Data-enabled Correlation Analysis between Wildfire and Climate using GIS

Weiwei Xie
Department of Computer Science
Mississipi State University
Mississippi State, MS
wx38@msstate.edu

Mingyan He
Information Engineering College
Dalian University
Dalian, China
mingyan321@126.com

Bo Tang
Department of Electrical and Computer Engineering
Mississippi State University
Mississippi State, MS
tang@ece.msstate.edu

***Abstract*—This paper presents a study of examining the statistical correlation between wildfire and weather by mining historical spatial and temporal wildfire and climate data. Large wildfires have been recently becoming more frequent, intense and destructive in the West of United States. The occurrence of wildfires can be determined by many human and natural factors, such as the availability of fuels, physical settings, and weather conditions, among which weather is of great interest and importance for wildfire forecasting. The availability of landscape fire data sets and weather data sets now enables the analysis of correlation between wildfire and weather which indicates the possibility of wildfire for given weather conditions in one region. This paper investigates the relation between wildfire and drought conditions in California and visualize the results using geographic information system (GIS) computing technology. Our data analysis findings show a high correlation between the normalized number of wildfires per forest unit area and drought severity, illustrating the potential of forecasting wildfire using weather data.**

Keywords—Spatial and temporal data analysis; data visualization; GIS; wildfire forecasting

I. Introduction

In recent decades, large wildfires have become more frequent, more intense, and more destructive in the West of United State. During the three-year period between 2011 and 2013, we can see the largest fire in Arizona's history (2011's Wallow Fire), both the largest and most destructive fires in New Mexico's history (2012's Whitewater-Baldy Complex and Little Bear Fires, respectively), the most destructive fire in Texas's history (2011's Bastrop County Complex Fire), and the third largest fire in California's history (2013's Rim Fire). Wildfires are also very costly. For example, in 2018's wildfire season, more than 8 million acres have been burned, costing a staggering $24 billion primarily from the destruction of homes and public facilities, as well as firefighting. Meanwhile, the wildfire suppression cost is becoming more and more expensive. From General Accounting Office, the total wildfire funding for 2008 was a record high of $4.46 billion, while it was $1 billion for 2003 and $300 million for 1990.

Given the increasing high cost of fighting wildfires, it is of great importance to understand the leading factors of wildfires for wildfire forecasting and resource management. More recently, data-driven wildfire analysis [1-3] has been receiving increasing attentions for real-time or near-real-time wildfire forecasting, which aims to explore the relation between the occurrence of wildfires and other potential factors.

The process of wildfire is complex, but it is always determined by its environment [4][5], including vegetation type, atmospheric components and motion. More specifically, in addition to ecosystem, wildfire can be also related to many factors in climate, including temperature, atmosphere moisture, atmosphere stability, winds, clouds and precipitation, and fuel moisture, over a long period of time. These factors, driven by climate, are combined to influence the growth of vegetation in a given place and its burning in wildfire.

It is well-known that the wildfire would be determined by the weather condition – drought [6]. Drought is the key factor to determine how many wildfires will be ignited in that year and how much land will be burned. Drought is also closely related to the temperature, one of main wildfire's factors. Drought evaporates the water from surface, and the dry land slow down the cooling process in the air, resulting in the increasing of temperature. With the global warming, the rise of global temperature also influences the timing of seasonal events: such as the seasonal timing of snowmelt which further determines the moisture or dry of land and hence the severity of fire season.

Thanks to the crowdsourcing and information technologies, both wildfire occurrences and weather conditions can be monitored and reported in real-time or near-real-time manners. Analysis of these spatial and temporal data enables us to investigate which weather factors significantly lead to the wildfire so that it is impossible to achieve real-time wildfire forecasting to reduce resource management costs in the government and public property loss due to the unexpected wildfire. This paper studies the use of these datasets that are publicly available and investigate the correlation between wildfire and drought

978-1-7281-7283-5/20/$31.00 ©2020 IEEE
DOI 10.1109/ICICT50521.2020.00013

conditions. In particular, we use the case of California and utilize geographic information system (GIS) to extract numerical features from raw datasets and visualize our geo-spatial results. To measure the drought condition, we employ a modified Palmer Drought Severity Index (PDMI) which incorporates a weighted average of the wet and dry index terms. Our qualitive and quantitative results show a strong correlation between PDMI and wildfire occurrence, which indicates the great potential of using PDMI for wildfire forecasting.

