# Spatially Explicit Model Predicting Residual Vegetation Patch Existence within Boreal Wildfires

**Yikalo H Araya[1], and Tarmo K Remmel[2]**

1 Geography, York University, Toronto, ON, yikalo@yorku.ca
2 Geography, York University, Toronto, ON, remmelt@yorku.ca

**Abstract**

Wildfires are frequent boreal forest disturbances, and particularly in Ontario, emulating them with forest harvesting has emerged as a legislated forest management goal. Since wildfires typically contain a considerable number of unburned residual patches, we present means for learning their characteristics to improve the subsequent emulation of wildfires. We present a method for developing probability maps for the existence of residual vegetation within wildfire-dominated landscapes. We use the Random Forests ensemble learning approach to predict the occurrence and distribution of residual patches based on selected predictor variables. Satellite-derived data is partitioned into training and validation components using a holdout approach; the model is constructed and calibrated using the training data and evaluated with the validation data. The predictive power of the model is examined using a threshold-independent measure of model performance at five spatial resolutions (4, 8, 16, 32, and 64 m, hereafter described as $R_4$, $R_8$, $R_{16}$, $R_{32}$, and $R_{64}$ respectively). The predictive performance of the model ranges from good (at $R_{64}$) to excellent (at $R_4$) discrimination ability for one of the largest fire events (F01). The lowest predictive performance is observed for the smallest fire event (F02).

## Background and Relevance

Wildfire in boreal forests is usually intense and frequent, and consumes substantial forest cover but does not burn the entire landscape (Perera, Remmel, Buse, & Ouellete, 2009). Owing to the variation in the geo-environmental factors that affect fire spread, there are areas that partially or entirely escape fire, forming post-fire residual patches. A post-fire residual patch is conceptually defined as a mix of live (and dead) vegetation that forms a spatial continuum, ranging from undisturbed patches of live trees to a single stem (Swystun, Psyllakis, & Brigham, 2001). Understanding the existence and distribution of residual patches involves the need to assess the combined effects of various environmental factors. This lays the framework for assessing the ecological values of residual patches. Spatially explicit information about post-fire forest characteristics is also essential for developing land management policies in forested landscapes. Specifically in Ontario, mapping the characteristics of post-fire residual patches has become a primary requirement for emulating forest disturbances, emerging as a general forest management goal within disturbance driven landscapes (Perera et al., 2009).

The development of a framework for real world applications that emulates natural disturbances requires timely and spatially explicit information on residual occurrence. Such maps can be obtained using a predictive modeling approach by merging satellite-based information with ancillary data. Knowing site conditions at which residual patches are likely to occur (or not) forms a basic component of natural resource management and ecological research (Beauvais, Keinath, Hernandez, Master,

& Thurston, 2006). While there are several ways in which post-fire forest characteristics can be mapped, modeled and evaluated to learn about emulating natural disturbances, the incidences and distribution of post-fire residual patches within a fire-disturbed landscape have been poorly studied (Cuesta, Garcia, & Retana, 2009). In this study, a predictive model is developed to generate spatially explicit probability maps and provide information about the distribution of forest stands that escaped burning.

## Methods and Data

The occurrence of residual patches was studied using the Random Forest (RF) algorithm based on selected environmental variables. The algorithm was used to develop a spatially explicit predictive model, and hence generate predictive probability maps. The study was based on eleven fire events, each one ignited by lighting and none of the fires were suppressed. The fire events occurred in northwestern Ontario between 2001 and 2003, having footprint areas ranging from approximately 58 to 4225 ha. For the sake of ease of analysis, the 11 fire events were categorized into three groups based on their size (Table 1).

Table 1. The three categories of fire events.

| **Fire event group** | **Fire footprint extent** | **Fire footprint ID** |
|---|---|---|
| Large sized events | Fire footprint extent ≥ 3000 | F01, F06, F08, and F10 |
| Small sized events | Fire footprint extent ≤ 100 | F02, F03, and F09 |
| Medium sized | Extent > 100 and < 3000 | F04, F05, F07, and F11 |

The study used existing post-fire vegetation residual maps extracted from classified Ikonos images (Remmel & Perera, 2009). The classified Ikonos images were resampled into five spatial resolutions ($R_4$, $R_8$, $R_{16}$, $R_{32}$, and $R_{64}$) based on a non-overlapping block-majority filter for multiscale analysis. The use of RF for predictive model requires a response variable and explanatory variables. The response variable often incorporates the presence-absence data; hereafter described as residual and null-residual patches respectively. However, the vast majority of ecological data that are available today are consisting of presence-only datasets (Zaniewski, Lehman, & Overton, 2002). Yet, presence-only data are the most difficult element to integrate into statistical modeling. Additionally, models based on presence-only data do not provide a better performance (Pearce & Ferrier, 2000).

