Statistical Applications in Natural Resources

A Thesis

Presented in Partial Fulfillment of the Requirements for the

Degree of Master of Science

with a

Major in Natural Resources

in the

College of Graduate Studies

University of Idaho

by

Lindsay M. Grayson

Major Professor: Alistair M.S. Smith, Ph.D.

Committee Members: Jan U.H. Eitel, Ph.D. and Michael J. Anderson, Ph.D.

Department Administrator: Anthony S. Davis, Ph.D.

August 2015

AUTHORIZATION TO SUBMIT THESIS

This thesis of Lindsay M. Grayson, submitted for the degree of Master of Science with a Major in Natural Resources and titled "Statistical Applications in Natural Resources," has been reviewed in final form. Permission, as indicated by the signatures and dates below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor: ________________________________ Date:_____________
Dr. Alistair M.S. Smith

Committee
Members: ________________________________ Date:_______________
Dr. Jan U.H. Eitel

________________________________ Date:_______________
Dr. Michael J. Anderson

Department
Administrator ________________________________ Date:_______________
Dr. Anthony S. Davis

ABSTRACT

The improper use of statistics in research has become a plague to science; several publications show that more than 50% of papers have at least one statistical error. These may be through a lack of knowledge, misapplication of statistics, or misconduct. The epidemic has caused many journals to implement sections devoted to teaching basic statistical principles to educate both readers and authors. This thesis describes proper statistical techniques through case studies. It also seeks to elucidate the effects of performance assumptions of equipment through statistical analysis.

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my committee chair and advisor, Dr. Alistair M.S. Smith, for his support of my thesis research, guidance, and immense knowledge. It has been my privilege to have had this opportunity to be associated with him. I would also like to thank my committee members Dr. Jan Eitel and Dr. Michael Anderson for their effort, expertise, and suggestions. I am grateful for the funding and interest that the National Aeronautics and Space Administration award NNX11AO24G provided.

TABLE OF CONTENTS

Authorization to Submit Thesis ii

Abstract iii

Acknowledgements.. iv

Table of Contents v

List of Tables vi

List of Figures vii

Chapter 1:

Introduction... 1

Background... 3

Methods 8

Results/Discussion 11

Conclusions... 16

References... 18

Chapter 2:

Abstract 26

Introduction... 26

Background... 27

Methods 30

Results/Discussion 31

Conclusions... 33

References... 34

Chapter 3:

Abstract.. 39

Introduction.. 39

Methods 40

Results.. 42

Discussion.. 43

Conclusion 44

References.. 45

LIST OF TABLES

Chapter 1:

Table 1 – Common statistical tests for detecting outliers.................................... 21

Table 2 – Potential outliers indicating extreme fire years 22

Table 3 – PRISM descriptive statistics.. 23

Table 4 – PRISM data properties.. 24

Table 5 – Fire-climate regressions with PRISM data and area burned 25

Chapter 2:

Table 1 – Extreme fire classification.. 36

Chapter 3:

Table 1 – ANOVA with post hoc Tukey's HSD tests.. 47

LIST OF FIGURES

Chapter 2:

Chart 1 – Plot of the number of extreme wildfire events and total yearly burned area....................38

Chapter 3:

Figure 1 – Experimental set up....................48

Figure 2 – Box plot showing the time differences....................49

Chapter 1

The suitability of statistics to assess extreme fire years: observations via a case study evaluating fire-climate interactions in the northern Rocky Mountains

Introduction

Climate change is predicted in the northwestern United States to lead to a higher frequency of extended droughts and heat waves, increasing the likelihood that the region will experience disproportionate quantities of area burned in a given year (Pechony and Shindell 2010; IPCC 2013). These scenarios described are often as large (Fule *et al.* 2004), regional (Morgan *et al.* 2008), or extreme (Lannom *et al.* 2014) fire years. The prediction or immediate *a posteriori* assessment of such extreme fire years could aid in the rapid assignment of mitigation actions to maintain or restore critical ecosystem goods and services, especially if those ecosystems are vulnerable to regimes shifts or loss of ecosystem structure or function (Schefferm and Carpenter 2003; Smith *et al.* 2014). Many studies have sought to identify extreme fire years for a variety of reasons, including assessment of (i) atmospheric and climate anomalies are associated with those years (Johnson and Woschuk, 1992; Gedalof *et al.* 2005), (ii) fire-climate relationships occurring over across paleo- (Drobushev *et al.* 2014) and centennial-timescales (Fauria and Johnson, 2008; Morgan *et al.* 2008; Dillon *et al.* 2011; Lannom *et al.* 2014; Higuera *et al.* 2015), (iii) years that are more likely to lead to ecosystem regime shifts (Kasischke *et al.* 2010; Boiffin and Munson, 2013), and (iv) whether potential trends in the frequency and properties (e.g. areas burned with high severity) of such extreme fire years exist or have changing rates (Hanson and Odion, 2014). The majority of these studies use either geospatial datasets of area burned per year, whether fire atlas or satellite sensor derived, or paleoecology metrics derived from fire scars or cores (Morgan et al 2008; Dillon et al 2011; Drobushev *et al.* 2014).

Multiple definitions of extreme fire years or similar descriptors (e.g. big fire years, large fire years, regional fire years, and widespread fire years) already exist in the literature. Studies that have defined them include: (a) setting an upper percentile threshold on ranked lists of total area burned per year (Morgan *et al.* 2008), b) identifying outliers using a version of the Tukey's Range test (Dillon *et al.* 2011; Lannom *et al.* 2014), c) performing two sample t-tests to evaluate whether set years have significantly more burned area than other years (Miller *et al.* 2012), d) selecting the top five years of area burned within a temporal series (Gedalof *et al.* 2005), e) defining set thresholds (e.g., >50%) of area burned associated with a subset of years over short temporal series (Vivchar, 2011; Boiffin and Munson, 2013), f) non-Gaussian (e.g,., Weibull) parametric tests performed on fire scars, lake cores, or tree core data (Fule *et al.* 2003; Drobyshev *et al.* 2014), and g) years exceeding a set threshold of burned area within a region (Johnson and Wowchuk, 1992; Kasischke *et al.* 2010). In many cases, years have been described as large fire years without a qualifying test (Schoennagel *et al.* 2005; Fauria and Johnson, 2008).

This brief synthesis highlights that various statistical approaches have been performed to define such extreme, large, or regional fire years. However, these statistical approaches vary widely in their robustness and applicability, with many assumptions and limitations ignored and many tests applied inappropriately or without necessary post-hoc tests. Regardless of the particular research question under investigation, there exists a clear need to propose standardize approaches to derive such extreme fire years, or where appropriate (depending on the science question being asked) to not subset the data to develop relationships. The objectives of this study are to:

1. Identify extreme fire years for a region in the northwestern United States from 1889-2010 using a combination of fire atlas and satellite sensor data of area burned per year,
2. Independently identify anomalous climatic years over 1895-2013 using PRISM data of minimum and maximum temperature, accumulated precipitation, climate water deficit, and the Palmer Drought Severity Index,
3. Evaluate fire-climate relationships using both analysis of these subset years and analysis of the entire datasets, and
4. Demonstrate the impact of using appropriate and inappropriate statistical methods in assessing such questions.

Background: Statistical steps to identify extreme fire years

Identifying a distribution and creating residuals

Most studies seek to use parametric statistics to analyze their data as they are easier to use. Parametric statistics are those applied to any known probability distribution (e.g., Gaussian, Wiebull, Poisson). When considering the goal of identifying regional fire years from burned area datasets it is quite feasible depending on the regional extent analyzed that the data distribution can be matched or transformed to fit a known distribution. For example, regional area burned temporal series are often log-transformed to be normal (Higuera *et al.* 2015). Residuals are the difference of the observed value and an estimated value (e.g. data point and a regression line or data and sample mean) and are widely analyzed to assess the assumptions required for the various statistical analyses steps.

Testing for independence, normality, and homogeneity of variance

The first requirement to use most parametric tests is that ideally the data is independent and identically distributed, but in reality many tests can be tweaked to permit some degree of non-dependence. This is particularly the case as specific tests of independence are difficult to come by and most studies use logical arguments to determine independence of data (such as fitting two models, where one requires independence and the other allows modelling for different kinds of dependence). Importantly, conditional independence is necessary; namely, independence of the errors within the data. Usage of statistical approaches like semi-variograms and autocorrelation analyses are often applied to infer whether data is not dependent (Higuera *et al.* 2015). If dependence is suspected (e.g., from logic, very low p-values) additional tests for dependence can be performed.

The next step of pre-analysis is to assess for conditional normality; although depending on the data distribution and the sample size the Central Limit Theorem is often used to assume conformity. Alternatively, conformity can also be informally assessed by a bootstrap approach using quantile-quantile plots, where users compare the distribution of the estimates from a sample of the data with a parameter that exhibits a normal distribution. Although not widely used, there exist >40 available tests for normality (Dufour et al. 1998), with the most common being Shapiro-Wilk and Kolmogorov-Smirnov. A synthesis study by Razali and Wah (2011) determined that the Shapiro-Wilk test is the most powerful, with Kolmogorov-

Smirnov being the least favorable of four tested. The original Shapiro-Wilk test is valid up for sample sizes of 3-50, although a revised version that is commonly applied with most statistical software extends the range to 5000 (Roysten, 1995). The two other most common methods are the Anderson-Darling and the Lilliefors normality test, which is an adjusted version of the Kolmogorov-Smirnov Test that does not require the variance or expected value to be known (Razali and Wah, 2011). Graphical methods such as normal quantile-quantile plots, histogram, box-plots, and stem-and-leaf plots are commonly applied but these do not actually perform a test for normality. Likewise, skewness and kurtosis are also often applied but are less rigorous than formal normality tests.

