10.1071/WF23053
International Journal of Wildland Fire

Supplementary Material

Fighting wildfires: predicting initial attack success across Victoria, Australia

M. P. Plucinski[A,*], *S. Dunstall*[B], *N. F. McCarthy*[C], *S. Deutsch*[D], *E. Tartaglia*[B,D], *C. Huston*[B] *and A. G. Stephenson*[B]

[A]CSIRO, GPO Box 1700, Canberra, ACT 2601, Australia

[B]CSIRO Data61, Private Bag 10, Clayton South, Vic. 3169, Australia

[C]Research and Development team, Fire Risk, Research and Community Preparedness Department, Country Fire Authority, Burwood, Vic., Australia

[D]Department of Energy, Environment, and Climate Action, Data Insights, PO Box 500, East Melbourne, Vic. 8002, Australia

*Correspondence to: Email: matt.plucinski@csiro.au

Table S1. Confusion matrices for validation and training determined from the optimal datasets for the three IA success models threshold for each.

Grasslands

		Validation datasets (*n*=5087, threshold = 0.968)		Training datasets (*n*=20347, threshold = 0.973)	
		Predicted		Predicted	
		IA success	IA escape	IA success	IA escape
Observed	IA success	4000	957	14896	4880
	IA escape	33	97	122	449

Forests

		Validation datasets (*n*=1433, threshold = 0.939)		Training datasets (*n*=5731, threshold = 0.957)	
		Predicted		Predicted	
		IA success	IA escape	IA success	IA escape
Observed	IA success	953	395	3095	2300
	IA escape	22	63	54	282

Shrublands

		Validation datasets (*n*=511, threshold = 0.850)		Training datasets (*n*=2045, threshold = 0.892)	
		Predicted		Predicted	
		IA success	IA escape	IA success	IA escape
Observed	IA success	400	57	1465	397
	IA escape	20	34	57	126

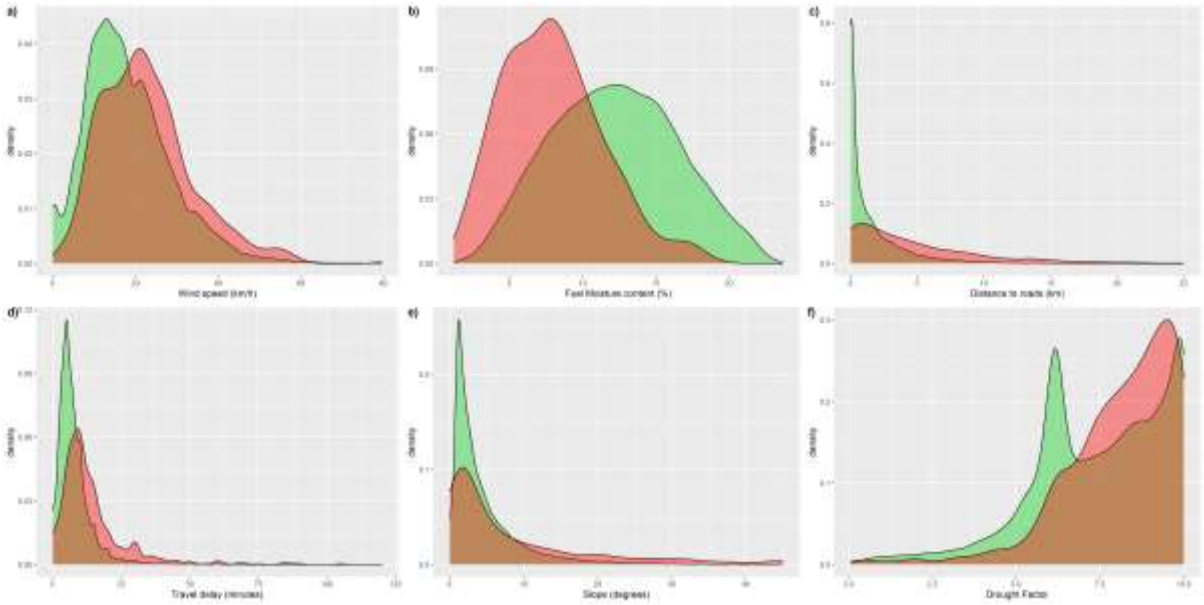

Figure S1. *The distribution of the variables included in the grassland IA model portioned by their IA outcome, with IA success fires shown in green (n= 24,733) and IA escape fires shown in red (n= 701).*

