Climatic Change
DOI 10.1007/s10584-017-2045-6

Fire weather and likelihood: characterizing climate space for fire occurrence and extent in Puerto Rico

Ashley E. Van Beusekom[1] · William A. Gould[1] · A. Carolina Monmany[2] · Azad Henareh Khalyani[3] · Maya Quiñones[1] · Stephen J. Fain[1] · Maria José Andrade-Núñez[4] · Grizelle González[1]

Received: 4 November 2016 / Accepted: 30 July 2017

Abstract Assessing the relationships between weather patterns and the likelihood of fire occurrence in the Caribbean has not been as central to climate change research as in temperate regions, due in part to the smaller extent of individual fires. However, the cumulative effect of small frequent fires can shape large landscapes, and fire-prone ecosystems are abundant in the tropics. Climate change has the potential to greatly expand fire-prone areas to moist and wet tropical forests and grasslands that have been traditionally less fire-prone, and to extend and create more temporal variability in fire seasons. We built a machine learning random forest classifier to analyze the relationship between climatic, socio-economic, and fire history data with fire occurrence and extent for the years 2003–2011 in Puerto Rico, nearly 35,000 fires. Using classifiers based on climate measurements alone, we found that the climate space is a reliable associate, if not a predictor, of fire occurrence and extent in this environment. We found a strong relationship between occurrence and a change from average weather conditions, and between extent and severity of weather conditions. The probability that the random forest classifiers will rank a positive example higher than a negative example is 0.8–0.89 in the classifiers for deciding if a fire occurs, and 0.64–0.69 in the classifiers for deciding if the fire is greater than 5 ha. Future climate projections of extreme seasons indicate increased potential for fire occurrence with larger extents.

This article is part of a Special Issue on 'Vulnerability Assessment of US Agriculture and Forests developed by the USDA Climate Hubs' edited by Jerry L. Hatfield, Rachel Steele, Beatrice van Horne, and William Gould.

✉ Ashley E. Van Beusekom
ashleyvanbeusekom@fs.fed.us

[1] USDA Forest Service, International Institute of Tropical Forestry, Río Piedras, Puerto Rico

[2] Instituto de Ecología Regional, CONICET-Universidad Nacional de Tucumán, Tucumán, Argentina

[3] Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, CO, USA

[4] Department of Environmental Science, University of Puerto Rico, Río Piedras, Puerto Rico

Published online: 17 August 2017

1 Introduction

Modeling fire likelihood and occurrence in tropical island systems is challenging due to the complexity of fuels, climatic variability, anthropogenic influences, and lack of data on the thousands of small fires that are characteristic of the region (Robbins et al. 2008). For similar reasons tropical evergreen systems as a whole have traditionally been left out of large-scale modeling efforts (Cochrane 2003). The major causes of wildfires in the Caribbean are anthropogenic, resulting from unintentional ignitions and intentional burning for agricultural or other reasons (Robbins et al. 2008). Certain sets of socio-economic factors such as unemployment, livestock density, population density, and tourism rates, have been linked to anthropogenic-caused wildfires in many areas (e.g., Ganteaume and Jappiot 2013; Oliveira et al. 2012; Rodrigues et al. 2014), presumably through ignition rate. Although tropical fires are generally smaller than temperate fires, an estimated 70% of global ignitions in a year occur in these regions (Dwyer et al. 1998) and fire activity in some of the most sensitive ecosystems in the tropics is increasing (van der Werf et al. 2008). With so many fires, a potential increase in average extent would result in large increases in emissions from tropical regions, regional emissions known to disproportionately affect global climate transiently (Forster et al. 2007). However, future projections of fire behavior in the tropics are the most uncertain globally (Krawchuk et al. 2009). These facts highlight the importance of studying wildfire dynamics in the tropics.

Recent satellite data have shown fire occurrence and burned area are the largest in areas such as the Caribbean with intermediate net primary production and precipitation, areas where neither fuel amount or fire season are very constrained (Pausas and Ribeiro 2013). Fires depend on three basic elements to occur: oxygen, fuel, and heat. Together, these components comprise the three sides of the “fire triangle”. Traditional fire danger or risk analysis relies on established relationships between weather patterns and fuel moisture. However, the stochastic and highly local nature of factors influencing fuel moisture, such as wind, exposure to sun, humidity, etc., make large scale analysis and projections difficult and at times misleading. During the summer months, fuel moisture levels are regularly sampled by state and federal fire managers and influence local fire dangers ratings and allocation of management resources. This analysis can provide valuable insight into potential fire behavior, but not its potential occurrence. Furthermore, detailed analysis of fuel availability is not yet available in many (dry or humid) areas of the Caribbean (Robbins et al. 2008), and an analysis of the mechanism of fire in the humid areas of the tropics is not yet quantified, making typical fire indicators ineffective (Taufik et al. 2017).

Our question is whether the likelihood of fire occurrence and its tendency to be large or small can be accurately classified using available weather and social-economic data without directly assessing fuels or fuel conditions. This allows for characterizing the likelihood of fire in given weather and climate scenarios in a particular region with the assumption that fuels are present and (due to weather conditions) flammable. The characterization furthers our understanding of drivers of fire in Caribbean ecosystems. While process-based models may be more effective tools for predicting future fire behavior, statistical models, such as machine learning models, are useful in identifying the most important factors controlling current behavior, and gaining insight into how future climate might change fire behavior (Aldersley et al. 2011). Such models can be the start to understanding how to modify existing process-based models for success in the Caribbean. If process-based models are to be applied in the Caribbean, it might be expected that they need to be modified to account for differing climate processes

(Taufik et al. 2017) and a high amount of anthropogenic links to fire (Luo et al. 2017). In other words, the fire regime is not strongly limited by lack of biomass fuel or ignition sources.

Decision tree methods have been used to analyze global and regional fire drivers in a variety of settings as they take into account the nonlinearity of the interactions of drivers (Argañaraz et al. 2015; Pourtaghi et al. 2016). The majority of these classifier models use numerous socio-economic, biological, infrastructural, geographic, and climatic variables as drivers, or predictors, of fire; many of which are complicated to calculate. We use random forest (RF) decision trees (Breiman 2001) with a simple framework of predictors constructed from daily minimum and maximum temperatures, precipitation, wind speed, recent fire history, and a socio-economic indicator (unemployment) to understand the climate space in which wildfire occurs and discuss implications of projected climate changes on the likelihood of fire.

