Trends in wildfire burn severity across Canada, 1985 to 2015

Luc Guindon, Sylvie Gauthier, Francis Manka, Marc-André Parisien, Ellen Whitman, Pierre Bernier, André Beaudoin, Philippe Villemaire and Rob Skakun

Supplementary Material 3

SM 3. Estimating pre-fire forest attributes

The methods to estimate the pre-fire values of the following six attributes are described below: Treed vsnon-Treed (a binary variable), total live aboveground biomass (t ha-1), crown closure (%), and percent of biomass in conifer species, in deciduous species, and in unknown species, where these three percentages sum to 100%.

SM 3a. Extracting reference forest attribute values

We extracted reference values of forest attributes from the photoplot (PP) database of Canada's National Forest Inventory (NFI) (Gillis et al. 2005). Each PP consists of a 2 × 2 km square located at each node of a 20 × 20 km grid laid across Canada's forests. For all the PPs, provincial and territorial inventory services provided forest inventory polygons along with a suite of forest attributes derived from a combination of photo-interpretation of aerial photography and modelling. We used a rasterized version of the NFI PP polygons at a 30 m resolution and the associated sampling grid based on 250-m pixel centroids from Beaudoin et al. (2018). This sampling grid provided around 64 reference samples per PP for a total of 376 274 candidate samples, along with the suite of the six required forest attributes, the date of the PP establishment.

SM 3b. Creating time-consistent training set samples

We processed and selected a proper sample subset of the reference pixels to create the random forest (RF) training data set through the application of three operations. The first operation ensured that there was no mismatch between the PP establishment date and the Landsat time series. For example, we want to avoid recently disturbed areas (less than 5 years) with a high biomass value, simply because the PP date was not correct. The comparison revealed a year mismatch for only 3% of the reference pixels, which were removed from our training set.

The second operation removed pixels with inconsistent year of PP establishment relative to *CanLaD* year for a more long-term perspective. We detected year mismatch using a pixel-level yearly biomass increase rate following a severe disturbance as detected in *CanLaD.* The biomass increase rate was calculated as NFI PP biomass (t/ha) divided by the time difference between *CanLaD* disturbance year and NFI PP establishment year. Using this approach, we removed 0.16% of pixels whose annual biomass increase rate was estimates to be above 5 t/ha/year.

The third operation eliminated all samples whose centroids fell within a 30-m buffer around PP polygon boundaries. We did so to avoid selecting mixed pixels overlapping neighbouring polygons, due to relative coregistration errors between Landsat pixels and NFI polygon boundaries. This procedure removed far more pixels in treed areas, where forest stand polygons are small and irregular, compared to pixels in non-treed areas, where polygons are generally large. The resulting sampling imbalance between treed and non-treed pixels was eliminated by gradually removing randomly selected non-treed pixels until the proportion of treed and of non-treed pixels matched the proportion found in the complete NFI PP database.

The initial reference set with 376 274 pixels was reduced to 375 667 then to 178 104 pixels after applying the second and third operations above. This procedure provided the final training set with reference attribute values and we associated to pixel centroids the spatially coincident values of 20 predictive variables.

SM 3c. Spectral predictive variables from Landsat composites

For building the training set, we first assigned to each reference pixel the values of spectral bands (B1,2,3,4,5,7) and spectral indices (NDVI,NDMI,NBR) extracted from one of the yearly Landsat composite from 1984 to 2015 of Guindon et al. (2018) that matched the year of the PP establishment.

In order to simplify the final mapping of the six estimated pre-fire attributes to *CanLaD* fires, we chose to use spectral predictive variables from Landsat composites created for only two target years, ca. 1985 and ca. 2000 (the use of "ca." acknowledges the use of gap filling across years in the Landsat composites). We selected spectral predictive variables from one of the two Landsat composites whose year was closest prior to fire year. We assumed negligible disturbances or growth within the time lag.

