition. The Yellowstone River runs to the northeast out of the canyon and the model does not capture this linear feature as a barrier to spread once it is beyond the lower canyon. Future models could attempt to estimate what effect, if any, linear barriers such as rivers and roads have on the probability of burning.

Figure 18 is zoomed in to the area close to the ignition and includes the numeric predicted probability values for drought and no drought conditions. The probabilities immediately to the north and northeast of the ignition increase about 60% in drought conditions (for example, from 0.51 to 0.84 in the cell immediately north of the ignition) while the probabilities in the cells a little further to the northeast more than double under drought – the cell in the upper right of the figure increases from a probability of 0.28 with no drought to 0.66 with drought. While these predictions are not expected to be exact, the model can quantify the likely average increase in probability of burning under severe or extreme monthly drought for use in strategic planning.

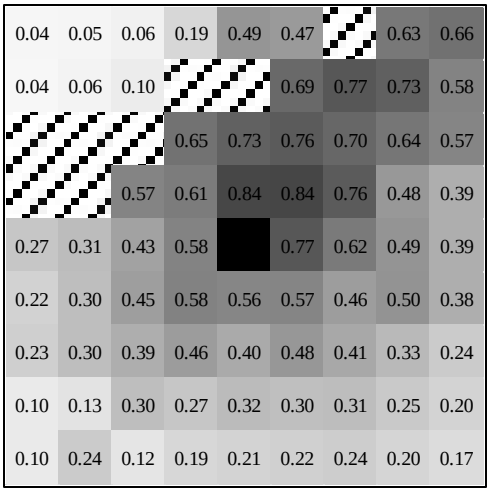


(a) No drought (b) Severe or extreme drought

Figure 17: Probability of burning given an ignition under (a) no drought or (b) drought. Probabilities are shaded: 



(b) No drought

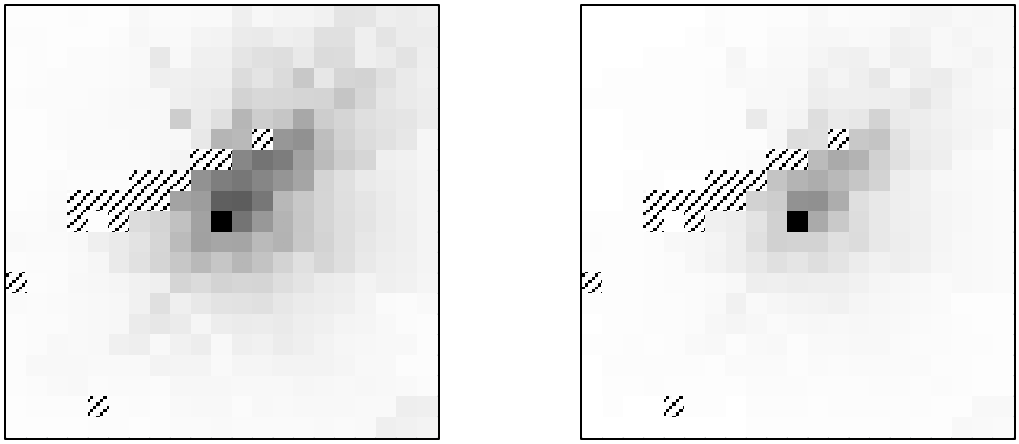


(b) Severe or extreme drought

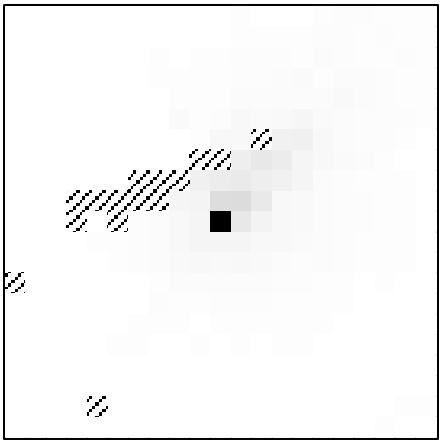
Figure 18: Probability of burning near the ignition under (a) no drought or (b) drought.

For the next series of graphics, I hold monthly drought constant at the 'no drought' level and varying the ENSO condition. Figure 19 shows the expected probability patterns for the selected ignition point under strong El Niño, moderate El Niño and strong La Niña and Figure 20 shows numeric predicted probabilities near the ignition.

Recall from the estimated coefficients for the autoregressive model that strong El Niño is expected to increase the conditional probability of burning while strong La Niña is expected to decrease the probability of burning. Quantitatively, strong El Niño condition are expected to increase the probability of burning by approximately 50% over those of a moderate El Niño year while a strong La Niña is expected to decrease the probability of burning by about 65%. Other areas of the grid near the ignition see similar changes in probability as the ENSO phase changes. Such information could be used to support staffing decisions or requests for additional resources as probability of burning changes with seasonal conditions in the park.

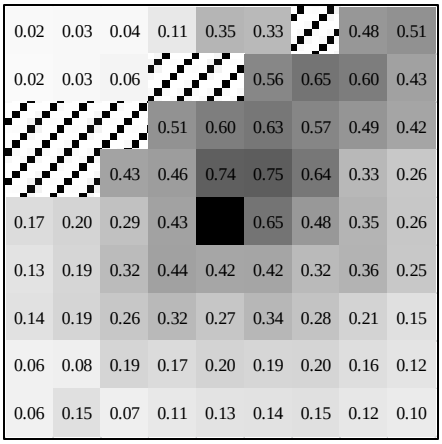


(a) Strong El Niño (b) Moderate El Niño

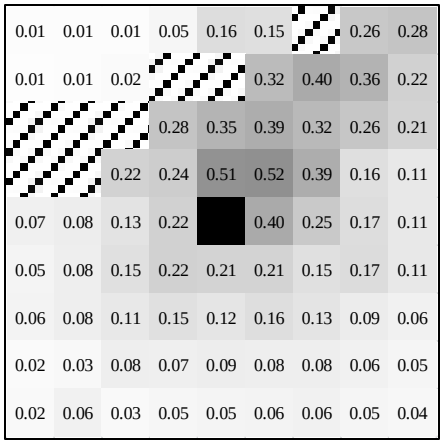


(c) Strong La Niña

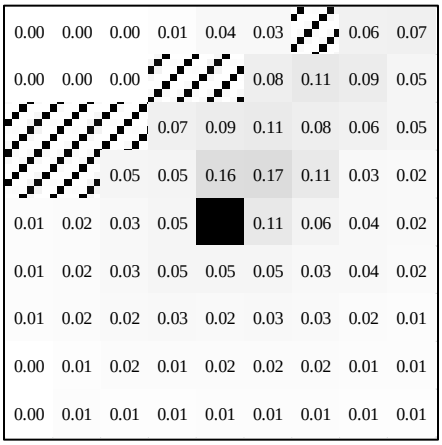
Figure 19: Probability of burning given an ignition under different ENSO conditions. Probabilities are shaded: 



(a) Strong El Niño



(b) Moderate El Niño



(c) Strong La Niña

Figure 20: Probability of burning near the ignition under different ENSO conditions. Probabilities are shaded: 

Test of Fire Progression Modeling with the Neighbor Model

This application is a test of concept to see if a modeling technique that incorporates the burn state of neighboring cells might be appropriate for estimating the final probabilistic size and shape of a large fire from fire progression data. A legitimate model for predicting fire progression would use fire perimeter progression maps to estimate a model rather than final fire perimeters as is the case for the neighbor model developed here. However, it is interesting to apply the current model to an actual fire event with fire progression to see if the technique might prove useful in the future for approximating probabilistic final fire sizes and spatial patterns for landscape fires permitted to burn over long periods as many are in Yellowstone National Park.

I used fire progression data from the Arnica fire which was ignited by lightning September 13, 2009 and burned an area around Arnica Creek north of the West Thumb Geyser Basin and southwest of Dryad Lake in Yellowstone National Park. Under warm, dry and windy conditions the fire grew from 4 acres on September 23 to about 9,300 acres by September 26. This and other descriptive information, including the final fire size (10,700 acres), can be found on the Inciweb Incident Information System website http://www.inciweb.org/incident/1901 (National Wildfire Coordinating Group 2013).

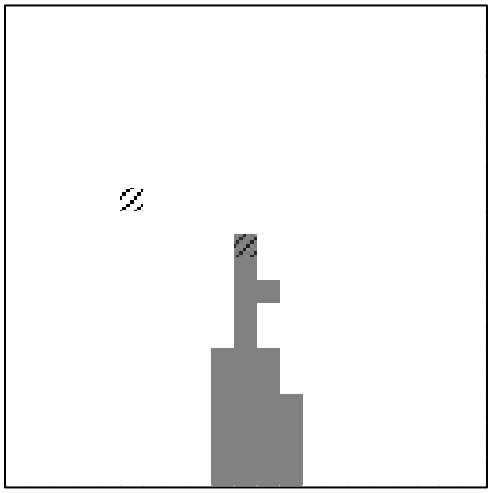
I used rasterized fire progression maps to predict the final probabilistic fire footprint at three different stages of the fire using the neighbor model. I supplied the model with information on which cells had burned and were unlikely to burn again ('cold black' where NeighborState = -1), which cells were actively burning (NeighborState = 1) and which cells had not yet burned (NeighborState = 0). This data set was available after the Arnica incident had ended, so I had to speculate as to which cells would be actively burning (1) or 'cold black' (-1) based on the

general direction of fire movement. If an analyst were working on an active fire they would have more accurate information available for their analysis.

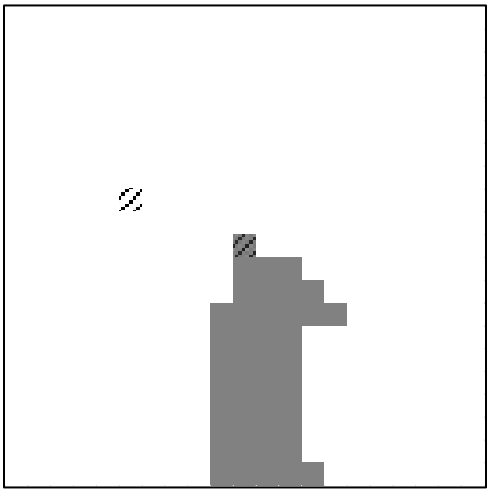
The first prediction uses the burn footprint from September 25 at 1500 (Figure 21). Cells in plain dark grey are actively burning while hatched grey cells are considered to be cold (non-burnable). For this first step only the ignition cell is considered cold. I estimated the final probabilistic footprint using the coefficients from Table 5 for the neighbor model and identifying each neighbor as not burned (0), not burnable (-1) or actively burning (1) based on the burn states shown in Figure 21(a). Figure 21(c) shows the estimated final footprint for the first prediction. The fire is most likely to burn areas labeled 'A' and the fire is generally expected to spread to the east of the ignition by the time it reaches its final size. Note that no weather streams or fire duration information were used to estimate this final probabilistic footprint. The next step in the actual fire progression map is shown in Figure 21(b). The neighbor model appears to have made an acceptable qualitative estimate of where the fire was most likely to spread: generally to the east and southeast of the ignition with the highest likelihoods being to the southeast of the ignition.

The second step of the prediction uses the burn footprint from September 26 at 1500 (Figure 22). Figure 22(c) shows the estimated final probabilistic footprint and Figure 22(b) shows the next step in the actual fire progression. Again, the fire is expected to move east and southeast of the actively burning area. A final set of predictions is shown in Figure 23. For all three fire progression steps the neighbor model makes an acceptable prediction of the probable path of the fire by indicating that, considering average large fire spread, the fire is expected to move generally east and southeast of the ignition. While this model was not specifically developed for estimating fire progressions, estimating a new model with fire progression data

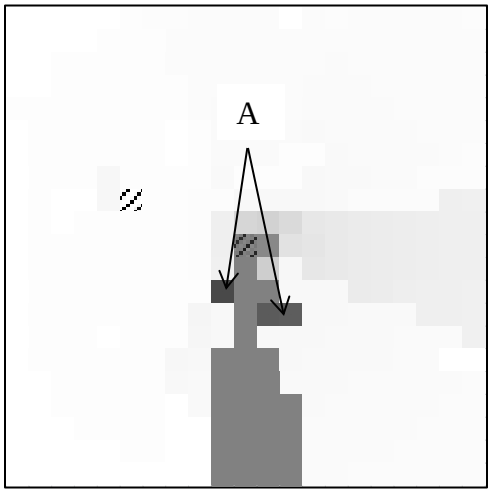
using the neighbor technique may have promise for an application like this. Potential improvements could be made by using actual fire progression data with an appropriate sampling scheme to develop a model specifically for long-term fire progression estimates.



(a) Starting point in fire progression

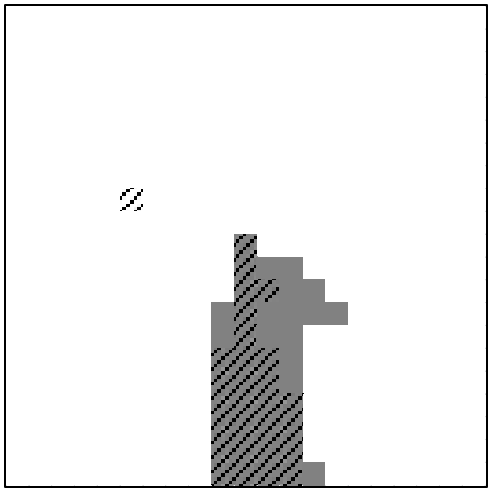


(b) Ending point for first progression

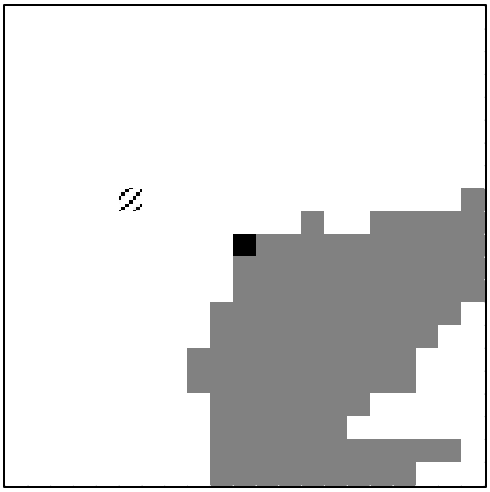


(c) Predicted final burn probability pattern given starting fire progression information in (a)

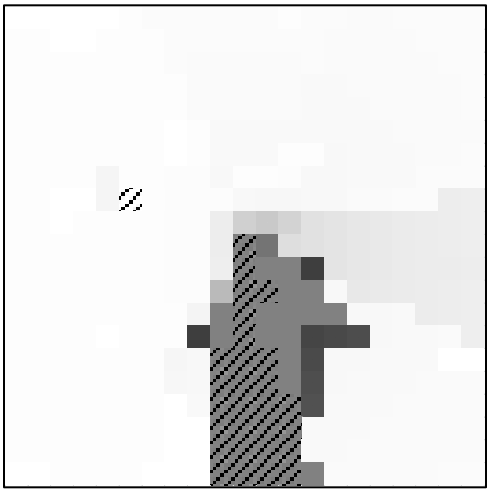
Figure 21: First fire progression prediction for Arnica fire. Panel (a) shows the data used in the model to make the prediction in (c): Hatched grey cells are considered 'cold black' and not burnable (-1); grey cells are burned (1) and white cells are not burned (0). Probabilities are shaded: 



