Natural Resources Canada / Ressources naturelles Canada

Canada

Predicting the unpredictable: Severe Wildfire Occurrence in Canada

Khurram Nadeem[1], Steve Taylor[1], Charmaine Dean[2], Douglas Woolford[2], Steen Magnussen[1] & Mike Wotton[3]

Canadian Forest Service[1], Western University[2], University of Toronto[3]

Rationale

- About 8000 fires occur in the protected area of Canada each year.
- Approximately 2% of these fires exceed 150+ ha in size; these account for most of the suppression costs and are the greatest threat to communities.
- Although statistical approaches to fire occurrence prediction (FOP) have evolved over the past 40 years, FOP has not been implemented at a national scale in Canada
- Fire occurrence prediction methods have recently been extended to large fire prediction in the US.

Objective

- Develop statistical models to predict severe large fire occurrences at a weekly to biweekly scale based on a Canada-wide gridded spatial resolution.
- Here, we focus on large fire occurrence in British Columbia (BC) to demonstrate our methodology.

Materials and Methods

Explanatory Data:

- We compiled a national dataset of ~ 30 variables important to fire occurrence (TABLE 1) on a 20 × 20 km grid resolution for the forested area of Canada (~8000 cells).
- Variable include vegetation greenness, fire weather, atmospheric stability, fuel conditions, human influences, and baseline fire risk.
- Weather variables are on a daily time step for a 30 years (1985-2015).
- Novel features of this study are use of the 30 years average baseline fire occurrence (Fig. 2) as explanatory variables, and the incorporation of NOAA reanalysis data in the weather data set.

National Daily Gridded Fire Weather Data Set: Daily noon weather observations were obtained from > 2000 MSC and provincial/territorial stations in Canada and ~200 NWS and RAWS stations within 60 km of the border (Fig. 1). We also obtained high resolution North American Regional Reanalysis (NARR) dataset from NOAA. Weather station data were interpolated to the grid cells daily using Thin Plate Spline smoother with elevation and NARR data used as covariates. Because the NARR dataset is generated via a physically based weather model, it provides a reliable estimate of daily synoptic weather patterns, which improves the interpolation.

Table 1 Covariates of Interest in FOP Modeling

Dynamic Variables		Static Variables	
Variable Name	Class/ Variable	Variable Name	Class/ Variable
	Vegetation Greenness	Elev_mean	mean elevation value (from 250 resolution)
NDVI	Normalized Difference Vegetation Index (interpolated and smoothed)	Elev_std	standard deviation of all values
	ATMOSPHERIC CONDITIONS		Vegetation
500MB_ANOM	500 mb anomaly		National Forest Inventory – 250 m attributes
Showalter	Showalter Index	V_mean	VEG - vegetated proportion
C_Haines	Continuous Haines Index	NV_mean	NONVEG - non-vegetated proportion
Tempr	Temperature	VT_mean	TREED treed proportion
Precip	Precipitation	VNT_mean	NON_TREED non-treed proportion
RHumid humidity	Relative humidity	Nleaf_mean	NLS_SPP – proportion of needle leafed species
WindSpd	Wind speed	Leaf_mean	BLS_SPP - proportion of broad leaved species
FFMC	Fine Fuel Moisture Code	Conifer_pct	percent of treed area being conifer
DMC	Duff Moisture Code	Decids_pct	percent of treed area being deciduous
DC	Drought Code	Mixed_pct	percent of treed area being mixed wood
ISI	Initial Spread Index		Ecumene
BUI	Buildup Index	Road_ln	sum of road segments in meters
FWI	Fire Weather Index	Population	population (areal interpolation o enumeration area polygons to grid cells –redistricting)
DSR	Daily Severity Rating (transformation of FWI)		Baseline Risk
Static Variables		IgnitionRate	average no. fire starts per day
	Topography	HumanRate	average no. human fire starts per day
ZONE_NAME	ecozone	LightnRate	average no. lightning fire starts per day

General Modeling Framework

The response variable in the first stage of model development is defined as $Z_{i,j}$, the ignition indicator for the i^{th} grid-cell on j^{th} study day. The overall predictive modeling framework is given as:

