International Journal of Wildland Fire 2017, **26**, 587–597
doi:10.1071/WF16135_AC

Supplementary material

An empirical machine learning method for predicting potential fire control locations for pre-fire planning and operational fire management

Christopher D. O'Connor[A,B], *David E. Calkin*[A] *and Matthew P. Thompson*[A]

[A]US Department of Agriculture Forest Service, Rocky Mountain Research Station, Forestry Sciences Laboratory, 800 East Beckwith Avenue, Missoula, MT 59801, USA.

[B]Corresponding author. Email: christopheroconnor@fs.fed.us

doi:10.1071/WF16135_AC

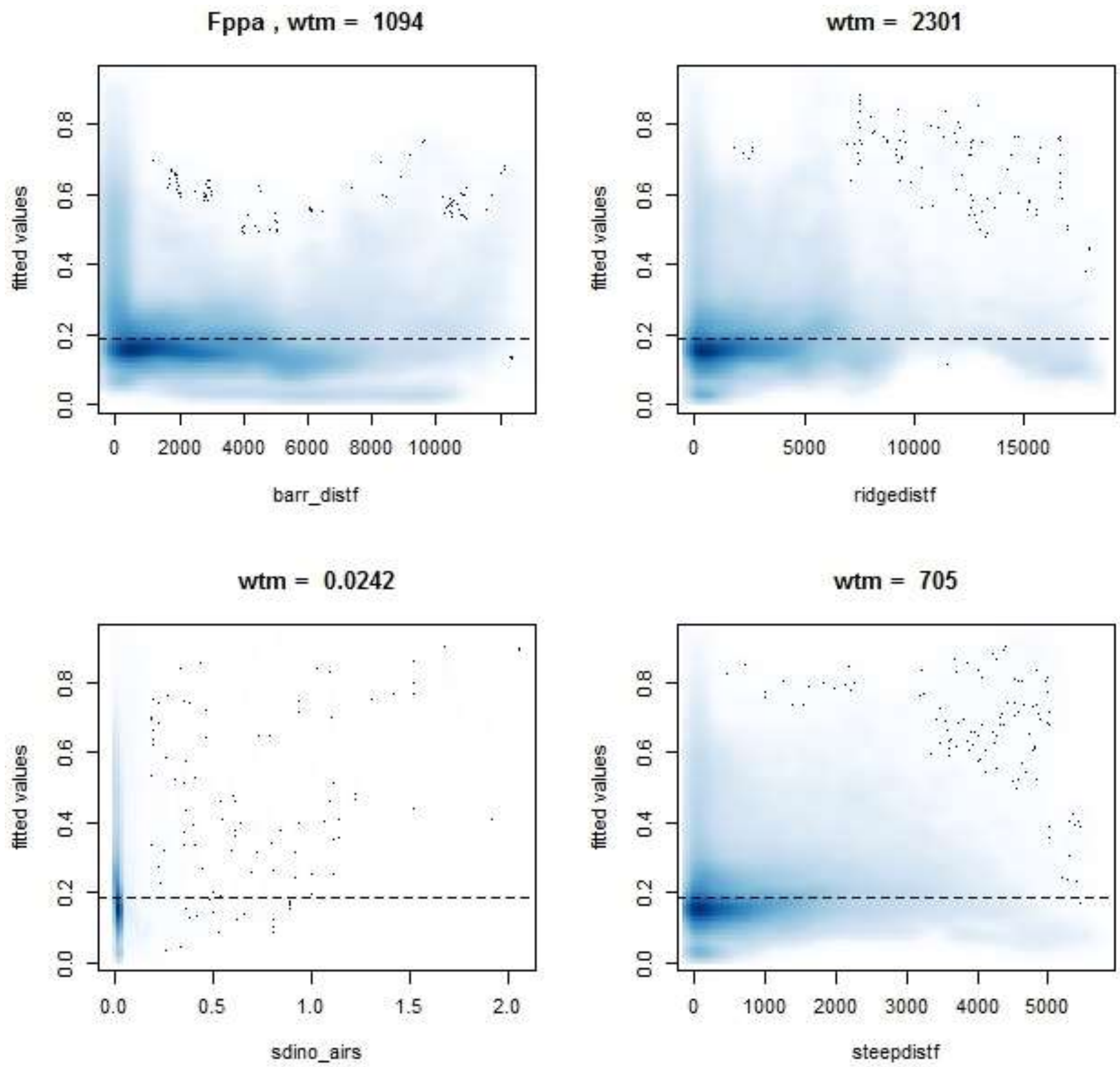


Fig. S1. Density smoothed scatter plots of model fitted fire perimeter presence or absence (fppa) in relation to each predictor variable. Y-axis represents probability of fire presence and X-axis represents predictor value. 'wtm' is the weighted mean value for each predictor. Dotted line is the average fire perimeter probability. 'barr_distf' is barrier distance (m), 'ridgedistf' is ridge distance (m), 'sdino_airs' is SDO with no air support, and 'steepdistf' is distance from steep slope (m).