Homogeneity of variance is often tested because goodness-of-fit values can be overestimated when it is not accounted for and as heterogeneity increases a corresponding rise in ANOVA Type I errors often occur (McGuinness, 2002). Homogeneity of variance can typically be evaluated using nested models, F-max test, Cochran's test, Levene's Test, or a Bartlett's Test (McGuinness, 2002; Jones et al. 2009). However, as outlined by McGuinness (2002) if pre-analysis tests are not performed the results of such tests can lead to increased errors. Data that do not exhibit homogeneity of variance is described as heteroscedastic were examples include datasets whether the variability between two variables changes, potentially due to an unknown third factor. For data with more than one factor, the assumption tests should be performed on the residuals or on a continuous predictor. If any of these three assumptions are violated, the data should be transformed and the transformed data re-tested; conversely a different distribution should be used (e.g., Wiebull). Subsequent analysis should then be performed on the transformed data or other distribution.

Tests to Detect Potential Outliers

The assessment of extreme fire years could be done without the use of statistical tests through the use of arbitrary decisions. However, we would contend that if the goal is to use a defensible statistical approach, then arguably the most appreciate suite of approaches to assess extreme values would be outlier detection and verification methods. Other approaches do exist, such as generalized extreme value theory where a tail of a distribution is fitted to values that are assumed to be extreme (CITE). Importantly, two steps are needed to assess for outliers: (i) identifying potential outliers followed by (ii) formal outlier tests conducted as part of a post-hoc analysis (described below). To identify potential outliers or extreme values within a normal dataset, common tests can include Tukey's Range Test, the Modified Z-score, and analysis of standard deviations about a mean (Table 1). Percentile breaks can be used on both parametric and non-parametric data.

The Tukey's Range Test identifiers potential upper outliers as all values > third quartile + 2.2 IQR and potential lower outliers for all values < first quartile – 2.2 IQR, where, IQR denotes the interquartile range (Tukey, 1977; Hoaglin et al. 1986). This method is commonly applied with the original multiplier of 1.5 times the IQR, however, analysis by Tukey and others subsequently determined that a multiplier of 2.2 was preferred (Hoaglin and Iglewicz, 1987). Another common test is the Modified Z-Score that defines potential outliers as having absolute calculated index values exceeding 3.5 (Table 1). The Modified Z-

Score is preferred over the original as it is a more robust test and the index values are easier to interpret as they extend over a larger range (Iglewicz and Hoaglin, 1993). Caution should be applied when using either percentile breaks or standard deviations above mean as a normal distribution will always tend to have some values above any set threshold. For example, a mean plus two standard deviation threshold will still have ~2.5% of the data above that value, where the data above that threshold could exhibit multiple orders of magnitude difference in size. This is similarly applicable to a percentile threshold as by default ~2.5% of the data points will be present following a 97.5% percentile threshold. Identification of potential outliers within non-parametric data distributions include the use of an adjusted box plot or other complex methods as described in the statistics literature (Gnanadesikan and Kattenring, 1972; Rousseeuw and Leroy, 2005; Rousseeuw and Hubert, 2011).

Post-hoc tests to test for actual outliers

To assess whether potential outliers in parametric data are actual outliers, various tests are available, where examples include the Grubb's Test, the Tietjan-Moore Test, the Generalized Extreme Studentized Deviate (ESD) Test, and the Dixon's Q Test. The Grubb's Test is an appropriate test for assessing single outliers within parametric data, where a generalization of the Grubb's Test a known number of for more than one outlier is provided by the Tietjen-Moore Test (Tietjan and Moore, 1972). The Generalized Extreme Studentized Deviate (ESD) Test is most appropriate when the precise number of outliers in unknown (Rosner, 1983). The Dixon's Q Test is only appropriate when seeking to evaluate a single outlier within a small (<40) sample size (Dixon, 1950). Although many non-parametric tests exist, the majority of commonly applied tests focus on cases where there are more than one set of data (e.g., Mann-Whitney Test, Wilcoxon-Paired Sample Test, Kruskal-Wallis Test, Spearman Test). There are limited examples of univariate non-parametric rank tests but these require repeated measures. Given the assessment of regional fire years within a burned area temporal series is a univariate problem, hereafter, the primary focus of this paper will be on parametric tests, where non-parametric statistics are referenced only as appropriate.

Correlation Analysis

The assessment of fire – climate relationships will require correlation or regression analysis. The most commonly applied correlation analysis methods are Spearman, Pearson, and Kendall, where none of these approaches require normal data (CITE). Although Pearson does not require normality it works better when the data is normal and of equal variances (CITE). The Spearman and Kendall tests are more robust to outliers. However, in each case these methods both rank the data from highest to lowest, where the highest is given a value of 1, the next a values of 2, and so on. Consequently, sensitivity in the data that would alert the user to patterns and trends is lowered. Furthermore, Spearman and Kendal are not appropriate for use on truncated data, such as where a sub-sample or percentile threshold has been applied (CITE).

Methods

Burned Area Data

Our study focused on annual burned area data from two Level III Ecoregions that prior studies have evaluated (Morgan et al., 2008; Dillon et al., 2011; Lannom et al., 2014). The current study uses both fire atlas data and MTBS fire perimeter polygons from the Columbia Mountains / Northern Rockies, Canadian Rockies, and Idaho Batholith ecoregions (Figure 1). The Fire Atlas data, which is described in detail in past studies (Holden et al., 2005; Morgan et al., 2008), extends from 1889 to 1983, after which point we replaced it with MTBS fire perimeter data from 1984 to 2010. These datasets report total burned area per year making it an ideal univariate case study for this paper. The uncertainties within these two datasets are well documented in the literature (Morgan et al., 2008; Kolden et al., 2012; Lannom et al., 2014; Sparks et al., 2015).

Climate Data

Climate data was acquired via the PRISM project (PRISM, 2015). Datasets included at monthly time-steps, average maximum temperature (°C), average minimum temperature (°C), accumulated precipitation (mm), the Palmer Drought Severity Index (PDSI), and the climatic water deficit (CWD, mm). Each dataset covers the range 1895 to 2013 and was calculated as areal averages from area encompassing the two ecoregions. Following past studies, we evaluated the August PDSI value as a proxy for soil moisture at the end of August, annual climate water deficit to account for the relative difference of precipitation minus potential evapo-transpiration. In the analysis of precipitation and temperature datasets, we followed past studies and evaluated different monthly date ranges of winter, spring, and summer (Morgan *et al.* 2008; Higuera *et al.* 2015). Autocorrelation analysis was performed between the precipitation and temperature data and between CWD and temperature data.

Statistical Tests

To demonstrate the impact of using appropriate and inappropriate statistical methods, this paper will demonstrate four scenarios to identifying extreme fire years from time series of annual area burned data:

1. Appropriate 1: (i) conduct normality test, (ii) if non-normal, transform and retest for normality, (iii) if normal detect potential outliers, and (iv) test potential outliers.
2. Appropriate 2: (i) conduct normality test, and (ii) if non-normal, but fits another distribution (e.g., Wiebull) performs appropriate parametric tests to detect potential outliers, and (iii) test potential outliers.
3. Inappropriate A: (i) no normality test conducted, (ii) detect, but not test potential outliers.
4. Inappropriate B: (ii) transform for normality or identify another distribution model, and (ii) detect but not test potential outliers.

In terms of the *Appropriate 2 Scenario*, it is not possible to test for true outliers on data where the distribution is not known as all the true outlier tests evaluate the values against a given distribution. The subsequent analysis of the appropriate scenarios will be used to answer the question to characterize the fire return interval of these extreme fire years for this study area.

To test for normality we performed the Shapiro-Wilk Test within SPSS version 22 (IBM, Armonk, NY). The data was not-normally distributed and therefore for the purposes of the *Appropriate Scenario 1* it was log-transformed and re-rested for normality. To test for potential outliers within the dataset we performed the following tests within the open source statistical software package, R (R Development Core Team, 2014) and as available were cross-checked in SPSS: Tukey's Boxplot Method, Modified Z-Score, Percentile, Standard Deviation (Z-Score), and Adjusted Box Plot. To test whether potential outliers are true outliers the Generalized Extreme Studentized Deviate (ESD) Test was performed in R. We acknowledge that parametric tests can be used on other distributions (e.g., Poisson, Poreta), but selected Gaussian and Weibull given their widespread usage in identifying large fire years.

