DEVELOPMENT OF A UWI DATA SET AND ITS USES

B.S. Autrey & J. Zicherman
Fire Cause Analysis, Inc.
935 Pardee St.,
Berkeley, CA 94710-2623
USA
joe@fcafire.com

ABSTRACT

Two interesting pursuits related to UWI fires are to give a spatial-temporal description of fire phenomenon (e.g. simulation, prediction, etc.) and to measure the risk posed by such fires. These two pursuits have a broad intersection because the data required in both cases are similar. There is a need for data with particular features: meteorological, topographical, vegetation, structure location, structure characteristics, ground temperatures, etc. Currently no such data exists. The construction of such a data product is explored in this paper.

OVERVIEW

In this paper we illustrate the construction of a data product to support fire research and risk modeling efforts. The need for such a data product has been broadly cited.[1,2,3] These citations are sometimes accompanied by descriptions of features the data product should have. For example, Rehm states that such a data product should have "detailed data on the topography, local meteorology, building layouts and elevations, three-dimensional distribution of fuels, and the material properties of both the natural fuels and the structures." Such descriptions serve well as guidance.

Sources of Cost

There are numerous challenges in constructing the requisite data product. Bringing the desired features into a single cohesive data product is costly. A data set providing a desired feature must be searched for, acquired, assessed, processed, and merged into the final data product. Often these data do not exist, so some of the features that are desired for the final data product are missing. There are varying formats and techniques for handling those formats. There are varying units for indexing space and time. Differing units must be transformed into common units for all component data sets. All component data sets share the characteristic that their scope is large but narrow in focus. For example, RAWS is a network of ~1,900 data collection stations for collecting meteorological data.[4] The fact that the network is ~1,900 stations makes the scope of the project large while the fact that the data collected is meteorological makes the scope of the project narrow. As a result, it is necessary to interact with multiple institutions, each with a narrow scope, in order to obtain the desired component data sets.

Guiding Principles and Ideas

There are a few guiding principles and ideas that are worth noting:

- Collect data associated with fire phenomenon: vegetation, meteorological, topographical, ground temperatures, etc.
- Collect data when it might help *define* risk: structures loss, natural resource loss, pollution, etc.
- Collect data when it might help *explain* risk: vegetation, meteorological, topographical, structure characteristics, structure location, etc.
- Disparate data sets can be linked together by their shared spatial and temporal attributes in order to build the data product.

The Old Fire

The Old Fire was chosen as the event of interest because of the high temporal resolution of the ground temperature data available for this fire. The event itself is described as follows:

> The Old Fire was a wildfire that started on October 25, 2003 in the San Bernardino Mountains… in California. It was one of at least a dozen wildfires burning around Southern California at this time… . Fanned by the Santa Ana winds, the fire burned 91,281 acres (369.4 km²), destroyed 993 homes and caused 6 deaths. The final cost of the fire was $42 million.[5]

DATA DESCRIPTIONS

A number of data sets will be used to construct the final data product. These data sets were chosen for some feature they would contribute to the final data product. The data sets were collected from a number of institutions. The data sets come in a number of types. The features that are provided by the data sets are ground temperatures, vegetation type, elevation, temperature, wind speed, wind direction, structure location, and structure status (damaged, destroyed, etc.).

Data Types

The data sets encountered assume one of the following data types:

- Pixels or Grid
- Points
- Polygon

Each data type is best explained by example.

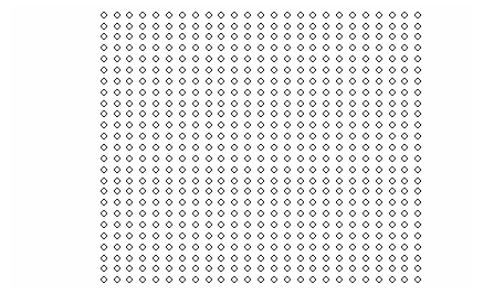

Figure 1: Example of a pixels or a grid data set.[6]

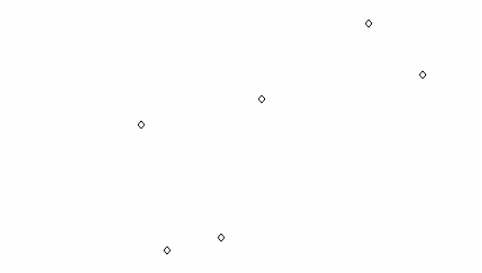

Figure 2: Example of a points data set.