(a) Starting point in fire progression

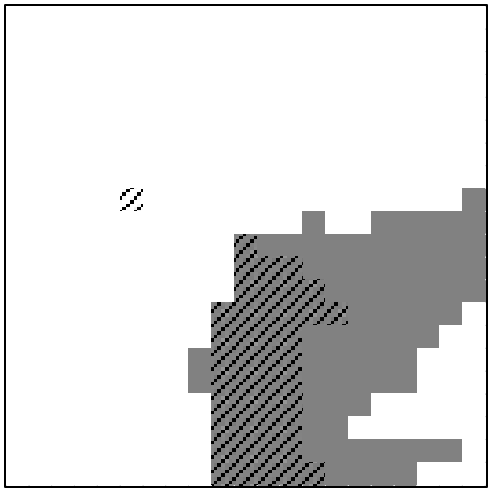


(b) Ending point for second progression

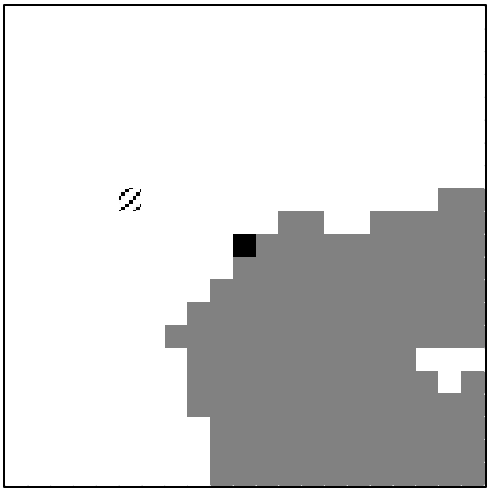


(c) Predicted final burn probability pattern given starting fire progression information in (a)

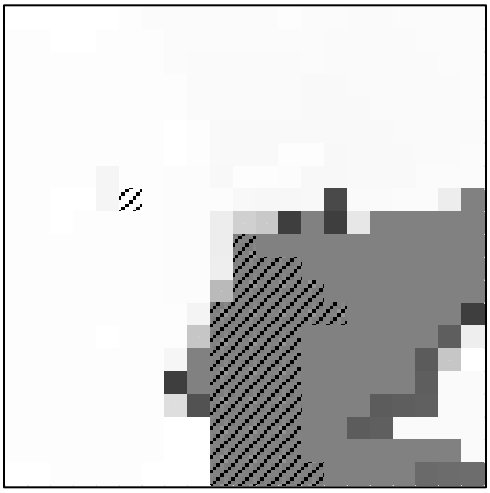
Figure 22: Second fire progression prediction for Arnica fire. Panel (a) shows the data used in the model to make the prediction in (c): Hatched grey cells are considered 'cold black' and not burnable (-1); grey cells are burned (1) and white cells are not burned (0). Probabilities are shaded: 



(a) Starting point in fire progression



(b) Ending point for third progression



(c) Predicted final burn probability pattern given starting fire progression information in (a)

Figure 23: Final fire progression for Arnica fire. Panel (a) shows the data used in the model to make the prediction in (c): Hatched grey cells are considered 'cold black' and not burnable (-1); grey cells are burned (1) and white cells are not burned (0). Probabilities are shaded:

CHAPTER 5

CONCLUSIONS AND SUGGESTIONS FOR FUTURE WORK

The techniques presented in this paper for predicting the conditional probability of burning for large fires addresses an important goal of a large fire risk analysis – estimation of large fire burn probabilities. The goal of this research was to develop statistical models of burn probability for large fires in two dimensions using data that are widely available or easily calculated and that achieve reasonable predictive ability. Both models were successfully estimated using variables that may be extracted directly or easily calculated from standard GIS layers and other sources. The selected cell size of 480 m was large enough to cover the entire park while maintaining a reasonable number of cells for statistical estimation and the coarseness of the raster lattice did not present any problems with estimation.

Two objectives were to estimate models from historical fire data that are appropriate for strategic program planning and to account for the average effect of weather streams and management actions on large fires without using detailed information on weather, fire duration or management tactics. The selected time period for the data set (1997-2000) had similar seasonal temperatures and precipitation to previous decades and the large fires occurred under short-term drought conditions that varied from no drought to extreme drought and under every category of the ENSO (El Niño – Southern Oscillation) phenomenon from strong La Niña to strong El Niño. I believe the selected time period was acceptable to represent a wide range of fire conditions for Yellowstone National Park and for developing and evaluating the models. I used distance and direction variables as proxies for weather streams, fire duration and management actions that affect fires. Based on the good predictive ability of both models, this

was a successful first attempt at capturing averaged information with proxy variables in a regression model on a landscape.

Another objective involved developing a variable to account for barriers to spread in a regression model. The view angle proxy for barriers to spread made a significant contribution to predicting conditional burn probability in both models and the negative coefficient supports the premise that an increasing view angle corresponds to a decrease in probability of burning. The view angle variable only captures barriers that are two-dimensional; future work on barriers to spread in regression models should consider methods to incorporate the effect of linear barriers to fire spread on a landscape.

The final objective was to investigate different methods for addressing the spreading, connected nature of large fire patterns on landscape within a regression model. The first method employed an autoregressive covariance structure. This supplied a mathematical relationship among all cells in a fire grid for use during model estimation. The resulting model performed well and produced reasonably smooth and connected conditional burn probabilities. The second method to address the spreading, connected nature of large fire patterns was to incorporate a variable that utilized the burn state of neighboring cells to affect the burn probability of a target cell. This model also performed well and produced connected burn probability patterns for fires in different fuel types and under different seasonal conditions. Both models suffer from instances of over-prediction (higher probabilities surrounded by areas of lower probability) near the edges of the fire grid, but the overall predictive ability of both models compared well with other fire probability models in the literature.

Future work could investigate weighted neighbor effects or a more sophisticated sector variable to further correct for over-predictions near the edges of the fire grid, particularly to the

north and northeast of the ignition. Additional future work could combine conditional probability model outputs with ignition model outputs to produce unconditional burn probabilities in support of wildland fire program planning (for example, see Preisler *et al.* (2004)). This would require using all ignitions to develop the conditional probability model instead of just modeling large fires (Brillinger 2003; Brillinger et al. 2003). It would also be beneficial to investigate grid alignment, the size of the fire grid (currently 441 cells) and cell size. Each of these parameters is expected to influence the contributions of variables, particularly topographic variables, and to influence predictive performance. Such studies could enhance our understanding of the effects of scale and orientation in spatial modeling and suggest ways to select optimal spatial parameters. Other variables should be investigated including: canopy data to enhance the description of forest and shrub fuels, other climate variables such as seasonal temperatures and precipitation, and human activity variables such as distance to roads and/or high use areas (Badia et al. 2011; Romero-Calcerrada et al. 2008). It would also be valuable to apply these techniques to other data sets to see how the performance of those models compares to the ones presented here.

The autoregressive model captures the diminishing effect of distance one expects to see in a population-averaged probability of burning for real fires on a landscape and has the advantage of modeling somewhat smoother probabilities outward from the ignition. The autoregressive model also does not require solving for final probabilities by iteration. The neighbor model performed slightly better (higher AUC) for this data set and may be appropriate to assess burn probabilities for fires when fire progression data is available, such as for ongoing wildland fires that are managed for resource benefits. With appropriate care in selecting the raster grid and fire window scale and in choosing a form for the working correlation matrix or

neighbor variables, either model could be used for large fire risk analysis in strategic planning when information is needed on probabilistic averages for shape and extent of large wildland fires.

REFERENCES

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. p. 267-281 in Proceedings of the Second International Symposium on Information Theory, Petrov, B.N. and F. Csaki (eds.), Budapest.

Albert, P.S. and L.M. McShane. 1995. A Generalized Estimating Equations Approach for Spatially Correlated Binary Data: Applications to the Analysis of Neuroimaging Data. *Biometrics* 51:627-638.

Amatulli, G., F. Perez-Cabello, and J.d.l. Riva. 2007. Mapping lightning/human-caused wildfire occurrence under ingition point location uncertainty. *Ecological Modelling* 200:321-333.

Anderson, K. 2010. A climatologically based long-range fire growth model. *International Journal of Wildland Fire* 19(7):879-894.

Anderson, K., G. Reuter, and M.D. Flannagan. 2007. Fire-growth modelling using meterological data with random and systematic perturbations. *International Journal of Wildland Fire* 16(2):174-182.

Badia, A., P. Serra, and S. Modugno. 2011. Identifying dynamics of fire ignition probabilities in two representative Mediterranean wildland-urban interface areas. *Applied Geography* 31:930-940.

Boychuk, D., W.J. Braun, R.J. Kulperger, Z.L. Krougly, and D.A. Stanford. 2009. A stochastic forest fire growth model. *Environmental and Ecological Statistics* 16:133-151.

Brillinger, D.R. 2003. Three Environmental Probabilistic Risk Problems. *Statistical Science* 18(4):412-421.

Brillinger, D.R., H.K. Preisler, and J.W. Benoit. 2003. Risk assessment: A forest fire example. p. 177-196 *in* Science and statistics: A Festschrift for Terry Speed, Goldstein, D. (ed.). Institute of Mathematical Statistics, Beachwood, OH.

Cardille, J.A. and S.J. Ventura. 2001. Occurrence of wildfire in the northern Great Lakes Region: Effects of land cover and land ownership assessed at multiple scales. *International Journal of Wildland Fire* 10:145-154.

Carmel, Y., S. Paz, F. Jahashan, and M. Shoshany. 2009. Assessing fire risk using Monte Carlo simulations of fire spread. *Forest Ecology and Management* 257:370-377.

Chou, Y.H., R.D. Minnich, and R.A. Chase. 1993. Mapping probability of fire occurrence in San Jacinto Mountains, California, USA. *Environmental Management* 17(1):129-140.

Chou, Y.H., R.D. Minnich, L. Salazar, J.D. Power, and R.J. Dezzani. 1990. Spatial autocorrelation of wildfire distribution in the Idylwild quandrangle, San Jacinto Mountain, California. *Photogrammetric Engineering and Remote Sensing* 56(11):1507-1513.

de Vasconcelos, M.J.P., S. Silva, M. Tome, M. Alvim, and J.M.C. Pereira. 2001. Spatial prediction of fire ignition probabilities: comparing logistic regression and neural networks. *Photogrammetric Engineering and Remote Sensing* 67(1):73-81.

Dickson, B.G., J.W. Prather, Y. Xu, H.M. Hampton, E.N. Aumack, and T.D. Sisk. 2006. Mapping the probability of large fire occurrence in northern Arizona, USA. *Landscape Ecology* 21:747-761.

Enloe, J. 2011. Historical Palmer Drought Indices. National Oceanic and Atmospheric Administration. *Date Accessed:* May 15, 2012, http://www.ncdc.noaa.gov/temp-and-precip/drought/historical-palmers.php

Fawcett, T. 2006. An introduction to ROC analysis. *Pattern Recognition Letters* 27:861-874.

Finney, M.A. 1998. FARSITE: Fire Area Simulator - model development and evaluation. U. S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Ogden, UT. 47 p.

Finney, M.A. 2005. The challenge of quantitative risk analysis for wildland fire. *Forest Ecology and Management* 211:97-108.

Finney, M.A. 2006. An Overview of FlamMap Fire Modeling Capabilities. p.213-220 in Fuels Management - How to Measure Success, Andrews, P.L. and B.W. Butler (eds.). U. S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Portland, OR.

Gonzalez, J.R., M. Palahi, A. Trasobares, and T. Pukkala. 2006. A fire probability model for forest stands in Catalonia (north-east Spain). *Annals of Forest Science* 63:169-176.

Hargrove, W.W., R.H. Gardner, M.G. Turner, W.H. Romme, and D.D. Despain. 2000. Simulating fire patterns in heterogeneous landscapes. *Ecological Modelling* 135:243-263.

He, H.S., B.Z. Shang, T.R. Crow, E.J. Gustafson, and S.R. Shifley. 2004. Simulating forest fuel and fire risk dynamics across landscapes - LANDIS fuel module design. *Ecological Modelling* 180:135-151.

Hernandez-Leal, P.A., M. Arbelo, and A. Gonzalez-Calvo. 2006. Fire risk assessment using satellite data. *Advances in Space Research* 37:741-746.

Krawchuk, M.A., S.G. Cumming, M.D. Flannigan, and R.W. Wein. 2006. Biotic and abiotic regulation of lightning fire initiation in the mixedwood boreal forest. *Ecology* 87(2):458-468.

Li, C. and M.J. Apps. 1996. Effects of contagious disturbance on forest temporal dynamics. *Ecological Modelling* 87:143-151.

Liang, K.Y. and S.L. Zeger. 1986. Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika Trust* 73(1):13-22.

Lozano, F.J., S. Suarez-Seoane, M. Kelly, and E. Luis. 2008. A multi-scale approach for modeling fire occurrence probability using satellite data and classification trees: A case study in a mountainous Mediterranean region. *Remote Sensing of Environment* 112:708-719.

Lozano, F.J., S. Suarez-Seoane, and E. Luis. 2007. Assessment of several spectral indices derived from multi-temporal Landsat data for fire occurrence probability modelling. *Remote Sensing of Environment* 107:533-544.

Mermoz, M., T. Kitzberger, and T.T. Veblen. 2005. Landscape Influences on Occurrence and Spread of Wildfires in Patagonian Forests and Shrublands. *Ecology* 86(10):2705-2715.

National Wildfire Coordinating Group. 2013. Inciweb, Incident Information System, Arnica Fire. *Date Accessed:* September 5, 2013, http://www.inciweb.org/incident/1901.

Pan, W. 2001. Akaike's Information Criterion in Generalized Estimating Equations. *Biometrics* 57(1):120-125.

Parisien, M.A., C. Miller, A.A. Ager, and M.A. Finney. 2010. Use of artificial landscapes to isolate controls on burn probability. *Landscape Ecology* 25:79-93.

Preisler, H.K., D.R. Brillinger, R.E. Burgan, and J.W. Benoit. 2004. Probability based models for estimation of wildfire risk. *International Journal of Wildland Fire* 13:133-142.

Redmond, K. 2007. Terminology of El Niño and La Niña Winters. Western Regional Climate Center. *Date Accessed:* May 15, 2012, www.wrcc.dri.edu/enso/ensodef.html

Roman-Cuesta, R.M., M. Gracia, and J. Retana. 2003. Environmental and human factors influencing fire trends in ENSO and Non-ENSO years in tropical Mexico. *Ecological Applications* 13(4):1177-1192.

Romero-Calcerrada, R., C.J. Novillo, J.D.A. Millington, and I. Gomez-Jimenez. 2008. GIS analysis of spatial patterns of human-caused wildfire ignition risk in the SW of Madrid (Central Spain). *Landscape Ecology* 23:341-354.

Rothermel, R. 1983. How to predict the spread and intensity of forest and range fires. Gen. Tech. Rep. INT-143. U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, Ogden, UT.

Scott, J.H. and R.E. Burgan. 2005. Standard fire behavior fuel models: a comprehensive set for use with Rothermel's surface fire spread model. Gen. Tech. Rep. RMRS-GTR-153. 72 p. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO.

Shoennagel, T., T.T. Veblen, W.H. Romme, J.S. Sibold, and E.R. Cook. 2005. ENSO and PDO variability affect drought-induced fire occurrence in Rocky Mountain subalpine forests. *Ecological Applications* 15(6):2000-2014.

USDA Forest Service. 2000. RERAP User's Guide: Rare Event Risk Assessment Process, Version 5.03. U.S. Department of Agriculture.

USDA Forest Service. 2012. FSPro Reference Guide. U.S. Department of Agriculture. *Date Accessed:* June 21, 2012, http://wfdss.usgs.gov/wfdss/pdfs/fspro_reference.pdf.

Vega Garcia, C., P.M. Woodard, S.J. Titus, W.L. Adamowicz, and B.S. Lee. 1995. A logit model for predicting the daily occurrence of human caused forest fires. *International Journal of Wildland Fire* 5(2):101-111.

Vilar del Hoyo, L., M.P. Martin Isabel, and F.J. Martinez Vega. 2011. Logistic regression models for human-caused wildfire risk estimation: analysing the effect of the spatial accuracy in fire occurrence data. *European Journal of Forest Research* 130:983-996.

Wei, Y., D. Rideout, and A. Kirsch. 2008. An optimization model for locating fuel treatments across a landscape to reduce expected fire losses. *Canadian Journal of Forest Research* 38:868-877.

Western Regional Climate Center. 2006a. Yellowstone Park, Wyoming Monthly Average Maximum Temperature (Degrees Fahrenheit). *Date Accessed:* May 24, 2012, http://www.wrcc.dri.edu/cgi-bin/cliMAIN.pl?wyyell.

Western Regional Climate Center. 2006b. Yellowstone Park, Wyoming Monthly Total Precipitation (inches). *Date Accessed:* May 24, 2012, http://www.wrcc.dri.edu/cgi-bin/cliMAIN.pl?wyyell.