In this study, a model based on presence-absence data is developed where the existing residual patches were considered as presence-data. However, information about the absence data is not readily available. Therefore, a computer simulation approach has been suggested to algorithmically generate null-residual patches. Yet, models designed based on presence-absence data can be affected by class imbalance (Evans & Cushman, 2009). In order to develop a model based on presence-absence data, a simulation algorithm was initially developed to extract null-residual patches. The algorithm was also designed to randomly generate null-residual patches in which the size, shape and orientation of the null-residual patches mimic the residual patches and hence class imbalance would be avoided. The explanatory variables used for the prediction are topographic variables (slope, ruggedness index – RI, and elevation), vegetation cover type and fire break features (water, wetland, and non-vegetated areas). The variables were obtained from different sources (digital elevation models and existing land cover maps), and were selected based on a prior ecological studies.

The relationship between response (residual and null-residual) and explanatory variables was modeled using RF as implemented in R (R development core 2007). A model based on RF was applied because RF: 1) is a nonparametric and nonlinear classifier; 2) adds an additional layer of randomness; 3) does not over-fit, 4) has high predictive performance, and is computationally efficient in both training and classification (Breiman, 2001). RF also provides error statistics, which is indicative of model fit, but not necessarily the predictive performance of a model. We choose to perform an external cross-validation (hold-out) approach to provide a statistically independent measure of model performance. Given the 11 fire events, the data records from an individual fire event was held-out for testing while the records from the remaining fire events were used for training.

A threshold-independent measure of model performance – receiver operating characteristics curves (ROC) - was used to assess the predictive power of the model. The ROC curve provides a graphical depiction of model's discrimination ability over a range of threshold values (Pearce & Ferrier, 2000). However, comparing ROC curves directly from the plot has never been easy; a single index that describes the discrimination ability of a model is required (Zweig & Campbell, 1993). The area under the resulting ROC curve, which is referred to as AUC, is then considered as an indicator of model's performance. The AUC provides a single measure of model's ability to distinguish between residual and null-residual patches, independent of a specific threshold value. We produced ROC plots for each of the fire events using R; for each of the ROC curve the AUC value was also computed.

## Results

The discrimination capability of the model for selected fire events (i.e., F01, F04, and F02; one for each category of Table 1) is graphically summarized in (Figure 1). A model that perfectly predicts the residual patches generates an ROC curve that follows the left axis and top of the plot, whilst a model with predictions that are no better than random produces an ROC curve that follows a 45° diagonal from the lower left corner to the upper right corner. A plot lying above and to the left of another plot indicates greater observed accuracy (Zweig & Campbell, 1993); such trend was evident in Figure 1 with changing grain sizes. The curves for F01 and F04 at $R_4$ were also closer to the perfect discrimination. However, it is not easy to assess and compare the predictive accuracy directly from the ROC curves. The AUC provides a summary measure of model's predictive accuracy; the ROC curve with the larger area is, on average, more accurate (Pearce & Ferrier, 2000). As a general rule, the AUC value includes: random guess (AUC = 0.5), low accuracy ($0.5 \geq AUC \leq 0.7$), reasonable accuracy ($0.7 \geq AUC \leq 0.9$), and high accuracy ($AUC > 0.9$) (Swets, 1988).

The model had the highest discrimination accuracy with an index value of 0.995 and 0.970 respectively for F01, and F04 (Table 2). Based on the rule of thumb set by Swets (1988), the RF model was evaluated as having reasonable to excellent discrimination ability for F01 and F04 across the gradient of scales; although low discrimination ability was exhibited for F04 at $R_{32}$ and $R_{64}$. The predictive model had also significantly higher discrimination ability (□ < 0.05) for F01 and F04 across the grain sizes considered. The results for F01 and F04 suggested that the occurrence and distribution of residual patches appears to be explained by the predictors incorporated in the model.

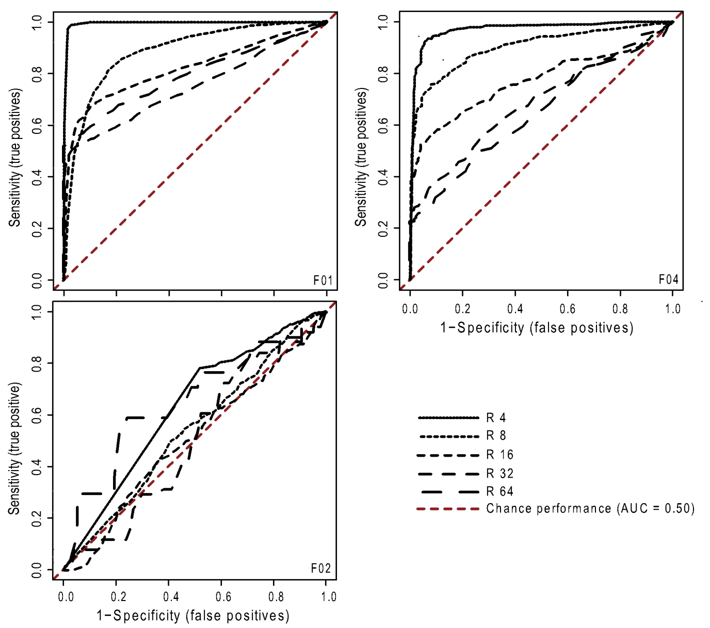


Figure 1. ROC Curves for selected fire events.