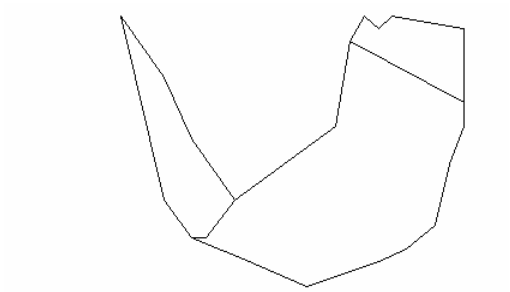

Figure 3: Example of a polygon data set.

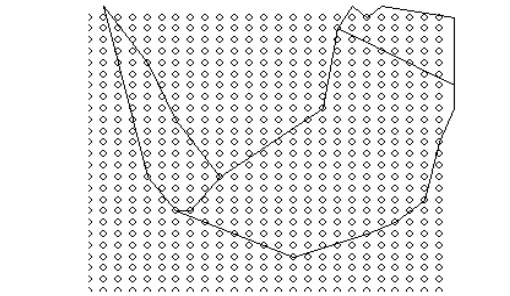

Figure 4: Example of pixels that are categorized according to their membership in a polygon.

Vegetation Data[7]

The vegetation data set is a collection of 2369 polygons. Each polygon corresponds to one of twelve vegetation classes: Agriculture (10), Basin Sagebrush (10), Chaparral (874), Coastal Sage Scrub (379), Developed (47), Meadow (32), Mixed Conifer (362), Pine Forest (15), Riparian (76), Unvegetated (157), Water (11), Woodland (391) and NA (5).