Wiitala, M. and D. Carlton. 1994. Assessing long-range fire movement risk in wilderness fire management. p.187-194 in Proceedings of the 12th Conference on Fire and Forest Meteorology. Sociey of American Foresters, Jekyll Island, GA.

Yang, J., H.S. He, and S.R. Shifley. 2008. Spatial Controls of Occurrence and Spread of Wildfires in the Missouri Ozark Highlands. *Ecological Applications* 18:1212-1225.

Yellowstone National Park. 1992. Yellowstone National Park Wildland Fire Management Plan. National Park Service. U.S. Department of the Interior, Mammoth, WY.

Yellowstone National Park. 2004. 2004 Update of the 1992 Wildland Fire Management Plan. National Park Service. U.S. Department of the Interior, Mammoth, WY.

Yellowstone National Park. 2010. Yellowstone Resources & Issues. National Park Service. NPS Division of Interpretation, Mammoth, WY.

Zeger, S.L. and K.Y. Liang. 1986. Longitudinal Data Analysis for Discrete and Continuous Outcomes. *Biometrics* 42(1):121-130.

Ziesler, P.S., D.B. Rideout, and R. Reich. 2013. Modelling conditional burn probability patterns for large wildland fires. *International Journal of Wildland Fire* 22(5):579-587.

APPENDIX A

SAS CODE

This appendix contains sample SAS code for estimating the two conditional probability models. Some file structure and data lists have been eliminated or shortened to preserve space.

A.1: Autoregressive Covariance model

```
data ynp;
    infile 'C:\...\<FILENAME>.txt' firstobs=2 delimiter=',' LRECL=2500;
    input WindowCellID $ FireName $ Dir $ Dsqroot Elev DeltaElev1 DeltaElev2 Slope
        DeltaSlope1 DeltaSlope2 Aspect $ Burn1SL8 $ Burn2SL8 $ Burn3SL8 $
        Burn4SL8 $ Burn5SL8 $ Burn6SL8 $ Burn7SL8 $ Burn8SL8 $ FuelModel $
        ViewAngle Burn1SL1 $ Burn1SL3 $ Burn2SL3 $ Burn3SL3 $ PalmerZ $ Sector
        $ ENSO $ Zone $;
proc genmod data = ynp descend;
    class Dir FireName WindowCellID FuelModel PalmerZ Sector ENSO Zone;
    model Burn = Dir Dsqroot Elev DeltaElev1 Slope ViewAngle FuelModel Sector PalmerZ
        ENSO Zone/ dist=bin;
    repeated subject=FireName / within=WindowCellID maxit=1000
    type=user( 1 0.6231 0.3883 0.242 0.1508 0.0939...<matrix eliminated for space
        reasons> ... 0.6231 1 );
run;
```

A.2: Neighbor Variable model

```
data ynp;
    infile  'C:\...\<FILENAME>.txt' firstobs=2 delimiter=',' LRECL=2500;
    input WindowCellID $ FireName $ Dir $ Dsqroot Elev DeltaElev1 DeltaElev2 Slope
        DeltaSlope1 DeltaSlope2 Aspect $ Burn1SL8 $ Burn2SL8 $ Burn3SL8 $
        Burn4SL8 $ Burn5SL8 $ Burn6SL8 $ Burn7SL8 $ Burn8SL8 $ FuelModel $
        ViewAngle Burn1SL1 $ Burn1SL3 $ Burn2SL3 $ Burn3SL3 $ PalmerZ $ Sector
        $ ENSO $ Zone $;
proc genmod data = ynp descend;
    class Dir FireName WindowCellID FuelModel PalmerZ Sector ENSO Zone Burn1SL1;
    model Burn = Dir Dsqroot Elev DeltaElev1 Slope ViewAngle FuelModel Sector PalmerZ
        ENSO Zone Neighbor/ dist=bin;
    repeated subject=FireName / within=WindowCellID maxit=1000;
run;
```

APPENDIX B

FIRE GRIDS

This appendix contains graphics of the 40 large fire grids used for estimating the conditional probability models. The graphics identify the ignition cell (black), cells inside the fire perimeter (grey), outside the fire perimeter (white) and barriers to spread (diagonal hash).

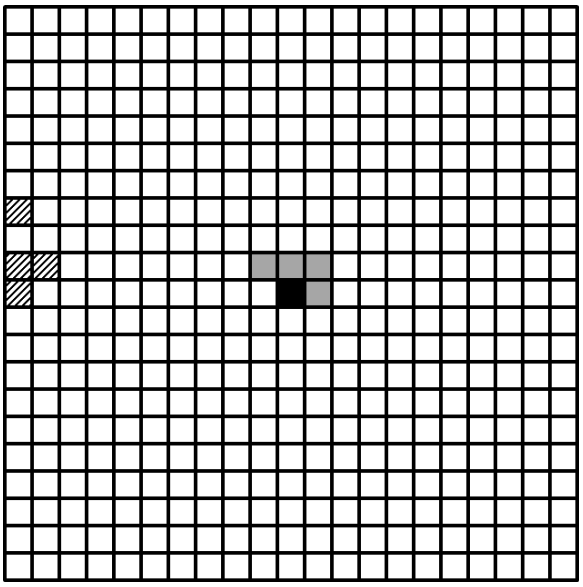


Figure A1: Washburn fire
1990; 374 acres (151 ha)

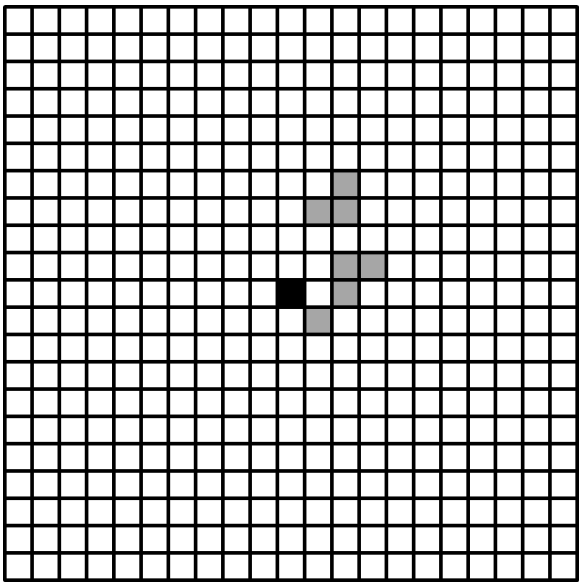


Figure A2: Pelican fire
1991; 322 acres (130 ha)

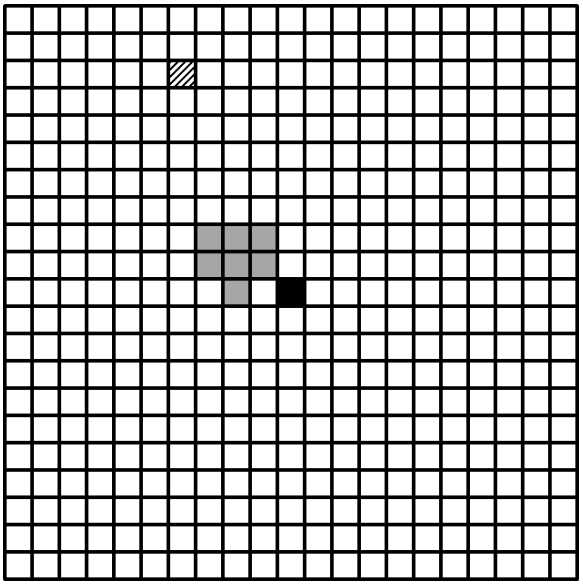


Figure A3: Plateau fire
1992; 389 acres (157 ha)

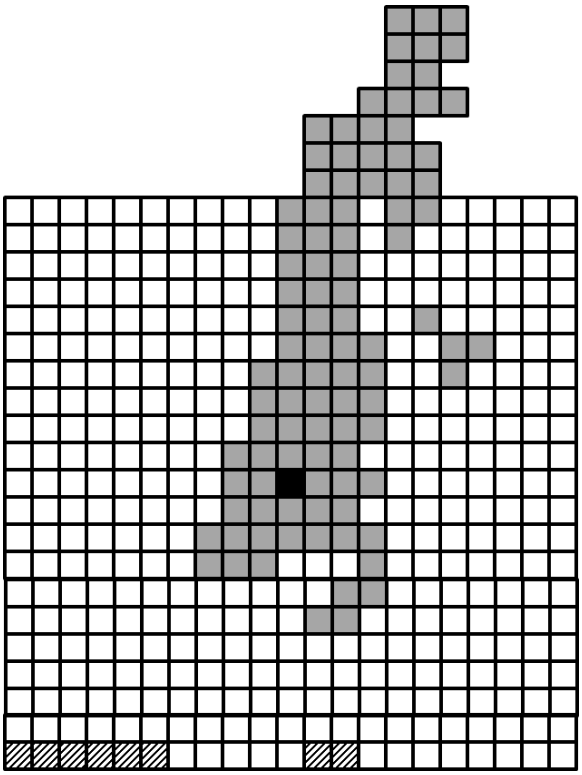


Figure A4: Tern fire
1994; 5266 acres (2131 ha)

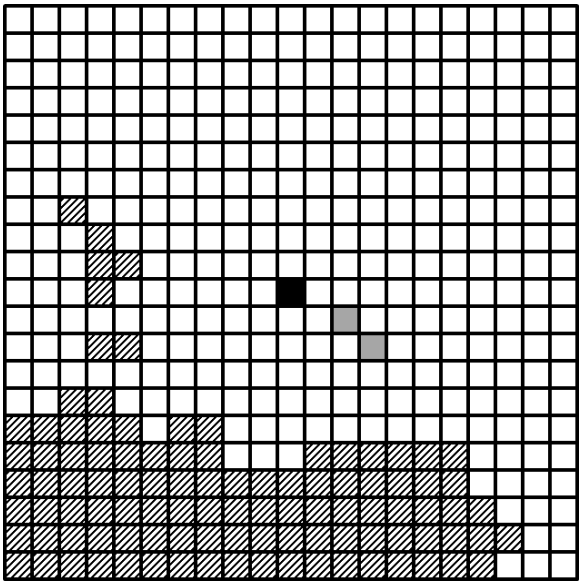


Figure A5: Pelican fire
1994; 95 acres (38 ha)

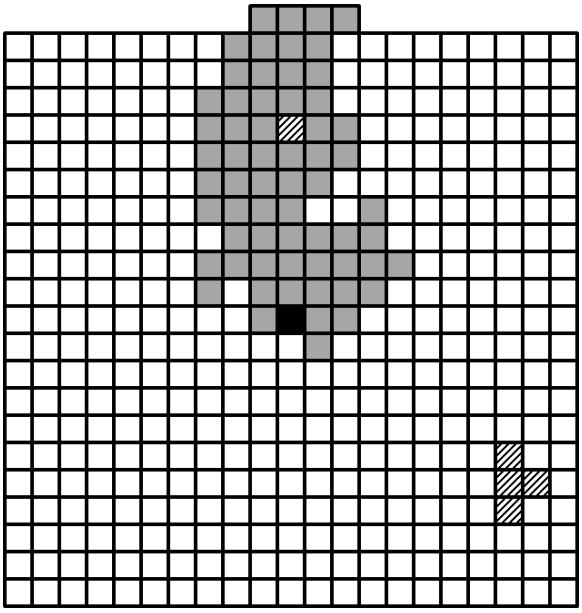


Figure A6: Raven fire
1994; 3648 acres (1476 ha)

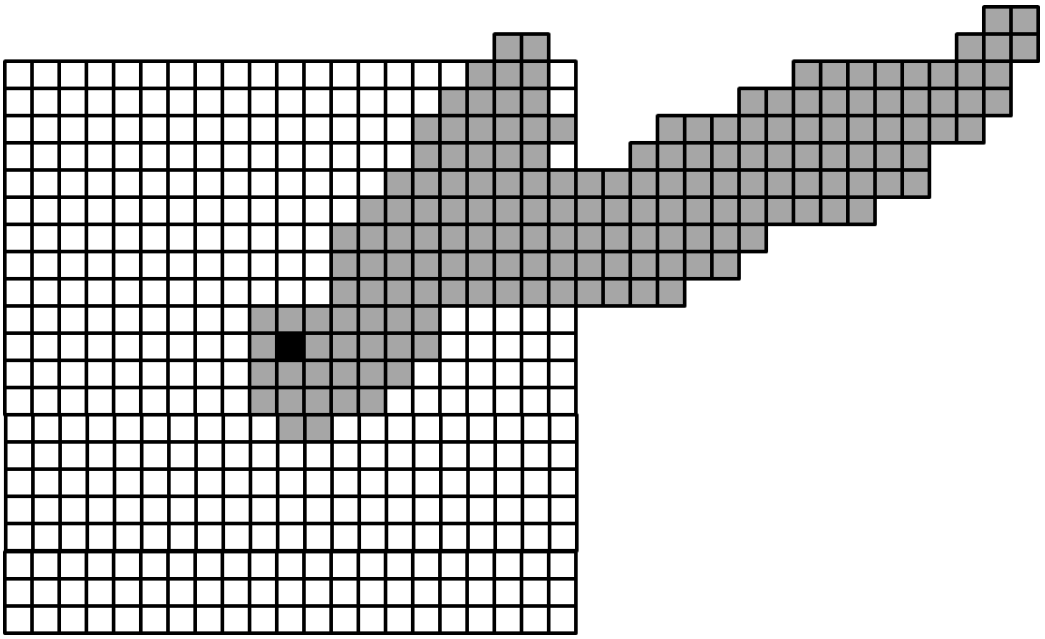


Figure A7: Robinson fire
1994; 9758 acres (3949 ha)

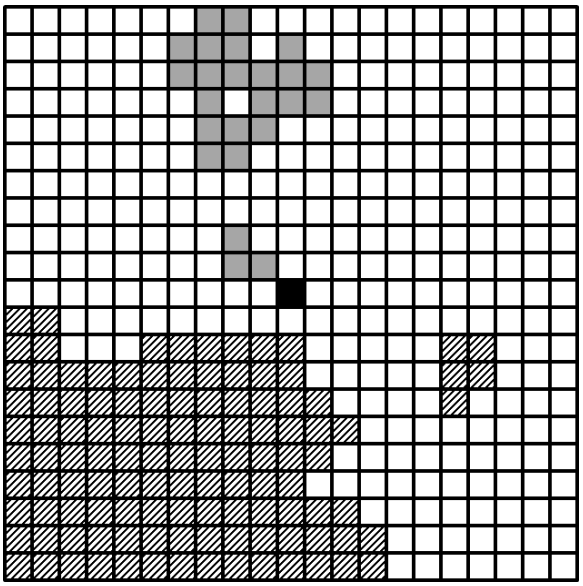


Figure A8: Pelican fire
1996; 1840 acres (745 ha)

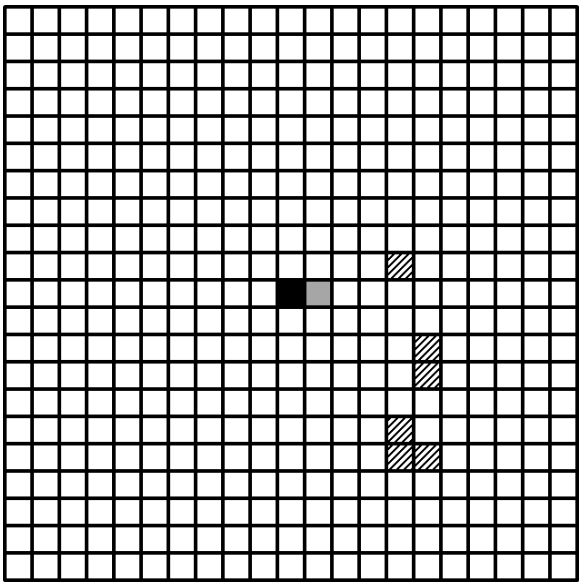


Figure A9: Sour fire
1998; 98 acres (40 ha)

