

GLOBALLY-APPLICABLE PREDICTIVE WILDFIRE MODEL – A Temporal–Spatial GIS Based Risk Analysis Using Data Driven Fuzzy Logic Functions

Gijs van den Dool^{a, b}

^aFaculty of Engineering, Science, and Mathematics, School of Geography, University of Southampton

^bCoreLogic, 7 rue Drouot, 75009, Paris, France, gvandendool@corelogic.com

KEY WORDS: Wildfire, GIS, Fuzzy Logic, Data Driven

ABSTRACT:

This study (van den Dool, 2017) is a proof of concept for a global predictive wildfire model, in which the temporal–spatial characteristics of wildfires are placed in a Geographical Information System (GIS), and the risk analysis is based on data-driven fuzzy logic functions. The data sources used in this model are available as global datasets, but subdivided into three pilot areas: North America (California/Nevada), Europe (Spain), and Asia (Mongolia), and are downscaled to the highest resolution (3-arc second).

The GIS is constructed around three themes: topography, fuel availability and climate. From the topographical data, six derived sub-themes are created and converted to a fuzzy membership based on the catchment area statistics. The fuel availability score is a composite of four data layers: land cover, wood loads, biomass, biovolumes. As input for the climatological sub-model reanalysed daily averaged, weather-related data is used, which is accumulated to a global weekly time-window (to account for the uncertainty within the climatological model) and forms the temporal component of the model. The final product is a wildfire risk score (from 0 to 1) by week, representing the average wildfire risk in an area. To compute the potential wildfire risk the sub-models are combined using a Multi-Criteria Approach, and the model results are validated against the area under the Receiver Operating Characteristic curve.

1. INTRODUCTION

Wildfires are one of the most unpredictable, devastating, natural catastrophes, with severe environmental consequences, causing damage to properties, infrastructure, and costing lives, like the fires in Pedrogao Grande in early June 2017, in Portugal or the Lake (CA) fire in October 2015. By using the Global Fire Emissions Database (GFED4) in combination with the 500m MODIS burned area maps with active fire data, from the Tropical Rainfall Measuring Mission, Giglio et al. (2013) estimates that every year on average of 3.5 million square kilometres of land is affected by wildfires.

The objective of this study was to provide a framework in which these globally available datasets can be analysed, and combined, to create a realistic hazard score without the use (or dependency) of local datasets or high-resolution information layers. The proposed framework consists of three components:

- The topographical database: Shuttle Radar Topography Mission (SRTM) data in 3 arc-seconds, which is 1/1200th of a degree of latitude and longitude, or about 90 meters (295 feet), has been used to find the topographical indicators. The SRTM data is available globally and provide a mid-range resolution suitable for modelling wild fires, and higher than most of the global wild fire models.
- The fuel database: to estimate the fuel availability a composite score of land cover, fuel loads, biomass, and biovolume is constructed. The fuel database is in the final step downscaled to the same resolution as the topographical database to allow cell-by-cell analysis.
- The climatological database: one of the most used indices to find the fire potential in an area is the Keetch-Byram Drought Index (KBDI) developed by Keetch and Byram (1968). The KBDI is forecasting the possibility of wildfire based on soil moisture by calculating the water gain or loss within the soil. The index (in this study) is calculated by using modelled

surface air temperature and precipitation amounts from Department of Civil and Environmental Engineering of the Princeton University (Sheffield et al., 2006).

Within the three components, fuzzy logic operators are used to determine the membership values, and the membership values are translated into scoring factors. In the final step, the topographical and fuel availability scores are combined, and the weekly KBDI is used to rescale the score to give a weekly fire risk.

2. BACKGROUND

Modelling forest or wildfires started roughly 40 years ago, with the first fire following model developed by Rothermel in 1973 (Wells, 2007). Rothermel (1972) used a mathematical approach to predict the fire spreading, with 13 fuel models (to describe the fuel conditions in the field). The model calculates the spread rate, intensity, and flame length under any combination of slope steepness, wind speed, and moisture content, to predict the fire behaviour (Scott and Burgan, 2005).

