ADVANCES IN FOREST FIRE RESEARCH

2022

Edited by
DOMINGOS XAVIER VIEGAS
LUÍS MÁRIO RIBEIRO

Leveraging a wildfire risk prediction metric with spatial clustering

Ujjwal KC[1]; Jagannath Aryal*[2]

[1]*CSIRO|Agriculture and Food, St Lucia, Australia, {Ujjwal.KC@csiro.au}*
[2] *Faculty of Engineering and IT, The University of Melbourne, Melbourne, Australia; {Jagannath.Aryal@unimelb.edu.au}*

**Corresponding author*

Keywords

Wildfire, risk metric, risk characterization, clustering, data-driven approach

Abstract

Fire authorities have started widely using operational fire simulations for effective wildfire management. These fire simulation outputs, when aggregated on a massive scale, create an opportunity to apply the evolving data-driven approach to closely estimate wildfire risks even without running computationally expensive simulations. We explored this opportunity in one of our previous works where we proposed a probability-based risk metric that gives a series of probability values for fire starting at a location under a given weather condition, to fall into different risk categories. The metric considered each fire start location as a unique entity, which could face scalability issues when the metric is used for a larger geographic area and make the metric hugely compute-intensive. As spatial clusters are significantly fewer than fire start locations, such spatial clusters may leverage the metric by reducing the computational requirements. In this work, we investigate if the spatial clustering of fire start locations based on historical fire areas can address the scalability issue without significantly compromising the accuracy of the metric.

1. Introduction

With an increased understanding of phenomena and advancements in computing technologies and observational sciences, natural disasters can be modeled and studied with greater detail (Razavi et al., 2012, Kaizer et al., 2015). Several wildfire models and tools have been developed that can estimate the wildfire behaviors and propagation accurately. Consequently, fire authorities have started widely using operational fire simulations for making better-informed decisions for wildfire management. These fire simulations when aggregated on a massive scale have created a unique opportunity to apply the evolving data-driven approach to closely estimate wildfire risks even without running a single computationally expensive simulation amidst the process being highly compute-intensive otherwise.

In one of our previous works (KC et al., 2022), we validated the application of the data-driven approach to facilitate rapid wildfire risk using a Bayesian probabilistic model. The wildfire risk metric detailed in the study gave a series of probability values for fire starting at a location under a given weather condition to fall into different risk categories. Despite the application of the model being computationally inexpensive, building the inference model was tedious and comparatively compute-intensive as each possible fire start location was considered a unique entity and the probability values conditioned on fire start location had to be calculated for each location. Such a consideration can face serious scalability issues when the geographical area undertaken for wildfire risk estimation is large and has a significantly large number of possible fire start locations. As such, in this work, we investigate if the spatial clustering based on historical data (the data used to build the inference model) could address this scalability issue of the risk metric without significantly compromising the accuracy of the metric. The concept of spatial clusters was envisioned to leverage the risk metric as the number of spatial clusters is significantly lower than the number of possible fire start locations in a geographical area, which could further reduce the overall computational requirements of the metric.

2. Methods and Procedures

2.1. Study Area

We chose Tasmania as the study area to validate the proposed idea of characterizing geographical regions into different risk zones using spatial clustering. The choice was made for several reasons - frequent occurrences of wildfires in the region, the prevalence of readily available high-quality land data sets for the region usable in operational wildfire simulation tools, and a well-studied and systematic grid configuration of fire start locations in the region (Service, 2019; KC et al., 2020a). Tasmania Fire Service (TFS) has maintained a grid configuration of 68,048 possible fire start locations at an interval of 1 km irrespective of land. Any start locations falling on the water bodies are shifted to the nearest land location.

2.2. Wildfire Simulation Tool – Spark

We used the Spark framework (Miller et al., 2015) to run wildfire simulations. Spark offers a flexible platform to simulate the progression of wildfires and their behaviors in different vegetation types. Each wildfire simulation in Spark requires input data sets for fire behavior models, maps of land classification, fuel load, topographical data sets, and weather data to produce output metrics such as total area burnt by fire, the intensity of the fire, and the number of urban cells burnt. All the calculations in Spark are parallelized using the OpenCL framework. All the simulations and their outputs used in this study are available in (KC et al., 2021).