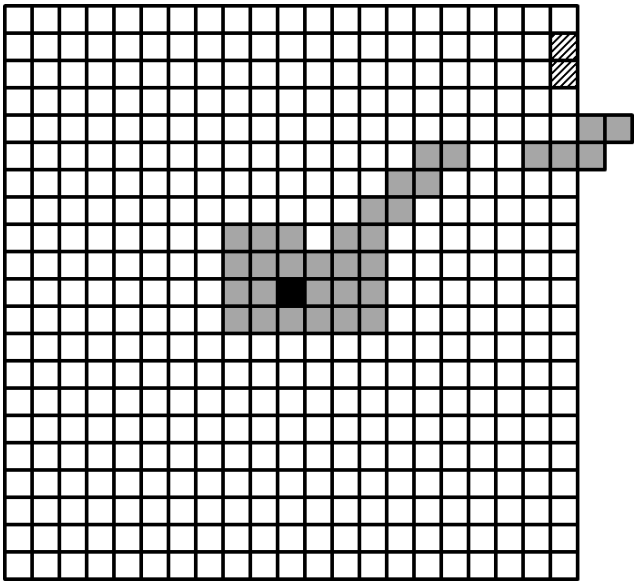


Figure A10: Unlucky fire
2000; 1986 acres (804 ha)

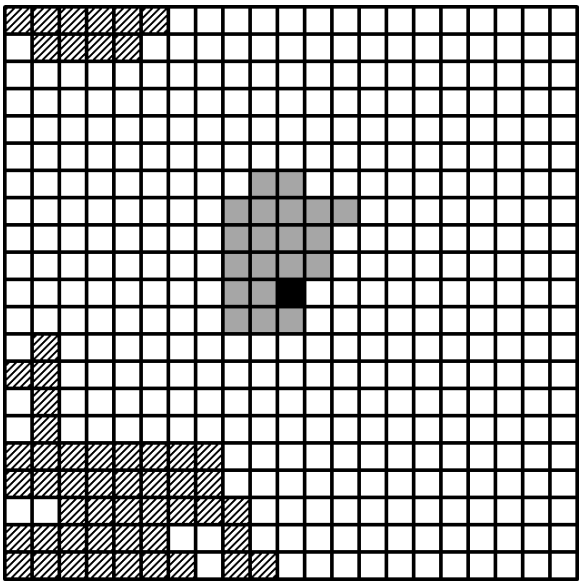


Figure A11: Moose fire
2000; 1136 acres (460 ha)

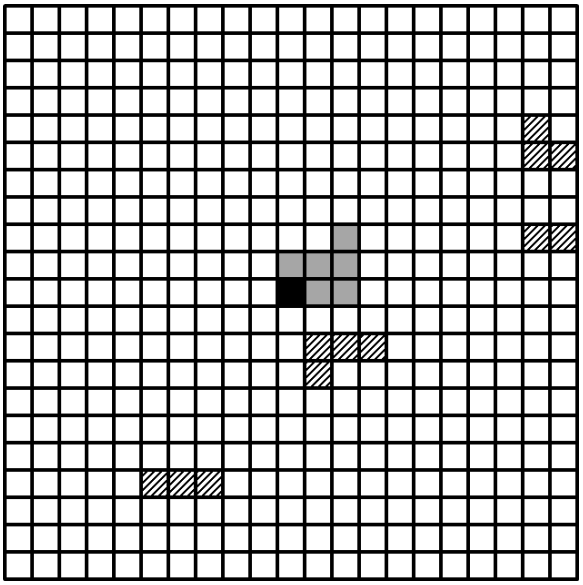


Figure A12: Boundary fire
2000; 339 acres (137 ha)

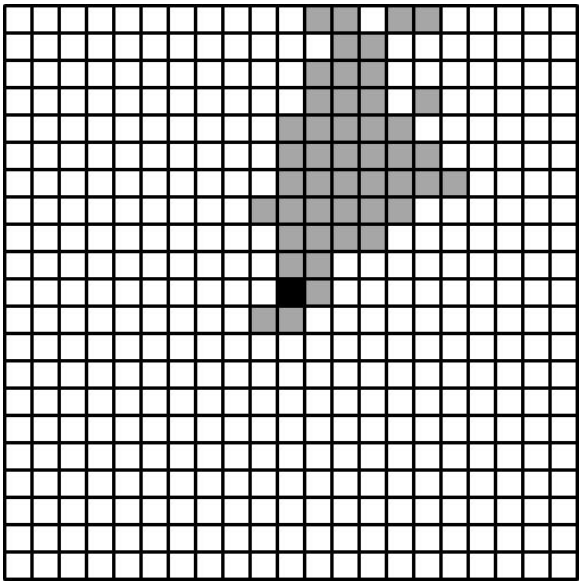


Figure A13: Plateau fire
2000; 2552 acres (1033 ha)

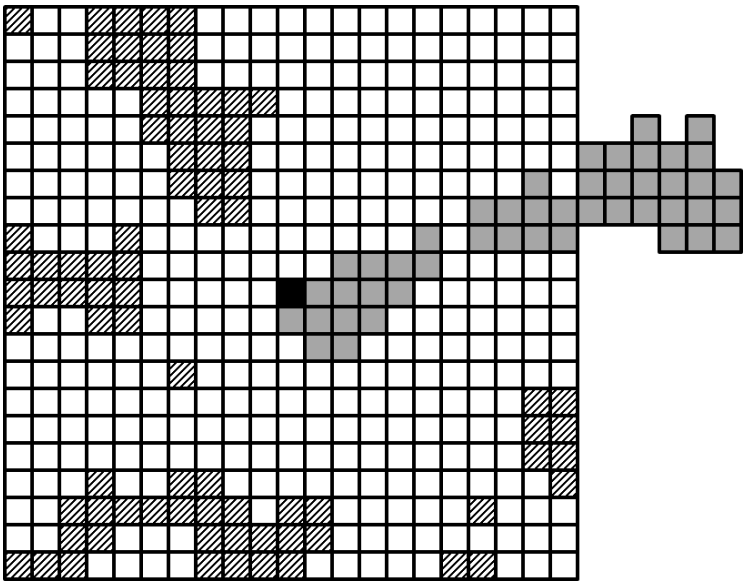


Figure A14: Arthur fire
2001; 2527 acres (1023 ha)

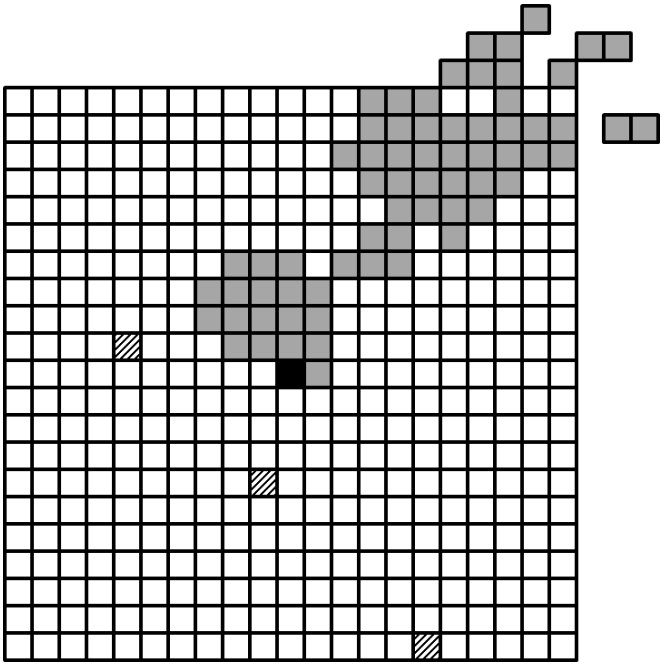


Figure A15: Sulphur fire
2001; 3976 acres (1609 ha)

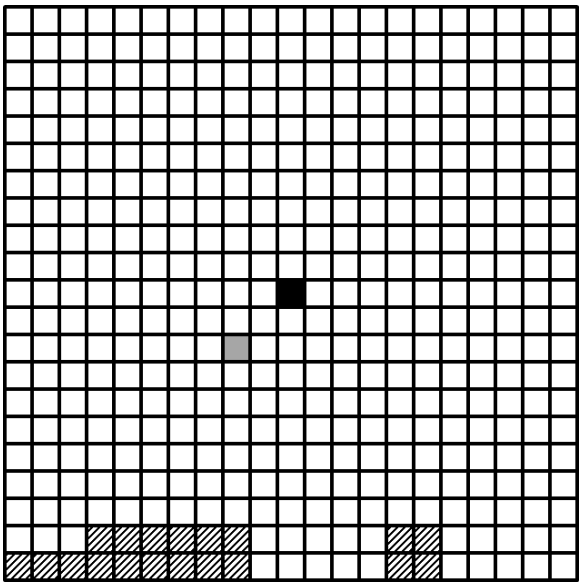


Figure A16: Stone fire
2001; 69 acres (30 ha)

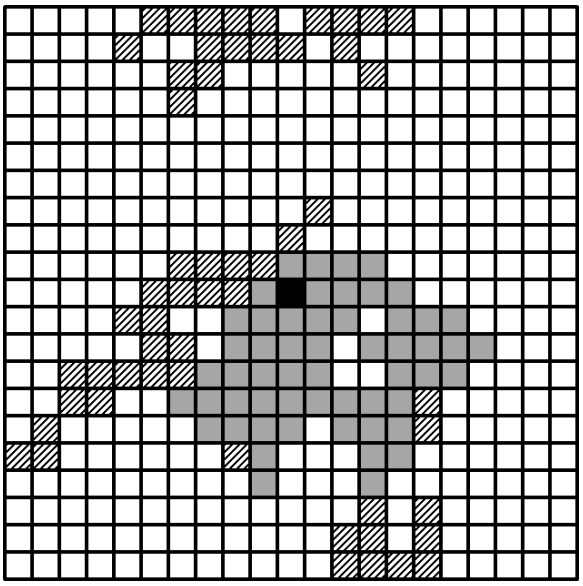


Figure A17: Falcon fire
2001; 3288 acres (1331 ha)

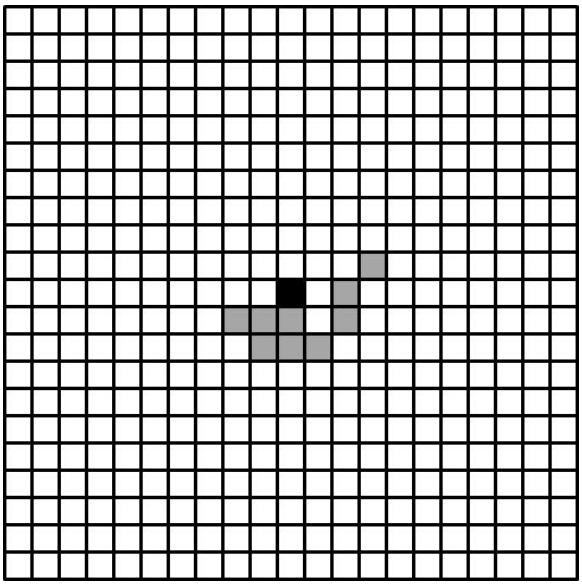


Figure A18: Little fire
2001; 456 acres (185 ha)

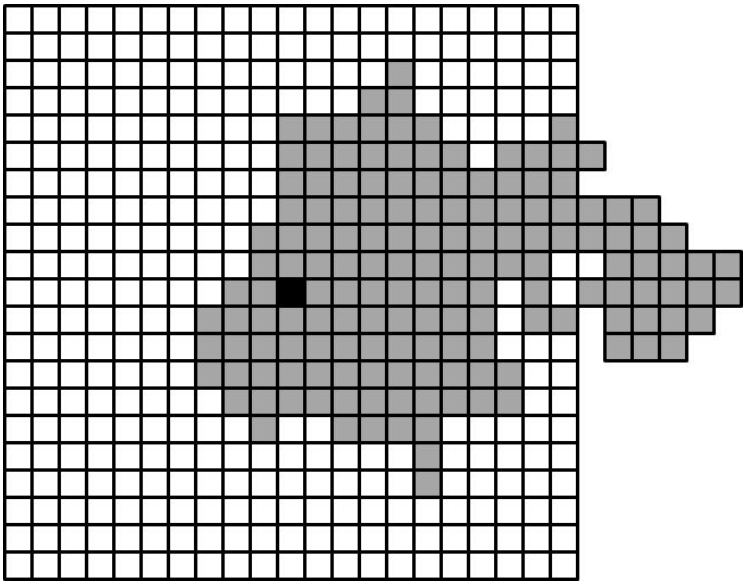


Figure A19: Broad fire
2002; 8775 acres (3551 ha)

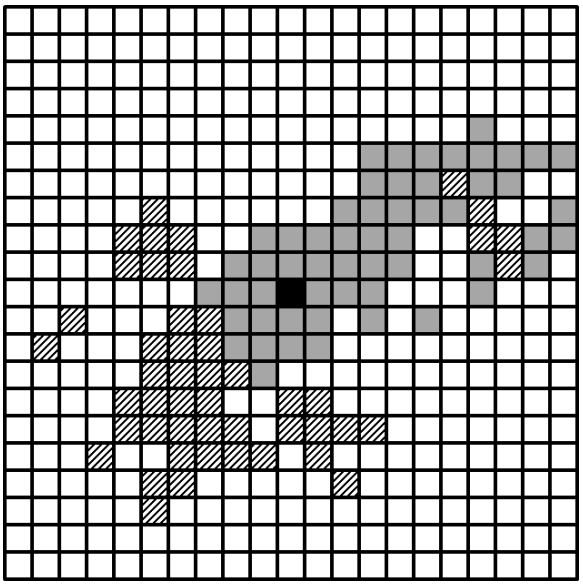


Figure A20: Phlox fire
2002; 3585 acres (1451 ha)

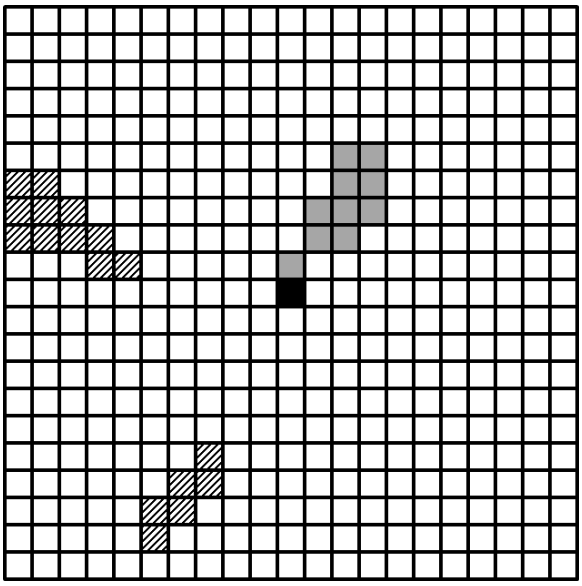


Figure A21: Baker's Hole fire
2003; 545 acres (221 ha)

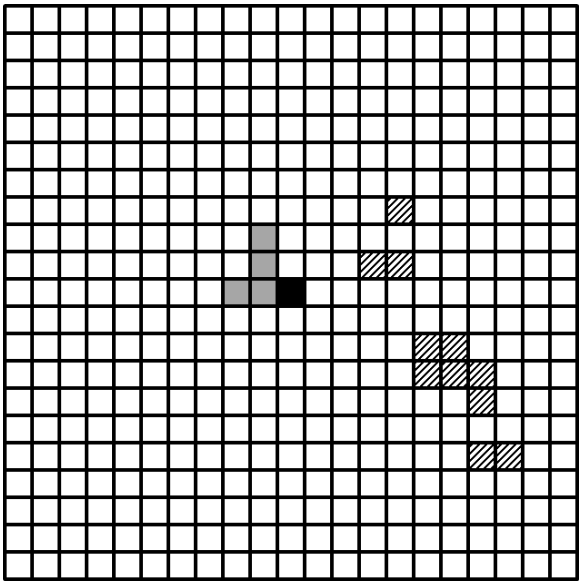


Figure A22: Amethyst fire
2003; 228 acres (92 ha)