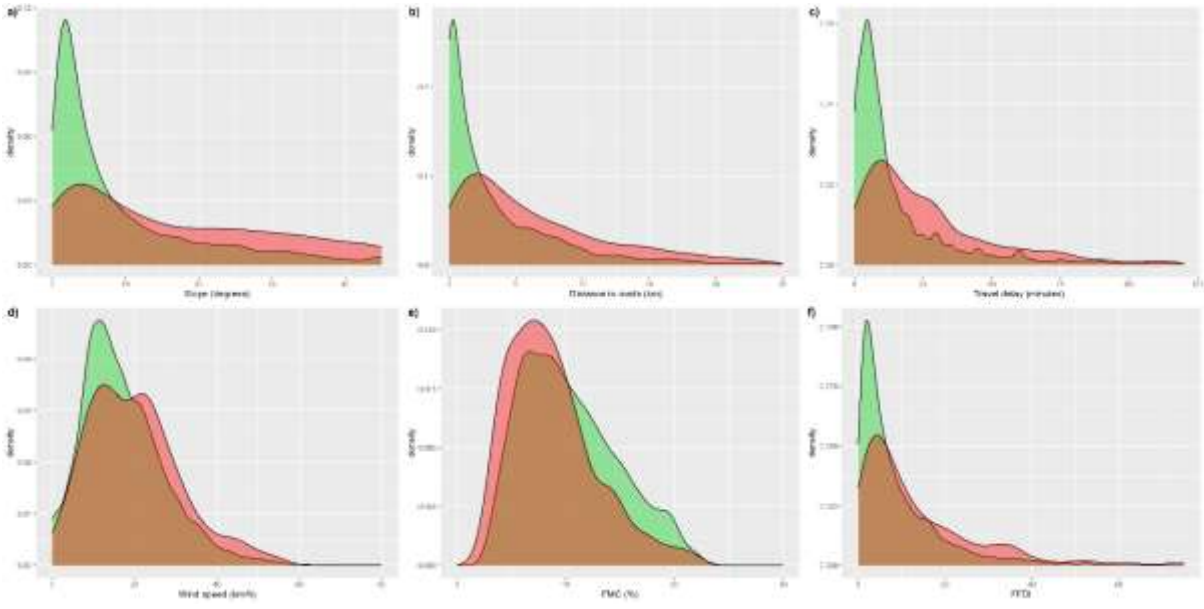

Figure S2. *The distribution of the variables included in the forest IA model and FFDI portioned by their IA outcome, with IA success fires shown in green (n= 6,743) and IA escape fires shown in red (n= 421).*