2.3. Fire simulations inputs

We chose four weather inputs - temperature, relative humidity, wind speed, and wind direction for this study following the experimental setup of one of our previous works (KC et al., 2020a, KC et al., 2020b). The ranges for these inputs considered for spatial clustering for risk zones are listed in Table 1. Five equally spaced discrete values of each weather input (except wind direction) were considered along with four distinct directions (east, west, north, and south) for wind directions. Wildfires grow aggressively under weather conditions characterized by high values of wind speed, temperature, and low values of relative humidity when the wind is pushing the fire away from the water bodies. All other static inputs to fire simulations were used as per the configurations and records maintained by TFS and the Tasmanian Government (Tasmania, 2021). All fire simulations were run for five hours, and the cumulative areas burnt by fire in the period were reported as a simulation output.

Table 1 - Range and discretization of the factors for fire weather

Parameters	Range	Labels with Interval
Air Temperature	[10, 40]	Low (L) [10,18] Medium (M)(18, 33) High (H) [33, 40]
Relative Humidity	[10, 90]	Low (L) [70,90] Medium (M) (30, 70) High (H) [10, 30]
Wind Speed	[10,60]	Low (L) [10,23] Medium (M) (23,48) High (H) [48, 60]

2.4. Wildfire risk zones assignment using spatial clustering

We employed spatial clustering enabled by k-means clustering (Likas et al., 2003) based on the values of area burnt by fires starting at the location to assign a risk zone to the fire start location. It should be noted that for any fire start location closer to any water bodies, the information on the fire burnt areas should be interpreted carefully. For example, under a given weather condition, a fire starting at a location with a water body to its east with the wind driving the fire towards the east ceases immediately giving an unburnt landscape, while any fire starting at the same location with the wind driving the fire in other directions not east, may burn a significant area of land. To overcome such circumstances, we adapted the clustering mechanisms based on the mean and median values of fire burnt areas for all locations. The characteristics of the clusters obtained from clustering mechanisms were interpreted to label all fire start locations in Tasmania under three risk zones - low, medium, and high.

3. Experimental setting

3.1. Computing environment for wildfire simulations

All the wildfire simulations used for this study were run using the cloud-based frameworks as designed in (KC et al., 2020) over the cloud infrastructure of Nectar Cloud (Nectar, 2018) and Google Cloud (Google, 2020). Several types of Cloud instances were used as this study does not include any time-related evaluation metrics.

3.2. Evaluation metrics

We compared the accuracy of the proposed clusters-based risk zone characterization against that of the baseline McArthur Forest Fire Danger Index (FFDI) (McArthur, 1967) and our previously proposed risk metric. We also compared the proportions of underfits (predicted area less than the true values) and overfits (predicted area more than the true values). The comparison was done for three and two risk zones as initially explained in our previous work (KC et al., 2022).

4. Results and Discussion

4.1. Risk zones characterization using spatial clustering

Figure 1 shows all possible fire start locations in Tasmania characterized as low, medium, and high-risk areas as given by the spatial clustering. Out of 68,048 fire start locations, about 66 % of the locations were characterized as low-risk zones while the compositions for the medium and the high-risk zone stood at about 24 % and 10 % respectively. Most of the fire start locations closer to water bodies were labeled as low wildfire risk zones while the fire-starting inwards were labeled as high-risk zones. The high-risk zone had a range of average fire area between 3,900 and 10,400 hectares while the low-risk zone had the same range between 2 and 1,700 hectares.

4.2. Comparison against the FFDI and the risk metric

Table 2 shows the comparison of the cluster-based risk metric against the baseline FFDI and the previous wildfire risk metric for three and two risk categories. The accuracy of the cluster-based metric for three distinct risk categories stood at about 70 % while the same for the FFDI and the previous metric stood at about 55 % and 75 % respectively. Similarly, for two distinct risk categories, the numbers stood at about 85 %, 88 %, and 77 % for the cluster based, FFDI, and previous metric respectively. As expected, the accuracy of the cluster-based risk metric is less than that of the previous risk metric. However, the accuracy is still better than that of the McArthur FFDI and is not considerably far from the accuracy of the previous metric. The cluster-based risk metric had more proportion of underfits than the previous metric which could be due to a wider range of fire areas for the low-risk zone compared to the original range of low-risk zone in our previous metric. Conversely, the overfit proportion with the cluster-based metric is marginally less than in the previous metric, which could minimize the overestimation of resources during wildfire management.