The remaining of this paper is organized as follows. Section II gives a brief summary and literature review of data-enabled wildfire analysis and prediction. Section III describes the wildfire and weather condition datasets that have been used in this study. Section IV presents the correlation analysis methods that we have employed to analyze the correlation between wildfire and drought condition. Section V shows our correlation analysis results using various correlation measures and data and results visualization using GIS tools. Section VI gives the conclusion and discussion of our future works.

II. Background and Literature Review

A. Background

Climate has a long effect on weather condition pattern. For example, in western North America, the climate, resulted by the circulation of warmer and cooler water in the eastern Pacific Ocean, leads to the high-pressure ridges and low-pressure troughs. The air movement across the land also forms all kinds of storms at specific locations, for example, the air movement over hills. The storm can bring precipitation sometimes, but the lighting without precipitation can ignite wildfires. Researchers have found that the ocean cycles in Pacific Ocean are roughly periodic, that is the oscillation of El Nino/Southern Oscillation (ENSO) and Pacific Decadal Oscillation (PDO). In ENSO, there are two extremes in a 3 to 7 years cycle of weather condition: El Nino and La Nina. While El Nino brings warm and wet weather condition to the southern US, and cold and dry weather condition to Northwest, La Nina flops these effects [6]. The PDO provides a period of 20 to 30 years weather condition pattern. Positive phases of PDO usually have warm springs and dry summers in the Pacific Northwest [7].

To understand the interaction between fire and climate, it is important to understand the history of fire with different types of vegetation and climate factors. The climate can not only affect the fire frequency, size and intensity, but also affect the ecosystem of fire areas, such as the vegetation types and structures. In this paper, we attempt to map the wildfire frequency and weather and climate characteristics in United State, and examine the relationship between wildfire frequency and one of climate factors, i.e., drought conditions, for different locations.

B. Literature Review

GIS has been widely used to analyze and visualize spatial and temporal data, including wildfire management. GIS [9][10] not only provides a mechanism to visually display information, such as the location of a wildfire and its associated weather conditions, but also offers useful tools for spatial and temporal analysis, such as overlay, aggregation, interpolation, segmentation, and so on. Studies in wildfire forecasting mainly rely on GIS-based computing technologies. For example, Riva et al. [10] employed an overlay approach to convert point wildfire data to area data and used kernel density interpolation to address spatial uncertainties to improve wildfire mapping accuracy at regional scale.

Meanwhile, the study of understanding the impact of climate to the wildfire can be traced back to few decades ago. Brenner [11] presented a strong evidence for a relationship between sea surface temperatures in the Pacific Ocean and wildfire occurrence in Florida, due to the El Nino-Southern Oscillation (ENSO). Heilman [12] employed empirical orthogonal function analysis and identify atmospheric circulation patterns associated with severe wildfires in six different regions of the United States. Recently, Goodrick and Hanley [13] found the relationship between monthly ENSO indices and area burned by Florida wildfires. Studies have also shown strong impacts of climate changes on wildfire. Westerling and Bryant [14] reported the increase in western U.S. forest wildfires to warmer spring and summer temperatures and reduced precipitation, which are projected to continue under climate change scenarios.

To understand the statistical relationship between two random variables, correlation analysis plays a key role in wildfire analysis. We have seen both linear and nonlinear correlation analysis have been widely used in literature [10-15]. Unlike previous research, this present paper aims to investigate the impact of drought conditions measured by PMDI on the wildfire activities using long-term historical wildfire and climate datasets.

III. Correlation Analysis Metrics

Correlation analysis is an important statistical method that has been widely used to evaluate the strength of relationship between two variables. To understand the type of relationship that exists between wildfire and drought severity, we let the random variable X denote the number of forest wildfire, and the random variable Y denote the Palmer Modified Drought Index (PMDI) which can be used to measure the drought severity. In this paper, the following two correlation measures between two quantitative variables X and Y will be calculated to determine the degree of their correlation.