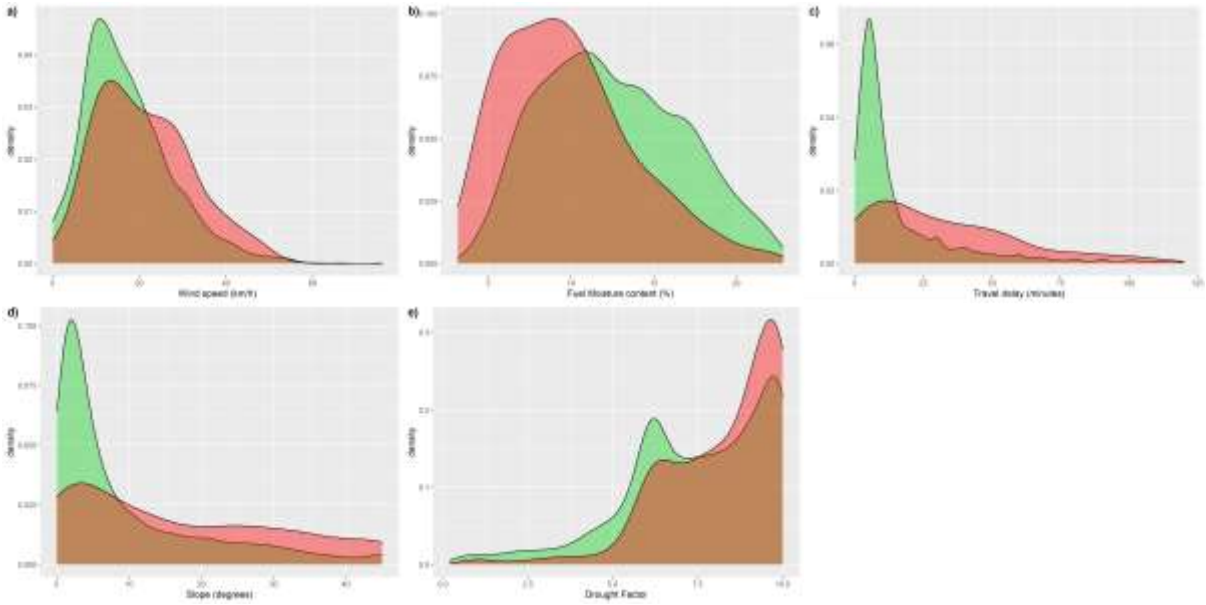

Figure S3. *The distribution of the variables included in the shrubland IA model portioned by their IA outcome, with IA success fires shown in green (n= 2,319) and IA escape fires shown in red (n= 237).*

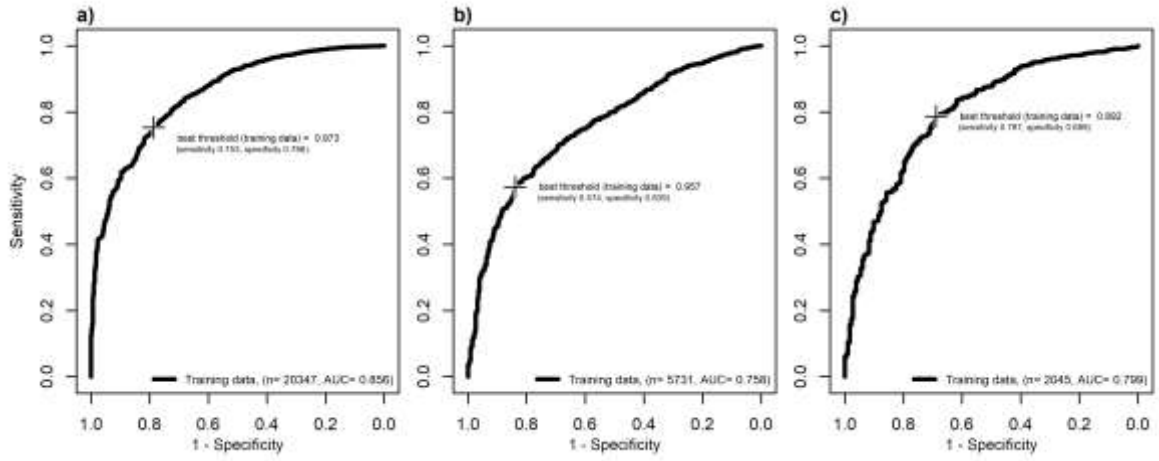

Figure S4. *Plots showing the receiver operating curves for training data for the initial attack models a) grasslands, (b) forests and (c) shrubland vegetation types. The dataset size and area under the curve (AUC) provided in the legends.*

Application of five-fold cross-validation for initial attack success models

The initial attack success models thus far were developed using training data, namely random selection of 80% of the fires from each type, with the remaining 20% forming the validation data. Figures S5-S7 give histograms showing a range of prediction distributions, using a different technique called cross validation. Here we use five-fold cross validation, which randomly splits the dataset into five different parts of 20% each. This yields five different models where for each model, one part is used for validation, with the remaining parts used for training. The five-fold cross validation resulted in similar rates of misclassification for escaped fires (21.4-25.2% for grassfires, 20.7-44.9 % for forest fires, and 25.9-35.8% for shrubland fires) to that of the logistic regression models (Figure 4).

