International Journal of Wildland Fire 25(4), 421–432
doi: 10.1071/WF15111_AC

Supplementary material

Ensemble lightning prediction models for the province of Alberta, Canada

Karen D. Blouin[A,B,D], *Mike D. Flannigan*[A,B], *Xianli Wang*[A,B] *and Bohdan Kochtubajda*[C]

[A]Department of Renewable Resources, University of Alberta, 751 General Services Building, Edmonton, AB T6G 2H1, Canada.

[B]Western Partnership for Wildland Fire Science, 751 General Services Building, Edmonton, AB T6G 2H1, Canada.

[C]Present address: Great Lakes Forestry Centre, Canadian Forest Service, Natural Resources Canada, 1219 Queen Street East, Sault Ste. Marie, ON P6A 2E5, Canada.

[D]Environment and Climate Change Canada, Meteorological Service of Canada, 9250-49th Street NW, Edmonton, AB T6B 1K5, Canada.

[E]Corresponding author: Email: karen.blouin@ualberta.ca

Table S1. Description of upperair parameters and indices

For the following equations, unless otherwise stated: Temperature (T) and dewpoint temperature (T_d) are in degrees Celsius; numbered subscripts refer to the atmospheric level (mb), and g is acceleration due to gravity

Convective Available Potential Energy (CAPE)	An indicator of atmospheric instability, CAPE is a numerical measure of the amount of energy (J/kg) available to a parcel of air if lifted through the atmosphere. The positive buoyancy of the parcel can be found by calculating the area on a thermodynamic diagram (Skew-T log-P) between the level of free convection, z_{LFC}, and the equilibrium level, z_{EQ}, where the environment temperature profile is cooler than the parcel temperature. CAPE < 1,000 J/kg indicates a lower likelihood of severe storms while values in excess of 2,000 J/kg and 3,000 J/kg indicate sufficient energy for thunderstorm and severe thunderstorms respectively. $$CAPE = \int_{Z_{LFC}}^{Z_{EQ}} g(\frac{T_{parcel} - T_{env}}{T_{env}})dz$$
Convective Inhibition (CINS)	CINS is the amount of energy in J/kg needed to overcome the negative buoyant energy exerted by the environment on an air parcel. Like CAPE, CINS can be determined from a Skew-T log-P diagram by taking the area between the cooler parcel temperature and warmer environment temperature profile beginning at the surface, Z_{sfc}, and going up to the level of free convection, Z_{LFC}. Given as a negative value, CINS of 0 to -50 J/kg represents a weak capping effect, -50 to -200 J/kg represent a moderate cap, and less than -200 J/kg represents a strong capping force (increased stability). $$CINS = \int_{Z_{sfc}}^{Z_{LFC}} g(\frac{T_{parcel} - T_{env}}{T_{env}})dz$$
Equilibrium Level (EQVL)	The height of the level of neutral buoyancy. This level is often near the tropopause.
Georges K-Index (KINX)	Taking the lapse rate, 850 mb dewpoint temperature, and 700 mb temperature-dewpoint spread into account, KINX provides an estimate of the likelihood of thunderstorms. KINX < 20 indicates zero likelihood, but as the value increases the likelihood of precipitation and thunderstorms increases. KINX > 35 indicates a high likelihood of numerous thunderstorms (George, 1960). $$KINX = (T_{850} - T_{500}) + T_{d\,850} - (T_{700} - T_{d\,700})$$
Haines Index (haines)	The Haines Index (Haines, 1988) is a lower atmosphere stability index. Although typically used as an indication of wildfire growth and extreme fire behaviour potential, haines also represents moisture and stability in the lower atmosphere. Ranging from 2 to 6, a Haines index of 2 represents a moist and stable lower atmosphere with very low fire growth potential while a value of 6 indicates a dry unstable lower atmosphere with a high risk of fire growth and extreme fire behaviour (Haines, 1988). The Haines Index can be calculated for low (950-850mb), mid (850-700mb) and high (700-500mb) elevations. The Haines index is the sum of the Stability Index and Moisture Index. To calculate the mid-level Haines Index: The Stability Term is based on the difference between the 850mb and 700mb air temperatures, where the Stability Term = **1** if $T_{diff} \leq 5°$; **2** if T_{diff} is between 6° and 10°; and **3** if $T_{diff} \geq 11°$. $$T_{diff} = T_{850} - T_{700}$$ The Moisture Term represents the temperature dewpoint spread at the 850mb level, where the Moisture Term = **1** if $T_{spread} \leq 5°$; **2** if T_{spread} is between 6° and 12°; and **3** if $T_{spread} \geq 13$. $$T_{spread} = T_{850} - T_{d\,850}$$

