CSIRO PUBLISHING

International Journal of Wildland Fire **2017**, *26*, 587–597
http://dx.doi.org/10.1071/WF16135

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

Abstract. During active fire incidents, decisions regarding where and how to safely and effectively deploy resources to meet management objectives are often made under rapidly evolving conditions, with limited time to assess management strategies or for development of backup plans if initial efforts prove unsuccessful. Under all but the most extreme fire weather conditions, topography and fuels are significant factors affecting potential fire spread and burn severity. We leverage these relationships to quantify the effects of topography, fuel characteristics, road networks and fire suppression effort on the perimeter locations of 238 large fires, and develop a predictive model of potential fire control locations spanning a range of fuel types, topographic features and natural and anthropogenic barriers to fire spread, on a 34 000 km^2 landscape in southern Idaho and northern Nevada. The boosted logistic regression model correctly classified final fire perimeter locations on an independent dataset with 69% accuracy without consideration of weather conditions on individual fires. The resulting fire control probability surface has potential for reducing unnecessary exposure for fire responders, coordinating pre-fire planning for operational fire response, and as a network of locations to incorporate into spatial fire planning to better align fire operations with land management objectives.

Additional keywords: boosted regression, fire responder safety, MaxEnt, operational decision support, pre-fire planning, risk analysis, spatial analysis.

Received 23 July 2016, accepted 17 January 2017, published online 13 February 2017

Introduction

Two of the most important challenges for pre-fire planning and operational fire management are positioning resources where they are most likely to be effective and identifying and avoiding locations with high potential for extreme fire behaviour and unacceptable risk to fire responder safety. The compressed decision space imposed upon fire managers requires reaction to rapidly evolving conditions, often with limited time to assess the full range of options to meet land management objectives or to develop backup plans if initial attempts at containment are unsuccessful. Further, in the United States, as large fires progress and increase in size and complexity, non-local Incident Management Teams (IMTs; primarily Type 1 and some Type 2) are frequently brought in to manage these events but often lack local knowledge regarding landscape control features. Pre-event documentation could provide out-of-area teams with critical information to facilitate strategic planning. Decisions on when and where to allocate fire management resources are limited by available information but often incorporate intuitive landscape features such as roads, ridgelines, water bodies and other potential constraints to fire spread recommended by the National Wildfire Coordinating Group (NWCG 2004).

Management decisions during a fire can be informed by deterministic fire modelling systems such as FARSITE (when used without spotting behaviour, Finney 2004) for 1–3-day fire growth projections; or probabilistic fire spread simulators such as FSPro (Finney *et al.* 2011*a*) for longer duration estimates (Calkin *et al.* 2011; Noonan-Wright *et al.* 2011). Additionally, the large fire behaviour simulator FSim (Finney *et al.* 2011*b*) can be utilised for pre-fire planning (Scott *et al.* 2013; Thompson *et al.* 2016*a*). However, accurate prediction of fire control features likely to result in successful containment remains a significant challenge. This is due in part to the coarse spatial resolution at which these modelling systems are typically run, which masks relatively narrow features like roads, such that simulated fire perimeters rarely align with actual fire perimeters. Initial attack modelling approaches (e.g. Fried and Fried 1996; Petrovic and Carlson 2012) generally consider rate of spread against rate of fireline production with little to no consideration of leveraging natural or anthropogenic barriers to spread – although Podur and Martell (2007) assume 25% of fireline is attributable to natural features. Fine-scale landscape features such as topographic variability and fuel heterogeneity are significant factors influencing the size and severity of fires

(Dillon *et al.* 2011*b*; Wu *et al.* 2013; Harris and Taylor 2015; Kane *et al.* 2015) and previous burn scars that affect fuel loading and structure have been shown to limit fire spread – though these effects vary with time since fire and weather conditions (Parks *et al.* 2015). Road networks have also been shown to limit fire spread, both through creation of fuel breaks and by facilitating placement of fire management personnel and resources (Narayanaraj and Wimberly 2011, 2012).

Wildfire risk modelling has also been used to identify locations of increased risk for high fire severity (Dillon *et al.* 2011*a*; Keane *et al.* 2013), potential for damage to human assets and natural resources (Haas *et al.* 2013; Scott *et al.* 2013; Thompson *et al.* 2015; Thompson *et al.* 2016*a*), and challenges for fire suppression activities (Rodríguez y Silva *et al.* 2014; Dillon *et al.* 2015; Duff and Tolhurst 2015; Mitsopoulos *et al.* 2016), including potential safety concerns for fire responders. The wildfire risk assessment tools developed by Scott *et al.* (2013) have been applied at spatial scales more appropriate for integrating landscape fire planning, which incorporates simulated response to fire, with operational incident management that must consider real time burning conditions, available resources and acceptable risk to assets and fire responders. Aligning operational fire response with land management objectives requires bringing together the 'what' defined by landscape management objectives with the 'how' defined by real time fire conditions on the ground.

Our objective with this study is to generate a decision space for active fire management and pre-fire planning that highlights potential fire control locations that can be mapped onto landscape management objectives to facilitate greater integration between land management and fire operations. We use a spatial database of historical fire perimeter locations to model interactions between topographic features (slope position and soil characteristics), fuel types, road networks and hydrologic features that resulted in cessation of fire progression. We then test the accuracy of the resulting predictive model of potential fire control locations and discuss strengths and limitations of model outputs. We conclude with future research directions that could improve model performance and a discussion of potential applications for fire management professionals.

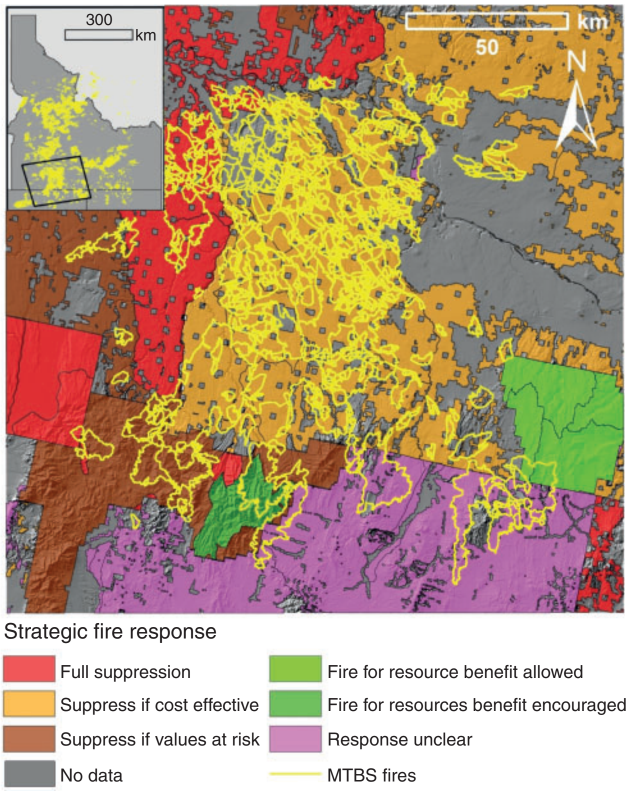

Fig. 1. Study area and strategic fire response designations overlapping the states of Idaho, Nevada and Utah, USA. Yellow polygons are fire perimeters >400 ha recorded from 1984 to 2014 in the Monitoring Trends in Burn Severity (MTBS) database (MTBS 2014). Shaded relief is developed from the LANDFIRE (2010) 30-m digital elevation model. Owner-identified strategic fire response classifications from Thompson *et al.* (2013) were updated in 2014 (C. Stonesifer, pers. comm).

Methods

The study area was an ~34 000 km^2 landscape located in the northern Rocky Mountains bordering Idaho, Nevada and Utah with relatively high density of large (>400 ha) fires recorded from 1984 to 2014 in the national Monitoring Trends in Burn Severity database (MTBS 2014) (Fig. 1). A subset of 238 fires with perimeters completely contained within the study area was used for development of the fire perimeter prediction model. The landscape was selected to include diverse vegetation types ranging from sagebrush steppe and riparian cottonwood and willow ecosystems at the lowest elevations (709 m/2326 ft above sea level, asl) and transitioning through pinyon–juniper woodlands, whitebark pine, mixed-conifer and aspen to subalpine forests at the highest elevations (3295 m/10 810 ft asl). Land ownership is partitioned between federal, state and private management with a prevailing strategic fire response of 'fire suppression with consideration of cost and values at risk' (Thompson *et al.* 2013) dictated by rangeland and timber management values. Approximately 8% of the landscape is designated for managed fire for resource benefit; thus some form of active fire suppression can be assumed for the majority of fire perimeters (Fig. 1).

Fire perimeter response variables

We buffered fire perimeters by 90 m to account for co-registration errors between actual and mapped fire perimeters and then rasterised the buffered features to 30-m pixel resolution. A previous study using MTBS data for fire perimeter classification in wilderness areas used a spatial buffer of 375 m to classify fire interactions (Parks *et al.* 2015). Changing the size of the fire perimeter overlap buffer results in a tradeoff between model classification accuracy and predictive power. To account for fine-scale spatial features such as roads and steep ridge tops, we opted for a relatively small buffer that emphasised potential operational application of the predictive fire perimeter model. On landscapes with fewer roads or other fine-scale barriers to fire spread, a larger buffer may be more appropriate. Raster values were classified into binomial values such that '1' represented pixels resting on a fire perimeter and '0' represented pixels on the interior of a fire perimeter. Pixels on the exterior of buffered fire perimeters were excluded. A binomial response variable was developed from the initial raster classification of fire perimeter presence or absence. To incorporate the effect of overlapping fire perimeter locations through time, we developed a second response variable that incorporated the number of instances of a pixel or its neighbour recording a fire perimeter. This method used a summed 3 $\times$ 3-cell focal statistic to add additional weight to locations that resulted in more than one fire perimeter over the 30-year interval. The focal statistic also included locations within 90 m of a mapped fire perimeter.