Results

Identifying distributions

Using the Shaprio-Wilk Test the raw burned area data did not exhibit a normal distribution ($p<0.001$). We also tested and determined that the burned area data somewhat fitted a Weibull distribution ($p=0.209$) under shape parameter of 0.49 and a scale parameter of 18,103. However, the burned area data was determined to be normally distributed following a log-transformation ($p=0.536$). Excluding the years with zero burned area only slightly changed the significance ($p=0.526$). Using the Shaprio-Wilk Test the PRISM average maximum temperature ($p=0.427$) and accumulated precipitation ($p=0.625$) datasets both exhibited normal distributions. Table 3 presents the descriptive statistics of the PRISM data. Excluding the years with zero burned area increases the significance to 0.901 and 0.651 for the temperature and precipitation data respectively. Annual CWD and PDSI August values both exhibit normal distributions, with significance of 0.315 and 0.241 respectively.

Identifying and verifying outliers

In terms of the appropriate scenarios, at most only four years are identified as outliers and thus are potentially extreme, large, or regional fire years in this study region. In terms of the ESD test no outlier's were identified, although 1910 only marginally failed the test. The years identified as outliers by the ADJ Box Plot were 2007, 2000, 1919, and 1910. The 1910 fire, commonly referred to as The Big Burn (Egan, 2014), burned in excess of 1 million hectares, whereas the 1919, 2000, and 2007 fires consumed between 471,000 and 551,000 hectares (Table 4). This is in sharp contrast to some years identified when using the inappropriate scenarios that were as low as 100,000 hectares. When analyzing the data by separate ecoregion, no potential outliers on the normalized data were identified. Inappropriately identifying outliers on the non-normal data yielded 12 potential outliers in the Columbia Northern Rockies, 17 in the Idaho Batholith, and 9 within the Canadian Rockies ecoregion.

Fire-Climate Relationships

Table 5 presents linear regressions between the natural logarithm of area burned and various climate variables. In terms of single variable regressions, the average maximum temperature (July-August) and the total summer precipitation (June to August) produced the best regressions with r^2=0.34 and 0.30, respectively. With two independent variables used with the regression, the average maximum temperature (July-August) coupled with the annual total climate water deficit accounted for ~10% more variance with a r^2=0.43. Only a minimal improvement was achieved when incorporating three or four independent variables.

Discussion

Impacts of inappropriate statistics

The results from *Inappropriate Scenarios A and B* highlight two problems that can result by using incorrect methods: (1) any percentile or standard deviation threshold will produce an arbitrary separation of data where the separated data can contain either similar values or values ranging over several orders of magnitude; and (2) potential detected outliers are not necessarily true outliers. In terms of (1), a fixed 90th percentile will always give you the top 10% values from a ranked list regardless of the actual sizes of the values in that list. For example, data could conceivably be near-identical over an analysis period, but the fixed 90th percentile would nevertheless still report the top 10% as being outliers. Conversely, the top 10% of data entries could equally arise from a highly skewed distribution and exhibit a wide range of values. The latter is apparent in this case study, as the 90th percentile identifies fire years with area burned greater than 135, 347 ha at the lower end and 1,031, 812 ha at the upper end. In terms of (2), not testing potential outliers as actual outliers clearly leads to many extreme fire years being erroneously identified (Table 2).

Comparing the results of the Tukey's method applied to the raw data and the transformed data highlights a problem that can arise with the incorrect usage of these tests. Namely, they detect outliers based on the data points distance from the whole distribution; thus it will be expected that non-normally distributed data would not fall within the normal distribution and instead be erroneously identified as outliers. In the case study, the Tukey's approaches applied to the raw (non-normal data) identified 12-14 potential outliers as compared to no identified outliers on the log-normal transformed data.

Comparison with previously detected extreme fire years

The number of outliers identified via the appropriate scenarios (i.e. maximum of 4) is considerably lower than previously reported by other studies assessing extreme fire years within the broadly representative ecoregions of our case study. The four years that were identified as potential outliers in the current study where 1910, 1919, 2000, and 2007. Table 3 contains the average and ranges of all the climate data for the region and Table 4 is data just for the potential 4 outlier years. Comparing these tables demonstrate that although not verified statistical outliers, each of these years (1910, 1919, 2000, and 2007) exhibit values that in most cases exceed two standard deviations above the mean.

Morgan et al. (2008) identified 11 years between 1900 and 2003 using a 90th percentile approach. All the extreme fires years identified by the appropriate scenario in the current case study were also identified by Morgan et al. (2008) as regional fire years; were arguably the current study places more stringent criteria. The additional identified extreme year of 2007 from the current case study was outside the temporal series analyzed by Morgan et al. (2008). However, the other 7 regional fires years identified by Morgan et al (2008) were not determined to be outliers within the appropriate scenarios evaluated within the current case study. In comparison to the current case study, the data from Morgan et al. (2008) geographically constrained the ecoregion data to the State of Idaho and parts of Montana that intersected with the Level III ecoregions of the Northern Rockies and the Canadian Rockies. In each of the current case study and Morgan et al., (2008) no fires were included from the Snake River Plain Level III ecoregion and Morgan et al. (2008) only included a small number of fires from the Blue Mountains Ecoregion. We contend that these slight differences between the datasets would not overly impact the results. Although Lannom et al. (2014) evaluated a wider study region, this prior study identified 2007 as being a statistically extreme burned area year using the Tukey's 1.5 IQR test, with 2006, 2000, and 1988 as outliers. Dillon *et al.* (2011) identified 5 widespread fire years between 1984 and 2006 using the Tukey's 1.5 IQR test, where none of these years matched the two outliers (i.e., 2000 and 2007) identified by the appropriate scenarios in the current study. Dillon *et al.* (2011) analyzed the Northern Rockies ecoregion, which differenced from the current study as it encompassed the Level III ecoregion of the Middle Rockies that included the 1988 Yellowstone fires, which could account for some of the differences in the analyzed distributions.

Comparison with previously assessed fire-climate relationships

Table 5 highlights the relationships that were also evaluated within this study region in two prior studies (Morgan *et al.* 2008; Higuera *et al.* 2015). In Morgan *et al.* (2008) the analysis was conducted on the non-log transformed are burned data and only on the truncated dataset of identified regional fire years. Consequently, these prior results were likely not correct and this likely explains why the correlation values are significantly lower than that observed in the current study and in Higuera *et al.* (2015). Higuera *et al.* (2015) observed similarly strong, although slight better, regression coefficients with the current study. The slight improvements observed by Higuera *et al.* (2015) likely arise as rather than excluding data with no area burned; this study set these values to 10 hectares to accommodate the log transformation.

Conclusions

On balance, the adjusted box plot or a Z-score would be the most appropriate tests to evaluate extreme years of area burned, as detection of true outlier years may be too constrained for the purposes of elucidating fire-climate interactions. Therefore, future studies avoid descriptors that imply that they are assessing statistical outliers and they should not be treated as such in the subsequent analysis (i.e. excluding them or truncating the data to just include them). When using rank correlations such as Spearman's or Kendall's usage of arbitrary percentiles and similar descriptive-statistics thresholds should be avoided. When thresholds are used for other analysis, these should be selected that match ecologically relevant phenomena; such as annual burned area sizes associated with increased probability of ecosystem regime shifts (Smith *et al.* 2014).

A common science question in wildland fire science is to evaluate how inter-annual climate variability impacts annual area burned (Higuera *et al.* 2015). It could be argued that such an analysis does not require the formal detection of anomalously large years as fire-climate relationships should ideally be transferable regardless whether or not the year experiences extreme large quantities of annual area burned. However, there could be nonlinearities and thresholds in such relationships whereby extreme fire years as discussed in this paper do occur under novel conditions. However in such approaches, care should still be taken to ensure that the subset maintains sufficient data points and meets the assumptions for statistical regression analysis. In these studies the detection of outlier annual burned area years as presented in this paper would be prudent to isolate true outliers when conducting the regressions.

Acknowledgements

This work was funded by the National Aeronautics and Space Administration (NASA) under award NNX11AO24G.

References

Boiffin J, Munson AD (2013) Three large fire years threaten resilience of closed crown black spruce forests in eastern Canada, Ecosphere, 4, 5, 56.

Dillon GK, Holden ZA, Morgan P, Crimmins MA, Heyerdahl EK, Luce CH (2011) Both topography and climate affected forest and woodland burn severity in two regions of the western US, 1984 to 2006. *Ecosphere* 2, 12, Article 130. doi:10.1890/ES11-00271.1

Dixon, W.J. (1950). Analysis of extreme values. Ann. Math. Stat. 21, 4, 488-506.

Drobyshev I, Granstrom A, Linderholm HW, Hellberg E, Bergeron Y, Niklasson M (2014) Multi-century reconstruction of fire activity in Northern European boreal forest suggests differences in regional fire regimes and their sensitivity to climate, Journal of Ecology, 102, 738-748, doi: 10.1111/1365-2745.12235

Dufour, J. M., & Renault, E. (1998). Short run and long run causality in time series: theory. *Econometrica*, 1099-1125.