A. Pearson's Product Moment Correlation Coefficient

The Pearson's correlation coefficient [15] is used to quantify linear relationships between two random variables. Let x and y be the quantitative measures of two random variables X and Y, each of which has n samples. Mathematically, the Pearson's correlation coefficient r can be given by:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

where

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ and } \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

In other words, the correlation coefficient r can be also written as the ratio of the sample covariance of two variables to the product of their standard deviations s_x and s_y:

$$r = \frac{Cov(x, y)}{s_x s_y}$$

From the above definition, it is known that the correlation measurement r ranges between -1 and +1, where $r>0$ indicates a positive relation, i.e., a higher value of one variable is associated with a higher value of the other variable, and $r<0$ indicates that a negative relation, i.e., a higher value of one variable is associated with a lower value of the other variable. The value $r=0$ indicates absence of any association between two variables.

B. Spearman's Rank Coefficient

The Spearman's correlation coefficient [15] evaluates the monotonic relationship between two random variables. The monotonic relationship means that two random variables could have different changing rates. It computes the correlation between the rank of two variables. Given n measurements for each of two variables, we first calculate their ranks denoted by x' and y', and the Spearman's rank correlation coefficient r is given by:

$$r = \frac{\sum_{i=1}^{n}(x_i' - \bar{x}_i')(y_i' - \bar{y}_i')}{\sqrt{\sum_{i=1}^{n}(x_i' - \bar{x}_i')^2}\sqrt{\sum_{i=1}^{n}(y_i' - \bar{y}_i')^2}}$$

In other words, this is a rank-based version of the above Pearson's product moment correlation coefficient, ranging from -1 to +1. For a strong monotonically increasing correlation, we have positive values of all correlation coefficients at the same time. But unlike the Pearson's correlation coefficient, the Spearman's rank coefficient equals to 1 for both linearly and not linearly correlated variables. An alternative equation to measure the Spearman rank correlation is:

$$r = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}$$

where d_i is the difference between the two ranks of each observation.

IV. DATA COLLECTION

In this paper, we examine the impact of climate on the occurrence of wildfires from historical wildfire and climate data sets that are available online. In particular, we use wildfire data that are collected by the LANDFIRE project and the Fire Program Analysis system, and extract climate data from NNDC Climate Data Online and the U.S. Wind Climatology, which are detailed in the rest of this section.

A. Wildfire Data

LANDFIRE: Landscape Fire and Resource Management Planning Tool (LANDFIRE) project provides comprehensive data including landscape change, disturbance, vegetation, fuel and fire regimes across the United State, provided by the United States Department of Agriculture, Forest Service. LANDFIRE was initiated as the need for consistent national geospatial data to support prioritization of hazardous fuel reduction, ecological conservation activities, and strategic resource management initiatives, fire management planning, as well as stewardship of public and private lands, and natural resource management. Because the LANDFIRE databases and models were developed and implemented nationally over the past decade, the related products were very useful for a broad range of purposes. The existing well known products include fire management, climate change research, carbon sequestration planning, and eco-regional assessments [1]. Moreover, the LANDFIRE data is periodically updated by agencies for the entire United States, which ensures the availability of both current and historic data and continues to improve the quality of data products into the future.

Based on the LANDFIRE data, we can first map the vegetation, fuels and wildfires, which can help us to understand the relationship between wildfire and ecosystem (landscape structure, vegetation type, and land composition). Because the integration of vegetation, fuels and wildfire events in LANDFIRE, the mapping of historic fire pattern related to the ecosystems becomes easy. Using the biophysical layer, succession layer, and landscape layer, the fire severity and frequency can be calculated. Moreover, we can obtain the Fire Regime Condition Class (FRCC) to measure the degree of departure of current vegetation from the historical vegetation reference conditions. We show the FRCC map in Fig. 1, using FRCC mapping tools.

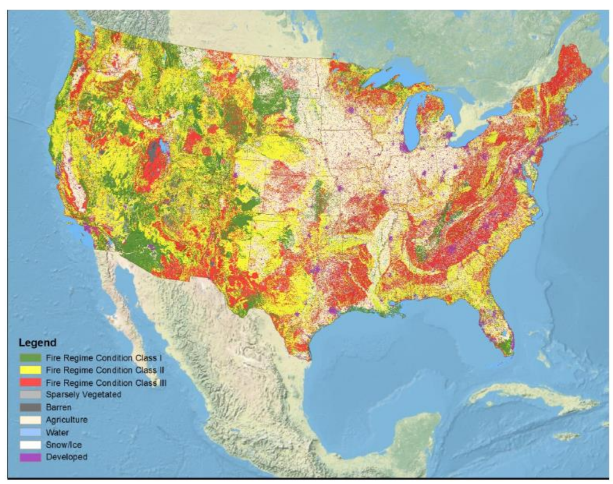

Figure 1. Existing vegetation as mapped from LANDFIRE data (from Sommers et al. 2009).