Ignition Process: $$Z_{i,j} \sim f(\boldsymbol{\theta}_y; \boldsymbol{x}_{i,j}) \quad \text{(1-a)}$$

Fire Severity Process: $$Y_{i,j} | Z_{i,j} = 1 \sim g(\boldsymbol{\theta}_y; \boldsymbol{x}_{i,j}), \quad \text{(1-b)}$$

where $Y_{i,j}$ is an indicator variable for a large fire (say >100 ha), $\boldsymbol{\theta}$ denote a vector of unknown model parameters and $\boldsymbol{x}$ is a p-dimensional covariate vector. Here, the fire severity process $g()$ is conditional on the event that at least one ignition has occurred (i.e. $Z_{i,j} = 1$).

We have two types of ignition processes in this study: human and lightening caused ignitions ($Z_{i,j}^{(H)}$ and $Z_{i,j}^{(L)}$, respectively). Development of an overall fire severity prediction (FSP) model is based on a two-step procedure:

a) Predicting $Z_{i,j}$ (i.e. occurrence of at least one ignition regardless of its source) by *combining* predictions of $Z_{i,j}^{(H)}$ and $Z_{i,j}^{(L)}$ from two separate models.
b) Conditioning on the predicted values of $Z_{i,j}$, developing a model to predict $Y_{i,j}$.

Random Forests (RF): We therefore opted for RF, a machine learning based algorithm, to *nonparametrically* predict $(Z_{i,j}, Y_{i,j})$ under (1). The RF based predictive models were trained as follows:

1. We split the whole dataset into a training set for model estimation and a test set reserved for model evaluation. The test set consisted of about 20% future data – years 2009-2014.
2. At the training stage, 100 *balanced* datasets (BD_i) were constructed using all the cases in the training set and a separate random sample of controls for each BD_i.
3. A RF model was trained on each BD_i, resulting in 100 RF fits. The predictions were then based on an ensemble of these 100 fits.

Fig. 1 Spatial Distribution of Weather Stations

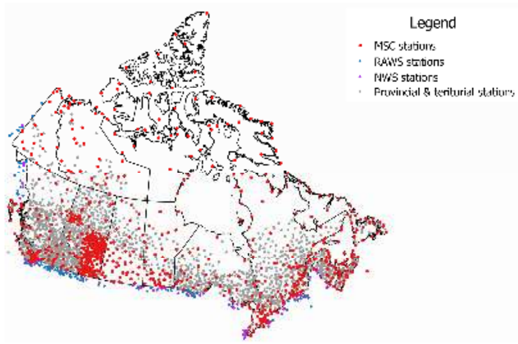
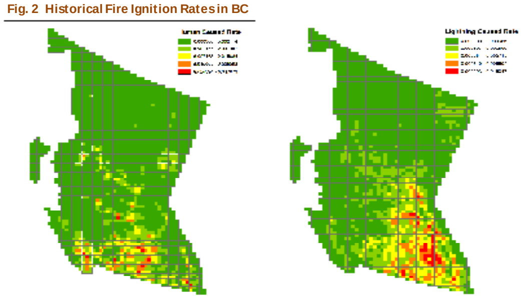


Fig. 2 Historical Fire Ignition Rates in BC



Results

Fig. 2 depicts spatial structure of the long term rate of human and lightening caused fires. Both ignition types reveal a strong spatial pattern where the human caused ignitions are highly associated with the presence of human infrastructure.

Variable Importance: We employed a built-in RF procedure of ranking variables in terms of their predictive ability. Table 2 reports importance ranks under three RF models trained for the response variables $Z_{i,j}^{(H)}$, $Z_{i,j}^{(L)}$ and $Y_{i,j}$. As evident from Fig. 2, various ignition rate variables turn to be leading predictors as they effectively account for spatial heterogeneity in the respective response variables. Sources of ignition are also strongly related to various fire weather indices (DSR, FWI, etc.). The initial spread index (ISI) turns out to be the most important variable in predicting large fires. Quite intuitively, this shows that the probability of an existing ignition turning into a severe fire is much higher in the presence of strong wind conditions.