Fire perimeter predictor variables

We developed a series of landscape feature-based predictor variables from national raster layers representing elevation, slope, aspect, fuel model and vegetation type (biophysical setting and existing vegetation) from the national LANDFIRE 1.3.0 database (LANDFIRE 2012); and vector features representing roads (US Census Bureau 2015) and rivers (USGS 2015). Vector layers were converted to rasters with a pixel resolution of 30 m and re-projected to the LANDFIRE (2012) USA contiguous Albers Equal Area Conic projection.

Elevation and slope rasters were further transformed to produce the Topographic Position Index (TPI) (Weiss 2001; Beier and Brost 2010) from the corridor designer extension for ArcGIS 10.2 (Majka *et al.* 2007). TPI calculations used a circular search radius of 200 m to identify ridges, valleys, flats and steep slope transitions. Ridgetop and canyon classifications required a minimum 12 m vertical offset from neighbourhood mean elevation. Slope and flat classifications used a minimum slope threshold of 6% (Jenness *et al.* 2013). We used Euclidean distance from TPI-defined ridges, valleys, flats or steep slopes as fire perimeter predictor variables.

We compiled road layers from county GIS data in the US Census TIGER system (US Census Bureau 2015). Primary (arterial routes) and secondary (paved) roads interrupt continuous surface fuels and were considered likely fire breaks (Narayanaraj and Wimberly 2011). Euclidean distance from major paved surfaces was used to develop the road distance surface. An additional road density metric (total road length per km^2) that also included tertiary (unpaved) roads was used as part of the Suppression Difficulty Index (SDI) compound fire suppression index detailed in the next section. Unpaved roads were also included in the cost surface calculation for their effect on relative ease of access.

Similar to roads, hydrography features from the United States Geological Survey Hydrography archive (USGS 2015) were classified by type, such that primary and secondary waterways (rivers) and lakes were considered potential fire breaks and perennial, intermediate or ephemeral streams were considered less likely to influence fire progression. Euclidean distance from lakes and rivers was used to develop a river distance surface.

We reclassified fire behaviour fuel models used by LANDFIRE and developed by Scott and Burgan (2005) to reflect total fuel loading of 1, 10 and 100-h fuels to generate a surface of total fuel loading (kg ha^{-1}). A second fuel model variable was developed from calculated variance of model codes derived from a 9 $\times$ 9-pixel focal statistic. The calculated level of fuel model variance was a function of LANDFIRE fuel model labelling conventions in which consecutive fuel number codes are used for similar fuel types and number codes separated by units of 10 are used for dissimilar fuel types. The expectation was that increased heterogeneity of fuel types, measured as fuel model variance, could be associated with cessation of fire progression. Although interactions with previous fires were not explicitly included in the suite of predictor variables, the LANDFIRE 1.3.0 national fuel model dataset incorporates fuel-altering disturbances at an annual time step from 1999 to 2012 (LANDFIRE 2012). True fuel characteristics before each fire are not known; however, the mosaic of fire effects that influenced past fires is captured in fuel loadings and fuel type transitions and serves as a proxy for direct fire–fire interactions.

Compound fire indices

In addition to direct physical landscape attributes, we developed four compound fire suppression variables: resistance to control (RTC), an index of fireline construction difficulty by fuel type (Dillon *et al.* 2015); travel cost, a measure of effort to transport personnel and resources to any location on the landscape; rate of fire spread, a simulated fire behaviour metric (Finney 2006); and SDI, a weighted index of relative accessibility in relation to potential fire intensity (Rodríguez y Silva *et al.* 2014). RTC is a component of the wildfire hazard potential calculation (Dillon *et al.* 2015) that is expressed as the inverse of fireline construction rate; that is, hours required for a 20-person hand crew to construct one chain (20.1 m) of fireline. Smaller values of RTC indicate lower resistance to control (greater fireline production rate). Two different RTC surfaces were developed from (1) the 2010 update to United States Forest Service Fire Program Analysis estimated hand crew line construction rates (Dillon *et al.* 2015) (Fig. 2*a*) and (2) a simplified series of four hand crew line construction rates from an independent dataset (Broyles 2011).

To generate a travel cost surface, we developed travel friction layers for road type, relative slope and river crossings that penalised movement away from primary paved roads. Linear road network features were converted to 30-m pixel rasters and then classified by ease of transport for personnel and resources (Table 1). A continuous road friction surface was developed from assigned weights to each road type such that roads capable of transporting large equipment at high speed had the lowest friction and unroaded areas had the highest friction (Table 1). The slope friction surface was classified into five classes based on Jenks natural breaks (ESRI 2015) such that slope friction doubled with each successive slope class. A water crossing friction surface was developed from waterway type (lake, river or stream) and intersections with different road types. We used an intersection analysis of all paved roads and water crossings to develop a bridge layer and then buffered bridge point locations by 120 m to account for errors in spatial co-registration between actual roads and rivers. Buffered bridge locations were converted to rasters and mosaicked onto the road layer so that bridges had the lowest resistance to travel (friction) when crossing waterways (Table 1).

Friction surfaces were summed and then mosaicked with a raster containing values only for primary and secondary roads, so that road friction values (0 or 1) replaced additive friction surface values, assuring that roads were local minima in the additive friction surface. We then calculated cost distance from

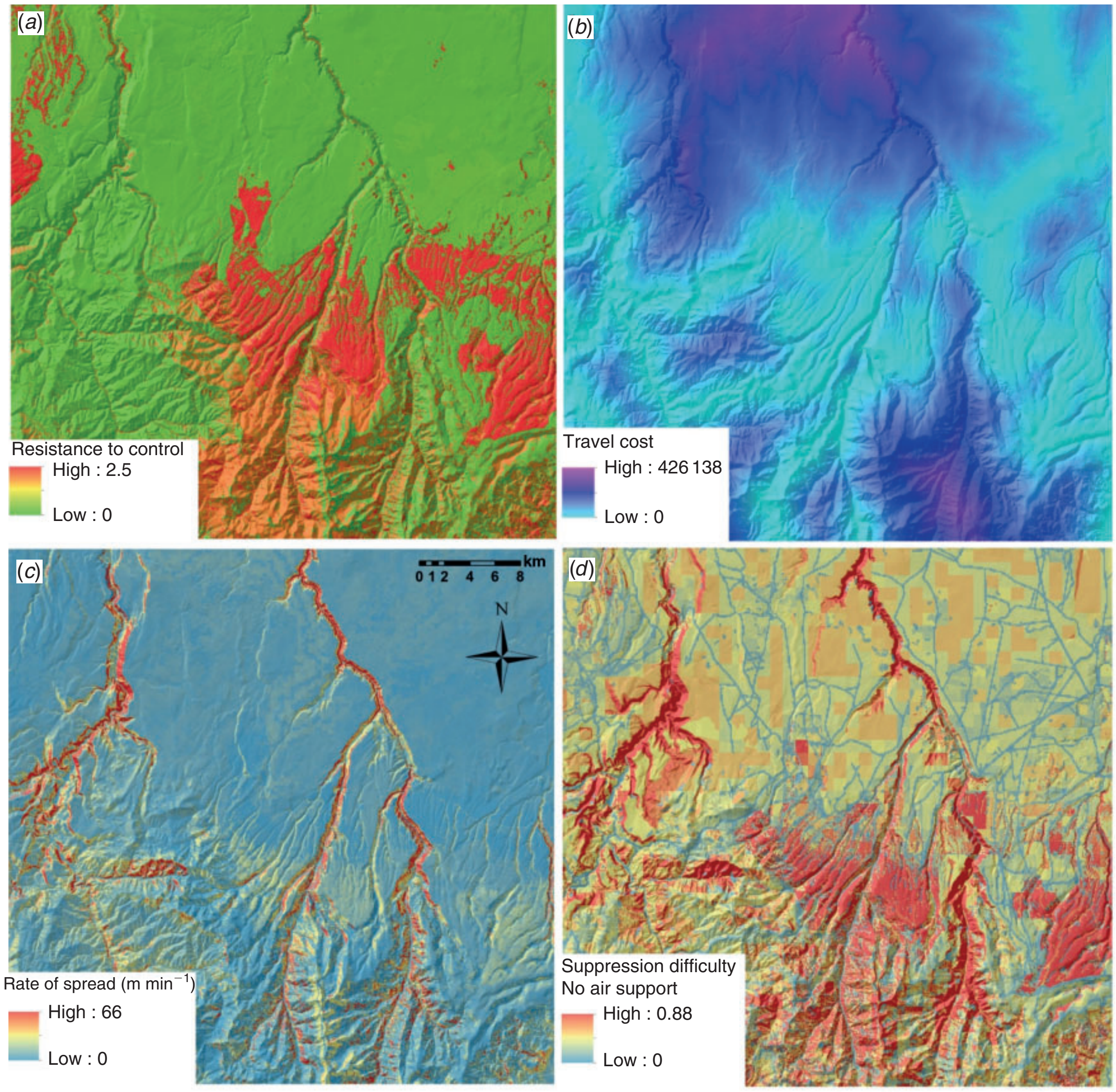

Fig. 2. Compound fire index variables used to develop fire perimeter prediction model. Resistance to control is calculated from USDA Forest Service Fire Program Analysis fireline construction rates (Dillon *et al.* 2015); travel cost is a function of access from primary roads; rate of fire spread is calculated from FlamMap 5.0 (Finney 2006). Suppression difficulty is adapted for use in the western United States from Rodríguez y Silva *et al.* (2014).

Table 1. Cost surface factors

Weighting factors used to develop friction surfaces incorporate relative ease of movement of machinery and personnel in relation to roads, hillslope and water crossings. Road friction factors incorporate road quality and ease of fire engine movement. Slope classes are generated from Jenks natural breaks and represent ease of movement on foot or bulldozer. Water crossing friction on bridges or unpaved roads assumes travel by motorised vehicle. Stream, river and lake crossings without roads assume movement on foot or by boat. Movement of personnel and resources by air is not incorporated into cost distance calculations

Road type	Road friction factor	Slope class (%)	Slope friction factor	Water crossing type	Water crossing friction factor
Interstate/primary	0	0–8	1	Bridge	0
Secondary (paved)	1	8–21	2	Unpaved road	1
Tertiary (unpaved)	4	21–39	4	Tertiary stream	5
No road	10	39–63	8	Secondary stream	10
		>63	16	River	15
				Lake	30

primary roads using the additive friction surface to weigh travel cost to all pixels on the landscape (Fig. 2*b*).