Fire Program Analysis (FPA) system: While LANDFIRE has a resolution of 30 meters, FPA's FOD (Fire Program Analysis fire-occurrence database) includes nearly 1.6 million geo-referenced wildfire records, representing a total of 113 million acres burned in the United States during the 20-year period from 1992 to 2011. In this database, the wildfire records were acquired from the reporting systems of federal, state, and local fire organizations. The data elements of records in this database include discovery date, final fire size, and a point location at least as precise as Public Land Survey System (PLSS) section (1-square-mile grid). Basic error-checking was performed and redundant records were identified and removed, to the degree possible. Using this database, we show the map of fire size in different states at this period in Fig. 2.

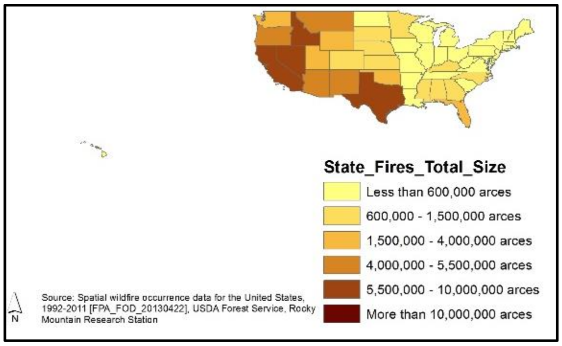

Figure 2. Total burning size during 1992 to 2011 in US.

B. Climate Data

<u>*NNDC Climate Data Online (CDO):*</u> The Climate Data Online dataset includes many measurements of climate, including drought, temperature, precipitation, and so on. In our project, we aim to measure the relationship between wildfire and drought, using the case of California. There are several indices to measure the drought. The common one is the Palmer Drought Severity Index (PDSI) that measures the duration and intensity of the long-term drought-inducing circulation patterns [6]. Instead of using PDSI, we use the measurement of Modified Palmer Drought Severity Index (PMDI). PMDI and PDSI have the same value during an established drought or wet spell but the PMDI incorporates a weighted average of the wet and dry index terms. The state of California includes seven climate divisions.

<u>*Wind Climatology:*</u> The U.S. Wind Climatology provides spatially and temporally continuous wind climatology for the contiguous U.S. on a monthly basis from January 1950 to present. It is a product of National Centers for Environmental Information in National Oceanic and Atmospheric Administration (NOAA). The wind speed and direction is described by U and V components. By drawing the direction of wind at a specific time, we can show how the wind influences on the wildfire development.

V. RESULTS ANALYSIS

This section summarizes our data-enabled correlation analysis results. For the given raw wildfire and climate data sets, we perform data preprocessing using GIS. For a time period, the number wildfires is extracted from these wildfire datasets and the PMDIs will be calculated from climate datasets to measure drought severity. Once these measurements of interest are extracted, we perform correlation analysis to study the degree of their relationship. Using GIS, we also visualize the impact of winds on the development of a wildfire. In particular, we use wildfires in California from 1992 to 2011 as a case study which can be easily extended to other places of different scales.

A. Data Preprocessing Results

The state of California includes seven climate divisions. We first extract the normalized number of wildfires in each division using aggregation tools in GIS, and computes their drought severity scores, i.e., PMDIs. The normalized number of wildfires is defined as the total number of wildfires per forest area. The wildfire and PMDI distribution are shown in Fig. 3 and 4 over these seven divisions. From these two maps, a clear reverse pattern between these two variables can be observed.