doi:10.1071/WF16135_AC

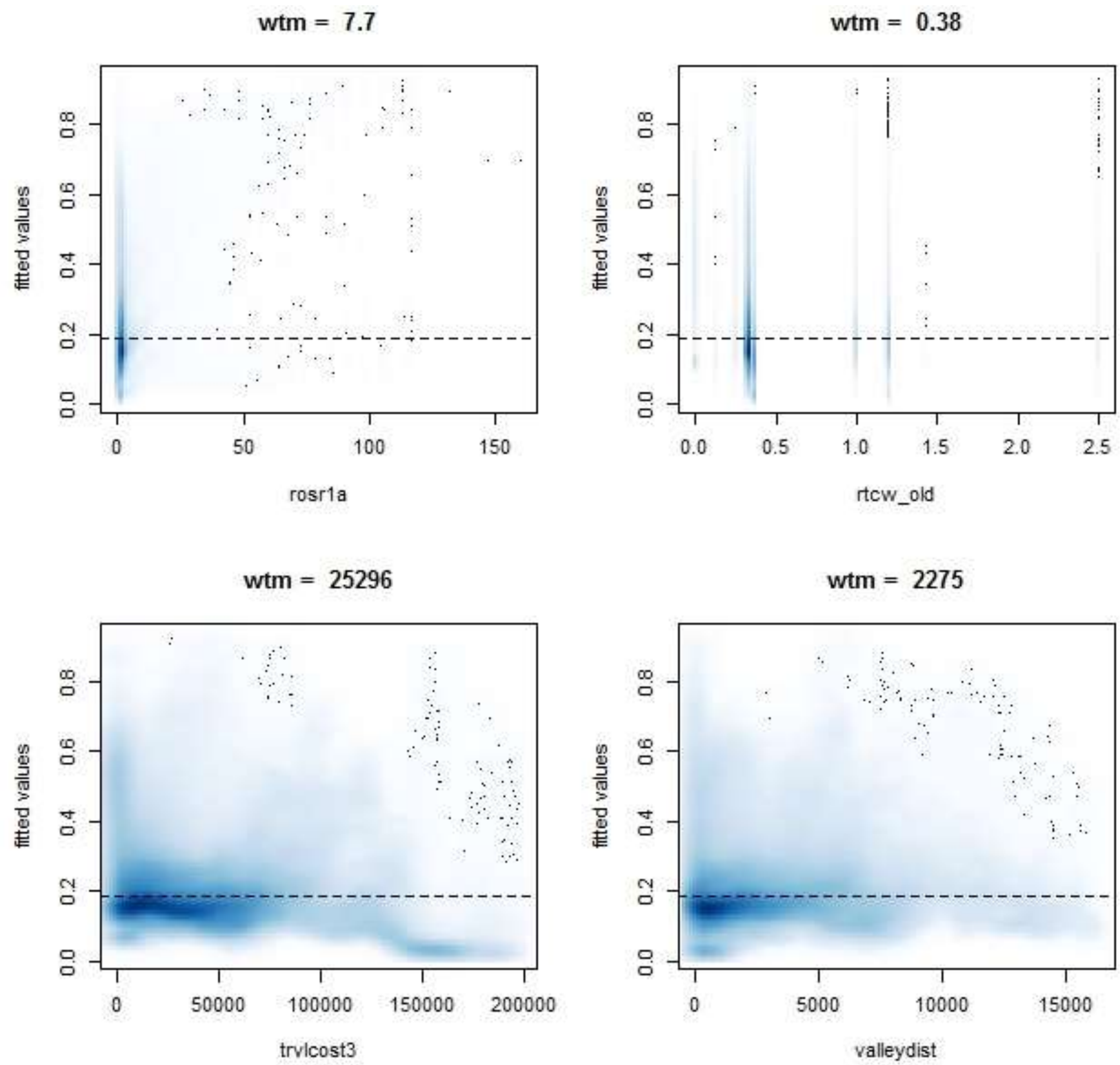


Fig. S2. Density smoothed scatter plots of model fitted fire perimeter presence or absence in relation to each predictor variable. Y-axis represents probability of fire presence and X-axis represents predictor value. Wtm is the weighted mean value for each predictor. Dotted line is the average fire perimeter probability. ‘rosr1a’ is rate of spread, ‘rtc_old’ is the calculated resistance to control, ‘trvlcost3’ is travel cost, and ‘valleydist’ is distance from valley bottom in metres.

doi:10.1071/WF16135_AC

Table S1. Spearman ranked correlation coefficients between predictor variables

Variable abbreviations are detailed in the manuscript text

	RidgeDist	ValleyDist	SteepDist	RTC	Trvlcost	RoS	Barr_dist	SDI
RidgeDist	1.000	0.597	–0.087	0.134	0.000	–0.022	–0.092	0.022
ValleyDist	0.597	1.000	–0.021	0.115	–0.009	–0.022	–0.085	0.026
SteepDist	–0.087	–0.021	1.000	–0.204	–0.065	–0.243	–0.019	–0.201
RTC	0.134	0.115	–0.204	1.000	0.018	–0.075	0.019	0.201
Trvlcost	0.000	–0.009	–0.065	0.018	1.000	0.100	0.735	0.248
RoS	–0.022	–0.022	–0.243	–0.075	0.100	1.000	0.071	0.299
Barr_dist	–0.092	–0.085	–0.019	0.019	0.735	0.071	1.000	0.149
SDI	0.022	0.026	–0.201	0.201	0.248	0.299	0.149	1.000

doi:10.1071/WF16135_AC

Table S2. Test of variable collinearity

Variance inflation factor (VIF) values greater than five suggest inflated model variance as a function of high correlation between variables

Variable	VIF
RidgeDist	1.428
ValleyDist	1.438
SteepDist	1.073
RTC	1.468
Trvlcost	1.743
RoS	1.197
Barr_dist	1.754
SDI	1.604

Section S1. Maxent modelling parameters and results