Rate of fire spread and other simulated metrics of fire behaviour have been used to quantify the amount of suppression effort necessary to stop fire progression (Mitsopoulos *et al.* 2016), as well as to explain topographic or fuel transition-derived cessation of fire progression (Anderson 1989; Finney *et al.* 2009). We used FlamMap 5.0 (Finney 2006; Finney *et al.* 2015) in the basic fire behaviour modelling mode to simulate fire surface spread rates across the calibration landscape (Fig. 2*c*). We used the Finney (2004) crown fire calculation method with gridded winds averaging 6.7 m s^{-1} (15 mph) at 6.1 m (20 ft) above ground surface and an azimuth of 164°. Wind speed and direction coincide with average conditions during the fuel moisture conditioning period from 2 July 2007 1000 h to 30 July 2007 1500 h at the Trail Gulch Remote Automatic Weather Station located near the centre of the simulation landscape (DRI 2015). Weather and wind files for the fuel moisture conditioning period were generated in Fire-Family Plus ver. 4.1 (Bradshaw 2013) and represented 90th-percentile fire weather conditions calculated from energy release component values (Riley *et al.* 2013). Initial live and dead fuel moisture content values were drawn from the 'low' dead fuel moisture and 'two-thirds cured' live fuel moisture values from Scott and Burgan (2005). Rate of surface fire spread was calculated in metres per minute and the resulting ascii text file was converted to an ArcGRID file for use as a rate of spread raster surface.

The final spatial index used to associate fire behaviour and suppression activities with fire perimeter locations was a modified version of the SDI developed for fire suppression in Mediterranean ecosystems and validated on forests in Chile and Israel (Rodríguez y Silva *et al.* 2014). SDI is an expert-informed fire risk and response mapping tool that combines weighted indices of potential access and egress from any point on a landscape with simulated fire behaviour (heat per unit area and potential surface flame length) under user-defined fire weather conditions. We used FlamMap 5.0 under the same operating parameters detailed above to calculate potential surface flame lengths and heat per unit area under 90th-percentile fire weather conditions. SDI takes into account road and trail density per square kilometre, road and fuel break length by relative fuel type area, foot access as a function of slope, soil erodibility, mobility within a fuel type and a fireline construction factor. Each component is ranked on a scale from 1 to 10 and combined such that the fire behaviour index proportional to fuel type area is divided by the sum of the accessibility parameter values. Look-up tables for weighted parameters are detailed in Rodríguez y Silva *et al.* (2014). In this calculation of SDI, return time for aerial suppression resources and retardant recharge were excluded due to insufficient data. Inclusion of a suite of accessibility factors and potential fire behaviour make SDI an appropriate metric to assess risk to fire responder safety as well as potential for successful suppression.

Model development pre-processing

We generated a random point layer over the fire-affected landscape and then extracted raster values for the response and predictor variables to each sample point location. This table of sampled points and associated variable values served as the data matrix for statistical model development.

We used a boosted regression tree (BRT) machine learning method to test the spatial relationships between fire perimeter locations and underlying landscape factors, potential fire behaviour and access for suppression resources. The BRT method is a flexible nonparametric machine learning approach that automatically identifies and models variable interactions and can handle large suites of candidate variables and sharp discontinuities common to rare events (such as fire perimeters) better than generalised linear or generalised additive models (De'ath 2007; Elith *et al.* 2008). Unlike random forest methods, boosting reduces variance through model averaging and can accommodate a range of response variable types by selecting the appropriate loss function (Elith *et al.* 2008). BRTs combine thousands of individual regression trees, used to quantify predictor and response variable relationships, with a forward stage-wise boosting procedure that builds upon each successive model iteration until a final model is reached where additional regression trees do not improve overall predictive power (Elith *et al.* 2008). The BRT method also incorporates a random subset of samples known as the 'bagging fraction' that generally improves model performance. After experimenting with several BRT learning rates and tree complexities, we found the parameters recommended by Elith *et al.* (2008) in the 'gbm' package in the R statistical environment (R Core Team 2015) to be a good balance of predictive performance and computational demands; we used a bagging fraction of 0.5, learning rate of 0.005 and tree complexity of 5. Loss functions for model pruning were fitted with a Bernoulli distribution for the binomial response model and a Poisson distribution for the perimeter count response model (range of 1–9 fire perimeters) (Friedman *et al.* 2000; Preisler *et al.* 2004; Ridgeway 2015). In a statistical analysis of wildfire risk, Preisler *et al.* (2004) recommended nonparametric binomial logistic regression as a robust method for handling point-based fire occurrence probabilities. The boosting procedure used here overcomes many of the limitations inherent in linear regression methods and has proven to be robust to pseudo-absence (zero-inflated) data and moderate collinearity between predictors (Friedman *et al.* 2000; Dormann *et al.* 2013). We used a correlation matrix to assess Spearman ranked correlation among predictors and a test of collinearity from the 'usdm' package in the R statistical environment (R Core Team 2015) to assess the effects of variance inflation factor on model results. A 10-fold cross-validation was used to assess relative contributions of each predictor variable and to assess overall model performance in correctly classifying the random 10% of data excluded from the training data for each model run. Area under the curve (AUC) and cross-validated error estimates of model deviance were used to assess overall prediction accuracy.

Density smoothed scatter plots of model-fitted values in relation to each predictor variable were used to inspect relationships between fire perimeter occurrence and predictor variable values. Plots were generated in the 'gbm' package in the R statistical environment (R Core Team 2015) with the 'smoothScatter' function.

As an additional check on the BRT model results, we used the MaxEnt package (Phillips *et al.* 2006; Phillips 2015) to assess differences and similarities between the BRT and presence-only projections of potential fire control locations. Presence-only regression methods (Ward *et al.* 2009; Elith *et al.* 2011) are specifically designed to assess potential suitability for a response condition when true absence data are not available. These methods rely on several assumptions about the sample distribution and are highly sensitive to user settings (Lobo *et al.* 2008; Merow *et al.* 2013; Merow *et al.* 2014). In addition, the scaling of AUC and potential suitability outputs from presence-only methods are not directly comparable to true logistic regression models such as BRT (Merow *et al.* 2013). Detailed methods and results from the MaxEnt analysis are presented in Section S1 (of the supplementary material available online).

Results

We developed fire perimeter models from 1 408 575 sampling points, of which 192 158 recorded a fire perimeter, resulting in a no perimeter to perimeter ratio of ~7:1. Binomial and perimeter count models did not produce significantly different results in terms of predictive variables selected or strength of variable fits; however, the binomial model had lower total model deviance. Results from the binomial model are presented here. Results from the perimeter count model are available upon request from the authors.

Two sets of predictor variables – travel cost and barrier distance ($r = 0.75$), and valley distance and ridge distance ($r = 0.60$) – were correlated at a level greater than 0.25. However in a test of variable collinearity, no two variables had a variance inflation factor (VIF) greater than 1.75. VIF values greater than 5 are considered significant enough to warrant removal from a predictive model (Dormann *et al.* 2013). A variable correlation matrix (Table S1) and VIF calculations (Table S2) are presented in the supplementary material.

Eight of the original 14 predictor variables significantly improved the power of the binomial fire perimeter model. Barrier distance, travel cost, SDI and distance from valley were the most influential variables (relative influences of 28.5%, 20.5%, 12.5% and 11.4%) (Fig. 3*a*). Additional variables ridge distance, rate of fire spread, steep slope distance and RTC each accounted for less than 10% of the final predictive model. Fire perimeters were strongly associated with locations representing a combination of close proximity to barriers (major roads and waterways) and valleys, minimal travel cost and low suppression difficulty. Proximity to ridge tops and steep transitions, slow modelled fire spread rate and low RTC were also significant but accounted for a smaller proportion of the final predictive model. Fitted distributions of fire perimeter presence or absence in relation to barrier distance, travel cost, SDI and valley distance demonstrate the strength of these relationships at low predictor values (e.g. high probability of a fire perimeter when barrier and valley distance, travel cost and SDI values are minimised), as well as the random effects of a large proportion of locations with no fire perimeter (Fig. 3*b*).

Scatter plots of the eight predictor variables and their relationships to fire perimeter presence or absence are presented in Fig. S1. Discontinuities between data points in the fitted RTC point distribution (Fig. S1) are a function of the limited number of fireline construction classification rates (Dillon *et al.* 2015; Table 3) (Fig. 3*b*). This limitation was further exemplified in calculated RTC values from the simplified four-class fireline construction rates (Broyles 2011) (not shown). Simplified RTC values did not significantly improve the predictive power of the fire perimeter model and were excluded from the final model formulation.

An AUC value of 0.686 suggests that the model correctly predicted ~69% of the random independent perimeter presence and absence observations used in the cross-validation assessment of model performance. The model was then projected across the landscape to produce a likelihood surface of potential fire perimeter locations that identified areas with high or low likelihood of a final fire perimeter (Fig. 4*a*). An example classification of the fire perimeter probability surface (Fig. 4*b*) highlights specific locations on the landscape with higher or lower probabilities of association with a future fire perimeter. In this example, thresholds of $P > 60\%$ and $40\% \leq P \leq 60\%$ were used to identify candidate primary and secondary control features respectively. Use of these features would be dependent upon weather conditions during fire operations being comparable to modelled conditions, as well as consideration of potential spotting behaviour not included in model simulations. After identifying candidate control locations, we used the continuous probability surface (Fig. 4*a*) to identify additional features with high probability of forming a fire perimeter to connect candidate control locations. The probability surface can also be used to highlight potential areas of high hazard where some mitigation effort before the fire season would be necessary to develop an effective network of control features (Fig. 4*b*). The continuous probability surface output provides a range of options for adapting control location probability thresholds to operational realities depending on weather conditions, available resources and landscape characteristics. Control location probability thresholds used here were appropriate for providing a range of options under 90th-percentile fire weather conditions.