Performance: Table 2 also reports Area under the ROC curve (AUC), sensitivity and specificity of various predictive models based on a 20% future test dataset. AUC values of 0.60-0.69, 0.70-0.79 and 0.80-1.0 respectively indicate low, moderate and strong predictive ability of a model.

Table 2 Random Forests Model Performance

Model/ Response	Top 5 Influential Variables	AUC	Sensitivity	Specificity
Human Caused Ignition: $Z_{i,j}^{(H)}$	Human Rate, Ignition Rate, DSR, DSR^2, FWI	0.831	0.888	0.774
Lightening Caused Ignition: $Z_{i,j}^{(L)}$	Temperature, Lightning Rate, BUI, BUI^2, DMC	0.854	0.854	0.823
Combined Ignition: $Z_{i,j}$	-	0.812	0.888	0.774
Large (100+ ha) Fire Event: $Y_{i,j}$	ISI, ISI^2, Ignition Rate, Road Length, Human Rate	0.673 (0.691)	0.676 (0.75)	0.671 (0.631)

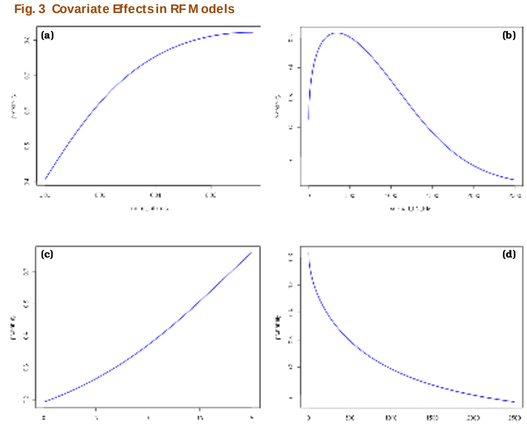
Note: Parenthesized values refer to predictions conditional on *true* ignition events.

The combined ignition model for $Z_{i,j}$ is quite powerful at predicting an ignition event. The predictive ability of the large fire model is low with an AUC of 0.673.

Covariate Effects: For each RF model, we also fitted a corresponding LASSO logistic model to examine the relationship of various covariates with the response variables. The plots in Fig. 2. show how the likelihood a given response variable is varies over the range of the covariate. Fig. a-b and Fig. c-d correspond to human caused ($Z_{i,j}^{(H)}$) and Large fire ($Y_{i,j}$) models respectively. Clearly, likelihood of a human caused fire increases (Fig2-a) as the spatial human cased ignition rate become strong (Fig. 2). The relationship with road length is interesting: human caused ignitions are less likely for area too close are too far from roads/human dwellings (Fig. 2-b).

For the large fires, the risk increases steadily as function of initial spread index (Fig. 2-c). Also, likelihood of a severe fire decreases as a grid-cell increasingly located away from a road. This likely indicates that severe fires activity does not take place in regions too remote for humans.

Fig. 3 Covariate Effects in RF Models



Summary

- Prediction of large and severe fires is an inherently difficult problem as indicated by our large fire model evaluation (Table 2).
- Prediction at the scale of Canada is a "Big Data" problem
- Machine learning methods can be effectively used to predict fire occurrence.
- The methodology is also very powerful in revealing the most influential dynamic and static covariates that affect fire ignition and spread activity.
- The most influential variables (Table 2) make intuitive sense. Notice that baseline human and lightning fire risk was important.
- A good overall predictive skill of RF models highlight the reliability of our weather interpolation algorithm as it retains the signal underlying the response and various covariates.
- Spread activity is often driven by conditions prevailing after the first day of ignition, i.e. future weather conditions. However, our current model was trained only on present or past data. Notice also that performance indicators (value in parentheses in Table 2) indicate that model performance marginally improved even when conditioning on true ignition events.

Future Direction

The modeling approach is being expanded to rest of Canada. The key challenges are:

1. Improving the existing predictive skill of the large/severe fire model.
2. Extending the current predictive approach to predict number of large fires two weeks into the future using weather forecast data.
3. Incorporating the model and output spatial maps and graphs in the Canadian Wildland Fire Information system.