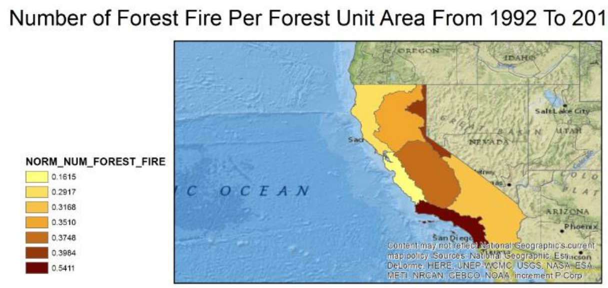

Figure 3. Number of forest fires normalized by the forest area for each climate division of CA from 1992 to 2011, reflecting the fire frequency.

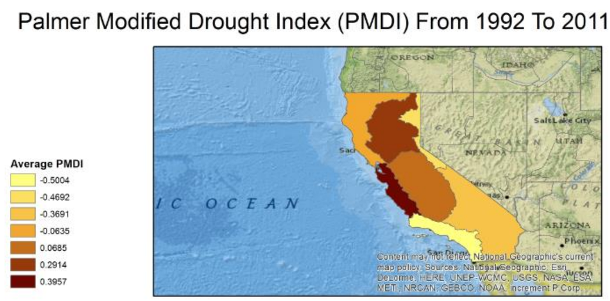

Figure 4. PMDI of each climate division in CA averaged over the years from 1992 to 2011. The negative PMDI values denote dry spells, and the positive values denote wet spells.

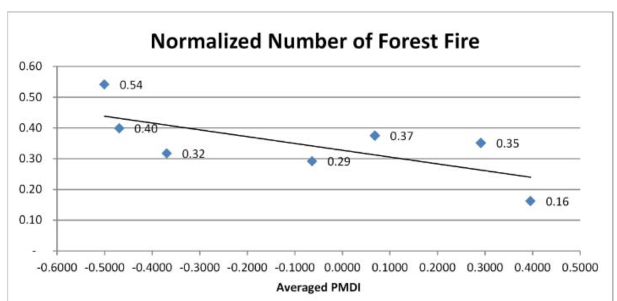

Figure 5. PMDI of each climate division in CA averaged over the years from 1992 to 2011. The negative PMDI values denote dry spells, and the positive values denote wet spells.

B. Correlation Analysis Results

While Fig. 3 and 4 qualitatively show a strong correlation between wildfire and PMDI, we further perform numerical analysis for correlation analysis. Fig. 5 draws the normalized number of forest fire in terms of averaged PMDI. Both Pearson's and Spearman's correlation coefficients are computed. According to their formulas, we compute the Pearson's correlation coefficient which has the value of -0.712 and the Spearman's correlation coefficient which has the value of -0.714. These numerical results indicate that these two variables are highly correlated in this case study.

C. Wildfire Development with Winds

It is known that fires grow with winds, but this is rarely visualized using historical wind and wildfire data. With the available wildfire and wind datasets, we show an example of October 2007 California Wildfires which is the most devastating wildfire in that year in Sothern California. Fig. 6 shows the a series of about 30 wildfires with a satellite photo showing the active fire zones, and Fig. 7 shows the pattern of strong Santa Ana winds from July to November. From these two figures, it can be seen that the wind reaches its peak in October which is consistent with the peak of wildfires. It also demonstrates that the wildfire appearing and disappearing is actually along with the development of wind.

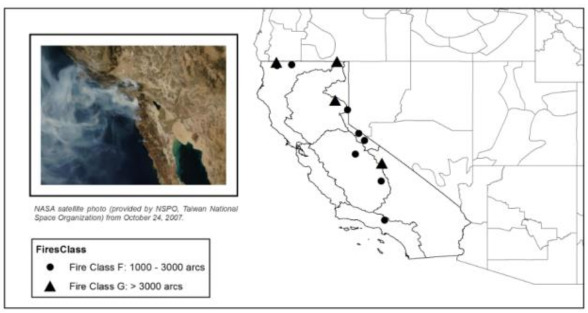

Figure 6. In 2007, a series of about thirty wildfires has taken place across Southern California on October 20, named as October 2007 California Wildfires. The left satellite photo shows the active fire zones and smoke plumes.