2 Methods

2.1 Study area and data

Puerto Rico is the smallest of the Greater Antilles Islands, located in the northeastern Caribbean Sea. The main island is approximately 8900 km^2 with a thin strip of coastal plains, 8–16 km wide, surrounding steep igneous upland. Orographic effects are a major control on temperature and precipitation (Daly et al. 2003). While temperature is fairly consistent temporarily and spatially, precipitation gradients are steep and highly varying (Fig. 1c–e). Puerto Rico follows the Caribbean weather pattern created by the easterly trade winds from the Atlantic Ocean with an early rainfall season from May through June and a late rainfall season from August to November. The island-wide dry season is from January to April, which corresponds with the fire season (Fig. 2a, d). The island has drier, more open forest in the south, and wetter, more closed forest in the north, east, and in the central highlands. The landscape is a complex matrix of wildlands, developed areas and agricultural lands (Gould et al. 2008; Van Beusekom et al. 2014).

Puerto Rico has a population density of 438 persons/km^2, similar to relatively urbanized settings such as New Jersey. Sixteen percent of the island is defined as urban and 48% of the island as sparsely populated rural, with the rest densely-populated rural (Martinuzzi et al. 2007). The island has legal divisions of 78 municipalities further broken into 901 barrios ranging in area of 0.1 to 64 km^2 (Fig. 1a). There is a large amount of wildland-urban interface, a setting of high fire risk across the Caribbean (Robbins et al. 2008), the tropics (Cochrane 2003), and globally (Mercer and Prestemon 2005).

Paleo-ecological evidence shows that fire frequency in Puerto Rico increased with settlement and fires now occur in humid areas that in pre-colonization had not been known to burn (Burney et al. 1994). The majority of fires are of anthropogenic origin; in 2013–2014, 40% of the fires occurred at night and were possibly somewhat intentional (Méndez-Tejeda et al. 2015). Small wildfires are common, with most fires too small to be seen with satellite data (Robbins et al. 2008). Intense surveys through local barrio managers recorded 34,628 fires from January 1, 2003 to December 31, 2011, 9 years, with burn extent data recorded for 92% of 8 years of these fires (81% in totality, no extent was recorded for 2008) (Figs. 1a, b and 2a, b). This is the first large historical fire data set that has been analyzed in Puerto Rico. Approximately 5% of the fires with recorded extents in this period were more than 5 ha

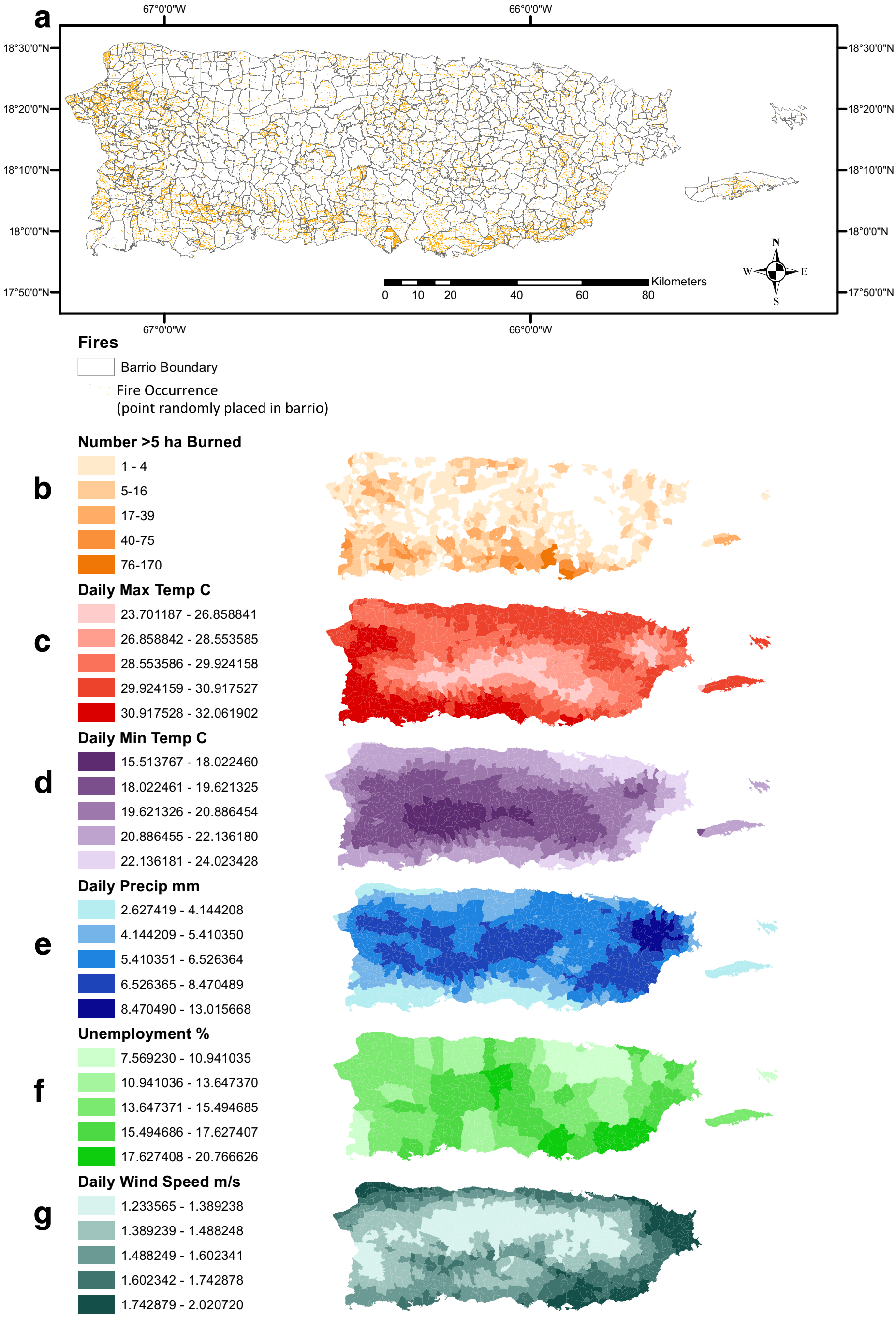

Fig. 1 Map of Puerto Rico with (a) barrio outlines and recorded fire occurrences in each barrio (location randomly placed inside the barrio) 2003–2011; (b) number of these fires recorded as greater than 5 ha; (c) mean daily maximum temperature by barrio 2002–2011; (d) mean daily minimum temperature by barrio 2002–2011; (e) mean daily precipitation by barrio 2002–2011; (f) mean monthly unemployment rate by municipality 2002–2011; and (g) mean daily wind speed by barrio 2009–2011