Most forest departments in countries with a history of wild/forest fires have deployed a model for risk assessment, and in general, all reviewed models are using the same basic parameters to estimate the worst-case scenario (as outlined by Rothermel). The increasing availability of digital data and Geographic Information Systems (GIS) has made it possible to create regional risk models, such as multicriteria risk evaluation or expert risk systems (Sirca, 2017).

All models reviewed, in this study, are either using very detailed local data sources or complex fuel models, in combination with sophisticated software to compute the wildfire risk in a region. The difference between the most commonly used strategies for predicting wildfire and the proposed methodology is the use (and

combining) of globally available data to determine the local wildfire risk.

Particularly, in areas where for example high-resolution data, or meteorological measurements, are not available the use of a more generic approach for calculating the fire hazard could help decision and policy makers to identify areas with a high-risk factor (van den Dool, 2017).

3. METHODOLOGY

There are several designs to model wildfire. The model designed to identify wildfire risk is sub divided in three components. The first two sub-models: Topographical Risk and Fuel Risk (Figure 1) are giving the initial risk levels. The result of the two models is combined to form the final risk score. A third sub-model (climate) is the time-dependency model and is used to propagate the wildfire risk by week by setting a drought scoring for the three study areas (see Appendix I for the information sources).

To summarise, the Basins are obtained from the HydroShed (Hydrological data and maps based on Shuttle Elevation Derivatives) dataset, the DEM, as pure elevation, and second level derivatives capture the Topography: aspect and slope. The Topographical Wetness Index (TWI) is a third level derivative and calculated by dividing the flow accumulation by the slope (Wilson and Gallant, 2000). The parameters in the topographical sub-model are fuzzified by using a trapezoid membership function (Zadeh (1983), Robinson (2003)) to express the relationship between the indicator and fire risk.

The Fuel sub-model contains land cover, biomass, biovolume, and fuel bed data, and are downscaled to the model resolution (3-arc second) by using walk-through tables providing the relationship between the land cover and vegetation layer.

Climatological indicators are not used in the first phase of the risk classifications but brought in later to highlight the most critical period, combining the average weekly temperature and precipitation to calculate the overall drought index (Alexander, 1990). The fire products from the MODIS are used to define the most critical conditions in each subset.

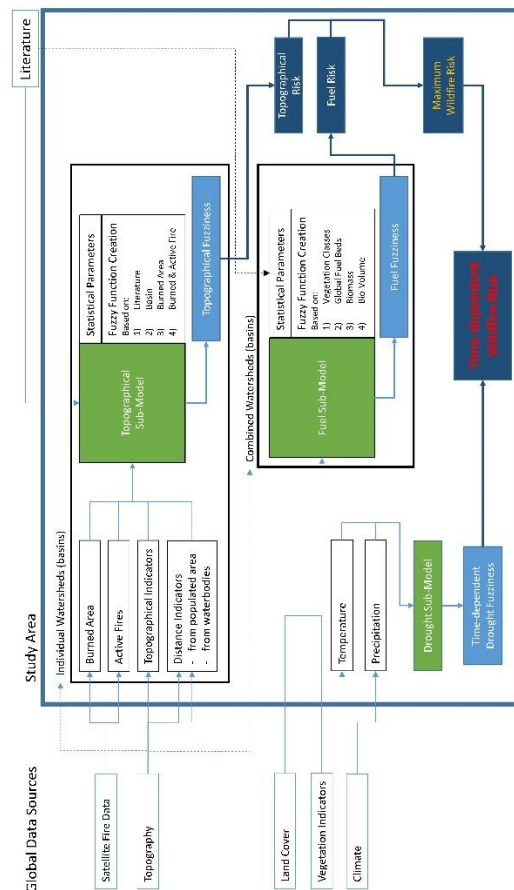


Figure 1: Schematic overview of the model components, and information flow

4. ANALYSIS & FINDINGS

This study aimed to test the possibility of creating a temporal-spatial predictive wildfire model, based on global datasets and fuzzy logic functions and data-driven operations using coarse datasets, e.g. fuel and climatological parameters, in combination with a less coarse DEM. The developed methodology has successfully been applied in three study areas, and the results are a realistic representation of the wildfire risk in those areas.

The individual model components (topographical derivatives, fuel loads, and meteorological indicators) are providing a robust framework for calculating a risk score. The creation of fuzzy membership functions with basic statistics is a proven method to analyse complex relationships and is well known as an efficient approach.