Maxent uses a presence-only generalised linear model approach that can account for a high density of pseudo-negative values where conditions are appropriate for a binary positive response, such as a fire perimeter location, but no information regarding true absence is available (Ward *et al.* 2009, Elith *et al.* 2011). In this case the national MTBS fire perimeter dataset excludes fires smaller than 400 ha and fires prior to 1984. The pseudo-absences caused by this incomplete data can hinder the performance of more traditional presence–absence based logistic regression machine learning methods such as boosted regression or Random Forest (Elith *et al.* 2008). Maxent is a flexible nonparametric machine learning approach that generates probability density surfaces for presence and background samples in covariate space (Phillips and Dudík 2008, Elith *et al.* 2011). Maxent is less influence by collinearity of predictor variables than stepwise regression, reducing the need to limit variables used in predictive modelling (Friedman *et al.* 2000, Dormann *et al.* 2013). The method also has a built in regularisation method that allows the model to be fit more closely to an individual landscape or to make predictions more generalisable to a common series of conditions. For these simulations, the regularisation parameter was set to “1”, which is considered a good balance between over or under-fitting results (Phillips 2015)

We used the maxent parameters recommended by Elith *et al.* (2011) in the ‘dismo’ package in the R statistical environment (R Core development Team 2015) to optimise predictive performance. We used five-fold cross-validation of the training dataset such that 80% of the training data were tested against a random 20% of withheld fire perimeter locations. We used 10 replicates of the maxent analysis to develop confidence intervals around average area under the curve (AUC) estimates of overall fire perimeter prediction accuracy.