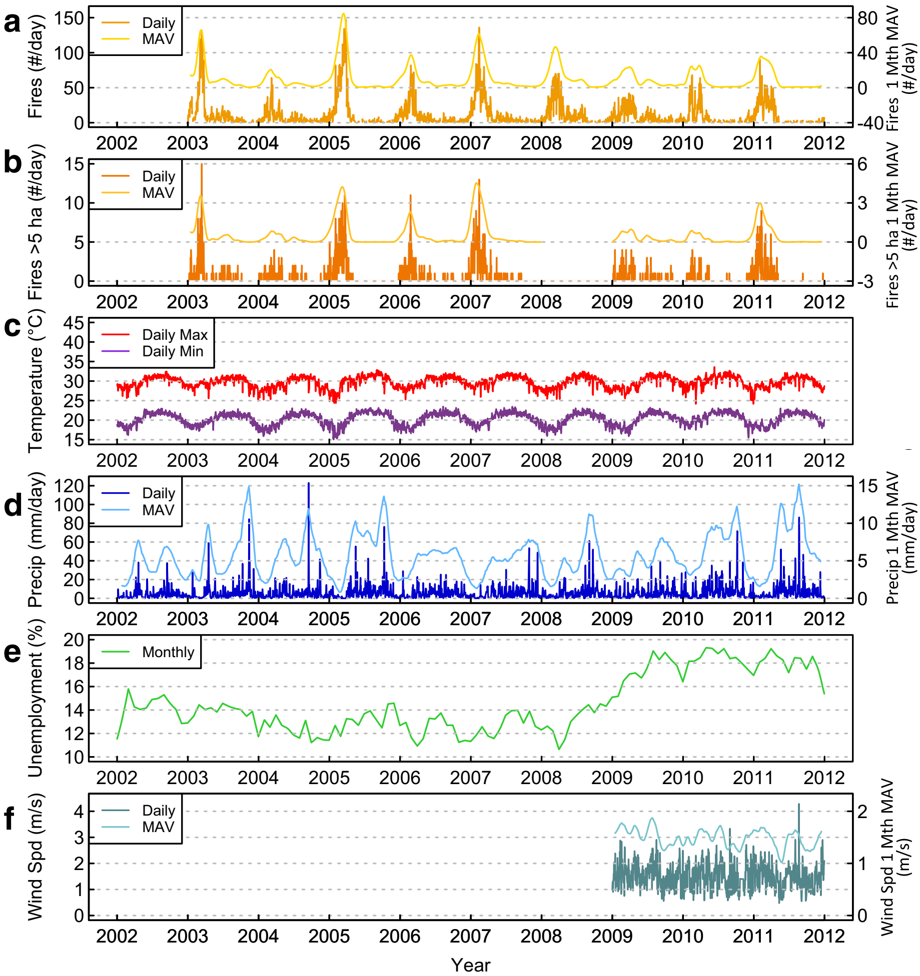

Fig. 2 Time series and 1 month moving average (MAV) (shown on different scales on right side axes) for (a) total recorded daily fire occurrences; (b) total recorded fires >5 ha (there no fire size data for 2008); (c) barrio mean daily maximum and minimum temperature; (d) barrio mean daily precipitation; (e) barrio mean monthly unemployment rate; and (f) barrio mean daily wind speed (starts in 2009)

(0.05 km^2). The median size was 0.4 ha and the maximum size was 1100 ha. Data of fire occurrence and extent were collocated by barrio, not by precise latitude and longitude.

Because the fire data is by barrio, data used for predictors in the RF classifiers were summarized to barrios. Daily maximum and minimum temperature and precipitation are recorded at National Weather Service Cooperative Observer stations and interpolated across the island as daily climate surfaces from 2002 to 2011, with climatically aided interpolation (CAI; Willmott and Robeson 1995) using the Parameter-Elevation Regressions on Independent Slopes Model (PRISM; Daly et al. 2003) as the basis of CAI (Henareh Khalyani et al. 2016). We use unemployment as a predictor indicating socio-economic conditions as it is reported frequently and spatially. Mean unemployment rate in Puerto Rico over 2002–2011 was 13%, more than 6% higher than the USA's unemployment rate average (http://www.bls.gov/). Unemployment rate is reported by municipality per month; we used it here interpolated to weekly averages (Figs. 1f and 2e). Daily wind speed is not measured on the interior of the island, but is modeled by the National Digital Forecast Database for eight 3-h values of wind speed daily spatially distributed over the island since 2009 (http://www.nws.noaa.gov/ndfd/). Model outputs were averaged for daily values of wind speed (Figs. 1g and 2f).

2.2 Classifiers

We use RF classifier models: non-parametric supervised learning to create a number of decision trees each based on a random sample of the data; a method widely successful in classifier problems (Fernández-Delgado et al. 2014). The fraction of individual trees assigning the set of inputted "predictors" to the positive class is the outputted probability the set is in that class, and a threshold is applied such that if the probability is above that threshold, the 'prediction' is the positive class. Data were pre-processed for the most potentially useful climatic and socio-economic predictors, then put into RF classifiers predicting classes for the two problems of (1) occurrence; for a day in a barrio with a set of predictors (the conditions) is there "no fire occurrence or yes fire occurrence?"; and (2) extent; given a fire on a day in a barrio with a set of predictors is there a "small fire or big fire?", for the 3287 days of record (2003–2011) and the 901 barrios. The data available for predictors were climate data and socio-economic data, as discussed in the previous section, as well as fire history data. Positive class is considered fire occurrence or a big fire. Each RF classifier was trained on up to 80% of the data available, and 20% was reserved for testing the classifier.