Lifted Index (LIFT)	LIFT is the difference between the temperature at 500mb (***T_{500}***) and the temperature resulting if a parcel of air (***T_{parcel}***) from the layer 500m above the surface (with average pressure, temperature, and dewpoint) were lifted to 500mb (Morales et al., 2007). As LIFT decreases, the atmosphere becomes more unstable. A LIFT of 10 or more indicates stable weather. As LIFT falls below 0 thunderstorms become possible, while values less than or equal to -4 indicate severe thunderstorm potential. $$LIFT = (T_{500} - T_{parcel})$$
Lifting Condensation Level Pressure (LCLP)	The pressure (mb) level at which a parcel of air lifted from the surface dry-adiabatically would become saturated. Lifting condensation level can often be observed as the cloud base and is easily found on a Skew-T log-P by lifting a near surface temperature and dewpoint value. Where ***T*** is in Kelvin and $\varkappa$ is Poisson's constant, 2/7. $$LCLP = p * (\frac{LCLT}{T})\hat{}(1/\varkappa)$$
Lifting Condensation Level Temperature (LCLT)	Temperature in Kelvin at the lifting condensation level (LCL). The LCL is the level at which a parcel of air lifted from the surface dry-adiabatically would become saturated. Note: for this equation temperate and dewpoint are in Kelvin. $$LCLT = [1/(\frac{1}{(T_d - 56)} + \frac{\ln\left(\frac{T}{Td}\right)}{800})] + 56$$
Precipitable Water (PWAT)	Total atmospheric water vapour (mm) within a vertical column cross-sectional area between the Earth's surface and the top of the atmosphere. Provides an idea of how much precipitation could fall as the result of a low pressure system or storm.
Severe Weather Threat Index (SWET)	More commonly referred to as SWEAT, the severe weather threat index (Williams et al., 2008) is used to analyze thunderstorm potential. Values greater than 300 indicate an increased risk of severe thunderstorm, while values of 400 and greater represent considerable risk of tornadoes. In the formula, ***TT*** is the Total Totals Index (if less than 49, set to zero), ***V*** refers to wind speed in knots and ***$\Delta V_{500\text{-}850}$*** is the change in wind direction (degrees) between the 500mb and 850mb levels. $$SWET = 12T_{d\,850} + 20(TT + 49) + 2V_{850} + V_{500} + 125[sin(\Delta V_{500-850}) + 0.2]$$
Showalter Index (SHOW)	Another atmospheric instability index that evaluates the potential for convective storm activity. The Showalter (1947) Index is the difference between the 500mb environmental temperature, **T_{500}**, and the temperature of a parcel lifted from 850mb to 500mb, **T_{parcel}**. As the Showalter Index decreases below zero the chance of convective activity, including precipitation and thunderstorms, is expected to increase. $$SHOW = (T_{500} - T_{parcel})$$