- Within the group of derivatives both slope and TWI are giving the strongest signal when combined with the MODIS data, while the relation between elevation/aspect and wildfire occurrence is less clear.
- The fuel score is a composite of four different information layers, each with product-specific limitations and resolutions, but the used method is creating a fuel score which is a good representation of the fuel availability in the study areas.
- The modelled surface air temperature and precipitation are reflecting the local conditions well and, especially in Spain and California, showing a strong correlation between high

temperatures and low precipitation amounts and the period burned areas are detected.

The combination of the different components is giving good results, but a field in which the model could be improved is the identification of false positives in the MODIS data products. More research is needed to classify the areas of false positives correctly, and develop a rule-set to prevent the model to calculate a score in these regions because false positives are likely to be in areas with a low fuel load.

To test the robustness of the model, and visualisation of the performance, a Receiver Operating Characteristic (ROC) curve is created (Figure 2), in which the two models are tested for sensitivity (true positive rate) and specificity (false positive rate). For each breakpoint (fire hazard class in bands of 0.1) the true/false negative rates are set to be half a standard deviation as the estimated error rate (Winner, 2009). The Mongolian results are less reliable due to the lack of fires in the study area (only two fire hazard classes) and is therefore not shown in figure 2.

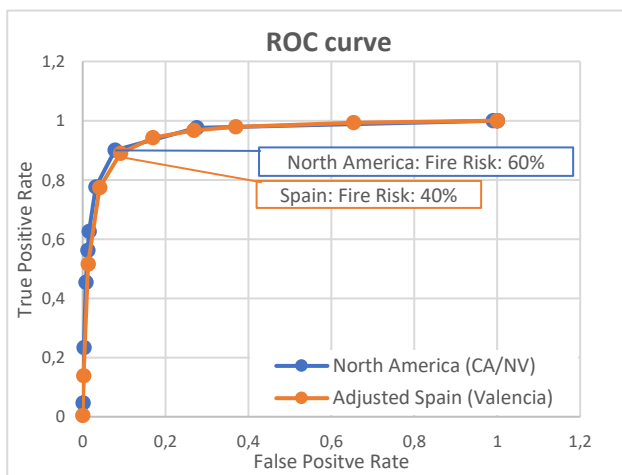


Figure 2: ROC curves for North America (CA/NV) & Spain (Valencia)

The models for the pilots in Spain and North America are behaving well, and the fire classification is sound for both models (AUC classification is Excellent with an AUC > 0.9). The susceptibility to wild fire between the pilot areas is tested by looking at the cut-off value which is closest to the upper left corner, a small distance to corner is better than large distances, the distance is a measurement of the predictive power of the model (Fawcett, 2006). The North America Model is (with a cut-point of 60%) more susceptible to wildfire than the study area in Spain (cut-point set to 40%).

Case Study: North America (Nevada):

In the Nevada region, as part of the North America dataset, the model could be validated against a Landscape-Scale Wildland Fire Risk/Hazard/Value Assessment for the Pershing County (Figure 2). Two catchments (169085 & 169119) are both in the report (WFA, 2009) and in the model. The risk classification in the large basin (169119) is matching the report well, the dominant features in this catchment are present in both maps, although the model results are higher than the rating in the report.