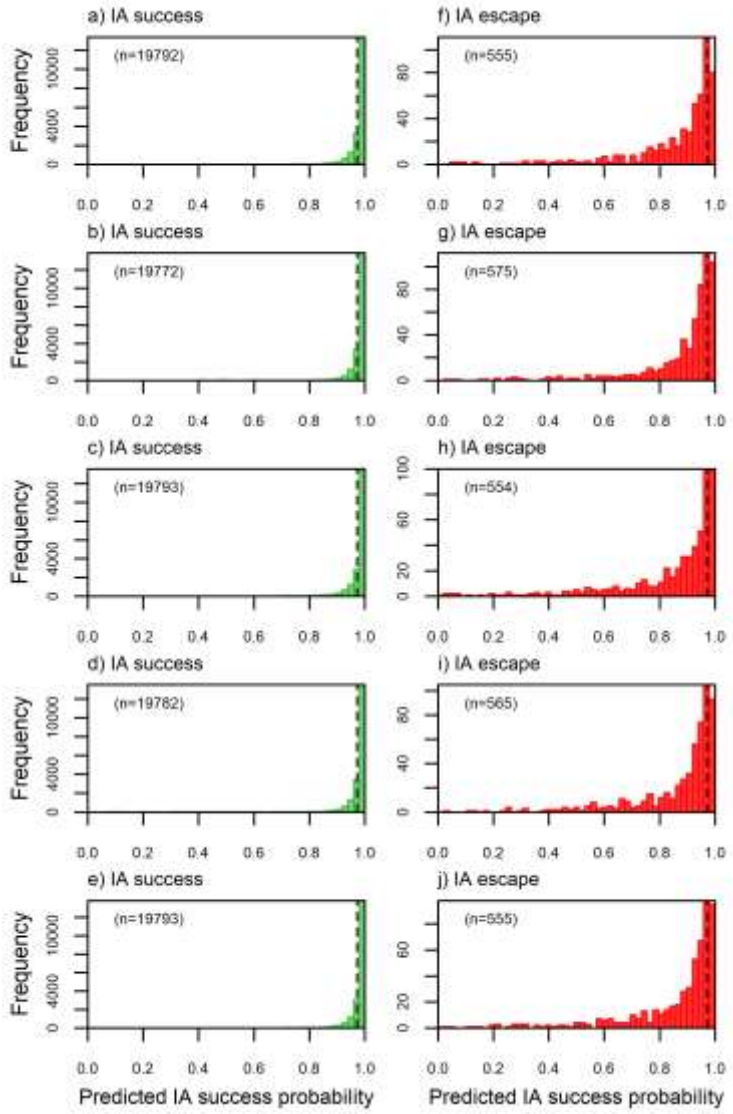

Figure S5. *Histograms of the distribution of predictions for fires in grassland fuels where IA has been successful (green) and fires that have escaped IA (red) in 5-fold cross-validation (each rows shows a separate fold). The dashed vertical lines show the optimal threshold determined from the main grassland logistic regression model (Table 2).*