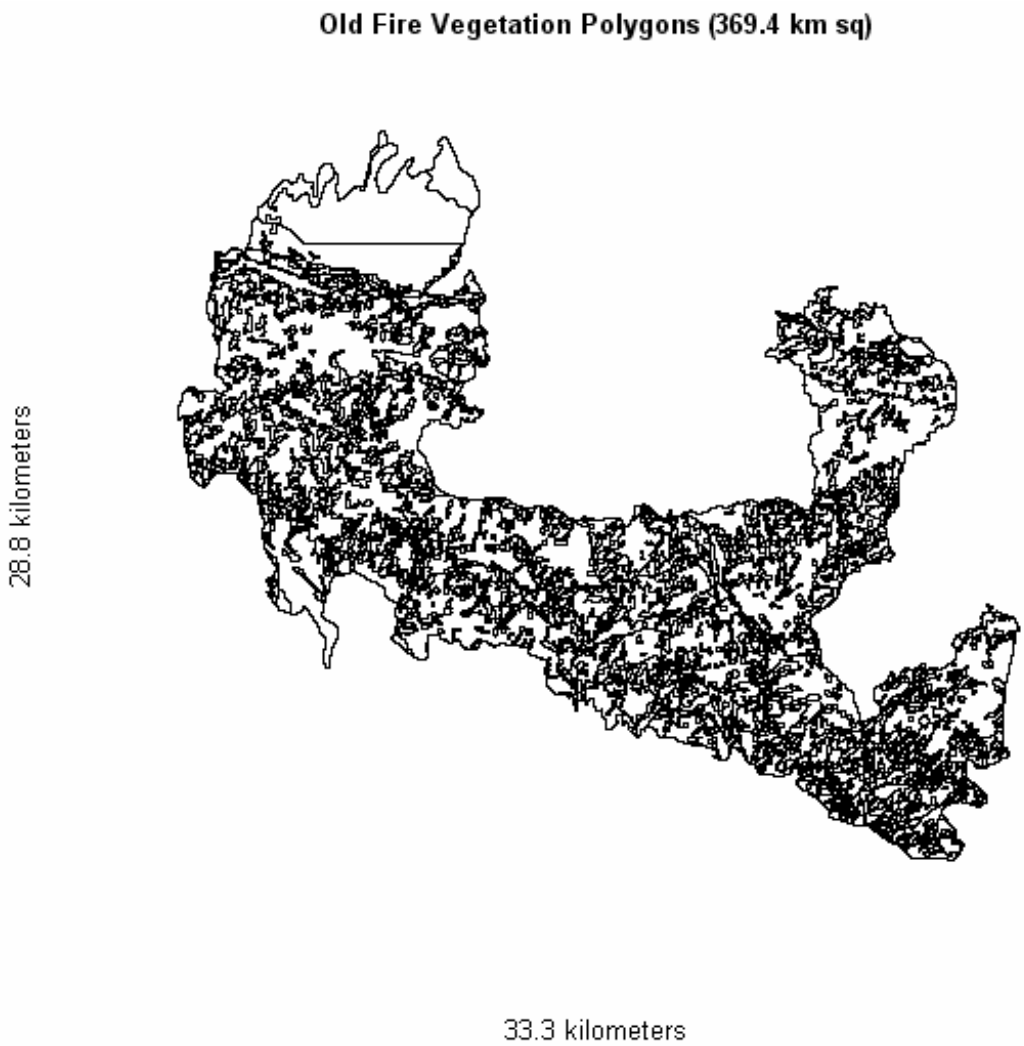

Figure 5: 2369 polygons are used to categorize the vegetation into 12 classes within the Old Fire perimeter.[8]

Meteorological Data[9]

The RAWS network provides meteorological data. There are 441 RAWS stations in California. Seventeen stations lie within fifty kilometers of the fire origin.

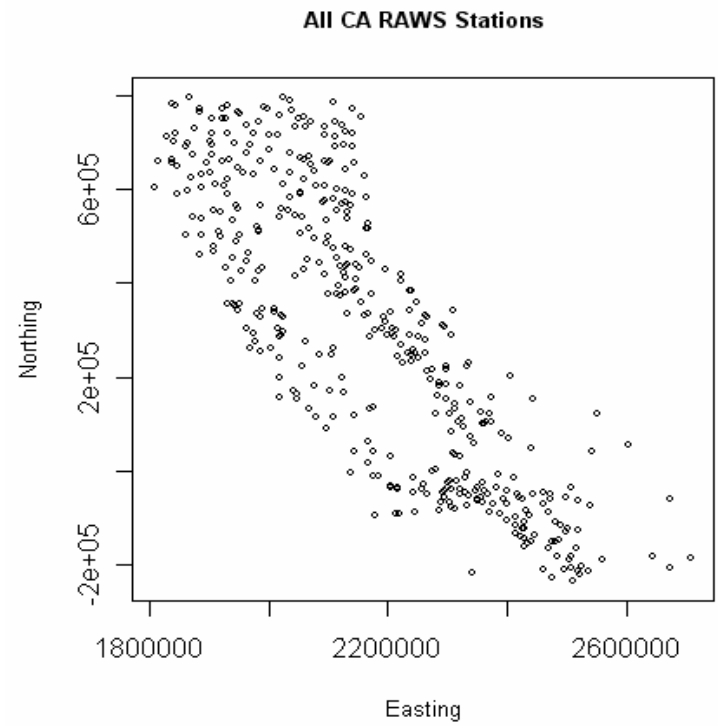

Figure 6: 441 RAWS stations in California.