Regularised training data gain, a statistic comparable to model deviance and measure of goodness of fit was 0.351. This measure of model concentration around the fire perimeter presence samples suggests that the modelled fire perimeter locations are a 42% improvement over random background data. In the five-fold cross-validation assessment of model performance, the AUC value of 0.728 suggests that the model correctly predicted approximately 73% of presence-only observations excluded from the training data (Fig. S3).

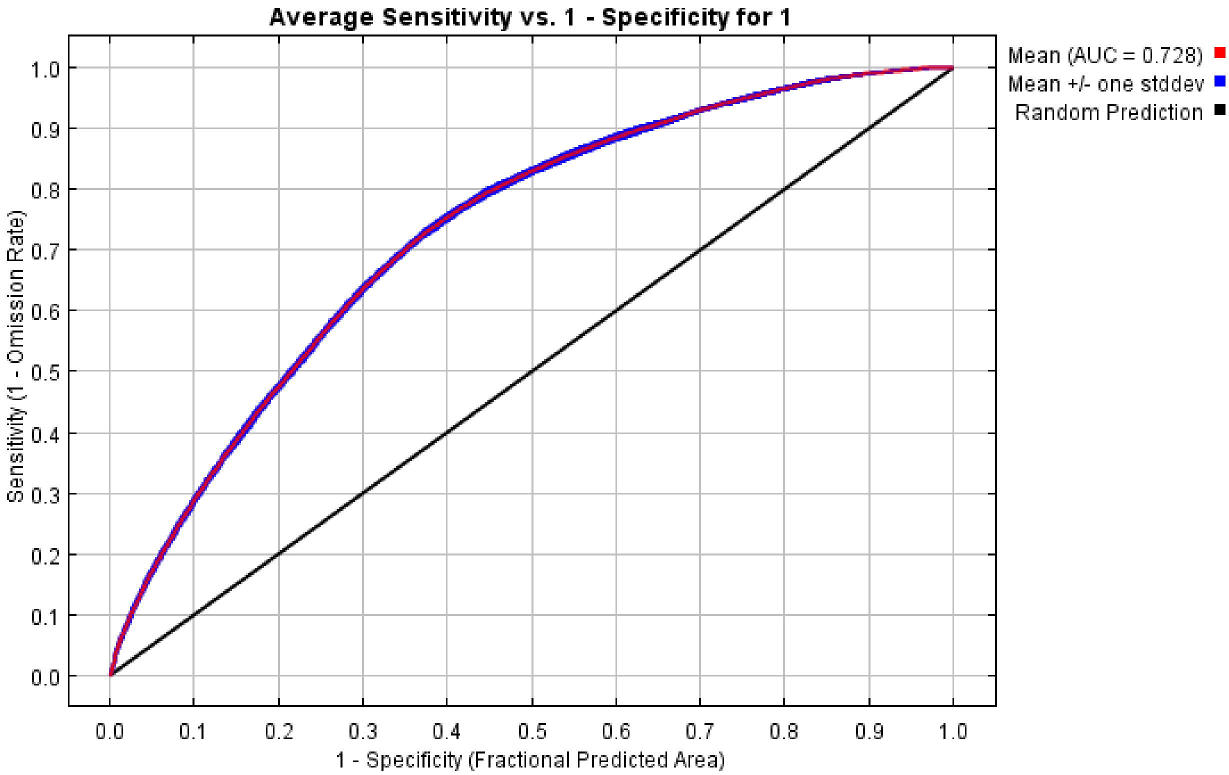
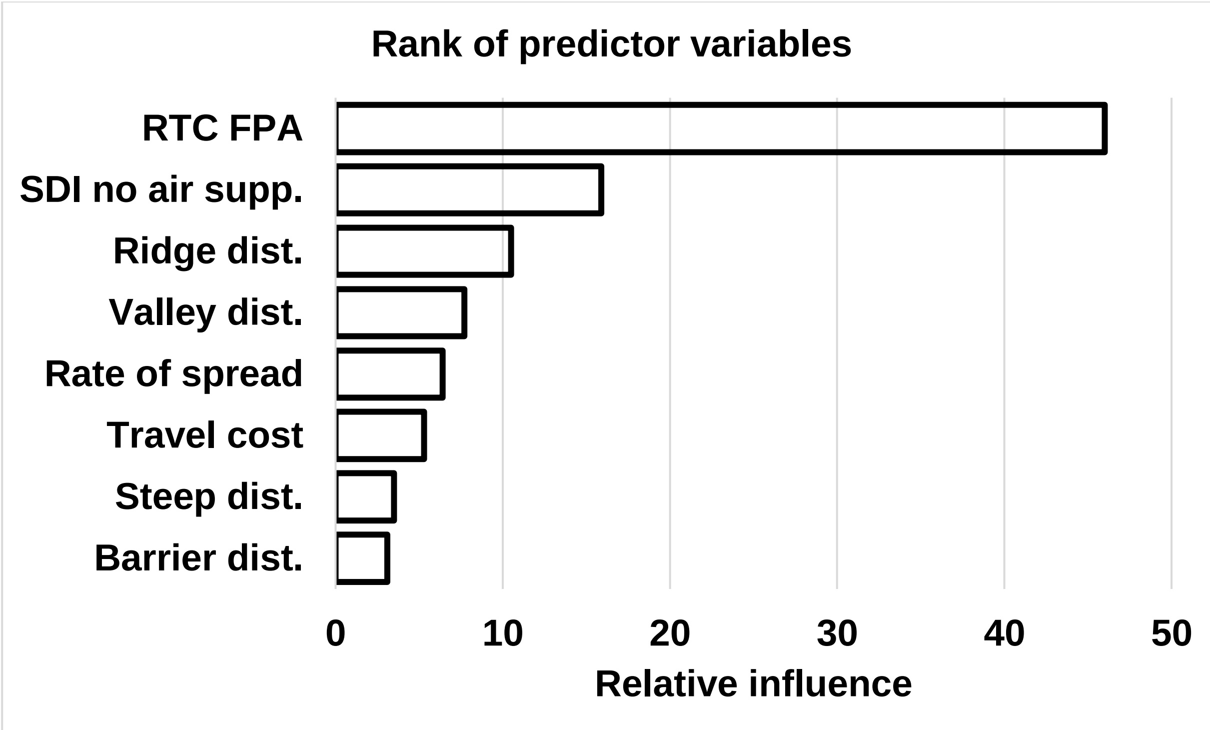