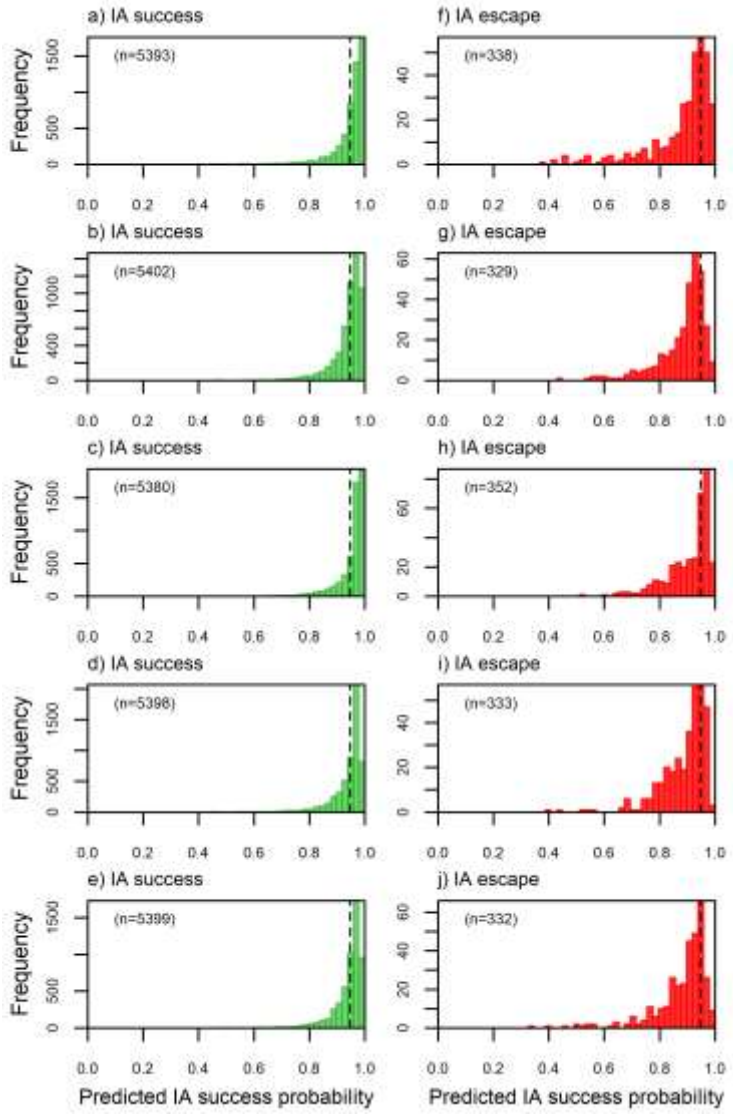

Figure S6. *Histograms of the distribution of predictions for fires in forest fuels where IA has been successful (green) and fires that have escaped IA (red) in 5-fold cross-validation (each rows shows a separate fold). The dashed vertical lines show the optimal threshold determined from the main forest logistic regression model (Table 2).*

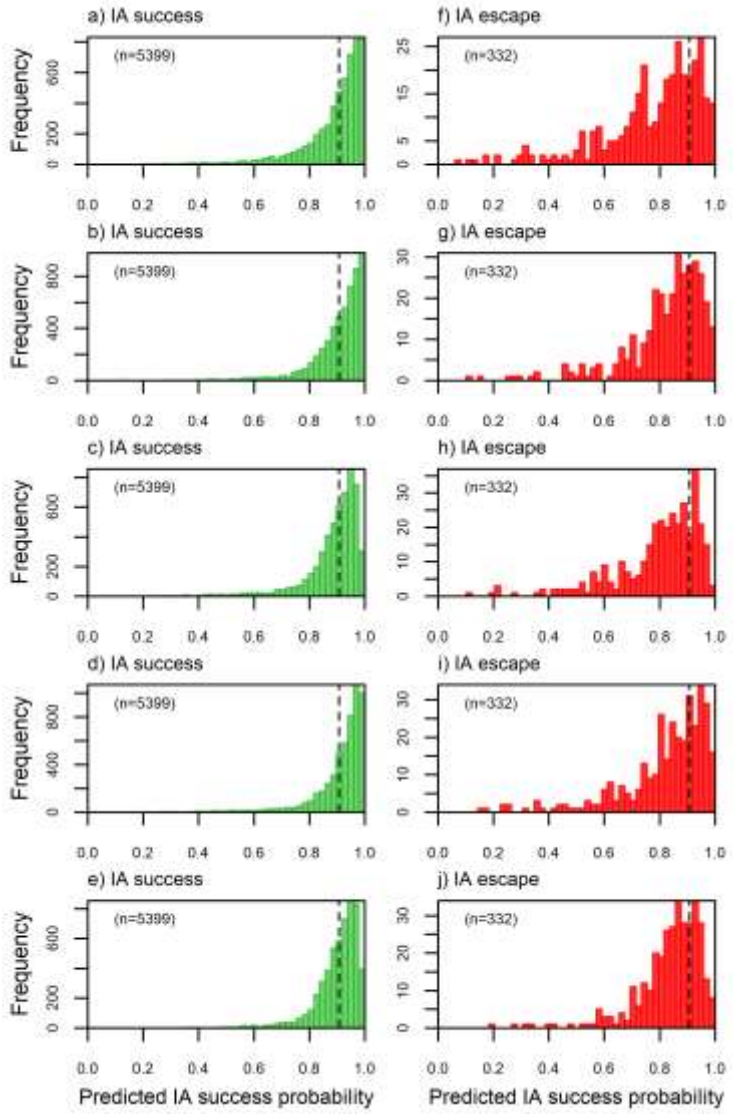

Figure S7. *Histograms of the distribution of predictions for fires in shrubland fuels where IA has been successful (green) and fires that have escaped IA (red) in 5-fold cross-validation (each rows shows a separate fold). The dashed vertical lines show the optimal threshold determined from the main shrubland logistic regression model (Table 2).*