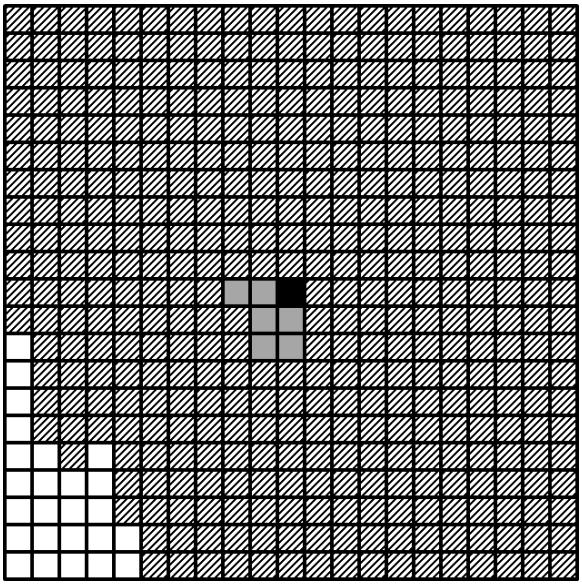


Figure A23: Frank fire
2003; 619 acres (251 ha)

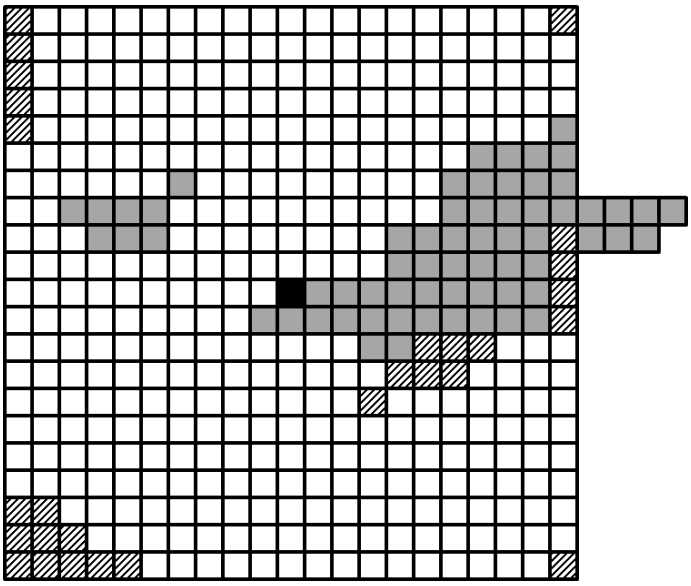


Figure A24: Grizzly fire
2003; 3702 acres (1498 ha)

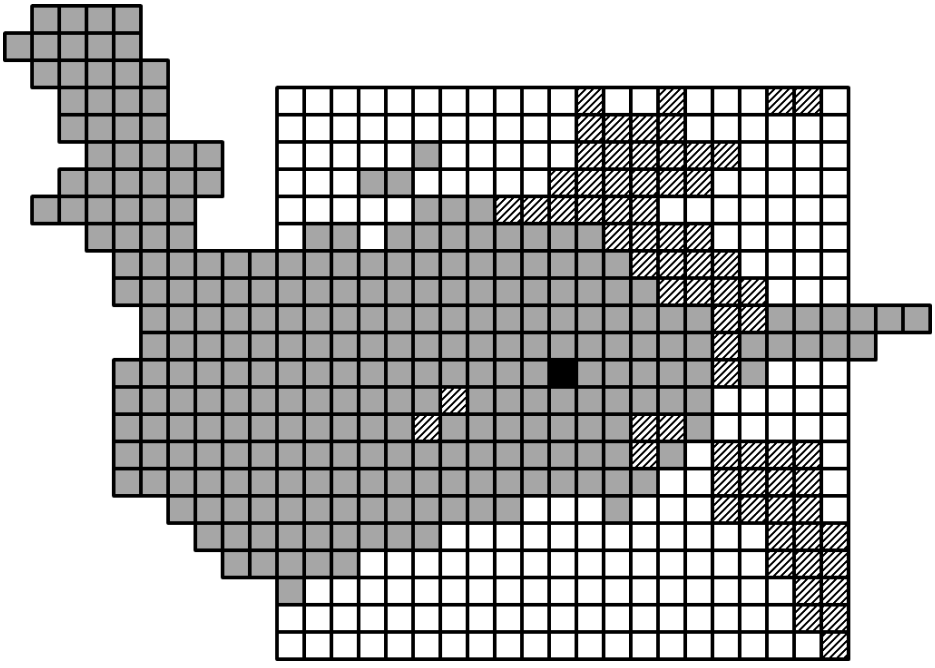


Figure A25: East fire
2003; 17294 acres (6999 ha)

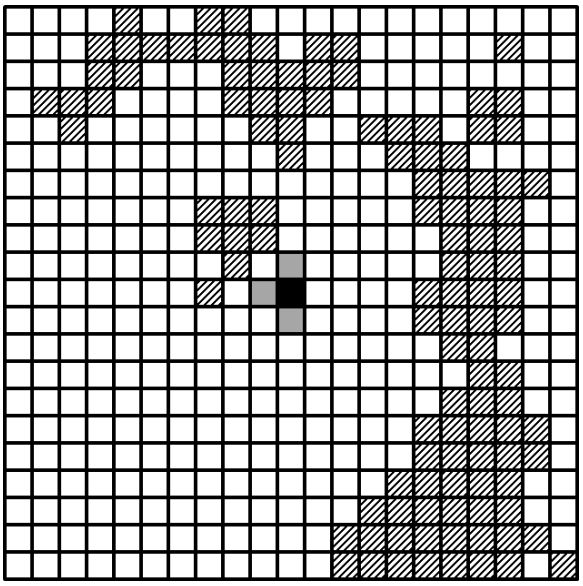


Figure A26: Tyson fire
2003; 192 acres (78 ha)

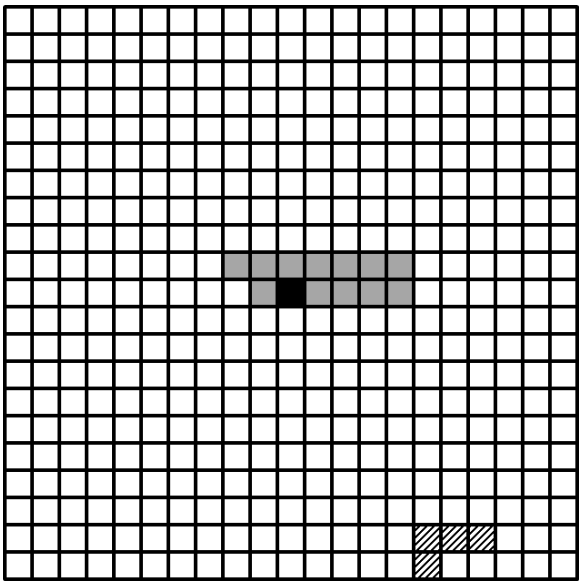


Figure A27: Union fire
2003; **incorrect acreage provided

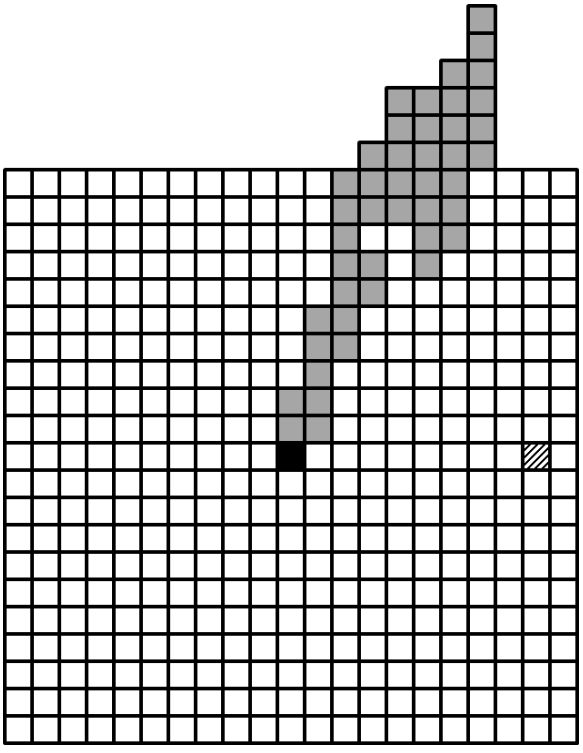


Figure A28: Rathbone fire
2003; 2751 acres (1113 ha)

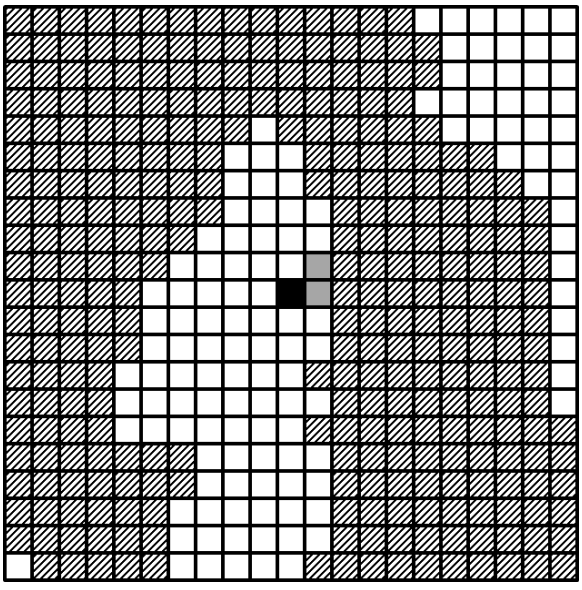


Figure A29: Promontory fire
2004; 160 acres (65 ha)

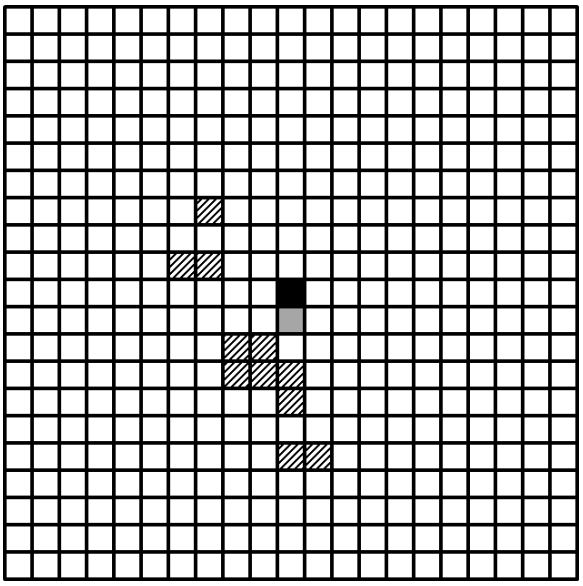


Figure A30: Broken Ankle fire
2004; 29 acres (12 ha)

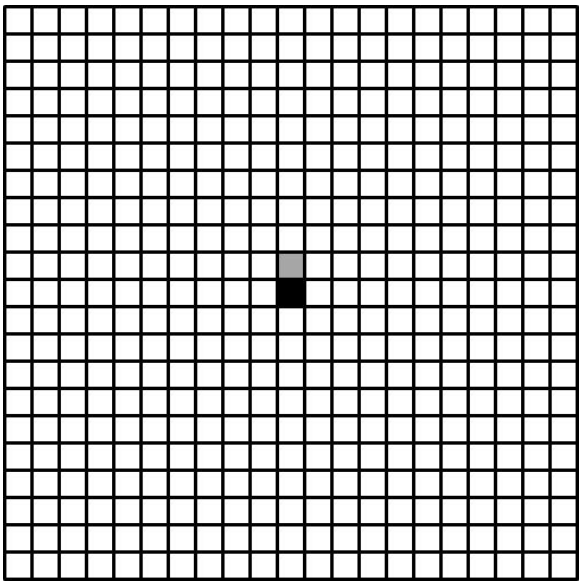


Figure A31: Chickadee fire
2005; 80 acres (32 ha)

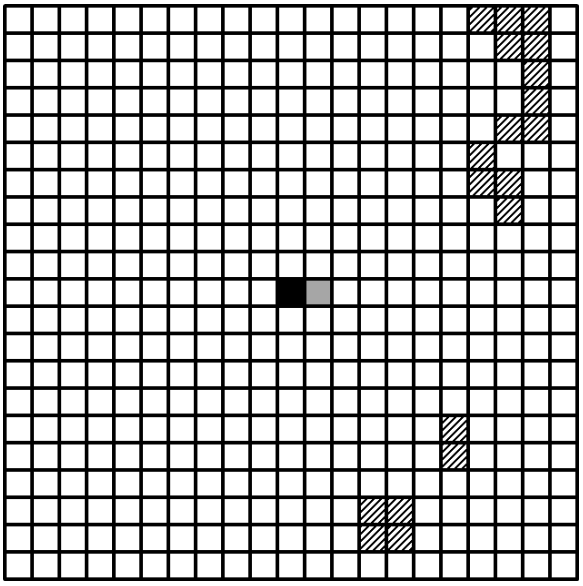


Figure A32: Elk Tongue fire
2005; 11 acres (4 ha)

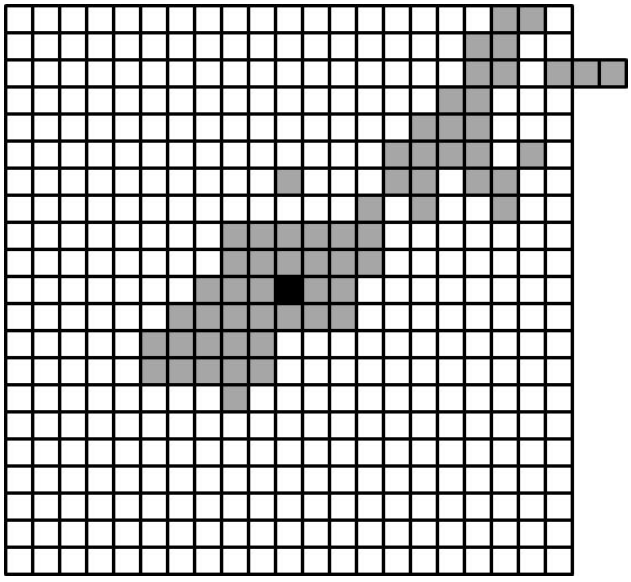


Figure A33: Magpie fire
2006; 3235 acres (1309 ha)

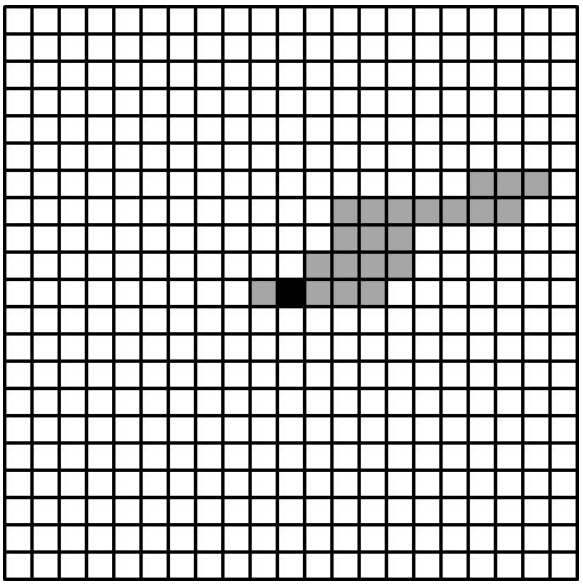


Figure A34: Stinky fire
2006; 1005 acres (407 ha)

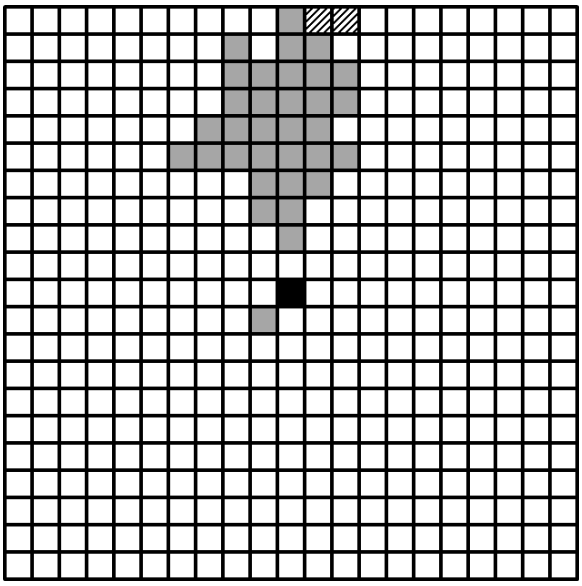


Figure A35: Owl fire
2007; 2066 acres (836 ha)