2.2.1 Data pre-processing

Predictors summarized by barrios means barrios of differing sizes will result in different resolutions to the predictors in each barrio; this problem is not too distressing. However, more egregiously, it will also mean fire ignitions have more potential area within which to occur for larger barrios, thus predictions in the RF classifier could be biased toward conditions in the largest barrios. To mediate biases, we used the middle two quartiles of barrio sizes for the fire ignition; five to 13 km^2 in size. There was no correlation between number of fire occurrences over the period of record and barrio size in the full or reduced dataset. The raw data possibility of a fire occurring in a barrio on a day was approximately the same in the full data set as in the data set reduced by barrio size, a 1% chance. For the fire extent problem, the smallest barrios were removed (those <1.6 km^2) as those are densely populated city-barrios and may be not indicative of a wildfire.

Two sets of potential predictors were then constructed. The first set was daily value summarized by barrio for the following attributes: maximum, minimum, and diurnal range temperature; precipitation and wind speed; and weekly unemployment rate. These attributes were then expanded into sets of mean values aggregated prior to a day, up until 24 weeks before the day of a fire, or:

$$A_{xbd} = \sum_{i=day\,d-x}^{i=day\,d} a_{ib}/(x+1) \;\; with\; 0 < x < 24{*}7, \qquad (1)$$

for each barrio b, each day d in years 2003–2011, and each weather or socio-economic attribute a. The upper limit of 24 weeks was chosen as it is near 6 months and considered representative of long-term behavior of the attribute. These predictors (A) are termed 'absolute predictors'. The second set of predictors (R), termed 'relative predictors', was absolute predictors divided by the average daily value over 2002–2011 of each attribute in each barrio, or:

$$R_{xbd} = \left(\sum_{i=\text{day } d}^{i=\text{day } d} a_{ib}/(x+1)\right) \Big/ \left(\sum_{i=\text{Jan } 1,2002}^{i=\text{Dec } 31,2011} a_{ib}/(3287)\right) \quad \textit{with } 0 < x < 24{*}7. \tag{2}$$

Note that for a = unemployment, only weekly values are used so $0 < x < 24$ in both equations and the denominator in Eq. 2 is divided by 120 instead of 3652. These two sets of potential predictors were tested for the most likely "top aggregate" predictor out of the range of aggregate means for each attribute, and that one was used in the final RF classifier. This was done by separating A_{bd} and R_{bd} into days with and without fire occurrence, and big and small fires, and testing which length of aggregation made the largest separation for the groups. Separation is defined as variance between the two groups divided by variance within the groups. This is the F-statistic used in ANOVA 1-way test where a large statistic rejects the null that the two groups are drawn from populations with the same mean values; however, we did not run this test as we cannot assert independence among the groups because of spatial correlation possibilities (Legendre 1993). Final classifiers were run with each attribute contributing either absolute or relative predictors depending on group separation results, with the top aggregate mean predictor, and the 'daily' observations A_{1bd} or R_{1bd} (since we do not know if the event happened before or after the A_{0bd} or R_{0bd} observations happened; this might be critical for a precipitation observation). A final of predictor describing fire history was added to each problem. This attribute is the area of the barrio multiplied by the number of days since a fire was last detected in it.

For the extent problem, the classes of big and small fires had to be defined. Most fires in Puerto Rico are too small to be considered large fires in other ecosystem models (Birch et al. 2015; Fang et al. 2015). Compounding the problem, the fire sizes are often reported as rounded values, so computing a cluster analysis for optimal classes would be impossible (e.g.; Yu et al. 2011). We defined classes by again looking at group separation balanced with considerations that enough cases would be in each class to build a reasonable classifier. Big fire size was defined as any fire more than 5 ha, with small fires defined as the complementary set. Arranging the problem such that an unpredicted middle class existed between big and small was tested, but results were not encouraging enough to warrant this removal of data, as our goal was to characterize the full effect (conclusive or not) of climate space on fire behavior. The raw data possibility of a fire extent greater than 5 ha was 5%.

2.2.2 Data imbalance

For both problems, classes were highly imbalanced. Large class imbalance is known to cause problems in machine learning due to the minority class concepts being underrepresented (He and Garcia 2009). Randomly over-sampling the minority class overfits the minority class, and under-sampling the majority class loses information in the majority class. We achieved best results by synthesizing a balanced data set using the method of ROSE (random over-sampling examples; Menardi and Torelli 2014), where new data is made in the neighborhoods of randomly chosen existing minority class data (i.e., the members of the estimated conditional density underlying the original data) and the majority class is kept as is until equal class balance. The size of the neighborhoods is tuned to achieve a data set that results in a good RF classifier.

2.2.3 Classifier fitting

Binary classification models were trained on the balanced data with 500 trees, by which point there are no significant gains in performance by increasing the number of trees. We evaluated classifier performance with the test data set (20% of the original data held out, thus imbalanced) and the metric of area under the receiver operating curve (AUC). The receiver operating curve (ROC) is the plot of true positive rate against one minus true negative rate; the area under the ROC is independent of the threshold probability used for class assignment. An AUC value of 0.5 indicates classifier performance is no better than random, whereas an AUC of 1.0 is perfect. The AUC is equivalent to the probability that a random positive case will be assigned a higher probability than a random negative case. This is considered one of the best metrics for unbalanced problems (Chawla 2005).

Importance measures on the predictors in the training set were computed by permuting the values of each feature and measuring how much the permutation decreased the accuracy of the balanced training classifier. However, two completely correlated predictors will both be deemed important even though one could be removed from the classifier, so predictor importance was assessed by leaving out subsets of correlated predictors (Gregorutti et al. 2017). Furthermore, predictor merit was also assessed by building classifiers with different sets of predictors and examining the resulting classifier AUC with the training data. These predictor sets were: all data available, all climate and socio-economic data (top aggregates and daily values, no fire history), all climate data, all top aggregate data, and climate top aggregate data. Because wind data is only available for three of the years of data, including it as a predictor necessitates a loss of information for the other predictors. Classifiers were run for the entire data set without wind data and for the 3 years of data including wind data.