To create those Landsat composites, we considered all available Landsat scenes from each target year plus or minus one year (i.e. 1984-1985-1986 and 1999-2000-2001), which provided a sufficient number of scenes required for the pixel- and median-based compositing approach as follows. For each target year, we ranked the available pixels according to their Tasselled Cap Brightness (TCB) values and calculated the TCB median value. For a pixel, once that TCB median value was found, the Julian day of that pixel was taken. We then simply used all the corresponding Landsat band values of that day (band 1,2,3,4,5 and 7) from which we then derived the usual Normalized Difference Vegetation index (NDVI), the Normalized Difference Moisture Index (NDMI) and the Normalized Burn Ratio (NBR) spectral indices (see table SM2). The use of median spectral values automatically removed residual cloud and shadow pixels generally located respectively at the higher and lower ends of the reflectance frequency distribution (Flood et al. 2013, Wulder et al. 2019).

SM 3d. Other predictive variables from various sources

Various site-related predictive variables related to topography, productivity, and climate were derived from various sources and were resampled to a raster with 30-m pixel size using bilinear interpolation to match the resolution of Landsat composites (Table S3.1). The simultaneous estimation of forest attributes was carried out using the r package RFsrc (Ishwaran and Kogalur 2014), a modified multivariate Random Forest procedure (Breiman 2001, Segal et al. 2011) with a 5-fold strategy (80%-20%) with 178 104 training points from Canada's National Forest Inventory (NFI) photoplot network (Gillis et al. 2005) across Canada. After testing for the number of trees, we sued 500 trees (Fig. 3.1) and opted to retain the Mtry pre-defined value. The importance of predictor is shown in Fig. S3.1 and key prediction results are presented in Table S3.2.

Table S3.1. List and description of 20 input predictive variables used in Random Forest and analysis were performed using a k-fold cross-validation strategy (5 folds, 80%-20%) based on the PP ID key to predict forest attributes.

Variable type	Variables	Comments
Multi-spectral data	-Yearly Landsat composite, 1984-2018 -Median Landsat composite, ca. 1985 and ca. 2000.	- Yearly Landsat Band 1, 2, 3, 4, 5, 7, NDVI, NDMI and NBR from 1984 to 2018. Used to get the spectral signatures matching the measurement year of the NFI photoplot to be included in the training set. - Median Landsat Band 1,2,3,4,5,7, NDVI, NDMI and NBR. Median mosaic using all available scenes over a three-year range, centered on 1985 or 2000, to be used to map the final RF model. Data were in Surface Reflectance using the LEDAPS and LaSRC algorithm (Ju et al. 2012, Masek et al. 2006) and were downloaded and processed by the USGS Earth Resources Observation and Science (EROS) Center Science Processing Architecture (ESPA). Landsat Level-2 Bulk Ordering and Downloading available through ESPA Bulk Downloader.
Topography and derivatives	- Elevation from Digital Elevation Model (DEM) - Slope from DEM - Landform Topographic Wetness Index from DEM	ASTGTM: ASTER Global Digital Elevation Model V002 https://lpdaac.usgs.gov/node/1079 Data were downloaded from https://e4ftl01.cr.usgs.gov/ ASTER Global Digital Elevation Model V002: DOI 10.5067/ASTER/ASTGTM.002 NASA/METI/AIST/Japan Spacesystems, and U.S./Japan ASTER Science Team. ASTER Global Digital Elevation Model. 2009, distributed by NASA EOSDIS Land Processes DAAC, https://doi.org/10.5067/ASTER/ASTGTM.002 Topographic Wetness Index - SAGA TWI (Conrad et al. 2015) http://www.saga-gis.org/saga_tool_doc/2.1.3/ta_hydrology_20.html TWI algorithm : Beven et al. 1979, Moore et al. 1991 and Böhner et al. 2006
Productivity	- Degree Days - Net Primary Productivity	-Degree days layer was derived using BioSIM (Régnière et al. 2014) as used in Guindon et al. 2014, 2018 - NPP layer described in Guindon et al. 2018 is based on

		the highest observed NPP over the 2000-2014 period and over a 10km area using the global 1-km pixel resolution MOD17A3 net primary productivity (NPP) time series (Running et al. 2015).
Climate	- Mean annual radiation - Mean annual temperature - Lowest temperature of any monthly minimum temperature - Precipitation over warmest quarter of the year - Total annual precipitation - Summer climatic moisture index	Data were kindly provided by Dan McKenney (McKenney et al. 2011). Customized spatial climate models for North America. https://dx.doi.org/10.1175/2011BAMS3132.1