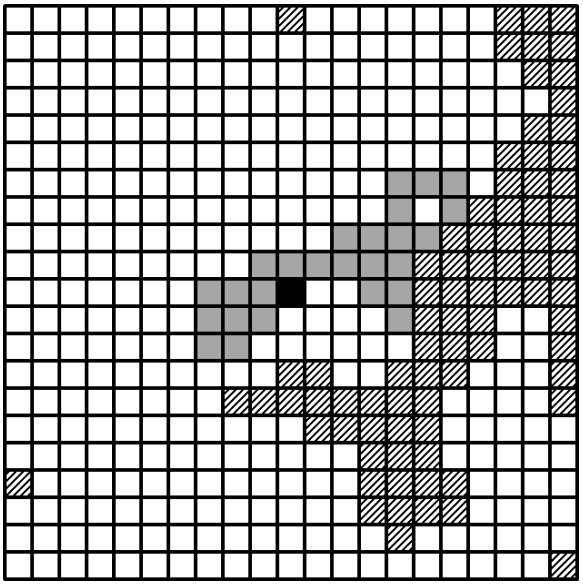


Figure A36: Beaverdam fire
2007; 1342 acres (543 ha)

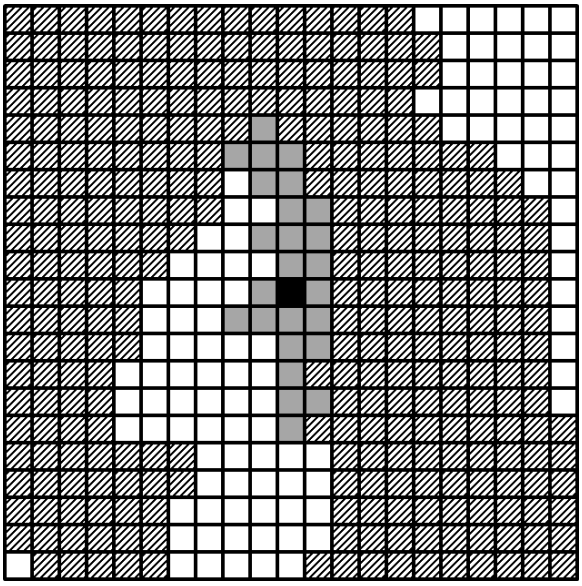


Figure A37: Promontory NW fire
2007; 1636 acres (662 ha)

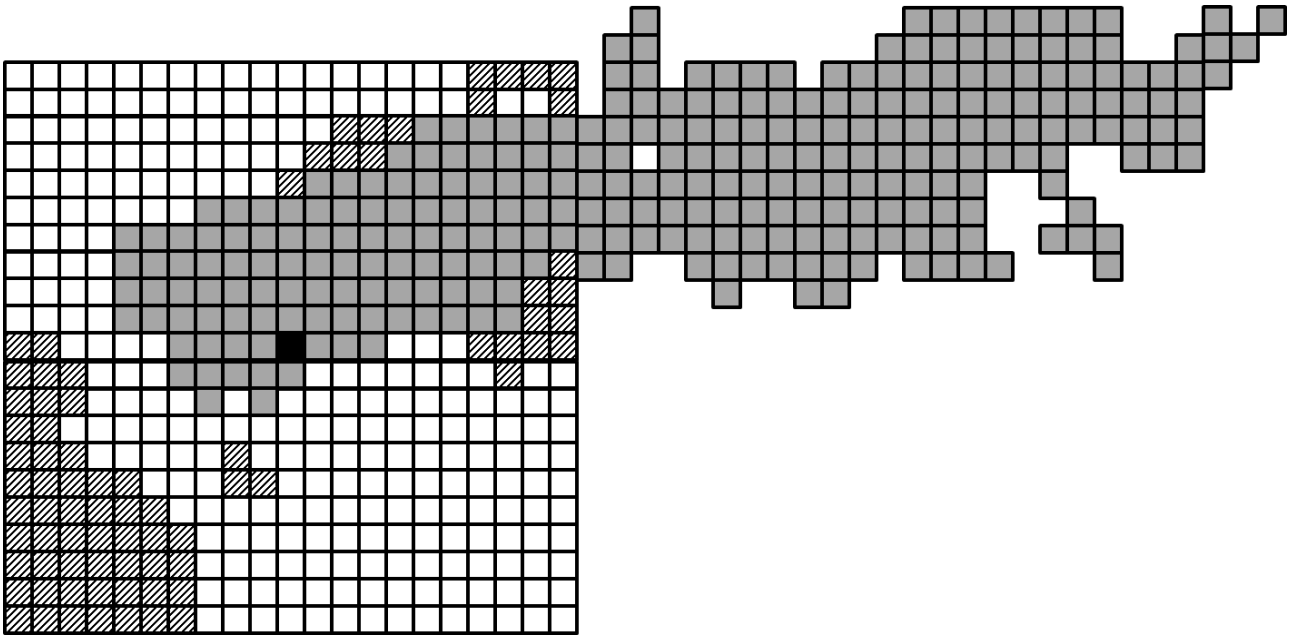


Figure A38: Columbine 1 fire
2007; 18337 acres (7421 ha)

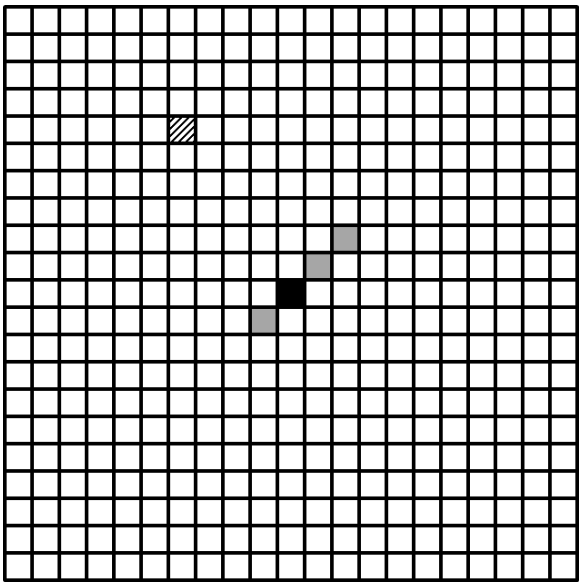


Figure A39: Cave fire
2007; 219 acres (89 ha)

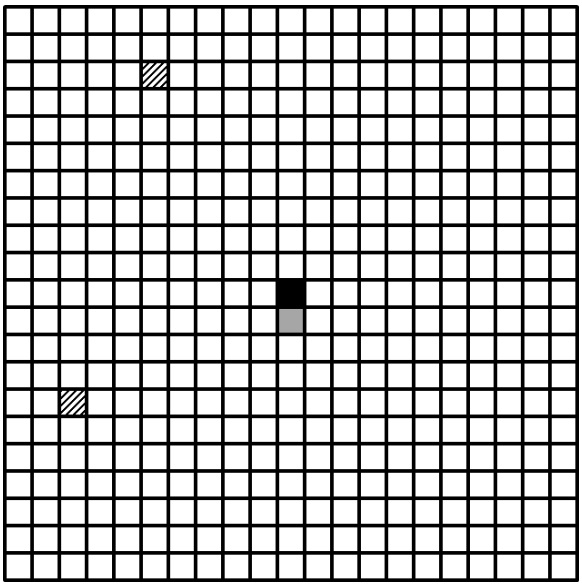


Figure A40: Norris RX fire
2007; 62 acres (25 ha)

APPENDIX C

VARIABLE LIST AND PSEUDOCODE

This appendix lists the variables and pseudocode used to build the data set. The pseudocode was used to extract variables from GIS coverages or to calculate new variables from the extracted data.

B.1: Variable List

Variable	Description
Burn	Not burned (0), burned (1)
CellID	Cell number from grid of park
WindowCellID	Cell number within 21 x 21 fire grid
Row	Cell row from grid of park
Column	Cell column from grid of park
WinCol	Column number within 21 x 21 fire grid
WinRow	Row number within 21 x 21 fire grid
FireName	Unique fire name for each fire; First four digits are the year; last two digits are the fire number
Dir	Categorical direction from ignition (N, NE, E, SE, S, SW, W, NW)
Distance	Distance from ignition (meters)
D1000	Distance from ignition (meters) / 1000
Dsqroot	Square root of D1000
Elev	Elevation of cell (meters)
DeltaElev1	Change in elevation between one neighbor and target cell (see Figure 4)
DeltaElev2	Average change in elevation between all eight neighbors and the target cell (see Figure 5)
Slope	Slope of cell (degrees)
DeltaSlope1	Change in slope between one neighbor and target cell (see Figure 4)
DeltaSlope2	Average change in slope between all eight neighbors and the target cell (see Figure 5)
Aspect	Categorical aspect of cell (AspN, AspNE, AspE, AspSE, AspS, AspSW, AspW, AspNW, AspFLAT)
FuelModel	Categorical – Scott and Burgan fuel model (GR1, GR2, GS1, GS2, SH2, TL, TU4, TU5)

Variable	Description
ViewAngle1	View angle of barrier between target and ignition cell
PalmerZ	Aggregated monthly drought variable Palmer Z (Drought, No Drought)
Sector	Categorical variable indicating membership in a ring of cells around the ignition 1=nearest ring , 2=middle ring, 3=outer ring
ENSO	Categorical variable indicating ENSO phase (SE, EN, N, LN, SL)
Zone	Categorical variable indicating management zone in park (NWZ, NCZ, MPZ, CPZ, SEZ, SWZ, SUZ, OUZ)
NeighborState	Neighbor cell burned (1), not burned (0), not burnable (-1)

B.2: Building the Yellowstone Data Set

The first step was to build a file to identify barriers to spread on the landscape. This was used to identify view angles of barriers between a target cell and the ignition cell.

Building the Barrier File

Scan park grid from upper left to lower right and identify separate barriers on the landscape. The cells of each contiguous barrier were collected together and identified as barrier"a" (a = 1 to number of separate barriers). The individual cells for each barrier were labeled from m = 1 to number of cells in the barrier. The final file was a list identifying a unique id (cellnum) for every barrier cell and an alphanumeric cellid that identifies the barrier number and the cell number within the barrier.

For example, a landscape with 2 barriers, 3 cells in the first barrier and 2 cells in the second barrier will have 5 entries:

Cellnum	BarrierCell
1	barrier1of1 (cell 1 in barrier 1)
2	barrier 2of1 (cell 2 in barrier 1)
3	barrier 3of1 (cell 3 in barrier1)
4	barrier 1of2 (cell 1 in barrier 2)
5	barrier 2of2 (cell 2 in barrier 2)

End of file

a = 0

cellnum = 0

For h = 1 to maxrow

For k = 1 to maxcol

Get fuel model for cell (h,k)

If fuelmodel(h, k) = “NB” (for non-burnable cell) then

check to see if this cell is already part of another barrier

If a = 0 then endif

else For z = 1 to cellnum

If ”c”of”b”(h,k) at cellnum = barriercell(h,k) then next

k

Next z

Endif

* Build barrier number a *

a=a+1

ring = 0

t = 1

cellnum = cellnum+1

”t”of”a” = cellid(h,k)

*record in barrier file

- cellnum under cellnum
- cellID from park grid under cellid
- “a” under barriernum

- “t” under barriercell
- “h” under row
- “k” under column*

For ring = 1 to maxring

u=0

top row of ring

i = -1 * ring

For j = (-1 * ring) to ring

If fuelmodel(h+i, j+k) = “NB” then

t = t+1

cellnum = cellnum+1

”t”of”a” = cellid(h+i, j+k)

*record in barrier file

- cellnum under heading cellnum
- cellid under heading cellid
- “a” under heading barriernum
- “t” under heading barriercell
- “h” under heading row
- “k” under heading column*

u = u+1

Endif

Next j

right side of ring

j = ring

For i = (-1 * ring)+1 to ring-1

If fuelmodel(h+i, j+k) = “NB” then

t = t+1

cellnum = cellnum+1

”t”of”a” = cellid(h+i, j+k)

*record in barrier file

- cellnum under heading cellnum
- cellid under heading cellid
- “a” under heading barriernum
- “t” under heading barriercell
- “h” under heading row
- “k” under heading column*

Endif

Next i

bottom row of ring

i = ring

For j = ring to (-1 * ring), step -1

If fuelmodel(h+i, j+k) = “NB” then

t = t+1

cellnum = cellnum+1

”t”of”a” = cellid(h+i, j+k)

*record in barrier file

- cellnum under heading cellnum
- cellid under heading cellid
- “a” under heading barriernum
- “t” under heading barriercell
- “h” under heading row
- “k” under heading column*

Endif

Next j

right side of ring

j = -1 * ring

For i = ring-1 to (-1 * ring) +1, step -1

If fuelmodel(h+i, j+k) = “NB” then

t = t+1

cellnum = cellnum+1

”t”of”a” = cellid(h+i, j+k)

*record in barrier file

- cellnum under heading cellnum
- cellid under heading cellid
- “a” under heading barriernum

- "t" under heading barriercell
- "h" under heading row
- "k" under heading column*

Endif

Next i

test if there was at least one NB in current ring

If u > 0 then Next ring

Else ENDFORLOOP (for rings)

Endif

end build of barrier"a" and move on

Endif

Next k, Next h

Building the Data File

- There is a series of data points for each fire
- All series will be collected into a single data set (append fires together into a single set)
- Each burnable cell from each fire will have a set of attributes
  - Some will come directly from the existing data set
  - Some will be new attributes calculated from the existing data set

For each FireID:

Identify the center of the fire window (ignition cell) with coordinates:

(R=Row from park grid for ignition cell,

C=Column from park grid for ignition cell)

Let NumCell = how many cells to move away from ignition

(i.e., for the 21x21 window NumCell=10)

For i = -NumCell to NumCell

For j = -NumCell to NumCell

Skip ignition cell

If i = 0 and j =0, next j

Else

** Identify the target cell CellID from park grid **

CellID = CellID(R+i, C+j) from park grid

** Identify the target cell Row from park grid **

Row = Row(R+i, C+j) from park grid

** Identify the target cell Column from park grid **

Col = Col(R+i, C+j) from park grid

** Identify the fire **

FireID = FireID from fire record

** Year of fire **

Year = Year from fire record

** Did the target cell burn in this fire? **

If the cell burned, then Burn = 1

Else Burn = 0

** Identify direction from ignition **

Let x = C – Col

Let y = R – Row

If $x \geq 0$ and $y \geq 0$ then ** NW quadrant – W, NW, N **

Angle= inverse sin(ABS(y)/sqrt($x^2 + y^2$))*(180/π)

If Angle$\geq$ 0 and $\leq$ 22.5 then Dir= “W”

Elseif Angle> 22.5 and < 67.5 then Dir= “NW”

Else Dir= “N”

Elseif $x \leq 0$ and $y > 0$ then ** NE quadrant – N, NE, E **

Angle= inverse sin((ABS(y)/sqrt($x^2 + y^2$)) *(180/π)

If Angle≥ 0 and ≤ 22.5 then Dir= “E”

Elseif Angle> 22.5 and < 67.5 then Dir= “NE”

Else Dir= “N”

Elseif x < 0 and y < 0 then ** SE quadrant – E, SE, S **

Angle= inverse sin((ABS(y)/sqrt(x^2 + y^2)) *(180/π)

If Angle≥ 0 and ≤ 22.5 then Dir= “E”

Elseif Angle> 22.5 and < 67.5 then Dir= “SE”

Else Dir= “S”

Else ** SW quadrant – S, SW, W where x is (+) and y is (-)**

Angle= inverse sin((ABS(y)/sqrt(x^2 + y^2)) *(180/π)

If Angle≥ 0 and ≤ 22.5 then Dir= “W”

Elseif Angle> 22.5 and < 67.5 then Dir = “SW”

Else Dir = “S”

** Fuel Model **

FuelModel = Scott and Burgan Fuel Model of target cell

** Linear Distance between Ignition Cell and Target Cell **

Distance = sqrt((480*i)^2 + (480*j)^2)

** Elevation of Target Cell**

Elev(R+i, C+j) = elevation of target cell

** Change in Elevation Method 1**

If Dir(R+i, C+j) = “N” then

DeltaElev1(R+i, C+j) = Elev(R+i, C+j) – Elev(R+i+1, C+j)