Visual comparison between predicted responses and the actual responses of the test set were made to see what each attributes response was, and how well the classifier represented the test set. Response curves are plots of partial dependence where the one predictor is varied over its range while the other predictors are held constant at their means. The probability of a positive outputted class is plotted for the classifier predicted response, and the true class is plotted for the data actual response.

3 Results

3.1 Data pre-processing for top aggregate predictors

Relative predictors (Eq. 2) achieved better separation between the group's positive fire occurrence and negative fire occurrence than did absolute predictors (Eq. 1). Conversely, in the extent problem, absolute predictors achieved better separation in positive large size and negative large size (small size). The top aggregates in separation of the occurrence problem were relative means of the previous 12 weeks for maximum temperature, 7 weeks for minimum temperature, 2.5 weeks for diurnal temperature range, 11 weeks for precipitation and wind speed, and 14 weeks for unemployment. The top aggregates in separation of the extent problem were absolute means of the previous >24 weeks for maximum and minimum temperature (the longest aggregate tested), 1 week for diurnal temperature range and unemployment, 9 weeks for precipitation, and 2 days for wind speed.

3.2 Classifiers

Importance measure experiments rated the attributes of top aggregate and daily precipitation as the most important predictors for each problem (relative predictors for the occurrence problem and absolute predictors for the extent problem). In the occurrence problem, after precipitation, top aggregate minimum temperature and fire history were important. In the extent problem, after precipitation, the top aggregate maximum temperature and wind speed were important.

The AUC metrics evaluated on the test data set (data not used in any way to train the classifiers) for the different classifiers are shown in Table 1. The classifiers using the reduced

Table 1 Random forest (RF) classifiers for fire occurrence and extent problems using different sets of predictors and the resulting AUC (area under the receiver operating curve) values

Problem	Years	Wind Available	Data Used					AUC
			Climate		Socio-economic		Fire History	
			TAgg[1]	Daily	TAgg[1]	Daily		
Fire Occurrence	2003-2011	no	■	■	■	■	■	0.89
			■	■	■	■		0.87
			■	■				0.86
			■		■			0.79
			■					0.79
	2009-2011	yes	■	■	■	■	■	0.83
			■	■	■	■		0.80
			■	■				0.81
			■		■			0.78
			■					0.79
Fire Extent	2003-2011	no	■	■	■	■	■	0.67
			■	■	■	■		0.66
			■	■				0.66
			■		■			0.67
			■					0.64
	2009-2011	yes	■	■	■	■	■	0.65
			■	■	■	■		0.68
			■	■				0.68
			■		■			0.68
			■					0.69

[1] TAgg is an abbreviation for Top Aggregate for each attribute: the mean of the attribute over a specific number of successive days prior that achieves the best separation between the two classes for each problem

data 2009–2011 (containing wind speed) are less stable that those with data 2002–2011, depending on the seed set for random number generation (Matsumoto and Nishimura 1998), especially in the case of the smaller extent problem. We show AUC values from typical classifiers out of multiple seed selections (Table 1). The AUC values for the fire occurrence classifiers characterize very good classifiers with an 80–90% chance of assigning a day in a barrio that has a fire as being more at risk than a day in a barrio that does not have a fire. The extent problem is more challenging, but the addition of wind data may be making the fire extent more predictable. Still, all classifiers are substantially better than random chance. The full classifiers with all available data have the highest AUC value in most of the problems, but the classifiers using only the climate data do comparably. We note as a side test, that the problems run with the opposite predictors, absolute predictors for the occurrence problem and relative predictors for the extent problem, achieved smaller AUC values as expected.

Smoothed lines of the scattered points of predicted and actual responses of the attributes are shown in Fig. 3 for the classifiers using all the predictors for the longest period available, 2003–2011, as the longer period classifier is considered a more reliable summary of fire behavior. Wind attribute responses are shown from the shorter period classifier over which wind was available. The predicted responses are probabilities of belonging to the positive class (again, a fire occurrence or a big fire) plotted against the value of the attribute, and the actual responses are the class plotted against the value of the attribute. We use local regression smoothing (LOESS) to turn both of these scatterplots into a curve of response for clarity. The predicted responses are fairly stable with different seed settings, but the actual responses coming from the smaller training set are less so; again the issue is worse with the smaller extent problem. Again, we show values from typical classifiers out of multiple seed selections.

4 Discussion

The AUC metric is very good for the occurrence problem with all RF classifiers, but is especially remarkable that such good RF classifiers for this Caribbean fire environment were built primarily off climate attribute predictors *alone* (Table 1). In other studies for regions with fewer anthropogenic ignition sources, fire classifiers depended more on predictors specifically describing humidity and fuel availability (California and western USA: Parisien and Moritz 2009; Colorado and Wyoming: West et al. 2015). Fire classifiers in other regions with a similarly large amount of wildland-urban interface depended more on predictors describing socio-economic conditions (Florida: Mercer and Prestemon 2005; Mediterranean Europe: Oliveira et al. 2012). Here, unemployment did not decide the fire behavior well, and if anything, low unemployment was associated with more fires, contrary to other RF classifiers with socio-economic indicators. One key difference in the Puerto Rico and the other study areas is the extremely large number of ignitions in the Puerto Rico (0.43 fires ignited/km^2/year versus 0.02–0.03 fires ignited/km^2/year in the Mediterranean). We hypothesize that the fire regime of our study area is not ignition-limited enough for the socio-economic indictor to exert influence on the fire regime, following the theory of Bradstock (2010) of limiting 'switches' controlling fire activity. The "ignition switch" not being a factor in Puerto Rico does not imply it is not a factor in other places in the Caribbean. A recent study on another Caribbean area with large wildland-urban interface but lower ignition rate (Columbian Caribbean 0.04 fires ignited/km^2/year: Hoyos et al. 2017) found climate attribute predictors much more important than socio-economic conditions, but fire ignitions still increased with socio-economic degradation.