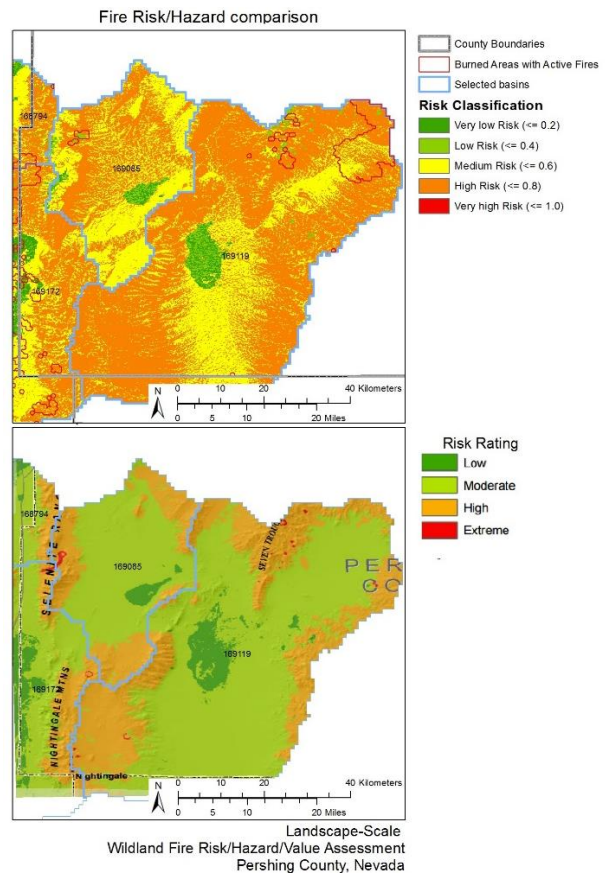


Figure 3: Fire Risk/Hazard comparison. Top: Wildfire Risk Classification for selected catchments in Pershing County, Nevada as a subset from the North America Study Area. Bottom: Landscape-Scale Wildland Fire Risk/Hazard/Value Assessment Pershing County, Nevada adapted from Wildland Fire Associates (2009)

The risk in the Selenite Range (upper left) is underestimated in the model, while the plains are receiving a higher classification, this effect is due to the small Burned Area with Active Fires (BAAF) region which is emphasising the local conditions under the BAAF area more than the landscape-scale model.

The high-risk features in the hazard assessment are not captured in the risk classification because they are buffered transmission lines and other right-of-way corridors determined to be of high-value; these are elements unique to this study and cannot be taken into consideration in a generic model. It was not possible to do a similar comparison for the Spain and Mongolian pilot areas because of local fire risk maps were not available.

5. CONCLUSIONS

The study has shown that it is possible to compute a realistic risk score by using coarse datasets, remote sensing products and GIS techniques. Using the proposed methodology can support local planning and establish a better understanding of the fire hazard, especially in regions where high-quality data is not available, and support decision makers and environmental agencies with the development of management policies and approaches to mitigate the fire risks in those regions.

The Multi-Criteria Analysis, in the current configuration, is setup without the consultation of (local) experts. Including expert opinions, in more detailed studies, might provide a less subjective process and an independent resource in the validation phase to adjust the hazard scores when the overall processes are not generating the correct score.

By using the BKDI as the temporal aspect for fire occurrences, it is possible to assess the potential risk of fire for each week. Although, this is an area where the model could be improved. In this study, only one scale factor is used for each study area; a better solution would be to calculate the BKDI for each grid cell and use the local maximum to compute the score and calculate the weekly hazard.

Recommendations:

- The dependencies in the model should further investigated in areas where local risk scoring maps are available (like in the Nevada part of the North American study area), or tested against established fire predictive models
- A more detailed study of the relationship between land cover, climate, and MODIS products might provide better insight in the detection of false positives of the burned area
- Including (local) expert knowledge in the MCA process might create more defendable weighting factors and an independent validation resource
- Including seasonality in the fuel-score might be another time-dependent element, which is not taken into consideration in this study. Neither is the effect of historic fire events on the vegetation; land cover, fuel scores, bio-volume and -mass are kept constant between years. Modelling land cover change due to forest/wildfires will give a better estimation of available fuel loads.

ACKNOWLEDGEMENTS

This paper is an extended abstract of the (unpublished) dissertation submitted in partial fulfilment of the requirements for the degree of M.Sc. in Geographical Information Systems by online distance learning provided by the University of Southampton, Faculty of Engineering, Science and Mathematics in the School of Geography. I would like to thank Dr. Jim Wright, and my colleagues at CoreLogic, for their patience and constructive feedback while I was working through the analysis and write up.

REFERENCES

Alexander, M.E. (1990) 'Computer calculation of the Keetch-Byram Drought Index-programmers beware', *Fire Manage Notes*, 51(4), pp. 23–25.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), pp.861-874.

Giglio, L., Randerson, J. and van der Werf, G. (2013). Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4). *Journal of Geophysical Research: Biogeosciences*, 118(1), pp.317-328.

Keetch, J.J. and Byram, G.M. (1968) 'A drought index for Forest Fire control', Res. Pap. SE-38. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southeastern Forest Experiment Station.