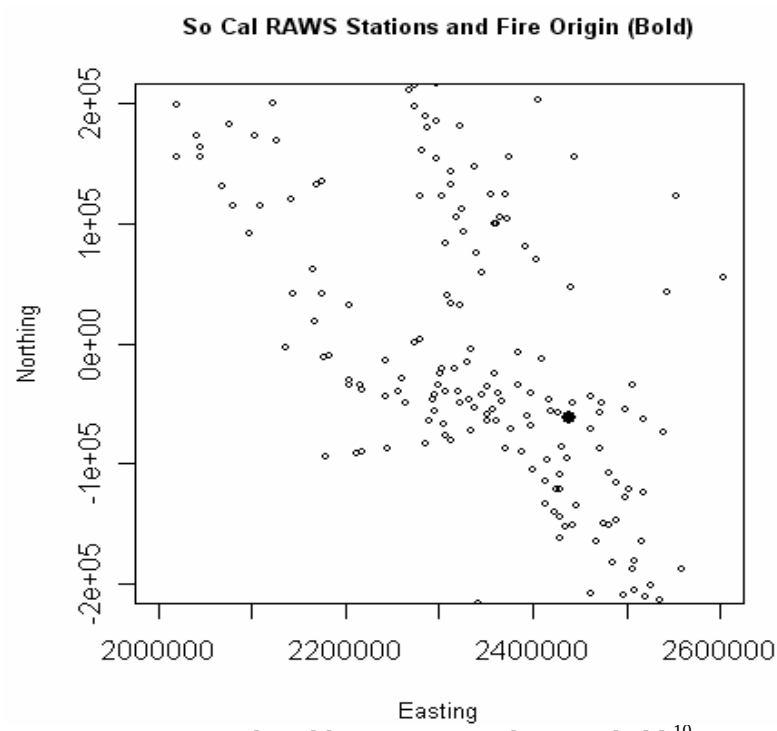

Figure 7: The Old Fire origin is shown in bold.[10]

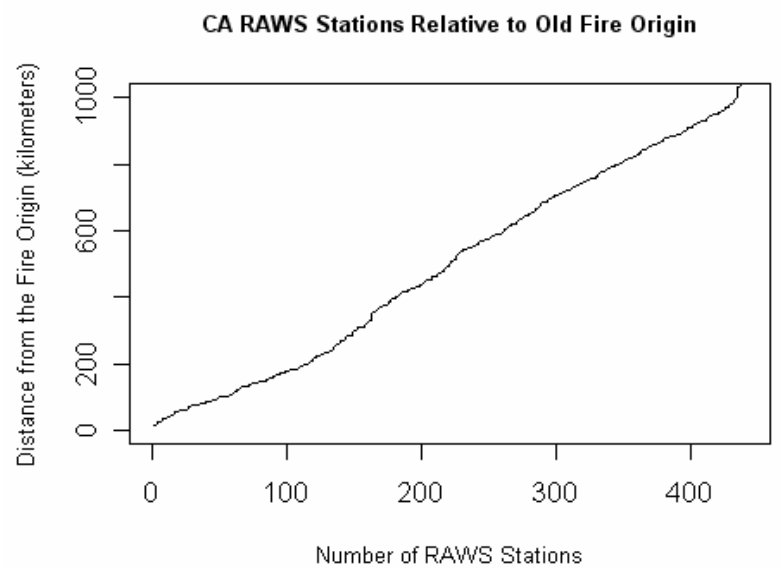

Figure 8: Number of RAWS stations and distance from the fire origin.

Data is collected hourly at each RAWS station. A data sample from Boonville, CA looks like this:[11]

```
Station                GOES ID   Elev Lat          Long
 ==========================================================
 CA BOONVILLE  CA4576D6 644 38:59:14 123:20:54
                 Day/Time  Tmp/Dew Pt/Wind Pcpn  Rh       Fuel Temp      Peak Wind Bat Volt Fuel
                                                                                          Moist
=========================================================================
 CA BOONVILLE  17/2150Z  81/ 47/1608/        5.81  RH 30  FT 90/ 19G15  13.5        FM 7 SR 462
 CA BOONVILLE  17/2050Z  1/ 46/1808/         5.81  RH 29  FT 91/ 17G14  13.5        FM 8 SR 512
 CA BOONVILLE  17/1950Z  78/ 52/1608/        5.81  RH 40  FT 88/ 19G11  13.6        FM 9 SR 504
  .....
```