A) B)

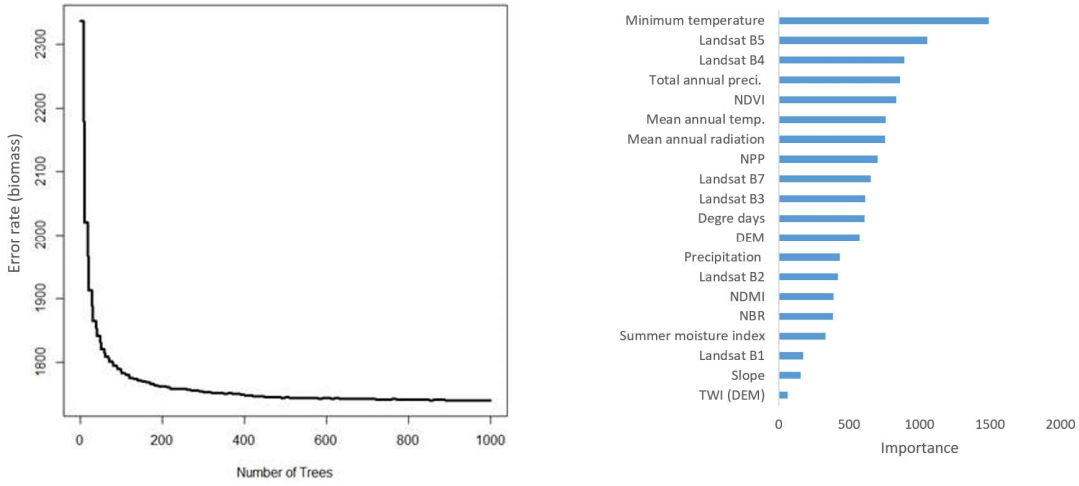

Figure S3.1. A) Test to optimise the number of trees in the Random Forest (RF) analysis. B) Importance of the variables presented in Table S3.1 in the final RF model.

Table S3.2. Random Forest prediction results for the six pre-fire attributes: min, max and mean predictions along with R2 and RMSE values.

	Min	Max	Mean	R^2	RMSE
Treed_Non_Treed *				0.88	0.8
Biomass (t/ha)	0	1103	60	0.66	48.7
Percent coniferous species (% of biomass)	0	100	45	0.60	28.4
Percent deciduous species (% of biomass)	0	100	16	0.62	18.5
Percent unknown species (% of biomass)	0	100	21	0.45	29.6
Crown Closure (%)	0	100	34	0.69	18.5

** In the case of the binary class Treed Non-treed, only overall accuracy and kappa value are presented instead of R^2 and RMSE*

SM 3e. Limitations

- For the pre-fire forest attributes mapped at 30-m resolution and across forested areas subsequently burnt according to *CanLaD* database, no distinction was made if the stands were mostly dead from the Mountain Pine Beetle infestation in B.C. (2000 and more), the Spruce Budworm infestation in Eastern Canada (early 1990-2000 then 2006 to now) or other insect disturbances.
- The training set was created using samples derived from the rasterized forest stand polygons within NFI photoplots because of the absence of a standardised ground plot database at the national scale. Although these polygons provided reference attribute values at a sub-optimal resolution and accuracy for our Landsat-scale analysis, we judged them adequate to stratify our analysis by broad cover types at the national level.
- Despite the observed good overall R^2 values, level of fit is expected to be lower at fine regional and local scales. Therefore, our predictions are more suitable for national and regional analysis.