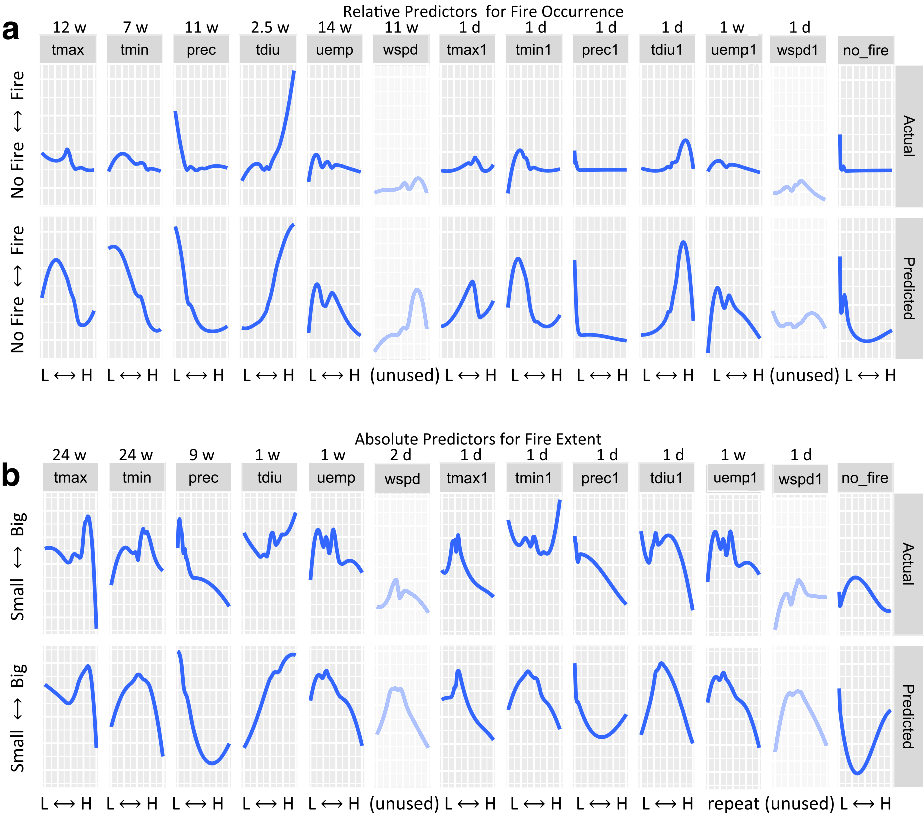

Fig. 3 Normalized local regression smoothed lines of predicted responses (predicted class plotted against predictor value) and actual responses for random forest (RF) classifier results using available data 2003–2011. Plot column headers give mean aggregate length (in weeks (w) or days (d)) for the top aggregates in class separation for each attribute of maximum temperature (tmax), minimum temperature (tmin), precipitation (prec), diurnal temperature range (tdiu), unemployment (uemp), wind speed (wspd). Suffixes of "1" are for the daily predictors consisting of the mean of the day and the day before due to unknown fire event time of day. The fire history attribute is days prior without fire multiplied by barrio area (no_fire). Low (L) to high (H) value of each predictor is on the x-axis. Figure a) shows the responses in the occurrence problem; and b) the responses in the extent problem. The wind predictor responses are from the reduced data classifiers with 3 years of data only

The usefulness of the long-period top aggregate means in the occurrence problem (2–3 months; Fig. 3a) suggests the analysis is determining the conditions needed for fire season, most importantly, a period of time, which is relatively drier than the rest of the year. The wetter, more closed forest in the north and east (Van Beusekom et al. 2014) does not have as dichotomous dry and wet seasons and thus the fire season is not as long (or non-existent, Fig. 1a). However, classifying fires with relative values of precipitation instead of absolute values is possibly more successful in identifying the fires on the fringes of forests or grasslands that are typically humid but very dry at times (Fig. 1a, e), when we do not have specific fuel data (Turco et al. 2017). The predictors are measuring humidity, which does affect fine fuels, by proxy. Cool and dry conditions, as well as greater diurnal fluxes indicate low humidity. The occurrence of a fire on a specific day in a specific location during fire season is driven by the low relative precipitation that day, as seen in the high importance of daily precipitation and in the reduction of AUC from the RF classifier using all climate data to just top aggregate climate data (Table 1). While a recent fire is a good predictor for another fire (thus identifying the barrio as in the fire season temporally), characterizing the weather for fire season, we can build a RF classifier almost as good (Table 1).

The difference in behavior of the RF classifier sets of the two problems, occurrence and extent, can be seen most obviously in the predicted responses of the top aggregates and daily maximum and minimum temperatures; a fire is more likely to occur at a time when the temperature has been relatively colder for several months (Fig. 3a), in the middle of the winter dry season, but more likely to be large if the absolute temperature of the barrio for the last half of the year is fairly warm (Fig. 3b). Absolute dryness of the dry season is the best predictor of larger fires, but a moderate amount of wind in the few days before may also be a factor. The largest winds may be associated with rain events, and therefore fire-reducing.

The causes of the lower performance in the extent problem are clear in the actual responses of the test set, which differ from the predicted responses notably in top aggregate maximum and diurnal temperature, and daily precipitation (Fig. 3b). The extent problem on the shorter time period, with wind, does better (Table 1) but this is due to the test set. The two RF classifiers are not substantially different in predicted responses for the shared attributes. This illustrates one of the difficulties with assessing RF classifier performance in randomly chosen test sets, which are necessarily smaller than the training set; and furthermore, with an imbalanced problem, the minority class will have very few test cases. Thus, unless the class space is very homogeneous, it is likely to get a few atypical cases in the test set minority class, greatly affecting the results. Fire growth is thought to be a self-organized criticality, so any success in classifying the size between small and bigger-small fires, points to a more orderly than normal system (Malamud et al. 1998). Furthermore, wind direction aligned with topography along with fire suppression response dictate fire spread (Finney et al. 2011); these are all complex relationships we cannot account for with the RF classifiers.