Ground Temperatures Data[12,13]

USDA's Pacific Southwest Research Station has developed an aircraft-based technology for collecting images for fire events called FireMapper. FireMapper generates high-resolution images that display radiance information collected in three bands: 8.1 – 9 micrometers, 11.4 – 12.4 micrometers and a broad band encompassing those two bands. Each pixel in the image corresponds to an area of 25 square meters or less. A number of fire properties can be calculated from the radiance information in each pixel. Ground temperatures are the property of interest for this project.

For the Old Fire, PSW has generated 44 grids on October 25th starting at 11:09 am and ending at 5:18 pm.[14] Thus, the data generated can be considered as a time series of grids.

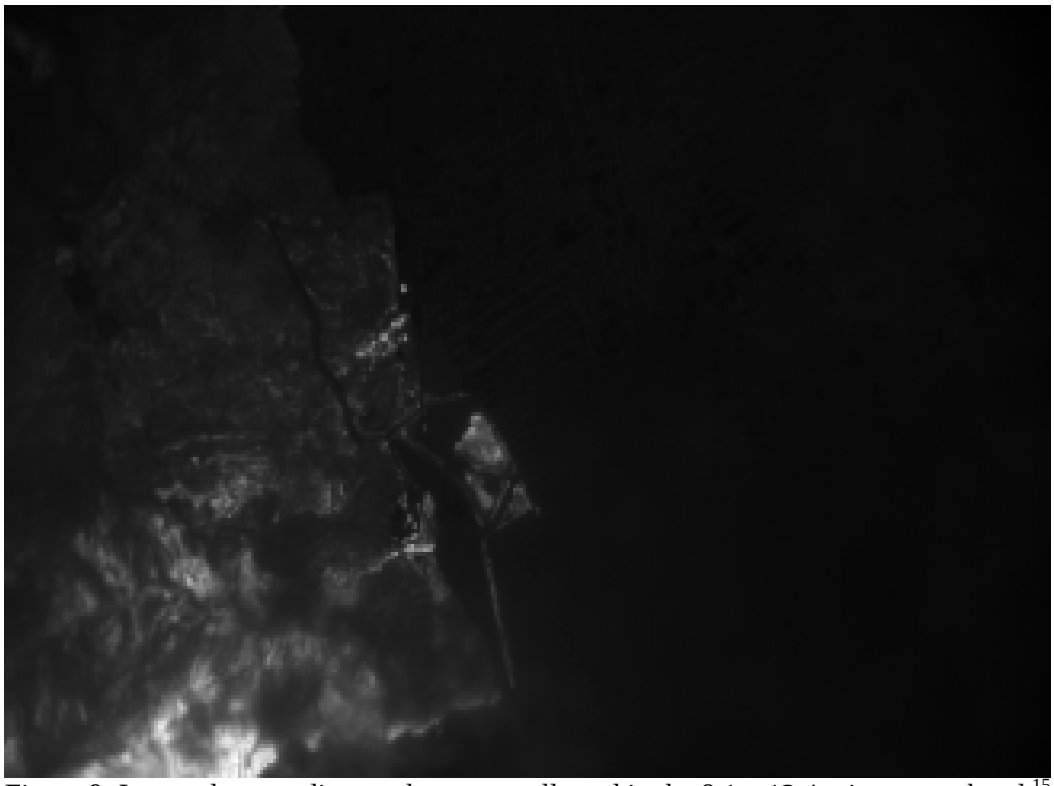

Figure 9: Image shows radiances that were collected in the 8.1 – 12.4 micrometer band.[15]

Structure and Parcel Data[16]

The San Bernardino Assessor's Office offers at least two types of records per assessor's parcel number, a list/label record and an NP450 record. Both records contain addresses. The addresses can be converted to a spatial coordinates like longitude and latitude. The NP450 records are more detailed and contain attributes for structures and parcels that are useful for this project. For example, the construction date can be used to determine if the structure existed before the fire event. This information is important for determining which structures were subject to fire conditions.