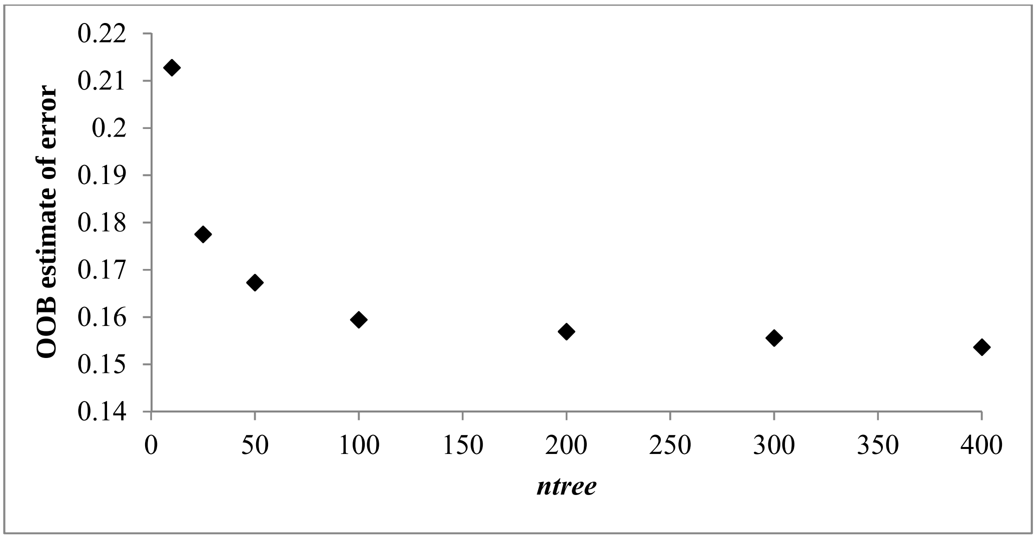

Fig. S1. The out-of-bag (OOB) error estimate trend for random forest models with different number of trees (*ntree*). The OOB estimate or error begins to balance around 100 trees and stabilises by 200 trees.

Table S2. Contingency matrix for model forecasts

		Event observed	
		Yes (1)	No (0)
Event forecast	Yes (1)	*A* *(hit)*	*B* *(false alarm)*
	No (0)	*C* *(miss)*	*D* *(correct non-event)*

Table S3. Forecast skill criteria

Formula inputs provided in Table S2

Name	Formula	Description
Critical Success Index Threat Score (CSI)	$CSI = \frac{A}{A+B+C}$	Taking into account false alarms and missed events, the CSI is sensitive to imbalanced data, often giving poor scores for rare events. This index was not used.
Equitable Threat Score (ETS)	$ETS = \frac{A - A_r}{A+B+C-A_r}$ where, $A_r = \frac{(A+B)(A+C)}{A+B+C+D}$	A more balanced score for rare events, the ETS has a range from -1/3 to one. A higher score indicates increased forecast skill while a score below zero indicates an unskilled forecast.
False Alarm Rate (F)	$F = \frac{B}{B+D}$	Also known as the probability of false detection, F gives the fraction of observed non-events that were forecasted as false alarms.
False Alarm Ratio (FAR)	$FAR = \frac{B}{A+B}$	The fraction of forecasted false alarms.
Hit Rate (H)	$H = \frac{A}{A+C}$	Also known as the probability of detection, H is a score from zero to one where one is a perfect forecast. This measure of skill is sensitive only to misses and does not take false alarms into account.
Post-Agreement (PAG)	$PAG = \frac{A}{A+B}$	The complement of FAR, PAG is the fraction of correctly forecasted events.
Proportion Correct (PC)	$PC = \frac{A+D}{A+B+C+D}$	Provides the overall skill of the model but is not a fair measure of skill for forecasting rare events.