The RF classifiers give evidence that with the current ignition sources, if the system is experiencing fire weather, there will be a fire. Fire season may be driven more by the relative weather conditions than the absolute weather, while absolute values may control the fire extent. Thus, the likelihood of fire occurrences and fire size is expected to increase under a more extreme dry season in this region. Regional climate projections encompassing Puerto Rico agree on a temperature increase for all seasons by the middle to end of the century, but models show precipitation changes of a drier wet season (Hall et al. 2013; Karmalkar et al. 2013), drier wet and dry seasons (Campbell et al. 2011), and a wetter wet season (Angeles et al. 2007). Downscaling of projections specifically to the island show high variability in outcomes based model selection and spatial variability related to orographic effects, with more extreme temperature increases than regional projections and increased dryness annually (Van Beusekom et al. 2016; Henareh Khalyani et al. 2016). Our RF classifiers suggest the temperature and annual dryness increases would make fires greater than 5 ha more likely, and fire occurrences would increase when relative dryness increases, e.g., a wetter wet season or drier dry season.

5 Conclusions

Global climate change has been projected to intensify behavior of fire in some areas and not others, with much uncertainty surrounding the projections (Krawchuk et al. 2009; Park

Williams and Abatzoglou 2016). Wildfire in tropical island systems, which are often densely populated and have a complex matrix of urban and rural land covers and uses, has effects on air quality, human health, greenhouse gas emissions, species distributions, ecosystem services and conservation. Understanding how the spatial and temporal distribution of fire occurrence and extent may change as climate changes requires an understanding of what specific climate variables best characterize fire occurrence and extent. In this study, with the use of machine learning classifiers and the data in Puerto Rico, we found that climatic variables alone can define a fire season in a location, with current relative conditions contributing to an increased likelihood of fire occurrence and current absolute conditions to an increased likelihood of larger size. Importantly, Puerto Rico exemplifies tropical island areas of intermediate net primary productivity and precipitation, and gives insights on how climate space may shape these fire regimes in the future.

Acknowledgements We thank Joel Figueroa and Iván Cruz from the Fire Department of Puerto Rico and Luis Rosa, Gary Votaw and Shawn Rossi from the National Weather Program, NOAA/National Weather Service, San Juan for providing data on fire occurrence. Eric Harmsen provided the initial processing of the daily wind speed data. This research was supported by grant DEB 1546686 from US National Science Foundation to the Institute for Tropical Ecosystem Studies, University of Puerto Rico, and to the International Institute of Tropical Forestry (IITF) USDA Forest Service, as part of the Luquillo Long-Term Ecological Research Program. The US Forest Service (Department of Agriculture) Research Unit, the Luquillo Critical Zone Observatory (EAR-1331841), and the University of Puerto Rico gave additional support. We thank Dr. Ariel E. Lugo for providing useful comments on an earlier version of the manuscript. All research at the International Institute of Tropical Forestry is done in collaboration with the University of Puerto Rico.

References

Aldersley A, Murray SJ, Cornell SE (2011) Global and regional analysis of climate and human drivers of wildfire. Sci Total Environ 409:3472–3481

Angeles ME, Gonzalez JE, Erickson DJ, Hernández JL (2007) Predictions of future climate change in the Caribbean region using global general circulation models. Int J Climatol 27:555–569

Argañaraz JP, Gavier Pizarro G, Zak M, Landi MA, Bellis LM (2015) Human and biophysical drivers of fires in semiarid Chaco mountains of Central Argentina. Sci Total Environ 520:1–12

Birch DS, Morgan P, Kolden CA, Abatzoglou JT, Dillon GK, Hudak AT, Smith A (2015) Vegetation, topography and daily weather influenced burn severity in central Idaho and western Montana forests. Ecosphere 6:1–23

Bradstock RA (2010) A biogeographic model of fire regimes in Australia: current and future implications. Glob Ecol Biogeogr 19:145–158

Breiman L (2001) Random forests. Mach Learn 45:5–32

Burney DA, Burney LP, MacPhee RD (1994) Holocene charcoal stratigraphy from Laguna Tortuguero, Puerto Rico, and the timing of human arrival on the island. J Archaeol Sci 21:273–281

Campbell JD, Taylor MA, Stephenson TS, Watson RA, Whyte FS (2011) Future climate of the Caribbean from a regional climate model. Int J Climatol 31:1866–1878

Chawla NV (2005) Data mining for imbalanced datasets: an overview. In: Maimon O, Rokach L (eds) Data mining and knowledge discovery handbook. Springer, New York, pp 853–867

Cochrane MA (2003) Fire science for rainforests. Nature 421:913–919

Daly C, Helmer EH, Quiñones M (2003) Mapping the climate of Puerto Rico, Vieques and Culebra. Int J Climatol 23:1359–1381

Dwyer E, Grégoire J, Malingreau J (1998) A global analysis of vegetation fires using satellite images: spatial and temporal dynamics. Ambio 27:175–181

Fang L, Yang J, Zu J, Li G, Zhang J (2015) Quantifying influences and relative importance of fire weather, topography, and vegetation on fire size and fire severity in a Chinese boreal forest landscape. For Ecol Manag 356:2–12

Fernández-Delgado M, Cernadas E, Barro S, Amorim D (2014) Do we need hundreds of classifiers to solve real world classification problems? J Mach Learn Res 15:3133–3181

Finney MA, Grenfell IC, McHugh CW, Seli RC, Trethewey D, Stratton RD, Brittain S (2011) A method for ensemble wildland fire simulation. Environ Model Assess 16:153–167

Forster P, Ramaswamy V, Artaxo P, Berntsen T, Betts R, Fahey DW, Haywood J, Lean J, Lowe DC, Myhre G, Nganga J, Prinn R, Raga G, Schulz M, Van Dorland R (2007) Changes in atmospheric constituents and in radiative forcing. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) Climate change 2007: the physical science basis. Contribution of working group I to the 4th assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, pp 129–234

Ganteaume A, Jappiot M (2013) What causes large fires in southern France. For Ecol Manag 294:76–85