In this investigative study, we were able to cut down the location-specific probability calculations from 68,048 to a clustering mechanism and a significantly fewer number of 3 (clusters). Such a provision in the risk metric could theoretically reduce the computational requirements of the calculations of location-specific probability values by a factor in thousands. The most important fact in the findings is that the change in the accuracy of the metric is a mere 8 % and 3 % for three and two distinct risk categories respectively. A comprise of less than 5 % in accuracy at such a scale for a significant reduction in the computational requirements outlines the spatial clustering as discussed in this paper to be a good alternative for fire start location risk characterization in the data-driven probability-based wildfire risk metric.

Table 2 - Performance comparison against the McArthur FFDI

Evaluation metrics	**Three Categories**			**Two Categories**		
	Cluster-based	Previous metric	McArthur FFDI	Cluster-based	Previous metric	McArthur FFDI
Accuracy	66.81	74.55	51.99	84.87	87.43	76.03
Underfits	24.9	15.88	38.87	13.31	10.27	13.49
Overfits	8.2	9.66	9.14	1.81	2.3	13.49

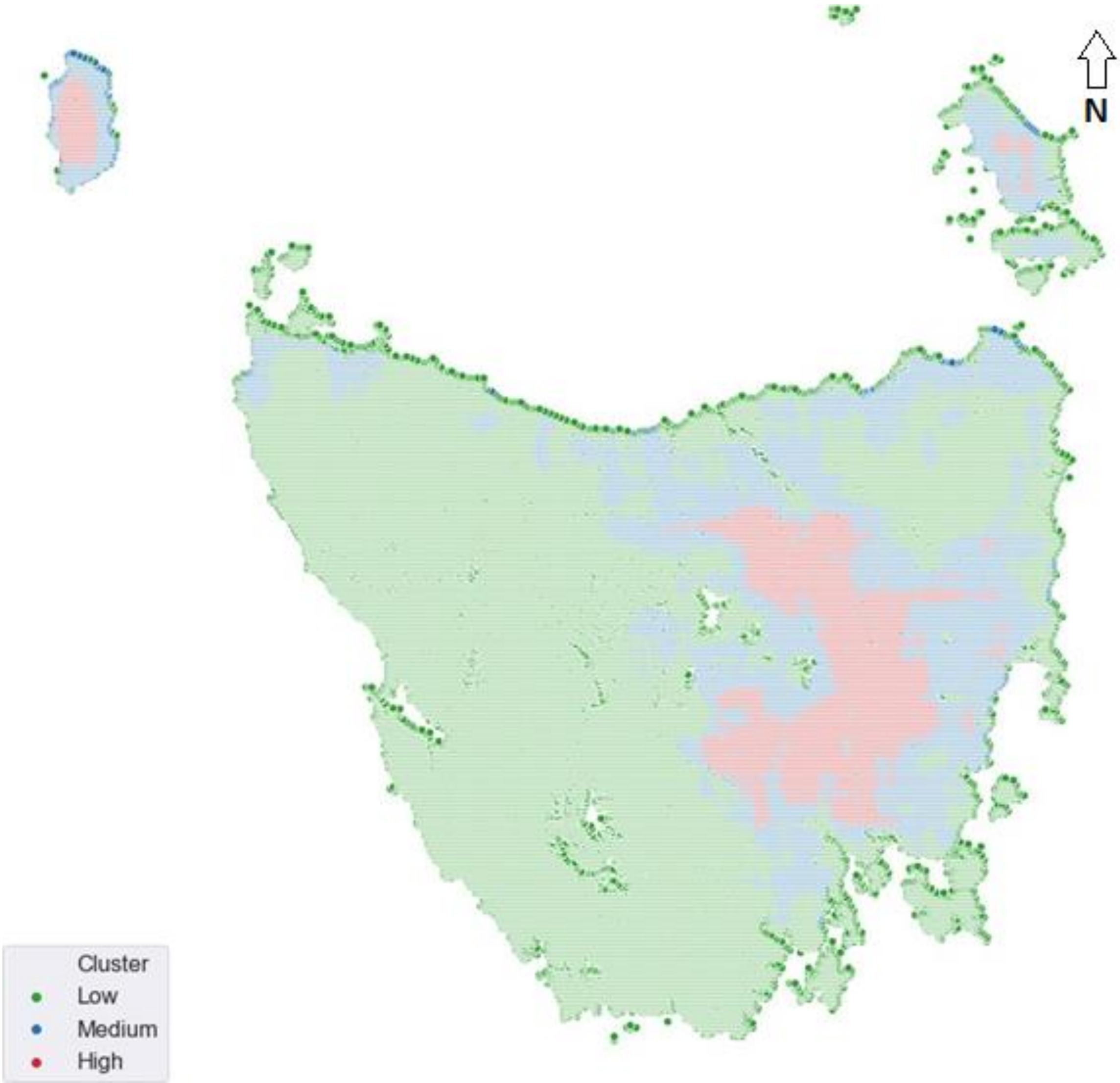

Figure 1 - Risk zone characterization of all possible fire start locations in Tasmania.

5. Conclusions and future research

The ever-evolving data-driven approach can be a computationally efficient alternative to rapidly estimate wildfire risks using several inference models. One such model was detailed in one of our previous works to build a probability-based risk metric, which was quite accurate in risk estimation. But the metric was subjected to scalability issues as each fire start location was considered a unique entity for location-related probability calculations. In this brief study, we investigated if spatial clustering could address the scalability issue of the metric without significantly compromising the accuracy of the metric. We found that the spatial clusters to characterize the risk of each start location could help solve the scalability issues by drastically reducing the number of calculations required by a theoretical factor in thousands with a mere compromise of about 5 % in accuracy. Such an inexpensive estimation of wildfire risk with a data-driven metric can help fire authorities to prioritize resource allocation