Fauria MM, Johnson EA (2008) Climate and wildfires in the North American boreal forest, Philosophical Transactions of the Royal Society B, 363, 2317-2329, doi: 10.1098/rstb.2007.2202

Flannigan MD, Logan KA, Mairo BD, Skinner WR, Stocks BJ (2005) Future area burned in Canada, Climatic Change, 72, 1-16, doi: 10.1007/s10584-005-5935-y

Gedalof Z, Peterson DL, Mantua NJ (2005) Atmospheric, climatic, and ecological controls on extreme wildfire years in the northwestern United States, Ecological Applications, 15,1, 154-174.

Hanson CT, Odion DC (2014) Is fire severity increasing in the Sieraa Nevada, California, USA? International Journal of Wildland Fire, 23, 1-8, doi: 10.1071/WF13016.

Hoaglin, D. C., and Iglewicz, B. (1987), Fine tuning some resistant rules for outlier labeling, Journal of American Statistical Association, 82, 1147-1149.

Hoaglin, D.C., Iglewicz, B., and Tukey, J.W. (1986). Performance of some resistant rules for outlier labeling, Journal of American Statistical Association, 81, 991-999.

Holden ZA, Smith AMS, Morgan P, Rollins MG, Gessler PE (2005) Evaluation of novel thermally enhanced spectral indices for mapping fire perimeters and comparison with fire atlas data. *International Journal of Remote Sensing* 26(21), 4801–4808. doi:10.1080/01431160500239008

Iglewicz B. and Hoaglin D. (1993). “Volume 16: How to Detect and Handle Outliers”, *The ASQC Basic References in Quality Control: Statistical Techniques*, Edward F. Mykyta, Ph.D., Editor.

IPCC (2013). The Fifth Assessment Report (AR5) of the United Nations Intergovernmental Panel on Climate Change (IPCC), Climate Change 2013: The Physical Science Basis, IPCC WGI AR5. Tech. rep., Intergovernmental Panel on Climate Change (IPCC).

Johnson EA, Wowchuck DR (1992) Wildfires in the southern Canadian Rocky Mountains and their relationship to mid-tropospheric anomalies, Canadian Journal of Forest Research, 1213-1222.

Kasichke, ES, Verbyla DL, Rupp TS, McGuire AD, Murphy KA, Jandt R, Barnes JL, Hoy EE, Duffy PA, Calef M, Turetsky MR (2010) Alaska’s changing fire regime – implications of the vulnerability of it boreal forests, Canadian Journal of Forest Research, 40, 1313-1324, doi: 10.1139/X10-098

Kolden CA, Lutz JA, Key CH, Kane JT, van Wagtendonk JW (2012) Mapped versus actual burned area within wildfire perimeters: characterizing the unburned. *Forest Ecology and Management* 286, 38–47. doi:10.1016/J.FORECO.2012.08.020

Lannom KO, Tinkham WT, Smith AMS, Abatzoglou J, Newingham BA, Hall T, Morgan P, Strand EK, Paveglio Y, Anderson JW, Sparks AM (2014) Defining extreme wildland fires using geospatial and ancillary metrics. *International Journal of Wildland Fire* 23, 322–337. doi:10.1071/WF13065

Miller JD, Skinner CN, Safford HD, Knapp EE, Ramierz CM (2012) Trends and causes of severity, size, and number of fires in northwestern California, USA, Ecological Applications, 22, 1, 184-203.

Morgan P, Heyerdahl EK, Gibson CE (2008) Multi-season climate synchronized forest fires throughout the 20th century, Northern Rockies, USA, Ecology, 717-728.

Pechony O, Shindell D.T (2010) Driving forces of global wildfires over the past millennium and the forthcoming century, Proc. Natl. Acad. Sci., 107(45), 19167–19170, doi:10.1073/pnas.1003669107.

R Development Core Team, 2007. R: a language and environment for statistical computing. R Founcation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org.

Razali NM, Wah YB (2011) Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests, Journal of Statistical Modeling and Analytics, 2, 1, 21-33.

Rosner, Bernard (May 1983), Percentage Points for a Generalized ESD Many-Outlier Procedure,Technometrics, 25(2), pp. 165-172.

Rousseeuw, P. J. and Hubert, M. (2011), Robust statistics for outlier detection. WIREs Data Mining Knowl Discov, 1: 73–79.

Rousseeuw, P. J. (1991), Tutorial to robust statistics. J. Chemometrics, 5: 1–20. doi: 10.1002/cem.1180050103

Royston, P. (1995). Remark AS R94: A remark on algorithm AS 181: The W-test for normality. *Applied Statistics*, 547-551.

Schoennagel T, Veblen TT, Romme WH, Sibold JS, Cook ER (2005) ENSO and PDO variability affect drought-induced fire occurrence in Rocky Mountain subalpine forests, Ecological Applications, 15, 6, 2000-2014.

Sparks AM, Boschetti L, Smith AMS, Tinkham WT, Lannom KO, Newingham BA (2015) An accuracy assessment of the MTBS burned area polygons for shrub-steppe fires in the northern Great Basin, United States, *International Journal of Wildland Fire*, 254, 70-78, doi: 10.1071/WF13206.

Tukey, J.W. (1977). Exploratory Data Analysis. Reading, MA: Addison-Wesley.

Vivchar A. (2011). Wildfires in Russia in 2000-2008: estimates of burnt areas using the satellite MODIS MCD45 data, Remote Sensing Letters, 2, 1, 81-90.

Table 1. Common statistical methods used in the assessment of outliers

Analysis Step	Statistical Test	Assumptions	Common Software
Normality	Shapiro-Wilk[1]	Sample Size: $3 \leq n \leq 5000$	SPSS, SAS, STATA, R: nortest
	Anderson-Darling[2]		SAS, STATA, R: nortest
	Lilliefors		SPSS, SAS, STATA, R: nortest
	Kolmogorov-Smirnov		SAS, STATA, R: nortest
Homogeneity of variance	Levene		SPSS, SAS, STATA, R: car
	Bartlett		SPSS, SAS, STATA, R
Detect Potential Outliers	Tukey's Boxplot Method	Normally Distributed Independent and Identically Distributed Homoscedastic	SPSS, SAS, STATA, R
	Modified Z-Score	Relatively Normally Distributed	SAS, STATA, R
	Percentile		SPSS, SAS, STATA, R
	Standard Deviation (Z-Score)	Normally Distributed	SPSS, SAS, STATA, R
	Adjusted Box Plot		SAS, R: robustbase
Formal Outlier Tests	Grubb's Test	Single Outlier Normally Distributed	SAS, STATA, R: outliers
	Tietjen-Moore Test	Known Number of Outliers	SAS, STATA, R: outliers
	Generalized Extreme Studentized Deviate (ESD)	Upper Bound for Number of Outliers	SAS, STATA, R: outliers
	Dixon's Q Test	Single Outlier	SAS, STATA, R: outliers

Table 2 Potential outliers as identified on both the original (non-paramedic data) and data transformed to be normally distributed.

		Inappropriate A: Analysis on raw non-normal data where all potentials are assumed to be outliers				Inappropriate B: Analysis on data transformed to be normal where all potentials are assumed to be outliers				Appropriate 1 and 2: Analysis on data transformed to be normal (or non-parametric tests used) where potential outliers are tested	
Year	Total Area Burned (ha)	90th Percentile	Tukey Outlier	Tukey Extreme	SD > 2	90th Percentile	Tukey Outlier	Tukey Extreme	SD > 2	Appropriate 1 ESD[1,2,3]	Appropriate 2 ADJ Box Plot
2007	551,488	y	y	y	Y	y			y		y
2006	135,347	y	y	y		y					
2005	82,806										
2003	210,168	y	y	y		y					
2000	471,056	y	y	y	y	y			y		y
1994	241,869	y	y	y		y					
1992	105,685		y								
1988	128,909		y	y							
1934	138,158	y	y	y		y					
1931	93,232		y								
1929	158,730	y	y	y		y					
1926	157,589	y	y	y		y					
1919	507,772	y	y	y	y	y			y		y
1910	1,031,812	y	y	y	y	y			y		y
1889	298,185	y	y	y		y					

[1] ESD is a formal outlier detection test (as opposed to identifying "potential" outliers)
[2] Analysis performed on transformed data
[3] Analysis performed on log-normal data

Table 3 Descriptive statistics of PRISM datasets (1895-2010) used within this study. Temperature data in units of °C and precipitation in units of mm. CWD – climate water deficit, PDSI – Palmer Drought Severity Index.

PRISM Variable	mean	s.d.	min	max	Potential Outliers
Annual Max Temperature (Jan - Dec)	26.2	1.4	22.1	29.7	1912 (22.1 °C)
Average Spring Max Temperature (Mar - May)	10.5	1.5	6.8	15.6	1934 (15.6 °C)
Average Summer Max Temperature (July - Aug)	23.5	1.0	21.3	26.9	1961 (26.9 °C)
Annual Total Precipitation (previous Aug - current July)	820.6	124.8	557.9	1157.8	1997 (1157.8 mm)
Total Spring Precipitation (June – Aug)	209.9	46.8	75.6	341.9	2011 (341.9 mm) 1924 (75.6 mm)
Total Summer Precipitation (June – Aug)	120.4	38.2	31.5	217.2	none
Annual Total CWD	975.7	71.4	777.7	1148.0	none
PDSI (Aug)	-0.1	1.6	-0.4	4.8	1983 (4.8) 1984 (4.7)

[1] Identified as outliers using the Adjusted box plot method.