There are approximately 800,000 APNs in San Bernardino County. The list/label record and NP450 record cost $0.013 and $0.09 per APN, respectively. It is not known which APNs are within the fire perimeter before a records order is placed with the assessor's office. Thus, records for all APNs in the county must be ordered and then the desired subset must be taken from that data set using the final fire perimeter (see *Figure 4*). Because this order is prohibitively expensive, the authors are seeking to arrange a process with the assessor's office that would allow the APNs within the fire perimeter to be identified *at the assessor's office* and then those records would be purchased. This process is expected to reduce the number of records purchased to a fraction of the original order, reducing the price of the order to a fraction of the original price.

Elevation Data[17]

The USGS provides elevation data sets called the National Elevation Dataset. This data has 10 m resolution.

UWI DATA PRODUCT

The final data product can be constructed from the component data sets. This requires a number of processing steps. One important processing step is illustrated in *Figure 4*. After the processing is done the resultant data can be stored in a database. The database, its contents, and some documentation constitute the final data product.

Major Processing Steps

The major processing steps are as follows:

- Convert addresses to spatial coordinates
- For each data set convert the spatial coordinates' units to units common to all data sets
- Where applicable, convert the temporal units to common units
- Generate a time series of 44 grids covering the area encompassed by the final fire perimeter
- Assign attributes to each pixel in the grid
- Design database schema
- Load the time series of grids, with attributes, into the database

The significance of these processing steps is obviated by the data product.

Description of the Data Product

Constructing the final data product can be conceptualized as populating a four-dimensional box with three spatial dimensions and one temporal dimension. These dimensions correspond to longitude, latitude, elevation, and time. The boundaries of the box are defined by the extremes of the fire event. The extremes are the minimum and maximum elevations within the final fire perimeter, the minimum and maximum longitude and latitude within the final fire perimeter, and the minimum and maximum times the ground temperatures data was collected. The objects that go into the box are grids. There are 44 grids; one for each time ground temperature data was collected. Each grid covers the same area, the area encompassed by the final fire perimeter, and has the same number of pixels. Each grid has 3.5M – 350M pixels corresponding to the required level of resolution, 10m – 1m, respectively.

Instead of including the component data sets in the final data product, the component data sets are used to generate attributes for each pixel in each grid. Only the grids and the attributes for each pixel in each grid will be included in the final data product. Ultimately, every pixel in every grid will have attribute information similar to this:

id	Grid Numb	Veg Type	Ground Temp (°C)	Elev (feet)	Air Temp (°F)	Wind Speed (MPH)	Wind Dir (°)	Structure Status	Time
3490	4	Chaparral	875	730	88	29	160	2	11:09
90218	23	Woodland	311	602	90	31	170	1	12:19

…

Table 1: Example of two pixels from two different grids from the final data product.

ASSIGNING ATTRIBUTES TO PIXELS

Up to this point we have not presented the details of assigning attributes to the pixels in each grid. It is worth sketching how this can be done, although the finer details are beyond the scope of this exposition.

RAWS Network and the Grids

The RAWS network provides meteorological data at each station location. The stations near the fire origin can be used to contribute meteorological features to the data product. The basic problem with adding meteorological features to the data product is that the data product is represented as a time series of grids with millions of pixels in each grid and there are only a few pixels in each grid that are very close to a RAWS station. The further a pixel is from a RAWS station, the less reasonable it is to expect that the data at that station will reflect the meteorological conditions at that pixel. One possible solution would be to use the data provided by the stations near the fire origin to estimate meteorological features for all pixels in each grid. The details of the estimation are omitted.