Rothermel, R.C. (1972) 'A mathematical model for predicting fire spread in wildland fuels', USDA Forest Service Research Paper, INT-115.

Robinson, V.B. (2003) 'A perspective on the fundamentals of fuzzy sets and their use in geographic information systems', *Transactions in GIS*, 7(1), pp. 3–30. doi: 10.1111/1467-9671.00127.

Scott, J.H. and Burgan, R.E. (2005) *Standard Fire Behavior Fuel Models: A Comprehensive Set for Use with Rothermel's Surface Fire Spread Model*. General Technical Report RMRS-GTR-153. Fort Collins, CO: United States Department of Agriculture, , Forest Service, Rocky Mountain Research Station. 72 p.

Sirca, C., Casula, F., Bouillon, C., García, B., Fernández Ramiro, M., Molina, B. and Spano, D. (2017). A wildfire risk oriented GIS tool for mapping Rural-Urban Interfaces. *Environmental Modelling & Software*, 94, pp.36-47.

Sheffield, J., G. Goteti, and E. F. Wood (2006) 'Development of a 50-yr high-resolution global dataset of meteorological forcings for land surface modeling', *J. Climate*, 19 (13), 3088-3111

van den Dool, G. (2017). *GLOBAL PREDICTIVE WILDFIRE MODEL - A Temporal-Spatial GIS Based Risk Analysis Using Data Driven Fuzzy Logic Functions*. M.Sc. University of Southampton.

Wells, G. (2007) 'The Rothermel fire-spread model: Still running like a champ', *JFSP Fire Science Digest* pp. 1–11.

Wilson, J.P. and Gallant, J.C. (2000) 'Digital terrain analysis', in Wilson, J.P. and Gallant, J.C. (eds.) *Terrain analysis: Principles and applications*. Wiley, pp. 1–27.

Wildland Fire Associates, (2009). *Landscape-Scale Wildland Fire Risk/Hazard/Value Assessment Pershing County, Nevada*. [online] Pershing County, Nevada: Nevada Fire Board, Nevada State Office. Available at: <http://forestry.nv.gov/wp-content/uploads/2013/12/Pershing-County-Assessment-Final.pdf> [Accessed 17 Apr. 2017].

Winner, L. (2009). *Applied Statistical Methods*. 1st ed. Department of Statistics, University of Florida.

Zadeh, L.A. (1983) 'The role of fuzzy logic in the management of uncertainty in expert systems', *Fuzzy Sets and Systems*, 11(1-3), pp. 199–227. doi: 10.1016/s0165-0114(83)80081-5.

APPENDIX I: DATA SOURCES

Layer	Product Name	Source	Web link	Accessed
MODIS - BA	Burned Area Monthly L3 Global 500m	NASA	https://modis.gsfc.nasa.gov/data/dataproduct/mod45.php http://modis-fire.umd.edu/pages/BurnedArea.php	19/04/2017
MODIS - AF	MODIS Active Fire Product	NASA	https://earthdata.nasa.gov/earth-observation-data/near-real-time/firms/active-fire-data	19/04/2017
DEM	SRTM 90m Digital Elevation Data	NASA	http://www.cgiar-csi.org/data/srtm-90m-digital-elevation-database-v4-1	19/04/2017
HydroSHEDS	Drainage directions Flow accumulation (number of cells) Drainage basins (watershed boundaries)	USGS	https://hydrosheds.cr.usgs.gov/hydro.php	19/04/2017
Land Cover	300 m annual global land cover time series from 1992 to 2015	ESA	https://www.esa-handcover-cci.org/	19/04/2017
Fuelbeds	Global Fuel Data set using the Fuel Characteristic Classification System	Pettinari	https://doi.pangaea.de/10.1594/PANGAEA.849808	19/04/2017
Biomass	BIOMASAR	Santoro	www.biomassar.org	19/04/2017
Biovolume	GEOCARBON	Thurner	https://www.bgc-jena.mpg.de/geodb/projects/Home.php	19/04/2017
Temperature	Global Meteorological Forcing Dataset for Land Surface Modeling	Princeton University	https://rda.ucar.edu/datasets/ds314.0/	19/04/2017
Precipitation				