Table 4. PRISM Properties of the four potential Adjusted Box Plot area burned years. Temperature data in units of °C and precipitation in units of mm. CWD – climate water deficit, PDSI – Palmer Drought Severity Index.

Year	**Total Area Burned (ha)**	**Annual Max Temperature (Jan - Dec)**	**Average Spring Max Temperature (Mar - May)**	**Average Summer Max Temperature (July - Aug)**	**Annual Total Precipitation (Aug - July)**	**Total Spring Precipitation (June – Aug)**	**Total Summer Precipitation (June – Aug)**	**Annual Total CWD**	**PDSI (Aug)**
1910	1031811.6	27.0	14.0	23.8	835.8	179.1	42.2	1115.0	-2.4
1919	507772.0	27.5	11.3	25.2	712.9	199.9	31.5	1045.4	-2.2
2000	471055.9	26.2	11.2	24.0	798.4	197.5	70.5	1051.9	-2.3
2007	551487.9	29.7	12.4	25.3	752.1	154.3	73.7	1120.2	-3.0

Table 5 Fire-climate relationships on all non-zero data predicting the natural logarithm of area burned. Regressions were all tested at the 0.05 level.

Independent Variable(s)	**r^2**
Annual Max Temperature (Jan - Dec)	0.20
Average Summer Max Temperature (July - Aug) [1]	0.34
Average Spring Max Temperature (Mar - May) [2]	0.14
Annual Total Precipitation (previous Aug - current July)	0.15
Total Summer Precipitation (June – Aug) [1, 2]	0.30
PDSI (Aug)	0.19
Annual Total CWD	0.25
Average Spring Max Temperature (Mar – May) **Annual Total CWD**	0.30
Average Summer Max Temperature (July – August) [1, 2] **Summer Precipitation (July - Aug)**	0.38
Average Spring Max Temperature (Mar – May) [2] **Average Summer Max Temperature (July – August)**	0.40
Average Summer Maximum Temperature (July – Aug) **Annual Total CWD**	0.43
Annual Max Temperature (Jan - Dec) **Annual Total CWD**	0.36
Average Spring Max Temperature (Mar – May) **Average Summer Maximum Temperature (July – Aug)** **Annual Total CWD**	0.46
Average Spring Max Temperature (Mar – May) **Average Summer Maximum Temperature (July – Aug)** **Annual Max Temperature (Jan - Dec)**	0.40
Average Spring Max Temperature (Mar – May) **Average Summer Maximum Temperature (July – Aug)** **Total Spring Precipitation (March – May)** **Total Summer Precipitation (July – Aug)**	0.44

[1] Relationships also assessed in Higuera et al. (2015).
[2] Relationships also assessed in Morgan et al. (2008).

Chapter 2

Assessing extreme fires: observations via a case study evaluating extreme fires in the intermountain West

Abstract

Wildfires play a significant role in natural, political, social, and economical trends around the world. As the effects of climate change become more severe and apparent, many believe wildfire events are becoming more extreme or frequent. However, the definition and evaluation of 'extreme' events is ambiguous at best and misleading at worst. This study assesses the classification of extreme wildfire events with a focus on physical metrics. Using a case study of the MTBS era intermountain west, we compared previously utilized statistical methods determining extreme fires with a novel approach, Mahalanobis distance. Fewer 'extreme' events were discovered with our methods (~1%); most of these were also identified as extreme events by the other analyses.

Introduction

Wildfires are a natural cycle throughout many ecosystems and ecosystem services. A combination of humans spreading into previously uninhabited areas and fire regime shifts means that wildfires are increasingly likely to affect human activity and interests. As climate change becomes more severe, the effects on fire regimes will likely become more pronounced. It is widely believed that extreme fires are becoming more common. However, there is little consensus as to what defined an 'extreme' wildfire event. There is also only a short time's worth of reliable satellite data available for fires. This short time series greatly hampers efforts to assess long term trends such as an increase in extreme wildfire events. Either more years will have to be added to the temporal series, or the increase in extreme events must be severe enough to overcome the brevity of the timescale.

Studies have sought to classify extreme wildfire events in several ways. Some focus on the effects an event has on natural cycles of the area or consider an extreme event to be the combination of several fires (Strauss *et al.*. 1989). Others evaluate a fire's economic, social, or political impacts on the region (Calkin *et al.*. 2005; Gerbert *et al.*. 2007; Liang *et al.*. 2008). This approach may well be most appropriate for social sciences. Physical metrics collected through the Monitoring Trends in Burn Severity (MTBS) project are more consistent, more accurate, and require only access to the database to utilize.

Background: Statistical steps to identify extreme fire years

Multivariate analyses and tests of their assumptions are generally more convoluted than their univariate counterparts (Tabachnick and Fidell, 2001). Testing for normality can become particularly difficult. It is easiest to first assess whether the variables are conditionally independent. This can be done through logical arguments or through statistical methods such as semivariograms or autocorrelation analyses. Should the variables be conditionally independent and both normal, the multivariate distribution will also be normal. However, if the data are not conditionally independent, multivariate normality must be tested as they will not necessarily follow a joint multivariate normal distribution. This can be done via several statistical tests (e.g. Mardia's test) or through graphical methods (Mardia, 1970). Where a dependent variable is involved, it can also be done by creating a regression of the variables and examining the residuals. If the residuals are normal, the joint multivariate distribution is normal as well. The residuals can also be used to determine possible outliers.

It is not uncommon for data sets, particularly large ones such as these, to not pass a goodness-of-fit test for a Gaussian distribution despite transformations. Large sample sizes make these tests, such as the Shapiro-Wilk test or the Kolmogorov-Smirnov test, very sensitive to even small departures from the distribution (Oztuna et al, 2006). Additional methods, particularly graphical ones such as histograms or Normal Q-Q plots, may be necessary to determine if the data are normal. When datasets with large sample sizes exhibit a significant digression from the distribution, yet form a tight line around a Q-Q plot and appears normal upon visual inspection, it may be possible to utilize parametric methods. Additionally, by the Central Limit Theorem, sufficiently large samples (as here, with 1956 samples) which depart from normality should not be an issue in parametric tests (Elliott and Woodward 2007, Altman and Bland 1995). Robust analyses may still be utilized if the data are 'normal enough', though this entails personal judgment and results should be taken with a degree caution (Pallant 2007, Ghasemi and Zahediasl 2012).

Lannom *et al.* (2014) developed a method for determining extreme fire events by selecting those fires which met three or four of the following criteria: 1) Over the 90th percentile in size, 2) Over the 90th percentile in the percent of the fire which burned severely, 3) Over the 90th percentile in fire durations, and 4) Less than the 10th percentile in distance to Wildlife Urban Interfaces (WUI). This process is representative of more aspects of fires than any single metric, such as area burned, would be.

Mahalanobis distance measures the distance between a point and a multivariate distribution, essentially an n dimensional expansion of how many standard deviations away from the distribution a point is, to locate multivariate extreme values (Mahalanobis, 1936). Values which are outliers of multiple variables may not be outliers of the multivariate dataset. It is particularly useful in that it accounts for correlations between variables and can be used when the variables have different scales and units. It can be utilized in its traditional form when the data follow a Gaussian distribution or expanded to identify outliers in other distributions (Ekstrom, 2011). When the data are normally distributed, the Mahalanobis distance follows a Chi-Square distribution (Mahalanobis, 1936). Then any point above the standard p-value for Mahalanobis distance (p=0.001) is statistically significant and an extreme value. Non-normal data which do not fit another known distribution may still be ranked through Mahalanobis distance.

Methods

Fire Data

The raw data used here are described in detail by Lannom *et al.* (2014). The study area involved two Level I ecoregions, the Northwest (NW) Forested Mountains and North American Deserts. Area burned, percent burned severely, and duration were collected using MTBS data for the study region. Issues in the wildlands fire area and burn severity data sets are well known and documented both in general and in these data in particular (Lannom *et al.*. 2014). However, duration is perhaps the most inconsistent fire metric and its problems rarely discussed (Moritz. 1997). First, it is often difficult to determine exactly when a fire started or has been completely extinguished. Second, the end date of a fire is frequently an artifact of management practices. It may refer to the end of active burning, the end of smoldering and hot spot activity, or the date when personnel are sent back. Additionally, the allocation of resources during 'active' fires makes it beneficial to maintain active status for a mitigation period after the fire is seemingly extinguished (Abt. 2009). Incident commanders of each event should be contacted to assess his or her method of determining the start and end dates of an event.

Care must also be taken in data configuration. Many fires have multiple polygons which must be joined so that each fire is represented only once. Failure to do so will artificially skew the data right, with many more small fires than actually exist and fewer large fires. This may also affect which fires are seen as extreme; some fires are comprised of several nearly equal size patches and will thus become many times larger when these uniqueness issues are resolved. Difficulties in obtaining accurate data may result in events which are missing metrics. It is imperative that these are recognized as no data and not as zeroes or another numerical placeholder. Failure to do so will again result in data artificially skewed right.