Elseif Dir(R+i, C+j) = “NE” then

DeltaElev1(R+i, C+j) = Elev(R+i, C+j) – Elev(R+i+1, C+j-1)

Elseif Dir(R+i, C+j) = “E” then

DeltaElev1(R+i, C+j) = Elev(R+i, C+j) – Elev(R+i, C+j-1)

Elseif Dir(R+i, C+j) = “SE” then

DeltaElev1(R+i, C+j) = Elev(R+i, C+j) – Elev(R+i-1, C+j-1)

Elseif Dir(R+i, C+j) = “S” then

DeltaElev1(R+i, C+j) = Elev(R+i, C+j) – Elev(R+i-1, C+j)

Elseif Dir(R+i, C+j) = “SW” then

DeltaElev1(R+i, C+j) = Elev(R+i, C+j) – Elev(R+i-1, C+j+1)

Elseif Dir(R+i, C+j) = “W” then

DeltaElev1(R+i, C+j) = Elev(R+i, C+j) – Elev(R+i, C+j+1)

Else

DeltaElev1(R+i, C+j) = Elev(R+i, C+j) – Elev(R+i+1, C+j1+1)

** Change in Elevation Method 2**

f = 0

sumelev = 0

if fuelmodel(indexrow-1, indexcol-1) = NB, then endif

else sumelev = sumelev + Elev(indexrow-1, indexcol-1)

f = f + 1

endif

if fuelmodel(indexrow-1, indexcol) = NB, then endif

else sumelev = sumelev + Elev(indexrow-1, indexcol)

f = f + 1

endif

if fuelmodel(indexrow-1, indexcol+1) = NB, then endif

else sumelev = sumelev + Elev(indexrow-1, indexcol+1)

f = f + 1

endif

```
if fuelmodel( indexrow, indexcol+1) = NB, then endif

        else    sumelev = sumelev + Elev( indexrow, indexcol+1)

                f = f + 1

endif

if fuelmodel( indexrow+1, indexcol+1) = NB, then endif

        else    sumelev = sumelev + Elev( indexrow+1, indexcol+1)

                f = f + 1

endif

if fuelmodel( indexrow+1, indexcol) = NB, then endif

        else    sumelev = sumelev + Elev( indexrow+1, indexcol)

                f = f + 1

endif

if fuelmodel( indexrow+1, indexcol-1) = NB, then endif

        else    sumelev = sumelev + Elev( indexrow+1, indexcol-1)

                f = f + 1

endif
```

if fuelmodel(indexrow, indexcol-1) = NB, then endif

else sumelev = sumelev + Elev(indexrow, indexcol-1)

f = f + 1

endif

If f = 0, then DeltaElev2(indexcell) = -9999

else DeltaElev2(indexcell) = Elev(indexcell) – (1/f)* sumelev

endif

** Slope of Target Cell **

Slope(R+i, C+j) = Slope of target cell

** Change in Slope Method 1**

If Dir(R+i, C+j) = “N” then

DeltaSlope1(R+i, C+j) = Slope (R+i, C+j) – Slope (R+i+1, C+j)

Elseif Dir(R+i, C+j) = “NE” then

DeltaSlope1(R+i, C+j) = Slope (R+i, C+j) – Slope (R+i+1, C+j-1)

Elseif Dir(R+i, C+j) = “E” then

DeltaSlope1(R+i, C+j) = Slope (R+i, C+j) – Slope (R+i, C+j-1)

Elseif Dir(R+i, C+j) = “SE” then

DeltaSlope1(R+i, C+j) = Slope (R+i, C+j) – Slope (R+i-1, C+j-1)

Elseif Dir(R+i, C+j) = “S” then

DeltaSlope1(R+i, C+j) = Slope (R+i, C+j) – Slope (R+i-1, C+j)

Elseif Dir(R+i, C+j) = “SW” then

DeltaSlope1(R+i, C+j) = Slope (R+i, C+j) – Slope (R+i-1, C+j+1)

Elseif Dir(R+i, C+j) = “W” then

DeltaSlope1(R+i, C+j) = Slope (R+i, C+j) – Slope (R+i, C+j+1)

Else

DeltaSlope1(R+i, C+j) = Slope (R+i, C+j) – Slope (R+i+1, C+j1+1)

** Change in Slope Method 2 **

f = 0

sumslope = 0

if fuelmodel(indexrow-1, indexcol-1) = NB, then endif

else sumslope = sumslope + Slope(indexrow-1, indexcol-1)

f = f + 1

endif

if fuelmodel(indexrow-1, indexcol) = NB, then endif

```
        else    sumslope = sumslope + Slope( indexrow-1, indexcol)
                f = f + 1
endif
if fuelmodel( indexrow-1, indexcol+1) = NB, then endif
        else    sumslope = sumslope + Slope( indexrow-1, indexcol+1)
                f = f + 1
endif
if fuelmodel( indexrow, indexcol+1) = NB, then endif
        else    sumslope = sumslope + Slope( indexrow, indexcol+1)
                f = f + 1
endif
if fuelmodel( indexrow+1, indexcol+1) = NB, then endif
        else    sumslope = sumslope + Slope( indexrow+1, indexcol+1)
                f = f + 1
endif
if fuelmodel( indexrow+1, indexcol) = NB, then endif
        else    sumslope = sumslope + Slope( indexrow+1, indexcol)
```

```
                f = f + 1

        endif

        if fuelmodel( indexrow+1, indexcol-1) = NB, then endif

            else    sumslope = sumslope + Slope( indexrow+1, indexcol-1)

                f = f + 1

        endif

        if fuelmodel( indexrow, indexcol-1) = NB, then endif

            else    sumslope = sumslope + Slope( indexrow, indexcol-1)

                f = f + 1

        endif

        If f = 0, then DeltaSlope2(indexcell) = -9999

            else DeltaSlope2(indexcell) = Slope(indexcell) – (1/f)* sumslope

        endif

    ** Aspect of Target Cell **

            Aspect = aspect of target cell

** Build a ray from the target cell to the ignition cell **

        *Row is target cell row; Col is target cell column*

        *R is ignition cell row; C is ignition cell column*
```

```
rayx = Col – C

rayy = Row – R


If rayx = 0, then endif

        else  slope = rayy / rayx


If rayx = 0 and rayy < 0, then                                  *N*

        For m = 1 to (ABS(rayy) – 1)

                cellray"m"  = cellid(R – m, C)

        Next m


Elseif rayx > 0 and rayy < 0, then                              *NE quad*

        If ABS(slope) >1, then                                  *NNE*

                xstep = ABS(1 / slope)

                ystep = -1

                p = 0.5

                q = 0

                Bigm = max (ABS(rayx) – 1, ABS(rayy) – 1)

                For m = 1 to Bigm

                        p = p + xstep

                        q = q + ystep

                        r = integer(p)

                        cellray"m"  = cellid(R + q, C + r)
```

```
        Next m
Elseif ABS(slope) = 1, then                                    *NE*
        xstep = 1
        ystep = -1
        p = 0
        q = 0
        Bigm = rayx – 1
        For m = 1 to Bigm
                p = p + xstep
                q = q + ystep
                cellray”m”  = cellid(R + q, C + p)
        Next m
Else                                                           *ENE*
        xstep = 1
        ystep = slope
        p = 0
        q = - 0.5
        Bigm = max (ABS(rayx) – 1, ABS(rayy) – 1)
        For m = 1 to Bigm
                p = p + xstep
                q = q + ystep
                r = integer(q)
                cellray”m”  = cellid(R + r, C + p)
```

```
            Next m
        Endif

Elseif rayx >0 and rayy = 0, then                                   *E*
        For m = 1 to (rayx) – 1)
            cellray"m"  = cellid(R, C + m)
        Next m

Elseif rayx > 0 and rayy > 0, then                                  *SE quad*
        If ABS(slope) < 1, then                                     *ESE*
            xstep = 1
            ystep = slope
            p = 0
            q = 0.5
            Bigm = max (ABS(rayx) – 1, ABS(rayy) – 1)
            For m = 1 to Bigm
                p = p + xstep
                q = q + ystep
                r = integer(q)
                cellray"m"  = cellid(R + r, C + p)
            Next m
        Elseif ABS(slope) = 1, then                                 *SE*
            xstep = 1
```

```
        ystep = 1
        p = 0
        q = 0
        Bigm = rayx – 1
        For m = 1 to Bigm
            p = p + xstep
            q = q + ystep
            cellray”m”  = cellid(R + q, C + p)
        Next m
Else                                                    *SSE*
        xstep = 1 / slope
        ystep = 1
        p = 0.5
        q = 0
        Bigm = max (ABS(rayx) – 1, ABS(rayy) – 1)
        For m = 1 to Bigm
            p = p + xstep
            q = q + ystep
            r = integer(p)
            cellray”m”  = cellid(R + q, C + r)
        Next m
Endif
```

```
Elseif rayx = 0 and rayy > 0, then                    *S*
    For m = 1 to (rayy) – 1)
        cellray"m"  = cellid(R + m, C)
    Next m


Elseif rayx < 0 and rayy > 0, then                    *SW quad*
    If ABS(slope) >1, then                            *SSW*
        xstep = 1 / slope
        ystep = 1
        p = - 0.5
        q = 0
        Bigm = max (ABS(rayx) – 1, ABS(rayy) – 1)
        For m = 1 to Bigm
            p = p + xstep
            q = q + ystep
            r = integer(p)
            cellray"m"  = cellid(R + q, C + r)
        Next m
    Elseif ABS(slope) = 1, then                       *SW*
        xstep = -1
        ystep = 1
        p = 0
        q = 0
```

```
            Bigm = rayy – 1
            For m = 1 to Bigm
                  p = p + xstep
                  q = q + ystep
                  cellray"m"  = cellid(R + q,C + p)
            Next m
      Else                                                    *WSW*
            xstep = -1
            ystep = slope * -1
            p = 0
            q = 0.5
            Bigm = max (ABS(rayx) – 1, ABS(rayy) – 1)
            For m = 1 to Bigm
                  p = p + xstep
                  q = q + ystep
                  r = integer(q)
                  cellray"m"  = cellid(R + r, C + p)
            Next m
      Endif

Elseif rayx < 0 and rayy = 0, then                            *W*
      For m = 1 to (ABS(rayx) – 1)
            cellray"m" = cellid(R, C – m)
```

```
        Next m

Else                                                    *NW quad*
        If ABS(slope) < 1, then                         *WNW*
                xstep = -1
                ystep = -slope
                p = 0
                q = - 0.5
                Bigm = max (ABS(rayx) – 1, ABS(rayy) – 1)
                For m = 1 to Bigm
                        p = p + xstep
                        q = q + ystep
                        r = integer(q)
                        cellray”m” = cellid(R + r, C + p)
                Next m
        Elseif ABS(slope) = 1, then                     *NW*
                xstep = -1
                ystep = -1
                p = 0
                q = 0
                Bigm = ABS(rayx – 1)
                For m = 1 to Bigm
                        p = p + xstep
```

q = q + ystep

cellray”m” = cellid(R + q, C + p)

Next m

Else *NNW*

xstep = 1 / slope * -1

ystep = -1

p = - 0.5

q = 0

Bigm = max (ABS(rayx) – 1, ABS(rayy) – 1)

For m = 1 to Bigm

p = p + xstep

q = q + ystep

r = integer(p)

cellray”m” = cellid(R + q, C + r)

Next m

Endif

Endif

- Check elements of cellray”m” to see if there are any matches with cell IDs in the barrier file

  If no matches

  ViewAngle = 0

Next j

Else

ViewAngle = 0

Identify the barrier and its cells:

*Sweep through all cells of the intercepted barrier to find the widest view angle of a barrier between the target cell and the ignition cell *

If there is only one cell in the barrier, skip to $$

b is cell number “b” of the barrier; c is cell number “c” of the barrier

Row is target cell row; Col is target cell column

For b = 1 to (number of cells in barrier – 1)

For c = b+1 to (number of cells in barrier)

xab = Col – bcol

yab = Row – brow

ab = sqrt($xab^2 + yab^2$)

xac = Col – ccol

yac = Row – crow

ac = sqrt($xac^2 + yac^2$)

xbc = bcol – ccol

ybc = brow – crow

bc = sqrt($xbc^2 + ybc^2$)

If ab = 0 then NEXT c

endif

If ac = 0 then NEXT c

endif

Angle = arccos(($ac^2 + ab^2 – bc^2$) / (2 * ac * ab)) * (180/π)

If Angle > ViewAngle then

ViewAngle = Angle

Endif

Next c

Next b

$$: If there is only one cell in the barrier...

xab = Col – bcol

yab = Row – brow

ab = sqrt($xab^2 + yab^2$)

ViewAngle = arcsin(0.5 / SQRT($ab^2 + 0.5^2$)

Next j

```
Next i
```

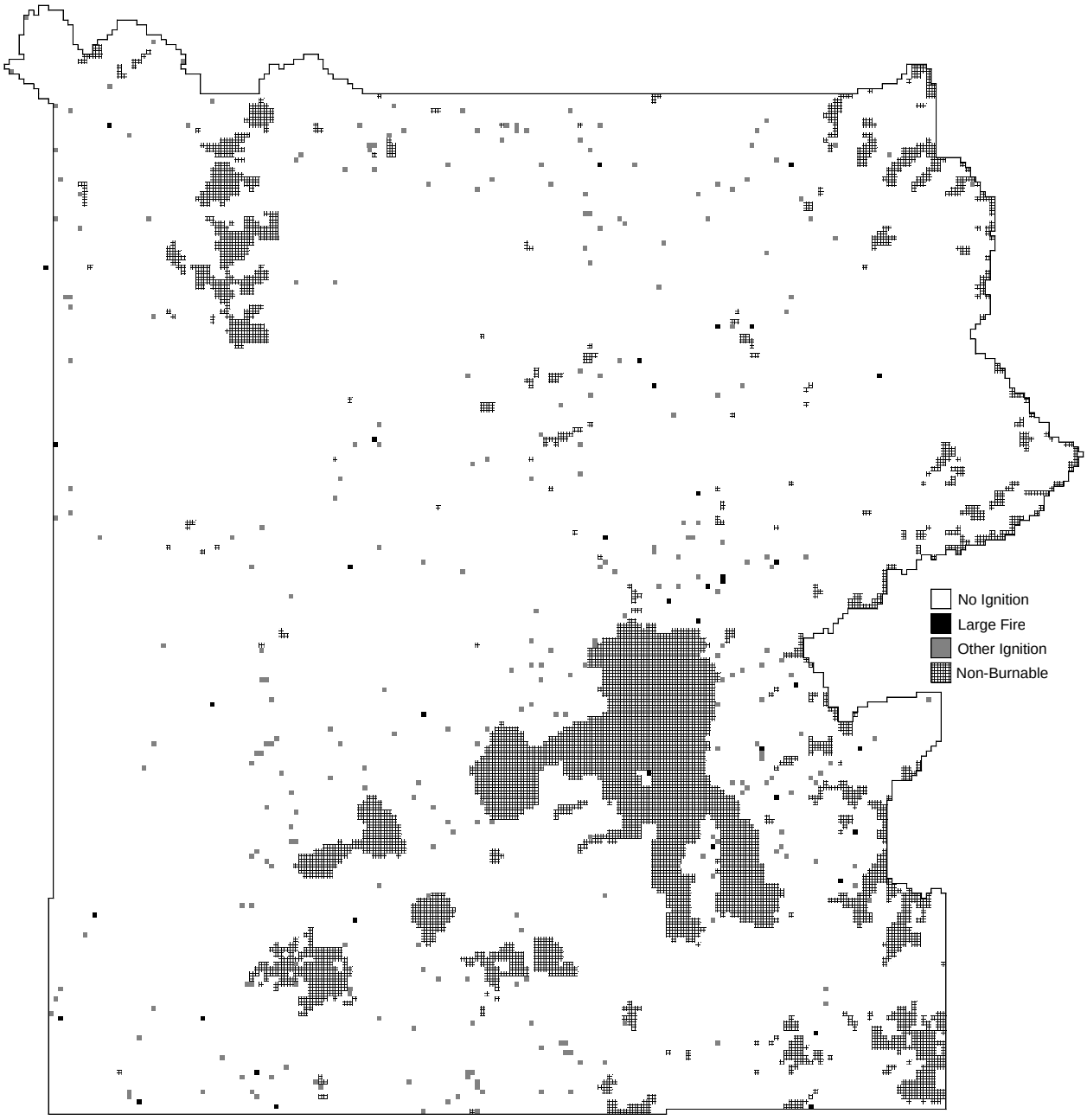
** Neighbor State, Sector, Palmer Z, ENSO and Zone variables were not calculated from the GIS coverages and were entered by hand**