REFERENCE

Beaudoin, A., Bernier, P.Y., Villemaire, P., Guindon, L., and Guo, X. 2018. Tracking forest attributes across Canada between 2001 and 2011 using a k nearest neighbours mapping approach applied to MODIS imagery. Canadian Journal of Forest Research **48**(1): 85-93. doi:10.1139/cjfr-2017-0184.

Beven, K. J., and Kirkby, M. J. 1979. A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. Hydrological Sciences Journal **24**(1): 43-69.

Böhner, J., and Selige, T. 2006. Spatial prediction of soil attributes using terrain analysis and climate regionalisation. SAGA - Analyses and Modelling Applications 115:13-28.

Breiman, L., 2001. Random forests. Machine learning, **45**(1): 5-32.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., and Böhner, J. 2015. System for automated geoscientific analyses (SAGA) v. 2.1. 4. Geoscientific Model Development **8**(7): 1991-2007.

Flood, N. 2013. Seasonal composite Landsat TM/ETM+ images using the medoid (a multi-dimensional median). Remote Sensing **5(**12): 6481-6500.

Gillis, M.D., Omule, A.Y., and Brierley, T. 2005. Monitoring Canada's forests: The National Forest Inventory. The Forestry Chronicle **81**(2): 214–221.

Guindon, L., Bernier, P.Y., Beaudoin, A., Pouliot, D., Villemaire, P., Hall, R.J., Latifovic, R., and St-Amant, R. 2014. Annual mapping of large forest disturbances across Canada's forests using 250 m MODIS imagery from 2000 to 2011. Canadian Journal of Forest Research **44**(12): 1545-1554. doi:10.1139/cjfr-2014-0229.

Guindon, L., Bernier, P.Y., Gauthier, S., Stinson, G., Villemaire, P., and Beaudoin, A. 2018. Missing forest cover gains in boreal forests explained. Ecosphere **9**(1): Article e02094. doi:10.1002/ecs2.2094.

Ishwaran, H., and Kogalur, U. B. 2014. RandomForestSRC: Random forests for survival, regression and classification (RF-SRC). R package version **1**(0). Ju, J., Roy, D. P., Vermote, E., Masek, J., and Kovalskyy, V. 2012. Continental-scale validation of MODIS-based and LEDAPS Landsat ETM+ atmospheric correction methods. Remote Sensing of Environment **122**: 175-184.

Masek, J. G., Vermote, E. F., Saleous, N. E., Wolfe, R., Hall, F. G., Huemmrich, K. F., and Lim, T. K. 2006. A Landsat surface reflectance dataset for North America, 1990-2000. IEEE Geoscience and Remote Sensing Letters **3**(1): 68-72. http://dx.doi.org/10.1109/LGRS.2005.857030.

McKenney, D. W., Hutchinson, M. F., Papadopol, P., Lawrence, K., Pedlar, J., Campbell, K., Milewska, E., Hopkinson, R.F., Price, D., and Owen, T. 2011. Customized spatial climate models for North America. Bulletin of the American Meteorological Society **92**(12): 1611-1622.

Moore, I. D., Grayson, R. B., and Ladson, A. R. 1991. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. Hydrological processes **5**(1): 3-30.

Régnière, J., Saint-Amant, R., and Béchard, A. 2014. BioSIM 10 – User's manual. 2014. Nat. Resour. Can., Can. For. Serv., Laurentian For. Cent., Québec (Quebec). Inf. Rep. LAU-X-137E.

Running, S., Q. Mu, and M. Zhao. 2015. MOD17A3HMODIS/Terra Net Primary Production Yearly L4Global 500 m SIN Grid V006. NASA EOSDIS LandProcesses DAAC. https://doi.org/10.5067/MODIS/MOD17A3H.006

Segal, M., and Xiao, Y. 2011. Multivariate random forests. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **1**(1): 80-87.

Wulder, M. A., Loveland, T. R., Roy, D. P., Crawford, C. J., Masek, J. G., Woodcock, C. E., ... and Dwyer, J. 2019. Current status of Landsat program, science, and applications. Remote sensing of environment **225**:127-147.