Embedding Attributes in the Grids

Clearly some judgment must be exercised when deciding how to represent the features provided by the component data sets as pixels, with attributes, in grids, i.e. some information is lost when this new representation is assumed. Rather that exploring this question we will simply state the approach that will likely be taken:

- The process illustrated in *Figure 4* is used to assign a vegetation class to each pixel in each grid.
- The pixel closest to a structure's location will be attributed with the structure's status (damaged, destroyed, etc.)
- Elevation data is of the form (longitude, latitude, elevation). Each pixel in each grid will be assigned the elevation that is closest to it.
- Ground temperature data is of the form (longitude, latitude, ground temp). Each pixel in each grid will be assigned the ground temperature that is closest to it.
- Estimates leveraging the RAWS data will be used to assign meteorological features to each pixel in each grid.

SUMMARY

Constructing a data product with features that are associated with fire phenomenon and risk is a prerequisite to describing fire phenomenon (simulation, prediction, etc.) and measuring risk. There are a number of challenges that contribute to the cost of constructing such a data product. The data product is composed from disparate data sets, each from a different institution. One must decide what data sets to collect. Once collected, data sets often have different units for indexing space and time. The data types can be typical and simple like points or more esoteric and complex like polygons. Ultimately, the data sets are used to construct attributes that are assigned to each pixel of each grid of the final data product.

AKNOWLEDGEMENTS

The authors would like to acknowledge the support for this work provided by the NIST Building and Fire Research Lab under Grant No. 60ANB7D6151 managed by Dr. William Mell.

REFERENCES

[1] Rehm, R., Evans, D., Mell, W., Hostikka, S., McGrattan, K., Forney, G., Bouldin, C. and Baker, E., *5th Symposium on Fire & Forest Meteorology*, Orlando, Florida, 16-20 November 2003.

[2] Riggan, J. and Hoffman, J., "FireMapperTM: A Thermal-Imaging Radiometer for Wildfire Research and Operations".

[3] Brillinger, D. R., Autrey, B. S., Cattaneo, M.D. 2008, "Probabilistic Risk Modeling at the Wildland Urban Interface: the 2003 Cedar Fire", *Environmetrics*.

[4] National Interagency Fire Center, "RAWS Fact Sheet (Remote Automated Weather Stations)", viewed 3 Dec, 2008, http://www.raws.dri.edu/.

[5] "Old Fire (2003)", viewed 3 December 2008, http://en.wikipedia.org/wiki/Old_Fire.

[6] R Development Core Team (2005), R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, URL http://www.R-project.org.

[7] The GIS listed data were collected and developed by the Department of Interior Burned Area Emergency Stabilization and Rehabilitation (BAER) Team.

[8] Nicholas J. Lewin-Koh and Roger Bivand, *Maptools.*

[9] National Oceanic and Atmospheric Administration's National Weather Service, viewed 3 December, 2008, http://raws.wrh.noaa.gov/rawsobs.html.

[10] "Old Fire: San Bernardino National Forest, Initial Start 10-25-03", viewed 3 December, 2008. http://www.incidentcontrol.com/oldfire/index.html.

[11] National Oceanic and Atmospheric Administration's National Weather Service, viewed 3 December, 2008, http://raws.wrh.noaa.gov/rawsobs.html.

[12] Riggan, P. and Hoffman, J., "FireMapperTM: A Thermal-Imaging Radiometer for Wildfire Research and Operations".

[13] Riggan, P., Tissell R., and Hoffman, J., "Application of the FireMapperTM Thermal-Imaging Radiometer for Wildfire Suppression".

[14] Pacific Southwest Research Station, "Fire Spread mapping of the Old Fire, San Bernardino NF, 25 October 2003, Set #1, Page 1, As viewed by the FireMapper™ Thermal-Imaging Radiometer", viewed 3 December, 2008, http://www.fireimaging.com/fires/2003/california/old/298/index.html.

[15] USDA Forest Service Image Specialist, Personal correspondence.

[16] San Bernardino Assessor's Office Business Systems Analyst, Personal correspondence.

[17] Image Peak Systems Corporation "CharTiff NED's USGS National Elevation Data Set", viewed 5 December, 2008. http://www.charttiff.com/neds/about.html.