The MaxEnt analysis of potential suitability of fire perimeter locations identified the same set of eight significant predictor variables used in the BRT model, although RTC surpassed barrier distance and travel cost as the primary determinant of suitable fire perimeter locations (Section S1).

Discussion

This modelling exercise is the first example of a quantitative geospatial tool for identifying potential fire control locations for use in pre-fire landscape planning and active fire management. The model identified significant associations between final fire perimeter locations and a series of landscape features and conditions relevant to fire suppression activities. Although many of the landscape components identified by the model are intuitive and would likely already be used by incident management personnel during active fire management, the power of the BRT approach is its ability to account for interactions between predictor variables and to weigh them according to their relative contribution. This approach evaluates each pixel on the landscape for its potential to function as a control feature in relation to every other pixel on the landscape. Thus, the mapped output leverages substantially more information than a traditional 'roads and ridges' approach, making it potentially more useful

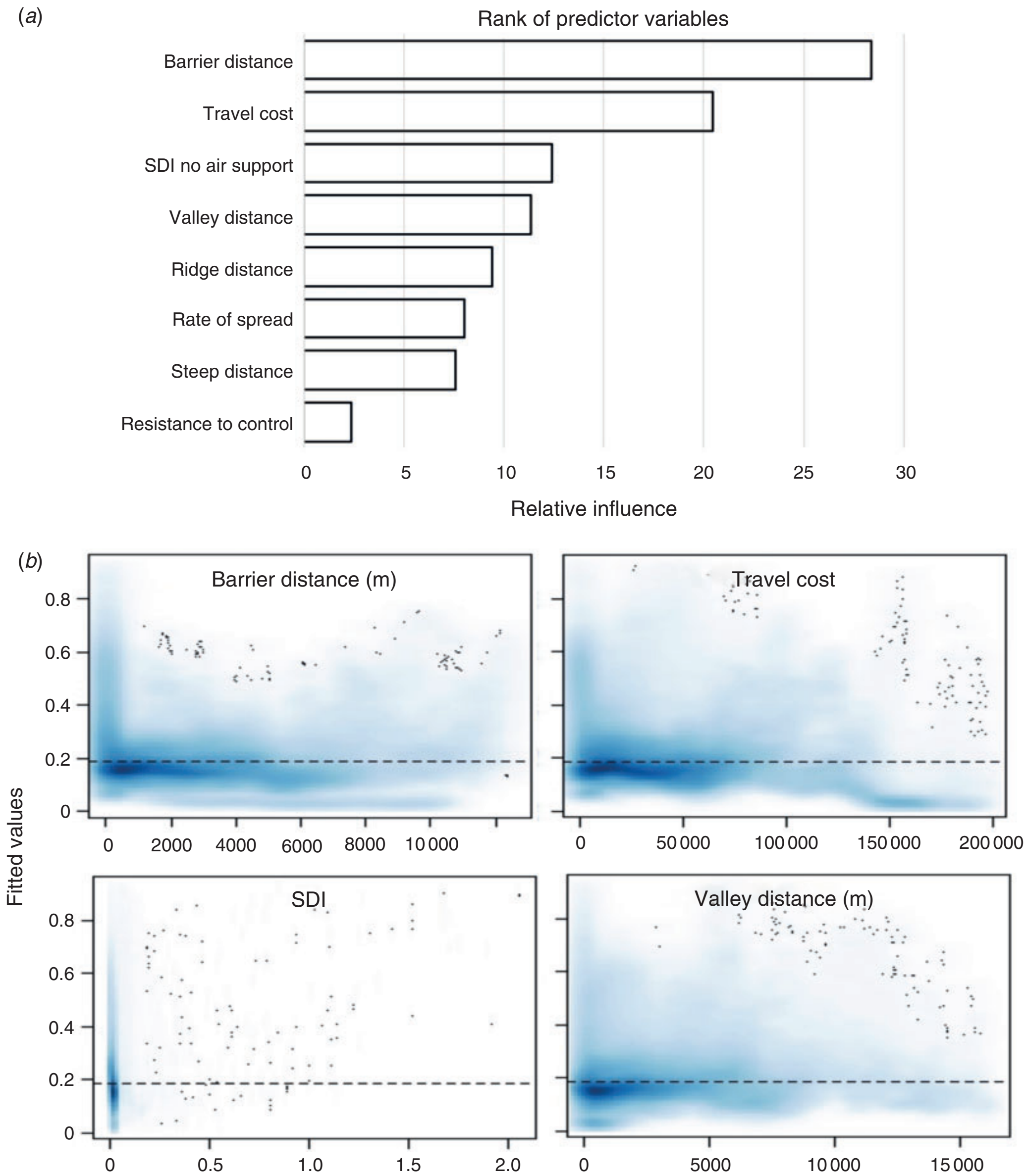

Fig. 3. Summary of predictive model inputs and relationships to observed fire perimeters. Eight significant predictors of final fire perimeter locations (*a*) are listed in order of relative contribution to improving model performance. Fitted distributions of the top four predictors (*b*) demonstrate the relationship between fire perimeter presence (*y*-axis) and value of a predictor variable (*x*-axis). Shading represents relative point density of the fitted perimeter predictor relationships. Suppression Difficulty Index (SDI) and travel cost are dimensionless index values; barrier distance and valley distance are in metres.

for ranking control features in advance as well as for incorporating potential exposure of fire responders.

Although several of the predictor variables contained information associated with fire behaviour, relative accessibility and ease of fuel modification, the different metrics used – for example, total road length per unit area (a component of SDI), linear distance from major road (a component of barrier distance) and cumulative penalty for travelling away from a paved surface (a component of travel cost) – made the information presented in each variable unique, such that model predictive performance improved without a subsequent increase in model variance. The relative contribution of each predictor variable would be expected to vary from one landscape to another depending on landscape features such as density and quality of road networks and topographic complexity, as well as ownership-determined fire management objectives and available resources.

Future work

Although we were able to predict fire perimeter locations with 69% accuracy on a large landscape with highly variable terrain

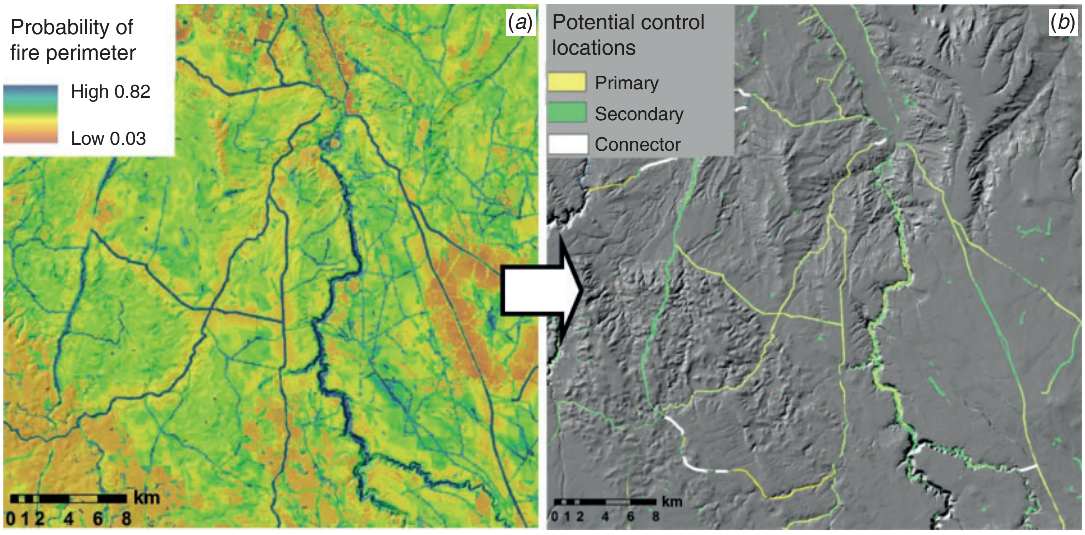

Fig. 4. Fire perimeter probability surface and ranked potential control features. Fire perimeter probability surface (*a*) is generated from the boosted regression tree model. Primary and secondary control locations represent probability >60% or 40% ≤ P ≤ 60% respectively. Potential control location connectors are 'next best' options for joining primary and secondary locations for operational fire management use (*b*).

and road access, additional work will be necessary to address the remaining 31% of fire perimeter locations that were not explained by the weather-independent model. Unexplained variance in the predictive model is likely related to some combination of (1) variability in fire weather conditions; (2) variability in fire suppression response; (3) inclusion of only fires greater than 400 ha in the model training and validation datasets; and (4) variability in topographic complexity and road density within the study area. In previous studies, extreme fire weather has been shown to significantly increase variability in fire size, severity and spread (Bradstock *et al.* 2010; Dillon *et al.* 2011*b*; Riley *et al.* 2013). Final fire perimeters are often associated with periods of relatively quiescent weather (Finney *et al.* 2009), or the opportunistic use of roads and other features for back burning after periods of major fire growth (Thompson *et al.* 2016*b*). We implicitly address the exclusion of fire weather by using a large number of individual fire perimeters, each formed under a different set of fire weather conditions. This variability in fire weather becomes the background signal that allows emergent properties from underlying landscape characteristics to explain consistent patterns of fire perimeter formation. Seasonal climate patterns and the underlying warming signal associated with changing climate are also embedded within these highly variable daily fire weather–fire perimeter relationships. To explicitly incorporate this variability into a fire weather metric, daily (or finer) time steps of weather conditions and fire perimeter locations could be used to test associations with specific landscape conditions, suppression tactics and fire perimeter expansion, similar to the methods used by Price *et al.* (2014) in tropical savannahs. Two challenges for this approach are the relatively few fires with daily fire perimeter information (GeoMac 2016) and the decentralised and inconsistent data recording systems used to track fire response actions taken (Short 2015). Holsinger *et al.* (2016) approached these limitations by assigning day of burn weather values to points within a fire perimeter and day after the end of fire progression weather values to points located on a final fire perimeter. Another way to incorporate fire weather thresholds into potential control feature predictions would be to conduct a sensitivity analysis that includes a range of fire weather scenarios to calculate rate of spread, flame length and potential energy per unit area in FlamMap simulations. Generation of a series of control feature options depending on fire weather conditions could prove useful for resource allocation during active fire management. A range of different fire weather scenarios could be used to develop a series of potential control location atlases depending on conditions during intended use. For example, during shoulder seasons a potential control location atlas based on 70th-percentile weather could be used to plan managed controlled burns and targeted fuel reduction treatments. During the height of fire season, more extreme 97th- or 99th-percentile fire weather conditions could be more appropriate to identify holding features under potentially extreme conditions. In a recent analysis of changing climate effects on area burned in the western USA, Abatzoglou and Williams (2016) found that over the past two decades, a trend of increasingly persistent drought conditions and extreme fire weather support the idea of incorporating more extreme fire weather conditions into incident-based fire modelling scenarios and response systems.