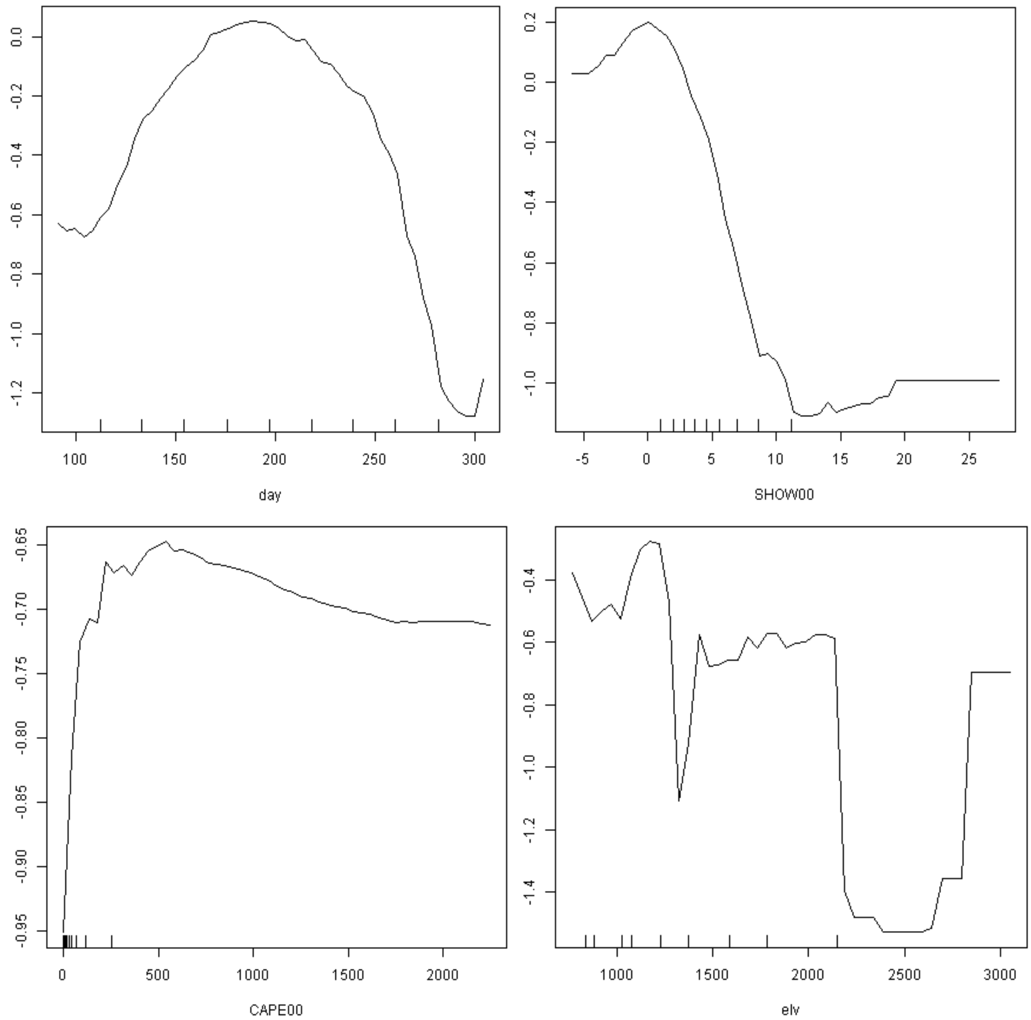

Figure S2. Partial dependence plots of Julian day, 00Z Showalter Index, 00Z convective available potential energy and elevation for daily lightning prediction in the Mountain and Foothills. The y-axes are the logit of the probability of lightning occurrence.

Partial dependence plots

Partial dependence plots can assist with understanding the contribution a variable makes to a given random forest model (Hastie et al., 2009) by providing a visualization of the relationship between a selected individual predictor variable and the binary output. By default, the plots focus on the first class (0); therefore the *which.class* argument was used to generate plots with a focus on lightning events (1).

The y-axis on the partial dependence plots provided is the logit of the probability of lightning occurrence. The ticks inside the x-axis show the deciles of the data distribution for the corresponding variables (Hastie et al., 2009). A lower density of ticks in a certain region of a plot indicates the data density is low in that range and the curves are less determined (Hastie et al., 2009). While the numbers on the y-axis vary from plot to plot, the general trends can be compared between variables by comparing the shape and range of the plots.