Statistical Tests

The assessments were performed using either two or three variables: area burned and percent severely burned or area burned, percent severely burned, and duration. Normality was assessed using the Shapiro-Wilk test, histogram, and Normal Q-Q plot via SPSS version 22 (IBM, Armonk, NY). Correlation was assessed with a Kendall tau test in SPSS. The data were not normally distributed and were significantly correlated. They were subsequently transformed and the residuals plotted. Normality tests via SPSS were performed on these residuals.

Results and Discussion

None of the data were normally distributed in their raw form. Additionally, there was significant correlation between area burned, duration, and percent severely burned (Kendall tau correlation, $p<0.001$ for all), thus assessments of univariate normality were not sufficient to determine distribution (Newson, 2002). Each variable was transformed to make it more normal. Residuals were then plotted to further determine goodness-of-fit to a Gaussian distribution. Statistical tests showed a significant difference from normal (Shapiro-Wilk,

$p<0.01$). However, the plotted data reasonably fit a Normal Q-Q plot and its histogram followed a normal distribution relatively well. It was thus used in a traditional Mahalanobis distance calculation with caution.

Standard outlier values found via Mahalanobis distance are at $p<0.001$. The ranking abilities of Mahalanobis distance used on non normal data are acceptable and cutoffs can be made arbitrarily. These were chosen as and $p=0.01$. Mahalanobis distance finds extreme values at both ends of the distribution. Since this study sought high severity, large fires, fires which were less than the sum of the mean and one standard deviation were removed as well as those with a percent severely burned less than the sum of the mean and one standard deviation. The two variable Mahalanobis distance (area burned and percent burned severely) found 7 outliers ($p<0.001$) and 15 extreme values. Three variable Mahalanobis distance found 10 outliers and 20 extreme values. Most of the fires identified by both the two variable and the three variable Mahalanobis distance were also found to be extreme fires by Lannom *et al.* (2014) (Table 1). Those that were not were almost exclusively medium sized fires that had not burned near WUI areas. The number of extreme events per year and total hectares burned per year were plotted (Chart 1). There was no significant trend in either.

Conclusions

Mahalanobis distance is a useful method in classifying and ranking extreme fires. It can be used on data that are normal or on those that are not normal while still providing relevant results. Subsequent analysis could include economic losses, distance to urban areas, encroachment into sensitive areas, and fire return interval. It is not currently possible to say that fires are becoming more extreme or that extreme fires are becoming more common. One of the biggest reasons for this is the short time series the data are over. A longer time series of data would allow long term trends to show. There is also an inherent lack of sample size when looking at time series of extreme events. If the threshold for 'extreme' is too low, the events can hardly be considered extreme. However, with a high threshold to truly qualify an event as extreme, very few events will be in the data set. More years of data are thus necessary to determine whether or not the frequency extreme wildfire events are increasing.

References

Abt, K. L., Prestemon, J. P., & Gebert, K. M. (2009). Wildfire suppression cost forecasts for the US Forest Service. Journal of Forestry, 107(4), 173-178.

Altman DG, Bland JM. Statistics notes: the normal distribution. Bmj. 1995;310(6975):298.

Calkin, D. E., Gebert, K. M., Jones, J. G., & Neilson, R. P. (2005). Forest Service large fire area burned and suppression expenditure trends, 1970–2002.Journal of Forestry, 103(4), 179-183.

Ekstrom, J. (2011a). Mahalanobis' distance beyond normal distributions.¨ UCLA Statistics Preprints, 624.

Elliott AC, Woodward WA. Statistical analysis quick reference guidebook with SPSS examples. 1st ed. London: Sage Publications; 2007.

Gebert, K. M., Calkin, D. E., & Yoder, J. (2007). Estimating suppression expenditures for individual large wildland fires. Western Journal of Applied Forestry, 22(3), 188-196.

Ghasemi A, Zahediasl S. Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. Int J Endocrinol Metab. 2012;10(2):486-9. DOI: 10.5812/ijem.3505

Liang, J., Calkin, D. E., Gebert, K. M., Venn, T. J., & Silverstein, R. P. (2008). Factors influencing large wildland fire suppression expenditures. International Journal of Wildland Fire, 17(5), 650-659.

Mahalanobis, Prasanta Chandra (1936). "On the generalised distance in statistics" (PDF). *Proceedings of the National Institute of Sciences of India* 2 (1): 49–55.

Mardia, K. V. (1970), Measures of multivariate skewness and kurtosis with applications. Biometrika, 57(3):519-530. Mardia, K. V. (1974), Applications of some measures of multivariate skewness and kurtosis for testing normality and robustness studies. Sankhy A, 36:115-128.

Moritz, M. A. (1997). Analyzing extreme disturbance events: fire in Los Padres National Forest. Ecological Applications, 7(4), 1252-1262.

Newson R. Parameters behind "nonparametric" statistics: Kendall's tau,Somers' D and median differences. Stata Journal 2002; 2(1):45-64.

Oztuna D, Elhan AH, Tuccar E. Investigation of four different normality tests in terms of type 1 error rate and power under different distributions. Turkish Journal of Medical Sciences.2006;36(3):171–6.

Pallant J. SPSS survival manual, a step by step guide to data analysis using SPSS for windows. 3 ed. Sydney: McGraw Hill; 2007. pp. 179–200.

Strauss, D., Bednar, L., Mees, R. 1989. Do one percent of the forest fires cause ninety-nine percent of the damage? Forest Science 35:2. 10p.

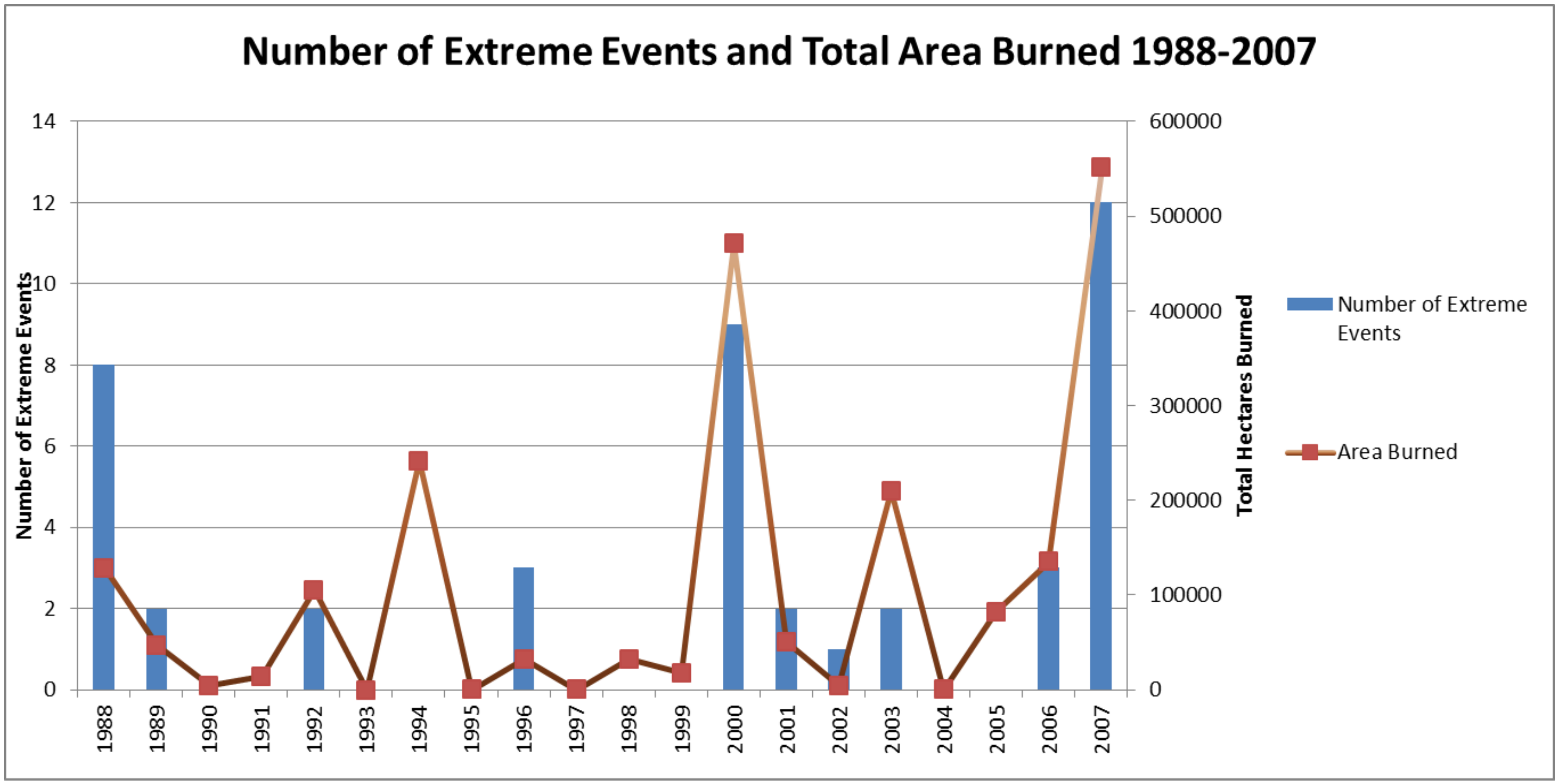
Tabachnick, B. G., & Fidell, L. S. (2001). Using multivariate statistics.