APPENDIX D

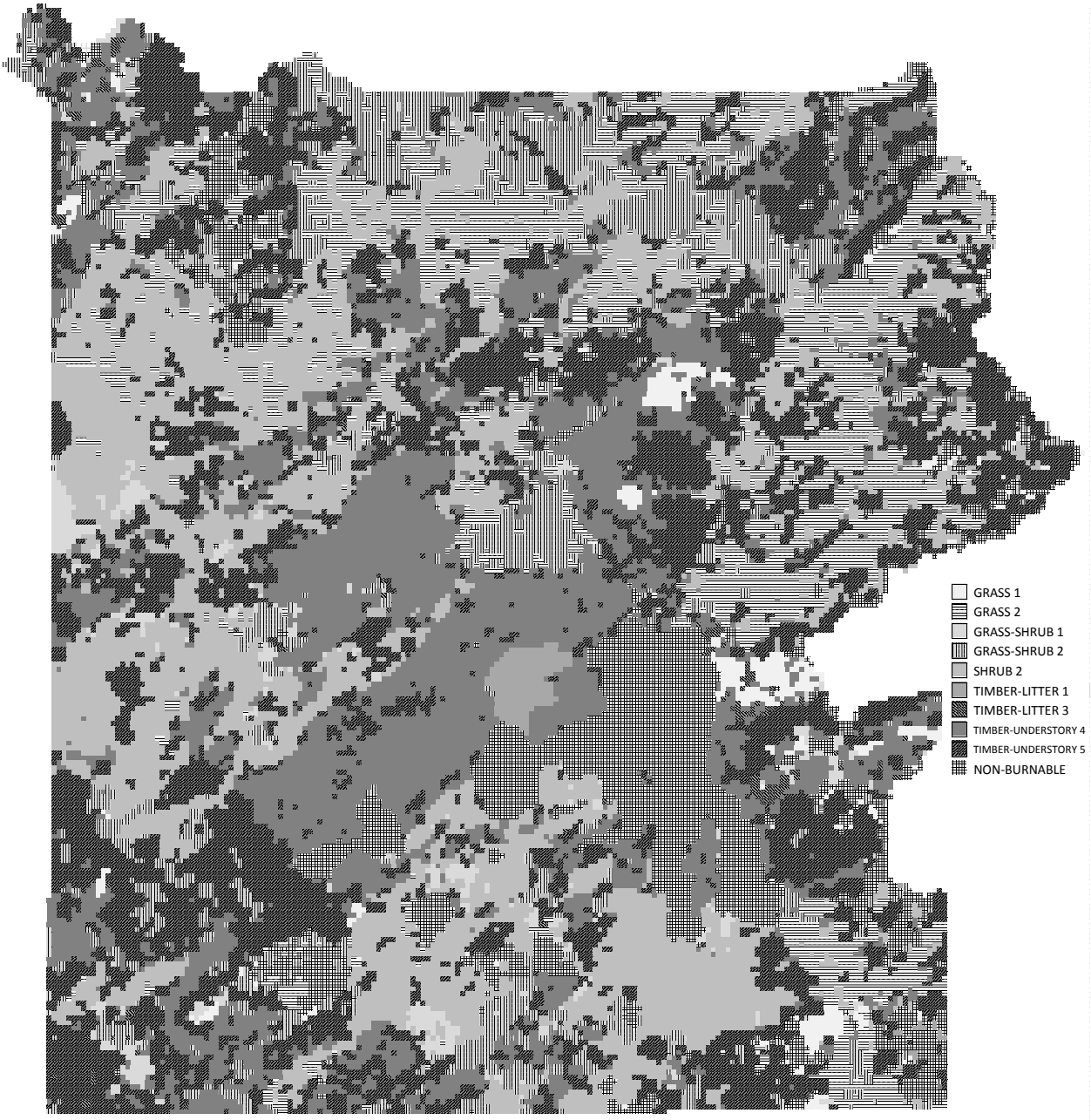
RASTER GRAPHICS OF PARK VARIABLES

This appendix contains graphics of the spatial variable set in raster format including ignition locations, fuel model, zone, slope, elevation and aspect.

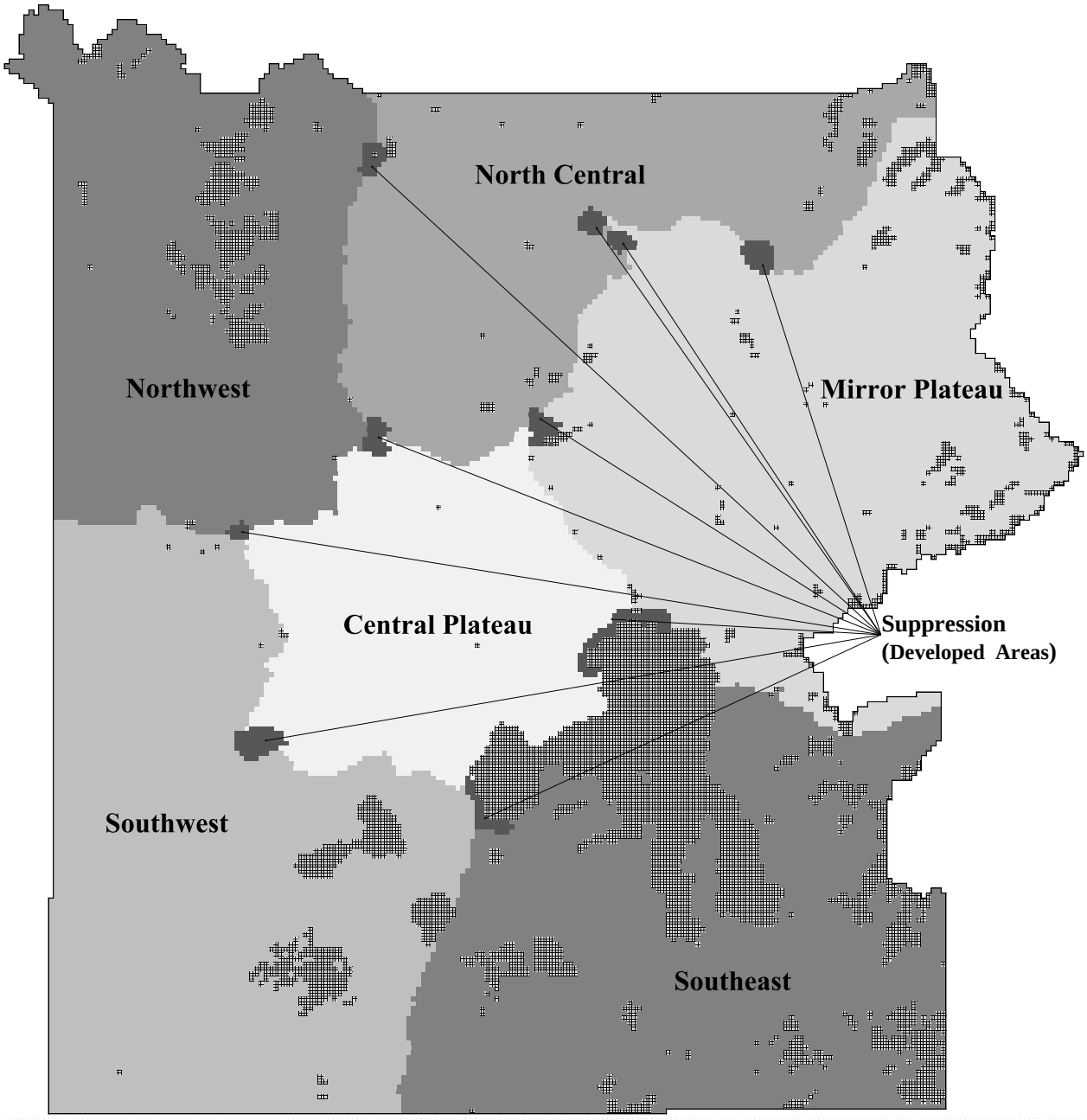
D.1: Ignition Locations



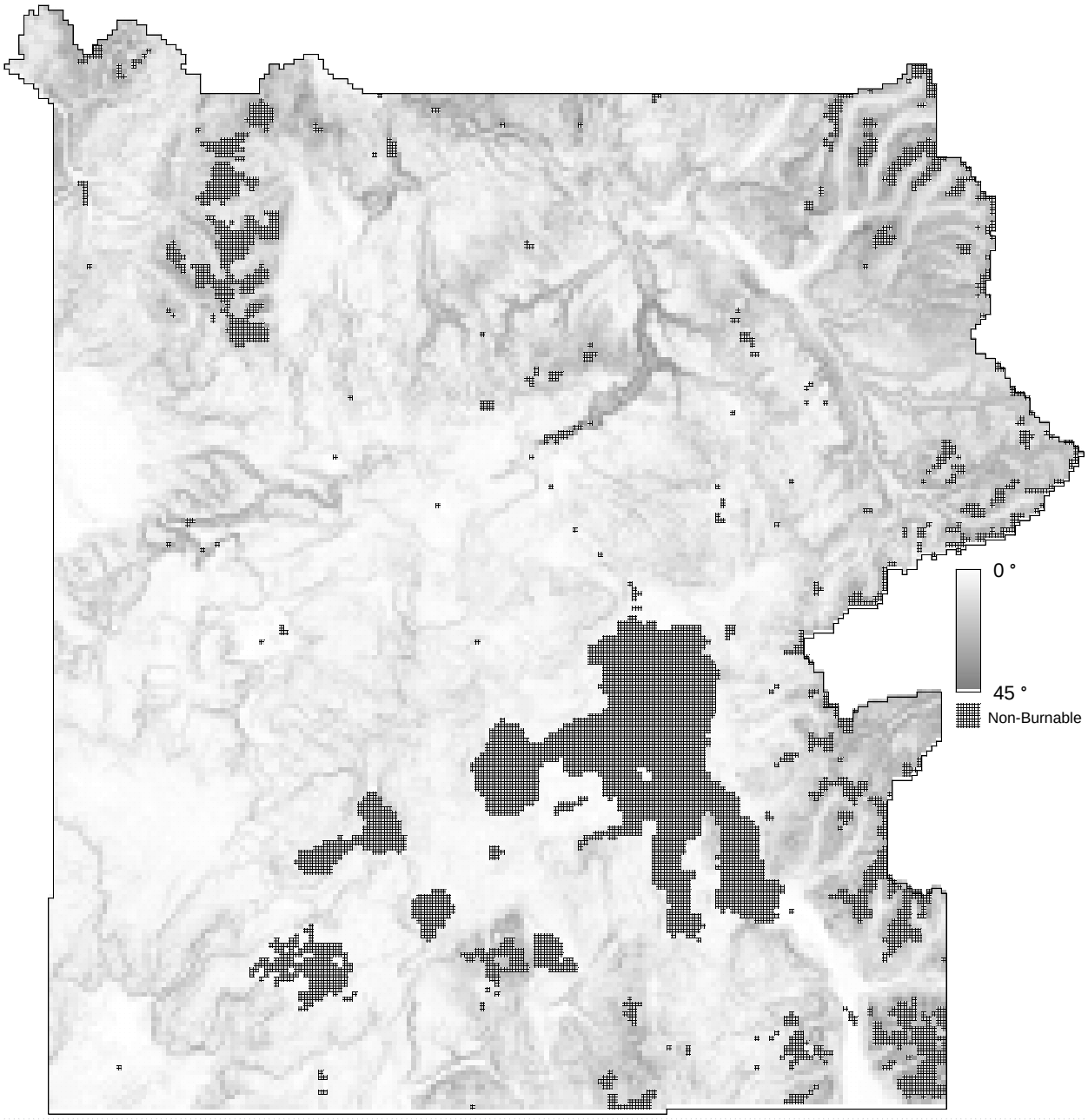
D.2: Fuel Models; from Scott and Burgan (2005)



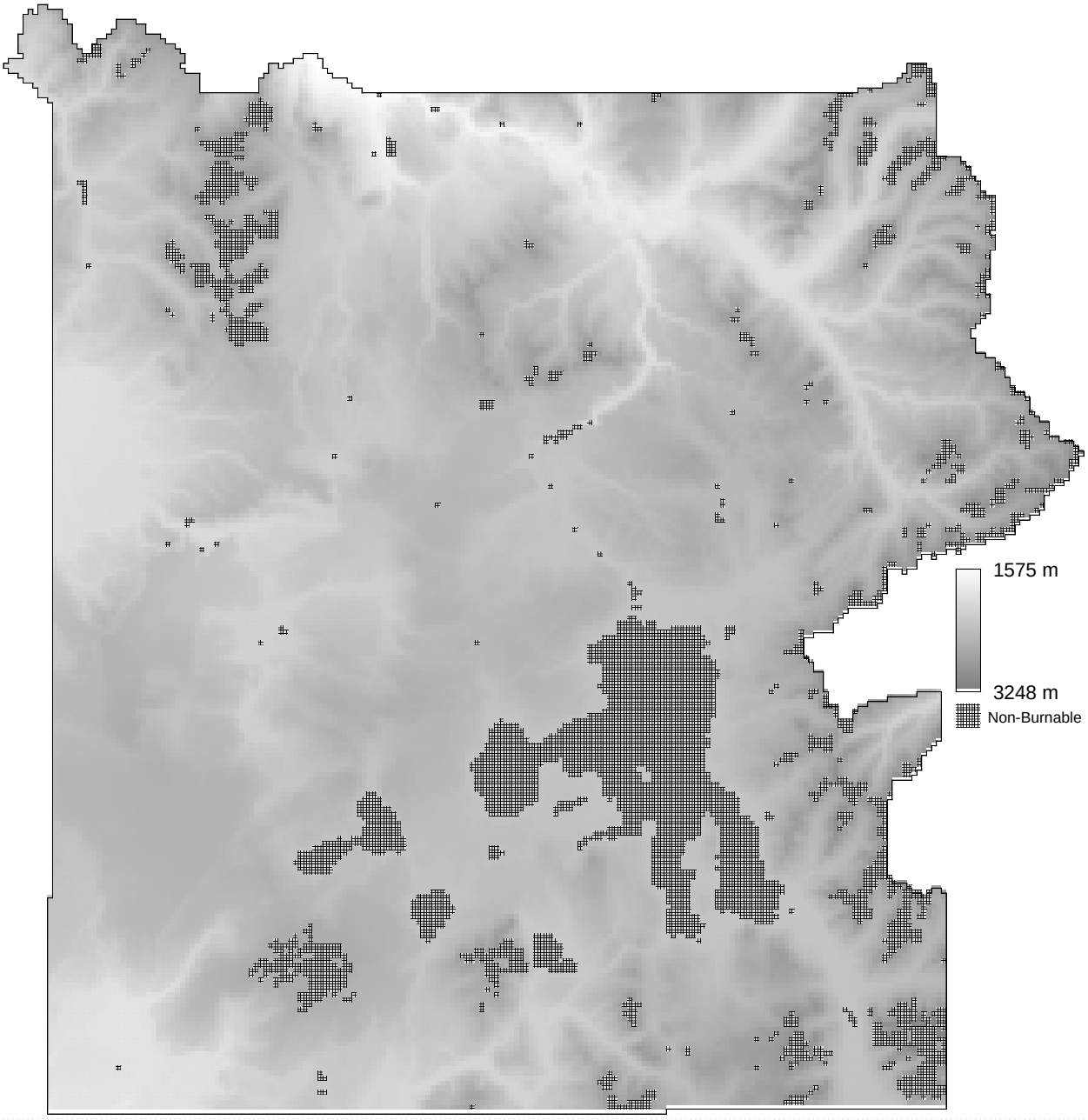
D.3: Zones; derived from old (1992) and new (2004) Yellowstone National Park Fire Management Plans



D.4: Slope; degrees from horizontal



D.5: Elevation; meters above sea level



D.6: Aspect; cardinal (N, E, S, W) and intercardinal (NE, SE, SW, NW) directions