The model did not account for specific suppression actions associated with each fire; however, the identification of suppression difficulty and travel cost as primary predictors of a fire perimeter align with the broadly defined suppression-driven strategic fire response objectives governing management actions over the majority of the landscape (Fig. 1). Proximity to identified high value resources and assets (HVRAs) also likely played a role in determining fire management response (Thompson *et al.* 2016*a*; Dunn *et al.* in press). Incorporation of HVRAs, either as a distance or density metric might also serve as a potential fire suppression effort variable for determining appropriate locations for future control features and for explaining variance in historical fire perimeter locations. The effects of aerial suppression resources, another important consideration for determining potential fire control locations, are not included in this model formulation primarily due to a lack of consistent data. The original formulation of SDI (Rodríguez y Silva *et al.* 2014) used helicopter and fixed-wing aircraft flying times between drops to incorporate potential effects of aerial resources. Although these data were not available for the modelled landscape, we suggest that a calculation of retardant drop cycling time, as well as the additional factors slope and fire weather conditions associated with retardant effectiveness could be included in an additional aerial resource effectiveness index.

An explicit fire responder exposure index could also help to prioritise potential control locations that align with the US Forest Service Life First Initiative (Tidwell 2016). Such an index might include factors associated with potential for injury during active fire engagement as well as with movement in dangerous landscape conditions, time required for medical evacuation, slope–wind alignment and other considerations that directly affect the exposure of fire responders to potential threats.

Stratifying landscapes by strategic fire response objectives could further refine candidate locations for operational fire management that more closely follow land management goals. For example, a modelled landscape containing a significant wilderness or fire-adapted forest component with sparse human population would likely express stronger associations between fire perimeters and natural fire boundaries such as fuel type transitions (e.g. continuous to sparse), elevational or ecological gradients (e.g. warm–dry to cool–moist), topographic transitions and hydrographic features. Spatial reconstruction of historical fire perimeters (Iniguez *et al.* 2008; Margolis and Balmat 2009; O'Connor *et al.* 2014) and observations from wilderness areas (Parks *et al.* 2014) demonstrate the role of ecological gradients in determining potential for fire spread.

Transferring the results from this analysis to fire management in other regions will require additional research to assess the sensitivity of the model to variable topographic complexity, road density and strategic fire response. Topography indirectly influences fuel distributions and ignition patterns, especially in steeply bisected landscapes (Parks *et al.* 2012). If fuel continuity is the primary driver of fire spread, constraints on this continuity can occur naturally in complex topography or with roads and other disturbance features on human modified landscapes. A catalogue of models developed from landscapes with low, moderate or high levels of road access and topographic complexity would account for interactions between topography, fuels and accessibility that influence fire spread at scales appropriate for local application. Fire managers could then select an appropriate model and input information from national data layers to produce a potential control location map tailored to a specific landscape.

As with the results from all simulation models, caution should be taken to validate model predictions with real world data. In this case potential control features identified from spatial modelling should be surveyed by experienced fire management personnel before any application in fire management operations or planning. In addition to general model testing, groundtruthing provides a basis for classifying primary and secondary control locations and identifying potential control connectors; for improving the control location modelling process; and for building confidence in the use of predictive modelling data for fire planning and management decisions.

Acknowledgements

The authors would like to thank Greg Dillon, Chris Dunn, Yu Wei, Karin Riley and Jessica Haas for help with model assumptions and appropriate formulations; Jon Rieck, Karen Short and John Hogland for help locating datasets and providing spatial analysis tools; and Margaret Evans and Scott Baggett for advice on statistical analyses. This manuscript was improved by suggestions from three anonymous reviewers.

References

Abatzoglou JT, Williams AP (2016) Impact of anthropogenic climate change on wildfire across western US forests. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 11770–11775. doi:10.1073/PNAS.1607171113

Anderson DH (1989) A mathematical model for fire containment. *Canadian Journal of Forest Research* **19**, 997–1003. doi:10.1139/X89-153

Beier P, Brost B (2010) Use of land facets to plan for climate change: conserving the arenas, not the actors. *Conservation Biology* **24**, 701–710. doi:10.1111/J.1523-1739.2009.01422.X

Bradshaw L (2013) FireFamily Plus 4.1. USDA Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory, Missoula, MT, USA. Available at http://www.firelab.org/project/firefamilyplus [Verified 24 January 2017]

Bradstock RA, Hammill KA, Collins L, Price O (2010) Effects of weather, fuel and terrain on fire severity in topographically diverse landscapes of south-eastern Australia. *Landscape Ecology* **25**, 607–619. doi:10.1007/S10980-009-9443-8

Broyles G (2011) Fireline production rates. USDA Forest Service, Fire management report 1151–1805. National Technology & Development Program. San Dimas Technology and Development Center (San Dimas, CA). Available at https://www.fs.fed.us/t-d/pubs//pdf/11511805.pdf [Verified 30 January 2017]

Calkin DE, Thompson MP, Finney MA, Hyde KD (2011) A real-time risk assessment tool supporting wildland fire decision making. *Journal of Forestry* **109**, 274–280.

De'ath G (2007) Boosted trees for ecological modeling and prediction. *Ecology* **88**, 243–251. doi:10.1890/0012-9658(2007)88[243:BTFEMA]2.0.CO;2

Dillon G, Morgan P, Holden Z (2011*a*) Mapping the potential for high severity wildfire in the western United States. *Fire Management Today* **71**, 25–28.

Dillon GK, Holden ZA, Morgan P, Crimmins MA, Heyerdahl EK, Luce CH (2011*b*) Both topography and climate affected forest and woodland burn severity in two regions of the western US, 1984 to 2006. *Ecosphere* **2**, 130. doi:10.1890/ES11-00271.1

Dillon GK, Menakis J, Fay F (2015) Wildland fire potential: a tool for assessing wildfire risk and fuels management needs. In 'Proceedings of the large wildland fires conference', 19–23 May 2014, Missoula, MT. (Eds RE Keane, M Jolly, R Parsons, K Riley) USDA Forest Service, Rocky Mountain Research Station, Proceedings RMRS-P-73, pp. 60–76. (Fort Collins, CO)

Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, Marquéz J. R. G., Gruber B, Lafourcade B, Leitão PJ (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**, 27–46. doi:10.1111/J.1600-0587.2012.07348.X

DRI (2015) Desert Research Institute Remote Automatic Weather Station data, Trail Gulch Weather Station. Weather dates 2–30 July 2007.

Available at http://www.raws.dri.edu/cgi-bin/rawMAIN.pl?idITRA [Verified 24 January 2017]

Duff TJ, Tolhurst KG (2015) Operational wildfire suppression modelling: a review evaluating development, state of the art and future directions. *International Journal of Wildland Fire* **24**, 735–748. doi:10.1071/WF15018

Dunn CJ, Thompson MP, Calkin DE (in press) A framework for developing safe and efficient large-fire incident response strategies and tactics for a new fire management paradigm. *International Journal of Wildland Fire*.

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. *Journal of Animal Ecology* **77**, 802–813. doi:10.1111/J.1365-2656.2008.01390.X

Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ (2011) A statistical explanation of MaxEnt for ecologists. *Diversity & Distributions* **17**, 43–57. doi:10.1111/J.1472-4642.2010.00725.X

ESRI (2015) 'ArcMap 10.2.' (ESRI Inc. Redlands, CA)

Finney MA (2004) FARSITE: Fire area simulator: model development and evaluation. USDA Forest Service, Rocky Mountain Research Station, Research Paper RMRS-RP-4 Revised. (Fort Colliins, CO) Available at https://www.fs.fed.us/rm/pubs/rmrs_rp004.pdf [Verified 30 January 2017].

Finney MA (2006) An overview of FlamMap fire modeling capabilities. In 'Fuels management – how to measure success: conference proceedings', 28–30 March 2006, Portland, OR. (Eds PL Andrews, B Butler) USDA Forest Service, Rocky Mountain Research Station, Proceedings RMRS-P-41. (Fort Collins, CO).

Finney M, Grenfell IC, McHugh CW (2009) Modeling containment of large wildfires using generalized linear mixed-model analysis. *Forest Science* **55**, 249–255.

Finney MA, Grenfell IC, McHugh CW, Seli RC, Trethewey D, Stratton RD, Brittain S (2011*a*) A method for ensemble wildland fire simulation. *Environmental Modeling and Assessment* **16**, 153–167. doi:10.1007/S10666-010-9241-3

Finney MA, McHugh CW, Grenfell IC, Riley KL, Short KC (2011*b*) A simulation of probabilistic wildfire risk components for the continental United States. *Stochastic Environmental Research and Risk Assessment* **25**, 973–1000. doi:10.1007/S00477-011-0462-Z

Finney MA, Brittain S, Seli RC, McHugh CW, Gangi L (2015) FlamMap: Fire Mapping and Analysis System (Version 5.0). USDA Forest Service, Rocky Mountain Research Station. (Missoula, MT) Available from https://www.firelab.org/document/flammap-software [Verified 24 January 2017]

Fried JS, Fried BD (1996) Simulating wildfire containment with realistic tactics. *Forest Science* **42**, 267–281.