Table 1 Showing those fires classified as 'extreme' for each method. Mahalanobis distance fires are ranked by severity. Those fires which are extreme but not outliers are denoted with an asterisk (*).

Fire Name	Year	Lannom *et al.* 2014	2 Variable Mahalanobis Distance	3 Variable Mahalanobis Distance	Area (ha)	% Burned Severely	Duration (days)
Davis	2003	1	5	7	8369.93	57.59	126
Canyon Creek	1988	2	2	5	67936.6	49.83	108
Lincoln Complex (Snowbank)	2003	3	11*	19*	14968.1	48.05	92
Dooley Mountain	1989	4			7586.68	42.60	93
Cascade Complex (Monumental)	2007	5	3	6	128348.8	37.27	107
Grizzly Complex (Winter)	2002	6			13754.1	38.08	112
Fool Creek	2007	7	13*	2	21817.5	45.04	155
North Fork	1988	8	1	1	228694	45.20	115
Mussigbrod Complex (Mussigbrod)	2000	9	8*	12*	26227.76	48.95	92
Little Blue	2000	10			1832.72	52.78	103
East Zone Complex (Raines)	2007	11	7	8	128983.3	29.47	112
Lake Creek	1988	12			229047.7	20.46	106
Murphy Complex	2007	13			229622	12.87	97
Flossie Complex	2000	14		16*	36669.1	29.80	115
Sawmill Complex (Wyman #2)	2007	15			27614.09	28.39	112
Canyon Ferry Complex (Cave Gulch)	2000	16			10925.1	41.56	66
Jungle	2006	17			9650.21	43.82	60
Red Bench	1988	18			13591.9	40.90	26
Trail Creek	2000	19			13576.3	34.56	60
Storm Creek	1988	20					
Tower	1996	21					
Ahorn	2007	22		3	18466.9	40.98	155
Fridley	2001	23	6	11*	11159.2	56.24	98
Sweet-Warrior Complex	1996	24			15982.9	39.07	91
Snowshoe	2001	25			9504.4	37.70	98
Red Eagle	2006	26	10*	15*	13177.9	50.97	83
Confluence Complex (Clear	2007	27			8440.11	8.63	117

Sage)							
Meriwether	2007	28			7634.34	29.12	104
Porphyry South	1992	29			52547.9	14.43	82
Poe Cabin	2007	30			24161.5	19.26	87
Burgdorf Junction	2000	31			27229.05	23.53	82
Canal	1989	32			7525.36	28.4	77
Upper Nine Mile Complex (Nine Mile)	2000	33			7970.15	19.76	85
Porcupine Creek	1992	34			7207.12	3.00	78
Burnt Flats	2000	35			3683.86	0	35
Rosa	1985	36			7867.31	0	2
Ringer	1986	37			9733.25	0	2
Coyote Butte	1996	38			12394	26.73	2
Clover	1988		4	4	143288.7	31.94	12
Diamond Complex (Diamond Peak)	2000		9*	10	109816	29.68	115
Corbin	1985		12*	9	12934.5	48.48	5
Skyland	2007			13*	16687.9	24.34	134
Canyon Creek (2)	1988		14*	14*	10720.1	1.07	75
Shower Bath Complex (Red Bluff)	2007			17*	20277.05	27.74	125
Sailor Cap	2006			18*	25373.3	34.96	8
Shower Bath Complex (Shower Bath)	2007			20*	20625.6	32.80	121
Sliver Creek	1988		15*				

Chart 1 The number do extreme events and total hectares burned per year. No significant trend was found.



Chapter 3

Accuracy of WAAS-enabled GPS-VHF warning signal when crossing a terrestrial geofence

Abstract

Geofences are geographic coordinate based virtual boundaries that can be static or dynamic, when combined with global positioning system (GPS) transmitters, they provide a powerful tool for monitoring the location and movements of objects of interest through proximity alarms. However, the accuracy of geofence alarms in a GPS-VHF transmitter receiver system has not been tested. To achieve these goals, a cart with a GPS-VHF transmitter was run on a straight path at three levels of cart speed, angles of the geofence to the track, and distances of the receiver from the track. A series of 81 trials showed that, at the $\alpha=0.10$ level, angle and receiver distance affect geofence alarm accuracy ($p=0.054$ and $p=0.000$, respectively). With the shortest receiver distance and largest geofence crossing angle resulting in the worst and best accuracies respectively.

Introduction

Geofences are virtual boundaries marked by global positioning system (GPS) coordinates, typically set either as a circle of a given radius from a central point, which may or may not be dynamically mobile, or as a polygon whose vertices are predetermined by an operator. Many disciplines now use GPS tracking to monitor objects or organisms of interest, especially in the United States, devices equipped with Wide Area Augmentation System (WAAS) differential correction are considered accurate to within 3 m [1]. Geofences coupled with GPS tracking have been deployed in remote monitoring of sites for security purposes, tracking patients with Alzheimer's disease, wildlife encroachment onto farmland, alerting to the escape of prisoners, ensuring children stay in a safe area, creating security boundaries for wireless signals, transportation management, and tagging animals covering large ranges in remote locations [2,3,4,5,6]. An emerging application area is forestry, where multi transmitter GPS systems may be useful for logging safety, boundary and silvicultural marking, controlling herbicide applications, and production and cost tracking [7, 8].

Despite its pervasive use, the accuracy of geofence alerts have never been investigated in a replicated, designed experiment. This objective of this study is to determine the temporal accuracy of a commercially available GPS-VHF unit's geofence alert system. Specific research questions we sought to address include an assessment of how dependent the accuracy is on (i) the speed of the tracked object, (ii) the angle of the object to the geofence, and (iii) the distance from the receiver.

Methods

A series of 81 trials involving a specialized personal recreational vehicle (PRV) designed to run on a customizable track were performed at the University of Idaho Forest Operations Laboratory in Princeton, Idaho (46.9135 N -116.8325W) to determine which factors affect the accuracy of a geofence crossing signal. Trials replicated speeds of 5, 10, and 15 kilometers per hour (kph), receiver angles of 30, 60, and 90 degrees, and receiver distances of 4, 100, and 400

m. A total of 81 trials were carried out, 27 in each variable class and 3 in each replicate. The track was leveled to less than 10 cm height difference along the 120 m extent and aligned to the magnetic West/East axis using a total station. The track was flagged at 10 m increments. A WAAS-enabled GPS-VHF transmitter (TT 15, Garmin, United States) was attached via its collar to the front bumper of the PRV, centered on both the bumper and the track approximately 15 cm above the track. The handheld receiver with extended antenna (Alpha 100, Garmin, United States) was attached via a cardboard holder and zip-ties at a height of 1.58 m to a PVC pole driven into the ground at a set 5 m distance perpendicular to the 80 m point along the track, as measured by a total station. The receiver and transmitter were separated only by unobstructed flat land to avoid signal interruptions or reroutes that could affect the transmission times. The geofence was centered on the track at 80 m, and geofence angles were sighted using a total station [9]. For each angle, the end points of the geofence were set to 20 m from the track (Figure 1). Because the Garmin Alpha handheld unit requires a polygonal or radius geofences, and not lines, the extraneous points were set 20 m out, 10 m before the start of the track. Points were added to the geofence in the receiver as an average of three waypoints at each location. A calibrated speedometer was mounted to the PRV to provide speed feedback to the driver and was validated through hand timing.

Trials began with an audio signal given when the front bumper of the PRV crossed the first 10 m mark, providing the vehicle 10 m to reach a consistent speed, and ended when the vehicle crossed the 80 m mark completing the 70 m run. A person standing at the transect through the track path manually timed the duration of the 70 m run, with another timing the first 40 m to verify speed over the course of the trial. A third person standing by the receiver timed duration from when the signal was heard to when the geofence crossing alarm sounded on the receiver. Trials were only run in the West to East direction, with the PRV removed from the track and transported via truck back to the starting end for each trial. The time difference between the manual timing at the track and the time recorded with the geofence alarm was computed for each trial as the global dependent variable, were all statistics are reported in seconds. All statistical tests were evaluated at a significance level of α=0.10.