and make better-informed decisions at various stages of fire emergencies to minimize the possible losses. We expect future works to verify the numbers around the factor by which the computational requirements of the probability-based risk metric are poised to get reduced with spatial clustering. Similarly, the studies around the influence of the number of spatial clusters on the accuracy of the metric could also be studied further.

6. References

(Google, 2020) Google (2020). Google cloud. https://cloud.google.com/. Accessed: 2020-05-12.

(Kaizer et al., 2015) Kaizer, J. S., Heller, A. K., and Oberkampf, W. L. (2015). Scientific computer simulation review. *Reliability Engineering & System Safety*, 138:210–218.

(KC et al., 2020a) KC, U., Garg, S., and Hilton, J. (2020). An efficient framework for ensemble of natural disaster simulations as a service. *Geoscience Frontiers*, 11(5):1859–1873.

(KC et al., 2020b) Ujjwal, K., Garg, S., Hilton, J., and Aryal, J. (2020b). A cloud-based framework for sensitivity analysis of natural hazard models. *Environmental Modelling & Software*, 134:104800.

(KC et al., 2021) KC, U., Garg, S., Hilton, J., and Aryal, J. (2021). Fire simulation data set for Tasmania

(KC et al., 2022) Ujjwal, K., Hilton, J., Garg, S., and Aryal, J. (2022). A probability-based risk metric for operational wildfire risk management. *Environmental Modelling & Software*, 148:105

(Likas et al., 2003) Likas, A., Vlassis, N., and Verbeek, J. J. (2003). The global k-means clustering algorithm. *Pattern recognition*, 36(2):451–461.

(McArthur, 1967) McArthur, A. (1967). Fire behavior in eucalypt forest. canberra, department of development. *Forestry and Timber Bureau*.

(Miller et al., 2015) Miller, C., Hilton, J., Sullivan, A., and Prakash, M. (2015). Spark–a bushfire spread prediction tool. In *International Symposium on Environmental Software Systems*, pages 262–271. Springer.

(Nectar, 2018)Nectar (2018). Nectar cloud. https://nectar.org.au/research-cloud/. Accessed: 2018-05-12.

(Razavi et al., 2012) Razavi, S., Tolson, B. A., and Burn, D. H. (2012). Review of surrogate modeling in water resources. *Water Resources Research*, 48(7).

(Service, 2019) Service, T. F. (2019). *STATE FIRE COMMISSION ANNUAL REPORT*.

(Tasmania, 2021) Tasmania (2021). List data. https://listdata.thelist.tas.gov.au/opendata/. Accessed: 2021-03-12.