Friedman J, Hastie T, Tibshirani R (2000) Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *Annals of Statistics* **28**, 337–407. doi:10.1214/AOS/1016218223

GeoMac (2016) Geospatial multi-agency coordination. USGS Geosciences & Environmental Change Science Center (Lakewood, CO). Available at http://www.geomac.gov/ [Verified 14 June 2016]

Haas JR, Calkin DE, Thompson MP (2013) A national approach for integrating wildfire simulation modeling into wildland–urban interface risk assessments within the United States. *Landscape and Urban Planning* **119**, 44–53. doi:10.1016/J.LANDURBPLAN.2013.06.011

Harris L, Taylor AH (2015) Topography, fuels, and fire exclusion drive fire severity of the Rim Fire in an old-growth mixed-conifer forest, Yosemite National Park, USA. *Ecosystems* **18**, 1192–1208. doi:10.1007/S10021-015-9890-9

Holsinger L, Parks SA, Miller C (2016) Weather, fuels, and topography impede wildland fire spread in western US landscapes. *Forest Ecology and Management* **380**, 59–69. doi:10.1016/J.FORECO.2016.08.035

Iniguez JM, Swetnam TW, Yool SR (2008) Topography affected landscape fire history patterns in southern Arizona, USA. *Forest Ecology and Management* **256**, 295–303. doi:10.1016/J.FORECO.2008.04.023

Jenness J, Brost B, Beier P (2013) 'Land Facet CorridorDesigner manual.' (Jenness Enterprises: Flagstaff, AZ). Available at http://www.corridordesign.org/ [Verified 24 January 2017]

Kane VR, Cansler CA, Povak NA, Kane JT, McGaughey RJ, Lutz JA, Churchill DJ, North MP (2015) Mixed severity fire effects within the Rim Fire: relative importance of local climate, fire weather, topography, and forest structure. *Forest Ecology and Management* **358**, 62–79. doi:10.1016/J.FORECO.2015.09.001

Keane RE, Morgan PM, Dillon GK, Sikkink PG, Karau EC, Holden ZA, Drury SA (2013) A fire severity mapping system for real-time fire management applications and long-term planning: the FIRESEV project. Final report January 2013. Project Number JFSP-09–1-07–4. JFSP Research Project Reports, Paper 18. (Joint Fire Sciences Program: Boise, ID)

LANDFIRE (2010) Digital Elevation Model, LANDFIRE 1.2.0. US Department of the Interior, Geological Survey. Available at http://www.landfire.gov/viewer/ [Verified 26 October 2015].

LANDFIRE (2012) 40 Scott & Burgan (2005) Fire Behavior Fuel Models, Slope, Aspect, Existing Vegetation Type, and Biophysical Settings layers, LANDFIRE 1.3.0. US Department of the Interior, Geological Survey. Available at http://www.landfire.gov/viewer/ [Verified 26 October 2015].

Lobo JM, Jiménez-Valverde A, Real R (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* **17**, 145–151. doi:10.1111/J.1466-8238.2007.00358.X

Majka D, Jenness J, Beier P (2007) CorridorDesigner: ArcGIS tools for designing and evaluating corridors. (CorridorDesigner: Minneapolis, MN) Available at http://www.corridordesign.org/ [Verified 24 January 2017]

Margolis EQ, Balmat J (2009) Fire history and fire–climate relationships along a fire regime gradient in the Santa Fe Municipal Watershed, NM, USA. *Forest Ecology and Management* **258**, 2416–2430. doi:10.1016/J.FORECO.2009.08.019

Merow C, Smith MJ, Silander JA (2013) A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* **36**, 1058–1069. doi:10.1111/J.1600-0587.2013.07872.X

Merow C, Smith MJ, Edwards TC, Guisan A, McMahon SM, Normand S, Thuiller W, Wüest RO, Zimmermann NE, Elith J (2014) What do we gain from simplicity versus complexity in species distribution models? *Ecography* **37**, 1267–1281. doi:10.1111/ECOG.00845

Mitsopoulos I, Mallinis G, Zibtsev S, Yavuz M, Saglam B, Kucuk O, Bogomolov V, Borsuk A, Zaimes G (2016) An integrated approach for mapping fire suppression difficulty in three different ecosystems of Eastern Europe. *Journal of Spatial Science*. doi:10.1080/14498596.2016.1169952

MTBS (2014) Monitoring Trends in Burn Severity. USDA Forest Service Remote Sensing Applications Center. Salt Lake City, UT. Available at http://www.mtbs.gov/dataaccess.html [Verified 20 October 2015]

Narayanaraj G, Wimberly MC (2011) Influences of forest roads on the spatial pattern of wildfire boundaries. *International Journal of Wildland Fire* **20**, 792–803. doi:10.1071/WF10032

Narayanaraj G, Wimberly MC (2012) Influences of forest roads on the spatial patterns of human- and lightning-caused wildfire ignitions. *Applied Geography (Sevenoaks, England)* **32**, 878–888. doi:10.1016/J.APGEOG.2011.09.004

Noonan-Wright EK, Opperman TS, Finney MA, Zimmerman GT, Seli RC, Elenz LM, Calkin DE, Fiedler JR (2011) Developing the US wildland fire decision support system. *Journal of Combustion* **2011**, 168473. doi:10.1155/2011/168473

National Wildfire Coordinating Group (2004) Fireline handbook. NWCG Handbook 3. National Interagency Fire Center, PMS 410–1, NFES 0065. (Boise, ID).

O'Connor CD, Falk DA, Lynch AM, Swetnam TW (2014) Fire severity, size, and climate associations diverge from historical precedent along an ecological gradient in the Pinaleño Mountains, Arizona, USA. *Forest Ecology and Management* **329**, 264–278. doi:10.1016/J.FORECO.2014.06.032

Parks SA, Parisien M-A, Miller C (2012) Spatial bottom-up controls on fire likelihood vary across western North America. *Ecosphere* **3**, 12. doi:10.1890/ES11-00298.1

Parks SA, Parisien M-A, Miller C, Dobrowski SZ (2014) Fire activity and severity in the western US vary along proxy gradients representing fuel amount and fuel moisture. *PLoS One* **9**, e99699. doi:10.1371/JOURNAL.PONE.0099699

Parks SA, Holsinger LM, Miller C, Nelson CR (2015) Wildland fire as a self-regulating mechanism: the role of previous burns and weather in limiting fire progression. *Ecological Applications* **25**, 1478–1492. doi:10.1890/14-1430.1

Petrovic N, Carlson J (2012) A decision-making framework for wildfire suppression. *International Journal of Wildland Fire* **21**, 927–937. doi:10.1071/WF11140

Phillips SJ (2015) 'A brief tutorial on MaxEnt.' (AT&T Research: Florham, NJ).

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**, 231–259. doi:10.1016/J.ECOLMODEL.2005.03.026

Podur JJ, Martell DL (2007) A simulation model of the growth and suppression of large forest fires in Ontario. *International Journal of Wildland Fire* **16**, 285–294. doi:10.1071/WF06107

Preisler HK, Brillinger DR, Burgan RE, Benoit J (2004) Probability based models for estimation of wildfire risk. *International Journal of Wildland Fire* **13**, 133–142. doi:10.1071/WF02061

Price OF, Borah R, Maier SW (2014) Role of weather and fuel in stopping fire spread in tropical savannas. *Austral Ecology* **39**, 135–144. doi:10.1111/AEC.12021

R Core Team (2015). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Available from http://www.R-project.org/ [Verified 19 January 2017]

Ridgeway G (2015) Generalized boosted models: A guide to the gbm package. Available at https://cran.r-project.org/web/packages/gbm/gbm.pdf [Verified 24 January 2017]

Riley KL, Abatzoglou JT, Grenfell IC, Klene AE, Heinsch FA (2013) The relationship of large fire occurrence with drought and fire danger indices in the western USA, 1984–2008: the role of temporal scale. *International Journal of Wildland Fire* **22**, 894–909. doi:10.1071/WF12149

Rodríguez y Silva F, Martínez J. R. M, González-Cabán A (2014) A methodology for determining operational priorities for prevention and suppression of wildland fires. *International Journal of Wildland Fire* **23**, 544–554. doi:10.1071/WF13063

Scott JH, Burgan RE (2005) Standard fire behavior fuel models: a comprehensive set for use with Rothermel's surface fire spread model. USDA Forest Service, Rocky Mountain Research Station, General Technical Report No. RMRS-GTR-153. (Fort Collins, CO).

Scott JH, Thompson MP, Calkin DE (2013) A wildfire risk assessment framework for land and resource management. USDA Forest Service, Rocky Mountain Research Station, General Technical Report No. RMRS-GTR-315. (Fort Collins, CO).

Short KC (2015) Sources and implications of bias and uncertainty in a century of US wildfire activity data. *International Journal of Wildland Fire* **24**, 883–891.

Thompson MP, Stonesifer CS, Seli RC, Hovorka M (2013) Developing standardized strategic response categories for fire management units. *Fire Management Today* **73**, 18–24.

Thompson MP, Haas JR, Gilbertson-Day JW, Scott JH, Langowski P, Bowne E, Calkin DE (2015) Development and application of a geospatial wildfire exposure and risk calculation tool. *Environmental Modelling & Software* **63**, 61–72. doi:10.1016/J.ENVSOFT.2014.09.018

Thompson MP, Bowden P, Brough A, Scott JH, Gilbertson-Day J, Taylor A, Anderson J, Haas JR (2016*a*) Application of wildfire risk assessment results to wildfire response planning in the Southern Sierra Nevada, California, USA. *Forests* **7**, 64. doi:10.3390/F7030064

Thompson MP, Freeborn P, Rieck JD, Calkin DE, Gilbertson-Day JW, Cochrane MA, Hand MS (2016*b*) Quantifying the influence of previously burned areas on suppression effectiveness and avoided exposure: a case study of the Las Conchas Fire. *International Journal of Wildland Fire* **25**, 167–181. doi:10.1071/WF14216

Tidwell TL (2016) Chief's letter of intent – 2016 wildland fire. File Code 5100. USDA Forest Service, Washington Office, 5 April 2016 (Washington, DC).