Table 2. AUC values for selected fire events.

| | Spatial resolutions | | | | |
|---|---|---|---|---|---|
| | **R4** | **R8** | **R16** | **R32** | **R64** |
| F01 | 0.995 | 0.886 | 0.816 | 0.749 | 0.793 |
| F04 | 0.970 | 0.902 | 0.771 | 0.688 | 0.643 |
| F02 | 0.629 | 0.537 | 0.507 | 0.507 | 0.648 |

However, the model's predictive performance was poor and statistically not significant for F02. One possible reason for the low prediction accuracy at F02 could be attributed to insufficient sample size; F02 is the smallest fire event and had less than 10 records (number of patches) in the evaluation dataset. Pearce, Ferrier, & Scotts (2001) found that the performance of a model for rarer species, with less than nine records in the evaluation dataset, was poorer than those with large number of records. Edwards, Cutler, Beard, & Gibson (2007) also noted that predictive models usually attain more accurate prediction with increased sample size. The results of our predictive probability maps also showed that the model was able to identify potential areas (unburnable areas, specifically wetlands) for residual patch occurrence. This supports a previous study on variable importance assessment where firebreak features (e.g., wetlands) were found to be the most informative predictors (Araya and Remmel, unpublished).

## Conclusions

Pearce & Ferrier (2000) stated that a model with good discrimination ability is the one that correctly discriminate between presence and absence in the evaluation dataset, irrespective of the reliability of the predicted probabilities. Our results support this view that a predictive model based on RF was flexible enough to identify the potential areas where residual patches are likely to occur. Specifically, high prediction probability is associated with the existences and abundance of firebreaks, particularly wetlands. For all the merits of RF in prediction, its interpretability is limited; it is a black-box and does not provide set of rules that are often obtained from standard classifications (e.g., CART) (Evans & Cushman, 2009). However, RF excels at identifying predictor variables and visually characterizing the relationship between predictor variables and predicted classes. Therefore, RF model was determined as a robust (ensemble) approach for learning complex and non-linear ecological relationship, and predicting residual patch distribution from presence-absence data, which is in agreement with previous studies undertaken based on RF models (e.g., Edwards et al., 2007; Evans & Cushman, 2009).

## References


Beauvais, G.P., Keinath, D.A., Hernandez, P., Master, L., & Thurston, R. (2006). *Element distribution modeling: a primer*. Retrieved from http://www.natureserve.org/prodServices/pdf/EDM_white_paper_2.0.pdf

Breiman, L. (2001). Random Forests. *Journal of Machine Learning, 45*(1), 5-31.

Cuesta, R. M., Garica, M., & Retana, J. (2009). Factors influencing the formation of unburned forest islands within the perimeter of a large forest fire. *Forest Ecology and Management, 258,* 71-80.

Edwards, T.C., Cutler, D.R., Beard, K.H., & Gibson, J. (2007). *Predicting invasive plant species occurrences in national parks: a process for prioritizing prevention*. USGS Utah cooperative Fish and Wildlife Research Unit, Utah State University, Logan, UT 84322-5290 USA.

Evans, J. E., & Cushman, S.A. (2009). Gradient modeling of conifer species using Random Forests. *Landscape Ecology, 24,* 673-683.

Perera, A.H. Remmel, T.K., Buse, L.J, & Ouellete, M.R. (2009). *An assessment of residual patches in boreal fires, in relation to Ontario's policy directions for emulating natural forest disturbance*. Ontario Ministry of Natural Resources, Ontario Forest Research Institute, Sault Ste. Marie, Ontario. Forest Research Report No. 169.

Pearce, J., Ferrier, S., & Scotts, D. (2001). An evaluation of the predictive performance of distributional models for flora and fauna in north-east New South Wales. *Journal of Environmental Management, 62*, 171-184.

Pearce, J., & Ferrier, S. (2000). Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling, 133*, 225-245

Remmel, T.K., & Perera, A.H. (2009). Mapping natural phenomena: boreal forest fires with nondiscrete boundaries. *Cartographica, 44*(4), 274-288.

Swets, J. (1988). Measuring the accuracy of diagnostic systems. *Science,* 240, 1285-1293.

Swystun, M.B., Psyllakis, J.M., & Brigham, R.M. (2001). The influence of residual tree patch isolation on habitat use by bats in central British Columbia. *Acta Chiropteroligca, 3*(2), 197-201.

Zaniewski, A.E., Lehmann, A. , & Overton, M.C. (2002). Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modeling, 157*, 261-280.

Zweig, M.H., & Campbell, G. (1993). Receiver-Operating Characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry, 39*(4), 561-577.