Gould WA, Alarcón C, Fevold B, Jiménez M, Martinuzzi S, Potts G, Quiñones M, Solórzano M, Ventosa E (2008) The Puerto Rico gap analysis project. Volume 1: land cover, vertebrate species distributions, and land stewardship. US Dep Agric For Serv Int Inst Trop For. IITF-GTR-39, 165 pp. Río Piedras, Puerto Rico

Gregorutti B, Michel B, Saint-Pierre P (2017) Correlation and variable importance in random forests. Stat Comput 27(3):659–678

Hall TC, Sealy AM, Stephenson TS, Kusunoki S, Taylor MA, Chen AA, Kitoh A (2013) Future climate of the Caribbean from a super-high-resolution atmospheric general circulation model. Theor Appl Climatol 113: 271–287

He H, Garcia EA (2009) Learning from imbalanced data. Knowl Data Eng IEEE Trans On 21:1263–1284

Henareh Khalyani A, Gould WA, Harmsen E, Terando A, Quinones M, Collazo JA (2016) Climate change implications for Tropical Islands: interpolating and interpreting statistically downscaled GCM projections for management and planning. J Appl Meteorol Climatol 55:265–282

Hoyos N, Correa-Metrio A, Sisa A, Ramos-Fabiel MA, Espinosa JM, Restrepo JC, Escobar J (2017) The environmental envelope of fires in the Colombian Caribbean. Appl Geogr 84:42–54

Karmalkar AV, Taylor MA, Campbell J, Stephenson T, New M, Centella A, Benzanilla A, Charlery J (2013) A review of observed and projected changes in climate for the islands in the Caribbean. Atmósfera 26:283–309

Krawchuk MA, Moritz MA, Parisien M-A, Dorn JV, Hayhoe K (2009) Global Pyrogeography: the current and future distribution of wildfire. PLoS One 4:e5102

Legendre P (1993) Spatial autocorrelation: trouble or new paradigm? Ecology 74:1659–1673

Luo R, Hui D, Miao N, Liang C, Wells N (2017) Global relationship of fire occurrence and fire intensity: a test of intermediate fire occurrence-intensity hypothesis. J Geophys Res: Biogeosci 122(5):1123–1136

Malamud BD, Morein G, Turcotte DL (1998) Forest fires: an example of self-organized critical behavior. Science 281:1840–1842

Martinuzzi S, Gould WA, Ramos González OM (2007) Creating cloud-free Landsat ETM+ data sets in tropical landscapes: cloud and cloud-shadow removal. US Dep. Agric. For. Serv. Int. Inst. Trop. For. IITF-GTR-32, 12 pp. Río Piedras, Puerto Rico

Matsumoto M, Nishimura T (1998) Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Trans Model Comput Simul TOMACS 8:3–30

Menardi G, Torelli N (2014) Training and assessing classification rules with imbalanced data. Data Min Knowl Discov 28:92–122

Méndez-Tejeda R, Santos-Corrada M, Ortiz-Morales S, Claudio-Vargas O (2015) Environmental and economic impact of Forest fires in Puerto Rico 2013-2014. Open J For 5:353

Mercer DE, Prestemon JP (2005) Comparing production function models for wildfire risk analysis in the wildland–urban interface. For Policy Econ 7:782–795

Oliveira S, Oehler F, San-Miguel-Ayanz J, Camia A, Pereira JMC (2012) Modeling spatial patterns of fire occurrence in Mediterranean Europe using multiple regression and random Forest. For Ecol Manag 275: 117–129

Parisien M-A, Moritz MA (2009) Environmental controls on the distribution of wildfire at multiple spatial scales. Ecol Monogr 79:127–154

Park Williams A, Abatzoglou JT (2016) Recent advances and remaining uncertainties in resolving past and future climate effects on global fire activity. Curr Clim Change Rep 2:1–14

Pausas JG, Ribeiro E (2013) The global fire–productivity relationship. Glob Ecol Biogeogr 22:728–736

Pourtaghi ZS, Pourghasemi HR, Aretano R, Semeraro T (2016) Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques. Ecol Indic 64:72–84

Robbins AMJ, Eckelmann C-M, Quiñones M (2008) Forest fires in the insular Caribbean. AMBIO J Hum Environ 37:528–534

Rodrigues M, de la Riva J, Fotheringham S (2014) Modeling the spatial variation of the explanatory factors of human-caused wildfires in Spain using geographically weighted logistic regression. Appl Geogr 48:52–63

Taufik M, Torfs PJ, Uijlenhoet R, Jones PD, Murdiyarso D, Van Lanen HA (2017) Amplification of wildfire area burnt by hydrological drought in the humid tropics. Nat Clim Chang 7(6):428–431

Turco M, Hardenberg JV, AghaKouchak A, Llasat MC, Provenzale A, Trigo RM (2017) On the key role of droughts in the dynamics of summer fires in Mediterranean Europe. Sci Rep 7:81–81

Van Beusekom AE, Gould WA, Terando AJ, Collazo JA (2016) Climate change and water resources in a tropical island system: propagation of uncertainty from statistically downscaled climate models to hydrologic models. Int J Climatol 36:3370–3383

Van Beusekom AE, Hay LE, Viger RJ, Gould WA, Collazo JA, Henareh Khalyani A (2014) The effects of changing land cover on streamflow simulation in Puerto Rico. JAWRA J Am Water Resour Assoc 50:1–19

van der Werf GR, Randerson JT, Giglio L, Gobron N, Dolman AJ (2008) Climate controls on the variability of fires in the tropics and subtropics. Glob Biogeochem Cycles 22:GB3028

West AM, Kumar S, Jarnevich CS (2015) Regional modeling of large wildfires under current and potential future climates in Colorado and Wyoming. USA Clim Change 134:565–577

Willmott CJ, Robeson SM (1995) Climatologically aided interpolation (CAI) of terrestrial air temperature. Int J Climatol 15:221–229

Yu YP, Omar R, Harrison RD, Sammathuria MK, Nik AR (2011) Pattern clustering of forest fires based on meteorological variables and its classification using hybrid data mining methods. J Comput Biol Bioinforma Res 3:47–52