US Census Bureau (2015) 'TIGER/Line 2015 Roads.' Available at ftp://ftp2.census.gov/geo/tiger/TIGER2015/ROADS/ [Verified 18 November 2015].

USGS (2015) 'USGS National Hydrography Dataset.' Available at http://viewer.nationalmap.gov/basic/?howTo=true [Verified 18 November 2015].

Ward G, Hastie T, Barry S, Elith J, Leathwick JR (2009) Presence-only data and the EM algorithm. *Biometrics* **65**, 554–563. doi:10.1111/J.1541-0420.2008.01116.X

Weiss AD (2001) Topographic position and landforms analysis. Conference poster for '21st Annual ESRI International User Conference', 9–13 July 2001, San Diego, CA. Available at http://www.jennessent.com/arcview/TPI_Weiss_poster.htm [Verified 24 January 2017]

Wu Z, He HS, Liang Y, Cai L, Lewis BJ (2013) Determining relative contributions of vegetation and topography to burn severity from LANDSAT imagery. *Environmental Management* **52**, 821–836. doi:10.1007/S00267-013-0128-3

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

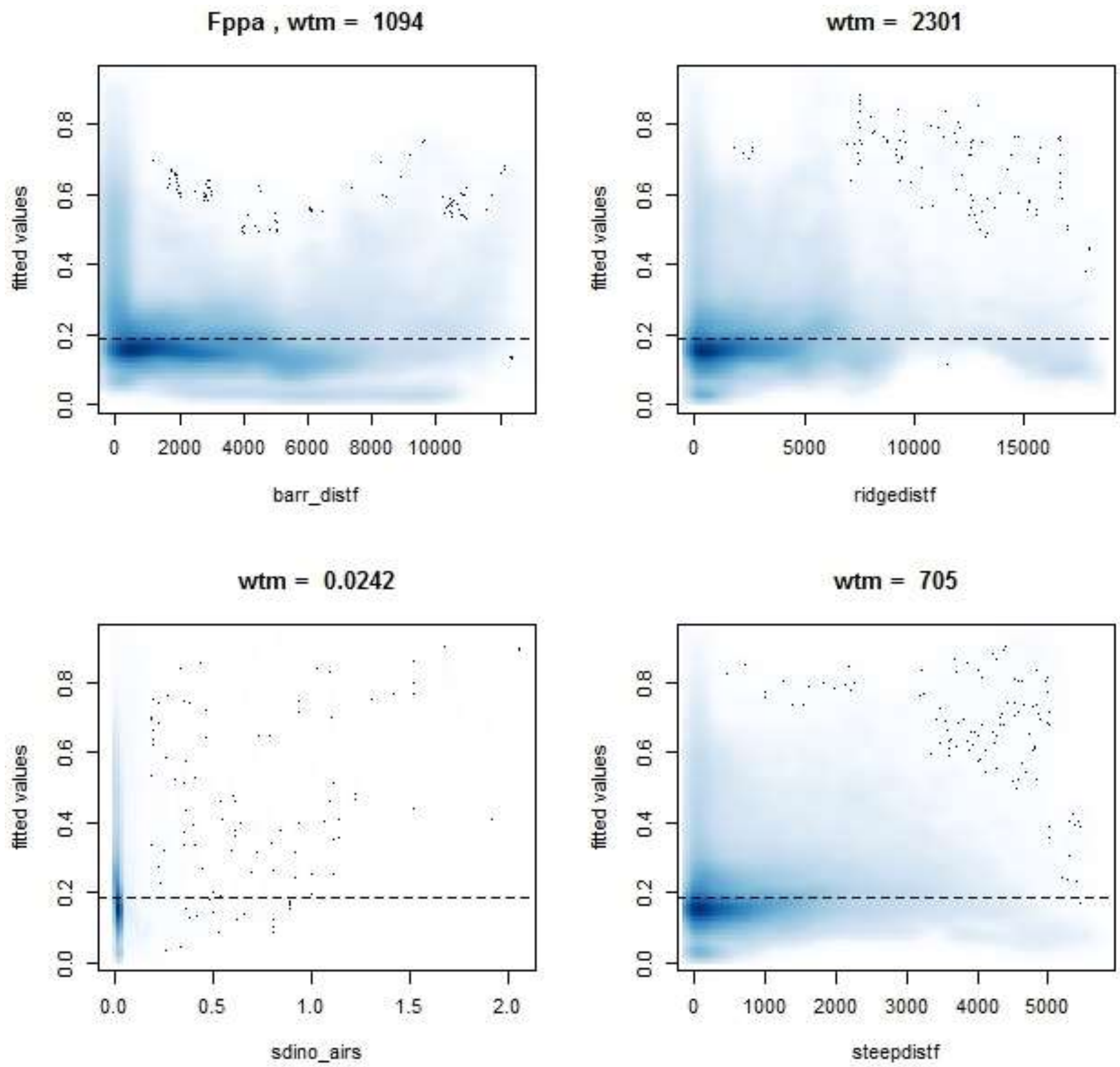

Fig. S1. Density smoothed scatter plots of model fitted fire perimeter presence or absence (fppa) in relation to each predictor variable. Y-axis represents probability of fire presence and X-axis represents predictor value. 'wtm' is the weighted mean value for each predictor. Dotted line is the average 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).

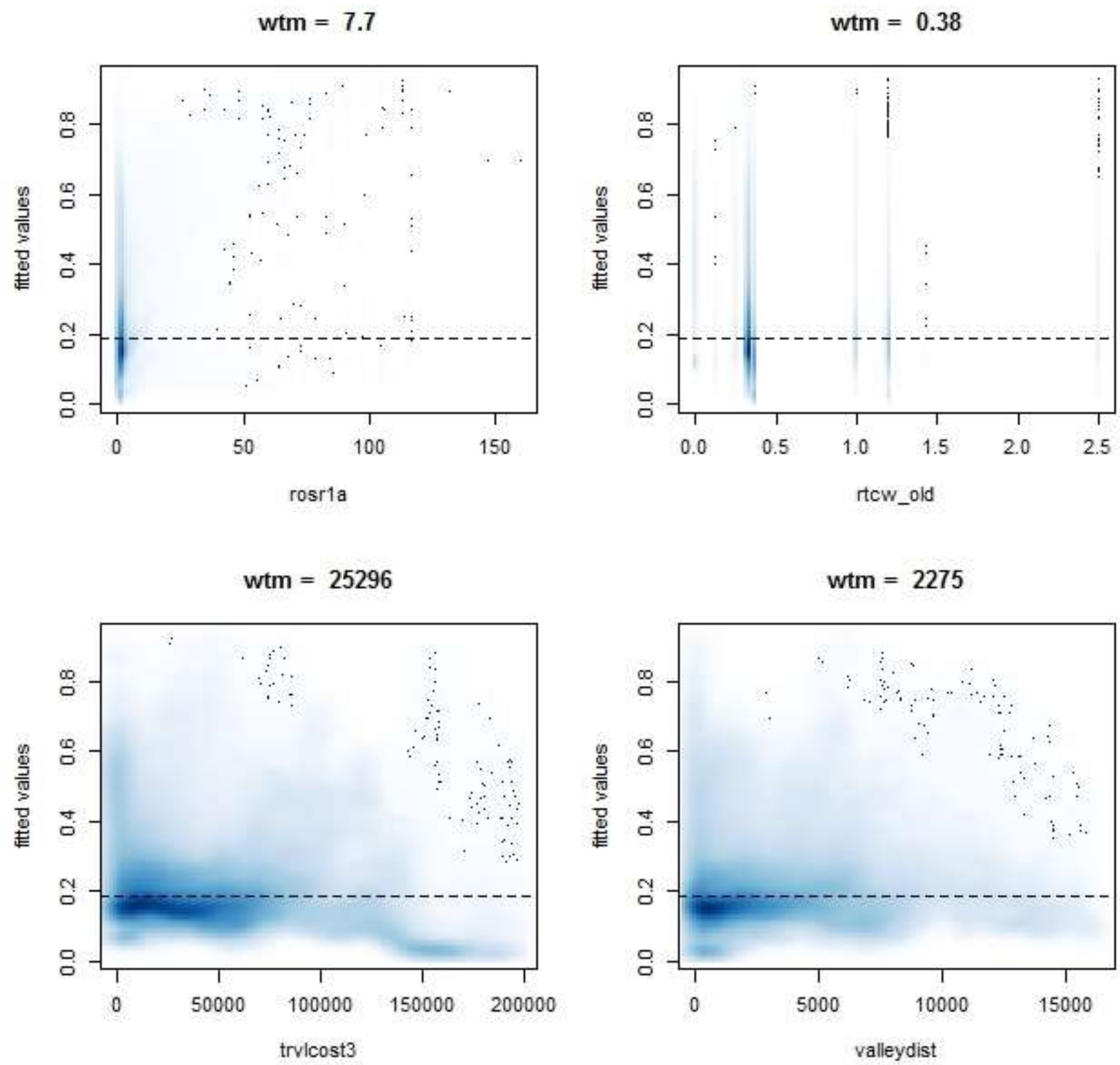

Fig. S2. Density smoothed scatter plots of model fitted fire perimeter presence or absence in relation to each predictor variable. Y-axis represents probability of fire presence and X-axis represents predictor value. Wtm is the weighted mean value for each predictor. Dotted line is the average fire perimeter 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.

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

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).