Fig S3. Average Area under the curve (AUC) calculation for 10 maxent runs. Maxent identified the same set of eight significant predictors of fire perimeter suitability as the BRT model, however, RTC surpassed SDI and travel cost as the primary determinant of suitable fire perimeter locations (Fig. S4). Response curves for RTC and SDI exhibit a similar pattern to those in the BRT model; however, the distance to ridge response curve exhibited a more complex stepped response such that fire perimeters were initially associated with a distance of 2-5 km from a ridge top, although on this landscape, the majority of fire perimeters were more than 15 km from a ridgetop. This makes intuitive sense in relation to the high density of fires occurring in sage lowlands but does not capture the effects of ridges on fire behavior.



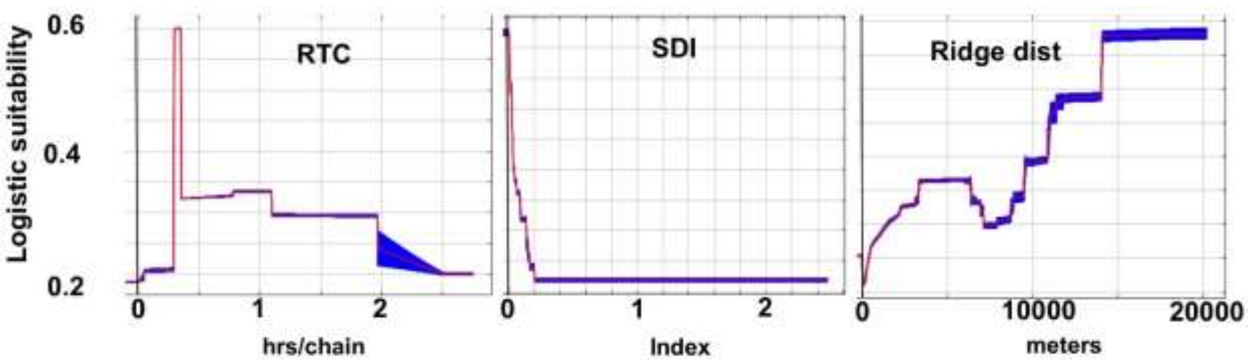


Fig. S4. Summary of maxent predictive model inputs and relationships to observed fire perimeters. Eight significant predictors of final fire perimeter locations are listed in order of relative contribution to improving model performance (a.). 'RTC FPA' is resistance to control using Fire Program Analysis data, 'SDI no air supp' is suppression difficulty index without air support, 'dist.' is shorthand for distance in metres. Response curves for the top three predictors (b.) demonstrate the relationship between the suitability for a fire perimeter (y-axis) in relation to the value of a predictor variable (x-axis). RTC is calculated in hours per chain, SDI is a dimensionless index value, and ridge distance is in metres.

The fire perimeter suitability surface produced by the maxent model is similar to that produced by the BRT (Fig. S5), however the raw output cannot be reliably classified into probability of true fire perimeter presence or absence, making it appropriate for a general overview of potential fire control locations but less suitable for developing operational fire management thresholds.

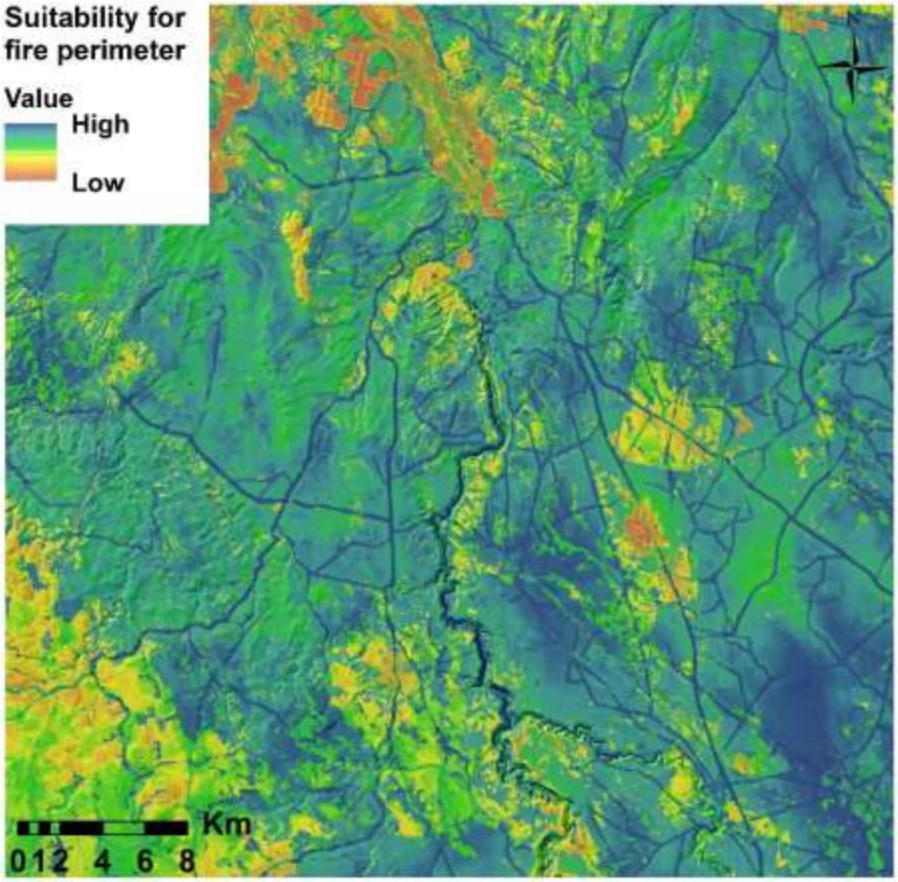


Fig. S5. Maxent model of suitability for fire perimeter locations developed from presence-only regression modelling. For model settings refer to text above. Modelled landscape is identical to that used in Fig. 4. Raw maxent suitability scores sum to one over the entire landscape and are considered a more faithful representation of true suitability that the pseudo-logistic probability output option (Merow *et al.* 2013).