Results

At the α=0.10 level, receiver distance and geofence angle significantly affected the time difference. Figure 2 shows that the time difference was largest for the 4 m distance (Mean=3.225, SE=0.3395) and relatively similar for the 100 m (Mean=1.762, SE=0.2785) and 400 m (Mean=2.153, SE=0.4050) distances. Time difference was approximately equal for 5 kph, 10 kph, and 15 kph at Mean=2.041 (SE=0.4515), Mean=2.639 (SE=0.2494), and Mean=1.904 (SE=0.1452), respectively. Geofence angle exhibited the largest variation in standard error of the variables. The trials at 30° were more accurate (Mean=1.631, SE=0.3438), but had a higher standard error than those at 60° (Mean=2.682, SE=0.3428) or 90° (Mean=2.272, SE=0.2040), which were about the same. This was supported by the comparisons of means done in Table 1. Trials run at the 30° geofence showed a lower mean time difference than those at 60° or 90°, though a larger standard error. The proximity of the 30° geofence to the approaching side of the track may have allowed the geofence alarm to be tripped earlier than for the other geofences whose angles did not result in such a close proximity. Because the geofence was only closer to the track and did not alter the true location of the crossing, the earliness of the alarm was more variable than in trials where geofence alignment placed a greater distance between the track and the approaching transmitter. The increase in time delay between trials at the 4 m receiver difference, compared with those at 100 m and 400 m is likely a resultant artifact in how the VHF signal is sent or received. It may thus be beneficial to conduct a more detailed analysis over a range of relatively short distances in order to determine when the artifact appears, and what minimum threshold distance may be needed in order to avoid it. It is possible that, given the rate at which VHF signals travel, the artifact is the only cause of the differences in mean timing of the distances, and that distance does not in fact play a role in the accuracy of the geofence signal above some minimum distance. The full factorial analysis of variance showed that the interaction between variables was not significant.

Discussion

The higher error observed at close distances has important implications for the applications of geofence alerts characterized by loud motorized machinery such as in construction sites or in natural resources applications including logging forests. In operational forestry an equipment operator could have the receiver in the cab of a motorized vehicle and the excess delay caused by the time for the signal transmission could allow the vehicle to travel a notable distance through a geofence boundary before being alerted. For example, at 15 kph, the approximately 1.5 second extra delay would cause an operator to move a maximum of 6.25 m past the geofence before being alerted. Not taking into account reaction time and the time needed to stop a piece of heavy machinery, this could easily result in the disturbance of a riparian or sensitive zone, or movement of a harvester well beyond a harvest unit boundary. In safety settings, these implications are even more critical. The earlier warning for the 30° geofence indicates that the crossing signal does not rely solely on the exact crossing moment, but rather is dynamically sensitive to the proximity of the GPS to the geofence. This sensitivity could trigger frequent false alarms when traveling near-parallel to the geofence.

The TT 15 transmitters currently must be mounted outside of an equipment cab and only allow for a vibration or an electric shock to alert that a geofence has been crossed. As the receivers are rather small and are unable to communicate with each other, this makes it currently difficult to implement on site. However, should the information be fed directly into the onboard computers that are currently prevalent in agricultural and forestry machinery, the preexisting monitors and hardware could easily also be modified to run GPS and geofence applications.

Conclusions

This study has shown that several factors must be considered when deploying geofence alert systems in high precision and short duration applications, such as those that occur in operational forestry. Geofence angle and the distance between the receiver and transmitter affect the accuracy of geofence alarms, though the latter must be investigated more thoroughly to determine if this is only so at very close distances. These known time delays could be built into geofence polygons to provide additional buffering for known sensitive and safety areas.

The outcomes of this research could be widely used in natural resources from monitoring polygons of sensitive site boundaries associated with harvest units to keeping personnel and equipment out of machine paths. Many forest operations sites are often poorly marked and new personnel are unfamiliar with all inherent risks on site. Accidents caused by thoughtless mistakes like walking through an active cable corridor could be avoided if an alert were sent that a worker was too close to the machinery. Additionally, accidental site boundary crossings could be prevented through the tracking of machinery and hand-fallers around a site marked with geofences. In particular, a forester or operator could easily upload a sale shapefile and use the polygon data as a geofence. Sensitive sites or riparian zones could be walked as well, with the stream classification used to determine how far from the walked path the geofence should be set. With transmitters on each piece of equipment and each person on site, anyone with a receiver would be warned if a geofence was crossed.

References

1. United States Department of Transportation. Specification for the Wide Area Augmentation System (WAAS). 2001.

2. De Lara E, LaMarca A, and Satyanarayanan M. (2008). *Location Systems: An Introduction to the Technology Behind Location Awareness*. Morgan & Claypool Publishers. p. 88. ISBN 978-1-59829-581-8.

3. Ensing EP, Ciuti S, de Wijs F, Lentferink DH, ten Hoedt A, Boyrce, MS, Hut, RA (2014) GPS Based Daily Activity Patterns in European Red Deer and North American Elk (Cervus elaphus): Indication for a Weak Circadian Clock in Ungulates, Plos One, 9,9, e106997.

4. Merill SB, Adams, LG, Nelson ME, Mech LD (1998) Testing releasable GPS radiocollars on wolves and white-tailed deer, Wildlife Society Bulletin, 26, 4, 830-835.

5. Yuce YK and Gulkesen KH. CaregiveNet: A novel social support intervention for locating and securing wandering Alzheimer's patients as soon as possible. *Wireless Communications and Mobile Computing Conference (IWCMC), 2013 9th International*. 2013, 1405-1411.

6. Douglas-Hamilton I, Krink T., Vollrath, F. (2005). Movements and corridors of African elephants in relation to protected areas, Naturwissenschaften, 92, 4, 158-163.

7. Sheth A, Seshan S, and Wetherall D. Geo-fencing: Confining Wi-Fi coverage to physical boundaries. *Pervasive Computing: Lecture Notes in Computer Science*. 2009, 5538, 274-290.

8. Tarnauca B, Puiu D, Nechifor S, Comnac V. Using complex event processing for implementing a geofencing service. 2013 *IEEE 11th International Symposium on Intelligent Systems and Informatics (SISY)* 391-396.

9. Witte TH and Wilson AM. Accuracy of WAAS-enabled GPS for the determination of position and speed over ground. *J of Biomech.* 2006, 38(8), 1717-1722.

Table 1.1: ANOVA with post hoc Tukey's HSD tests run at $\alpha = 0.10$

ANOVA with Post hoc Tukey's HSD Test		
Variable		p-value
Speed		0.209
5 kph	10 kph	0.363
5 kph	15 kph	0.947
10 kph	15 kph	0.219
Angle		0.054
30°	60°	0.044
30°	90°	0.300
60°	90°	0.609
Receiver Distance		0.000
4 m	100 m	0.001
4 m	400 m	0.000
100 m	400 m	0.910

Figure 1.1: Experimental set up. The track, in red, runs along the magnetic North South axis. Trials were exclusively run in the direction of the arrow (N-S).

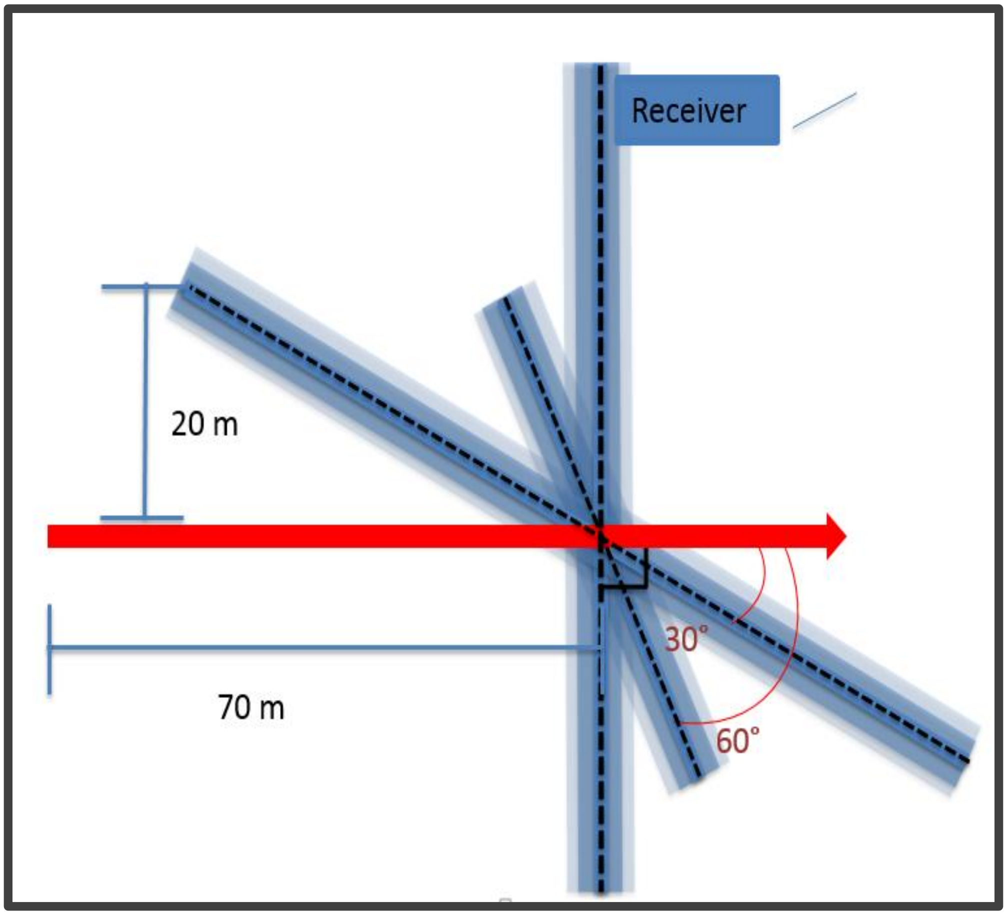


Figure 1.2: Box plot showing the time difference between the actual run time and the time as measured with the geofence crossing signal. Twenty-seven runs were completed in each variable class.