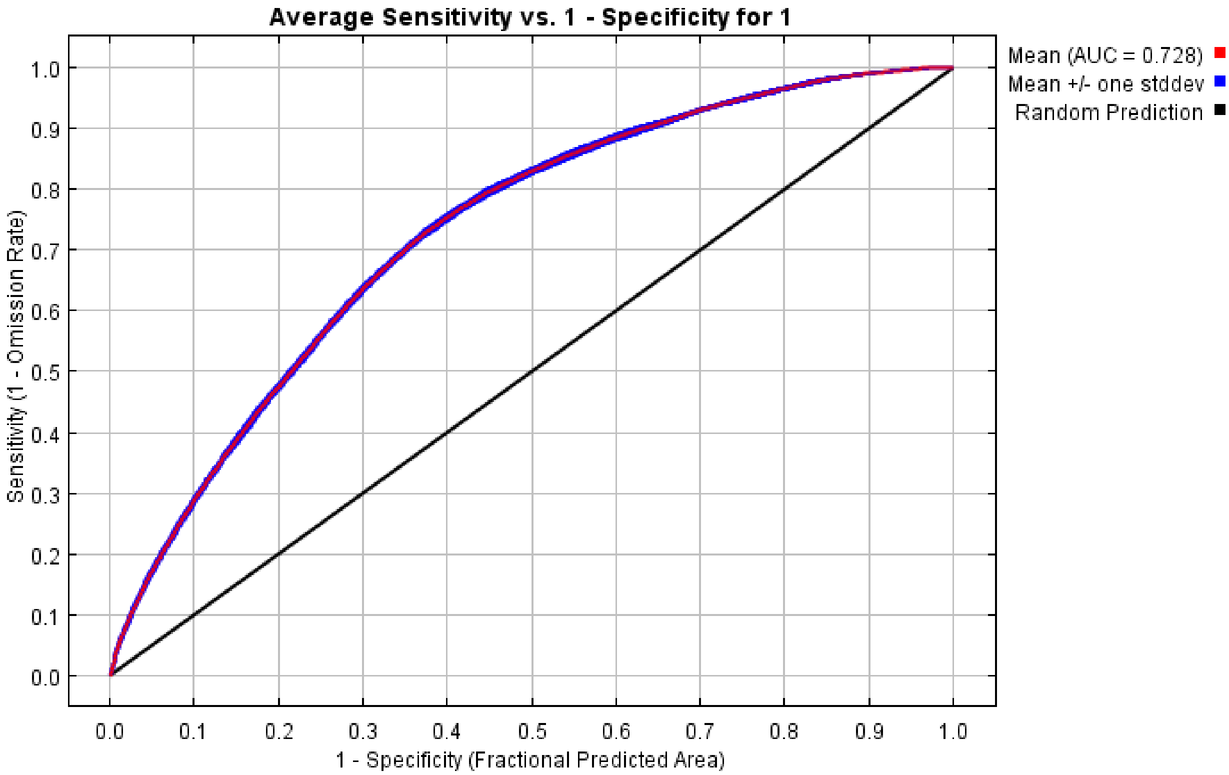

Fig S3. Average Area under the curve (AUC) calculation for 10 maxent runs. Maxent identified the same set of eight significant predictors of fire perimeter suitability as the BRT model, however, RTC surpassed SDI and travel cost as the primary determinant of suitable fire perimeter locations (Fig. S4). 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.

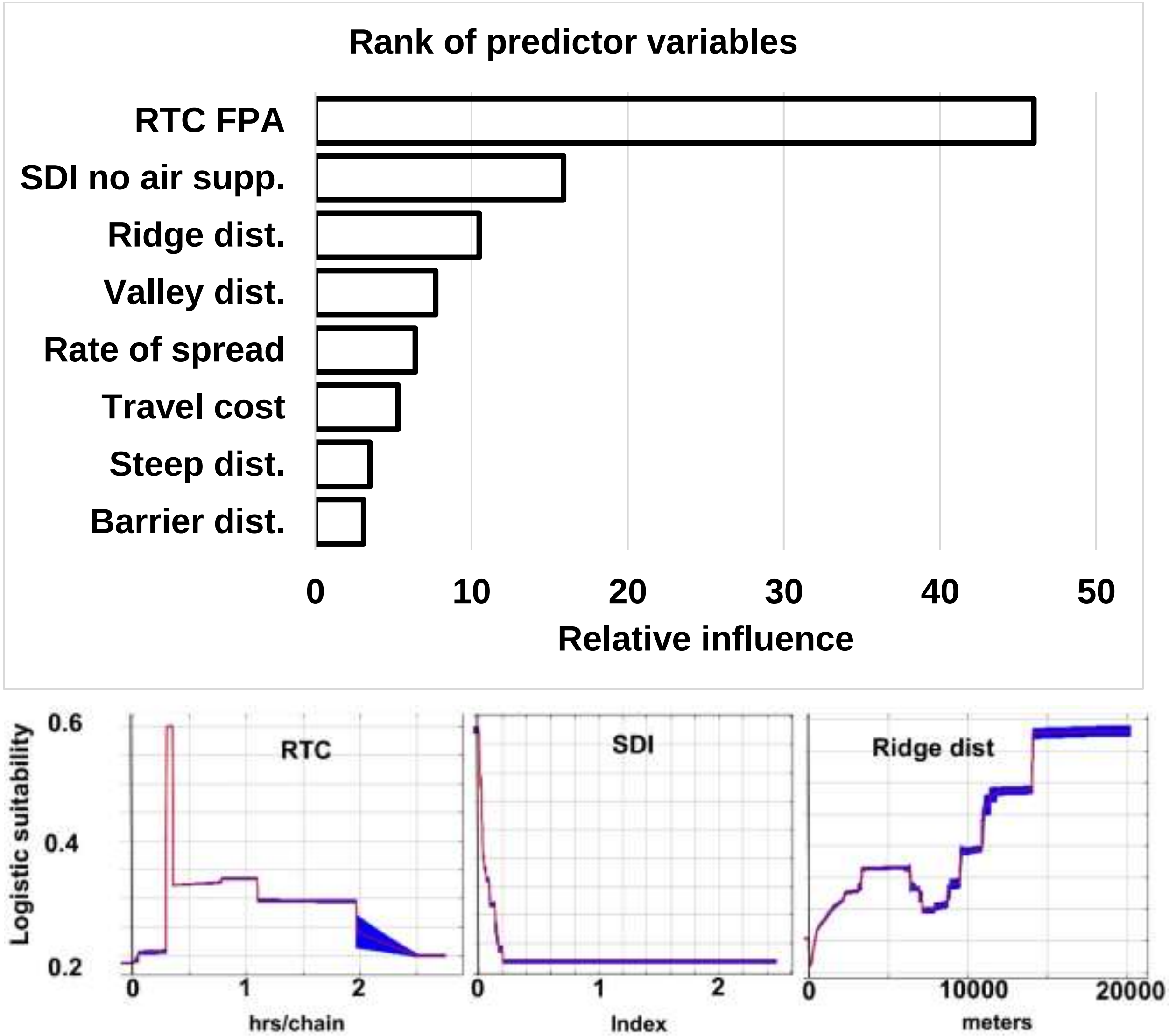

Fig. S4. Summary of maxent predictive model inputs and relationships to observed fire perimeters. Eight significant predictors of final fire perimeter locations are listed in order of relative contribution to improving model performance (a.). 'RTC FPA' is resistance to control using Fire Program Analysis 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.

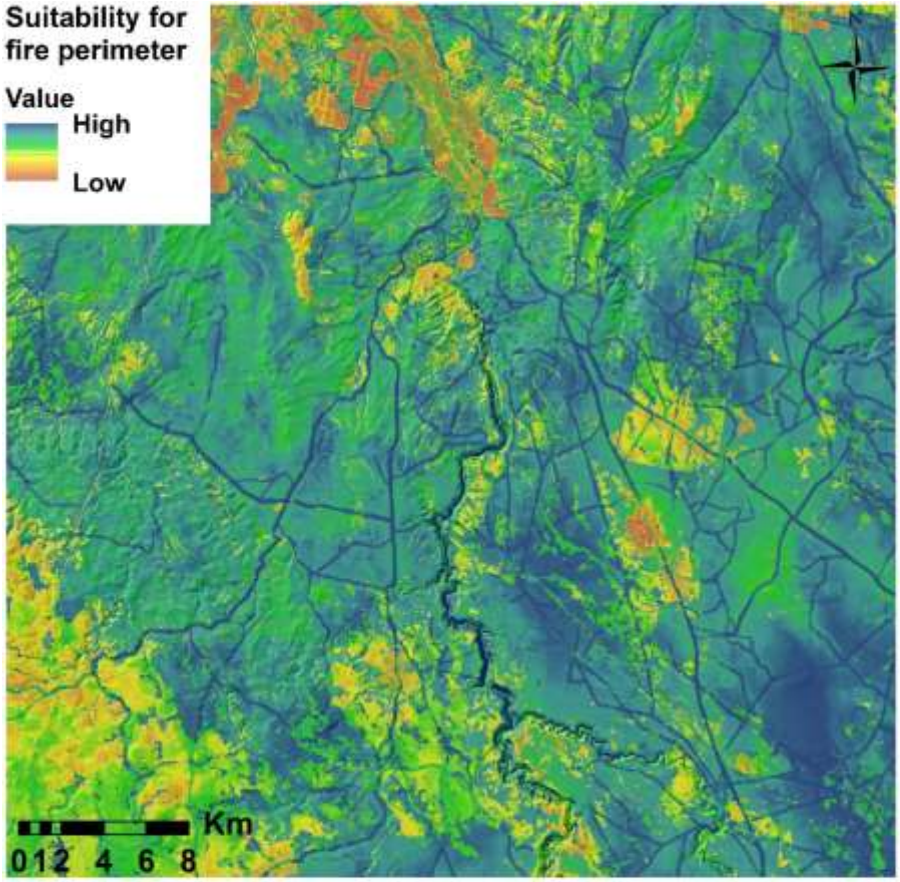

Fig. S5. Maxent model of suitability for fire perimeter locations developed from presence-only regression modelling. For model settings refer to text above. Modelled landscape is identical to that used in Fig. 4. Raw maxent suitability scores sum to one over the entire landscape and are considered a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