Figure S2 shows the daily Foothills models partial dependence plot for the variable Julian day, 00Z Showalter Index, 00Z convective available potential energy and elevation. The plot for Julian day has a unimodal distribution with the logit of predicted probability of lightning being highest when Julian day is ~150-230, and lowest for the days occurring at the tail ends (April and October).This trend was anticipated and clearly supports the seasonality of lightning with peak occurrence in the months of June, July and August (Burrows et al., 2002).

The 00Z Showalter index has a predominantly linear negative relationship with the logit of the probability of lightning occurrence. This trend was anticipated as a negative SHOW value indicates instability in the layer between the planetary boundary layer and the middle troposphere (Showalter, 1947) which is conducive to convection. Burrows et al., (2005) also found the Showalter index to be a good predictor of lightning identifying it as the top overall predictor for Canada and the northern United States. The Showalter Index is also recognized as an important instability measure for severe storms and tornadic activity (Dupilka and Reuter, 2011).

The plot for CAPE00 has a relatively linear positive relationship from zero to ~600 J/kg at which point the trend line plateaus. The positive relationship with the logit of the probability of lightning occurrence indicates higher values of CAPE00 correspond with higher probabilities of lightning occurrence. Again, this trend was anticipated as CAPE is a measure of the convective potential of the lower atmosphere. CAPE values less than 1000 J/kg indicate a relatively stable atmosphere (from a severe storm perspective) while values in excess of 3000 J/kg indicate sufficient energy for severe thunderstorms. The partial dependence plot doesn't seem to capture the upper end variations in CAPE. This could be due to the low number of observations available in the upper range. The thin-plate spline

(TPS) interpolation of CAPE may have also failed to capture the effects of solar heating of the south facing slopes of the mountains. The slopes facing the sun heat at a faster rate than the relatively flat areas to the east due to the angle of inclination. As the warm air rises into cooler ambient air, it has a positive buoyancy. This process may encourage strong updrafts that may become sufficient to support storm activity.

Partial plots are not provided for the remaining variables as the trends become increasingly difficult to interpret and are often obscured by noise. This is a typical response as only the highly relevant predictors for each models are likely to produce informative partial dependence plots (Hastie et al., 2009). In addition, variables with an additive effect also provide the clearest plots (Hastie et al., 2009). The remaining variables include latitude, longitude, SWET12, mslp, CINS00 and surface pressure at 00Z. Mean sea level pressure (mslp), and the Severe Weather Threat index (SWET) were also identified as top predictors of lightning occurrence ranking second and fourth overall by Burrows et al. (2005).

References

Burrows WR, King P, Lewis PJ, Kochtubajda B, Snyder B, Turcotte V (2002) Lightning occurrence patterns over Canada and adjacent United States from Lightning Detection Network observations. *Atmosphere–ocean* **40**, 59–80. doi:10.3137/ao.400104

Burrows WR, Price C, Wilson LJ (2005) Warm-season lightning probability prediction for Canada and the northern United States. *Weather and Forecasting* **20**, 971–988. doi:10.1175/WAF895.1

Dupilka ML, Reuter GW (2011) Composite soundings associated with severe and tornadic thunderstorms in central Alberta. *Atmosphere-ocean* **49**, 269–278. doi:10.1080/07055900.2011.607146

Haines D (1988) A lower atmosphere severity index for wildland fires. *National Weather Digest* **13**, 23–27.

Hastie T, Tibshirani R, Friedman JH (2009)' The elements of statistical learning: data mining, inference, and prediction' 2nd edn. (Springer: New York, NY)

Showalter AK (1947) A stability index for forecasting thunderstorms. *Bulletin of the American Meteorological Society* **34**, 250–252.