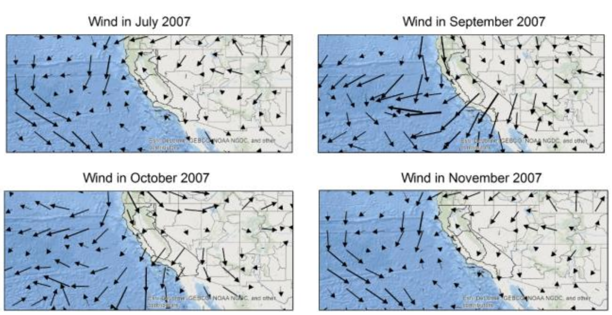

Figure 7. Wind development from July to November 2007

VI. CONCLUSIONS AND FUTURE WORKS

This paper presents a data-enabled correlation analysis to study the degree of relationship between wildfire and drought conditions from historical wildfire and climate datasets. GIS is used for spatial and temporal data preprocessing to extract the normalized number of wildfires and the PMDIs over a time period, and for data and result visualization by creating informative maps. Our qualitative and quantitative results show a highly correlated relationship between the normalized number of wildfires and the PMDIs. Using the wind dataset, we also visualize the impact of winds on the development of wildfires. While California is only considered in this study, our correlation analysis methods can be easily extended to other areas. Since a strong correlation between wildfire and weather conditions is demonstrated in this paper, for our future works, we plan to investigate the use of historical data to learn data-driven wildfire prediction models (e.g., regression approaches) for forecasting potential wildfires in one area based on its past climate and weather condition data.

REFERENCES

[1] Barber R, Huras M, Lohman G, Mohan C, Mueller R, Özcan F, Pirahesh H, Raman V, Sidle R, Sidorkin O, Storm A. Wildfire: Concurrent blazing data ingest and analytics. *International Conference on Management of Data*, pp. 2077-2080, 2016.

[2] Park J, Ko B, Nam JY, Kwak S. Wildfire smoke detection using spatiotemporal bag-of-features of smoke. *IEEE Workshop on Applications of Computer Vision*, pp. 200-205, 2013.

[3] Gunay O, Toreyin BU, Kose K, Cetin AE. Entropy-functional-based online adaptive decision fusion framework with application to wildfire detection in video. *IEEE Transactions on Image Processing*, 21(5):2853-65, 2012.

[4] Preisler HK, Westerling AL. Statistical model for forecasting monthly large wildfire events in western United States. *Journal of Applied Meteorology and Climatology*, 46(7):1020-30, 2007.

[5] Prestemon JP, Chas-Amil ML, Touza JM, Goodrick SL. Forecasting intentional wildfires using temporal and spatiotemporal autocorrelations. *International Journal of Wildland Fire*, 21(6):743-54, 2012.

[6] William T. Summers, Stanley G. Coloff, and Susan G. Conard. Synthesis of Knowledge: *Fire History and Climate Change*. 2011.

[7] Westerling AL, Gershunov A, Brown TJ, Cayan DR, Dettinger MD. Climate and wildfire in the western United States. *Bulletin of the American Meteorological Society*, 84(5):595-604, 2003.

[8] Mitchener LJ, Parker AJ. Climate, lightning, and wildfire in the national forests of the southeastern United States: 1989-1998. *Physical Geography*, 26(2):147-62, 2005.

[9] Romero-Calcerrada R, Novillo CJ, Millington JD, Gomez-Jimenez I. GIS analysis of spatial patterns of human-caused wildfire ignition risk in the SW of Madrid (Central Spain). *Landscape Ecology*, 23(3):341-54, 2008.

[10] De la Riva J, Pérez-Cabello F, Lana-Renault N, Koutsias N. Mapping wildfire occurrence at regional scale. *Remote Sensing of Environment*, 92(3):363-9, 2004.

[11] Brenner J. Southern Oscillation anomalies and their relationship to wildfire activity in Florida. *International Journal of Wildland Fire*, (1):73-8, 1991.

[12] Heilman WE. Synoptic circulation and temperature pattern during severe wildland fires. *Proceedings of the Northern Global Change Program*, vol. 214, 1996.

[13] Goodrick SL, Hanley DE. Florida wildfire activity and atmospheric teleconnections. *International Journal of Wildland Fire*. 18(4):476-82, 2009.

[14] Westerling AL, Bryant BP. Climate change and wildfire in California. *Climatic Change*, 87(1):231-49, 2008.

[15] Benesty J, Chen J, Huang Y, Cohen I. Pearson correlation coefficient. In *Noise reduction in speech processing*, pp. 1-4, Springer, 2009.