Application of Generalised Additive Models

In Figures S8-S10 we extend the logistic model using the Generalised Additive Modelling (GAM) framework. This allows the probability of initial attack success to depend on smooth functions of the predictors. Here we fit those curves using penalized b-splines, employing the same training dataset and variables for each vegetation type that were used in the logistic regression models of Table 2. The smooth curves are given in Figures S8 – S10, with prediction distributions given in Figure S11. The application of GAMs resulted in only modest improvements in the ability to correctly predict escaped fires (Figure S11) with one fifth being incorrectly classified in the validation dataset (18.5% for grassfires, 16.5% for forest fires and 25.9% for shrubland fires) compared to the logistic regression model (Figure 4).

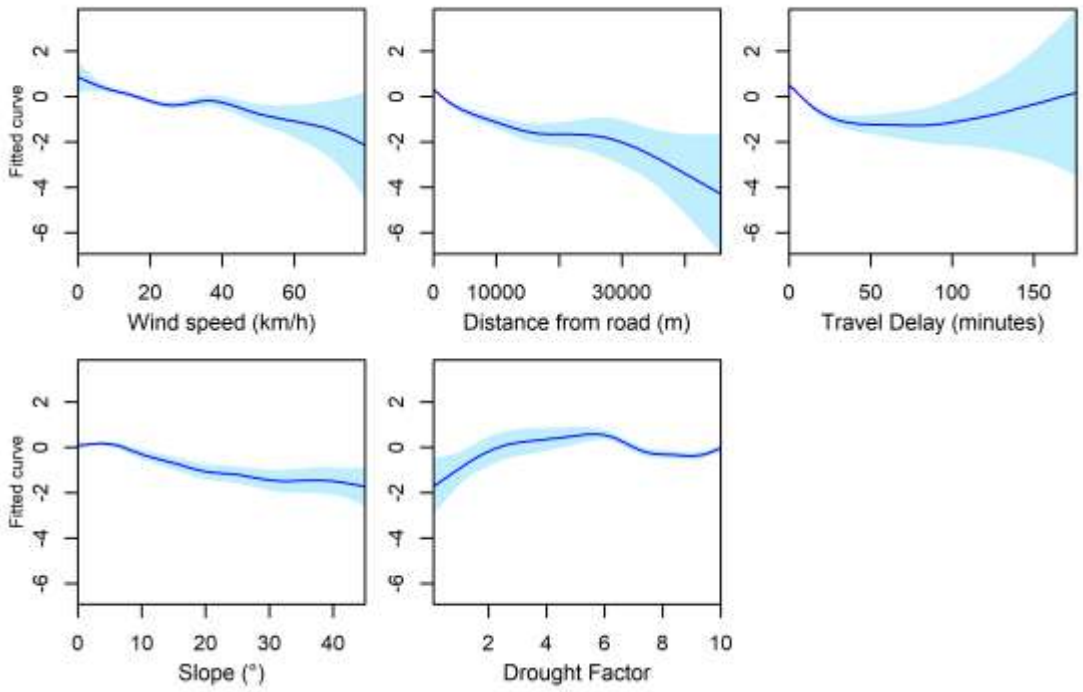

Figure S8. *The partial effects of selected explanatory variables in the grassland Generalised Additive Model on the probability of initial attack success. The y-axis represents the partial effect of each variable. The shaded areas indicate the 95% confidence intervals.*

There are some differences between the logistic regression models presented in the results and the GAMs, notably the trends of the GAM curves for travel delay (Figures S8 – S10). The GAM curves need to be interpreted carefully as the bulk of the data is typically concentrated towards the low end of the x axes (other than drought factor). The GAM for forests indicates that the probability of initial attack decreases more rapidly when the travel delay is longer than two hours (Figure S9), which is likely picking up on remote fires that were lower priority than other concurrent fires during mass ignition events and other remote fires that occurred during mild weather and detected late in the day where it is safer to have crews arrive early the next day. The curves for travel delays in grasslands and shrublands (Figures S8 and S10) were different and are likely influenced fewer fires having slow response times in these vegetation types.

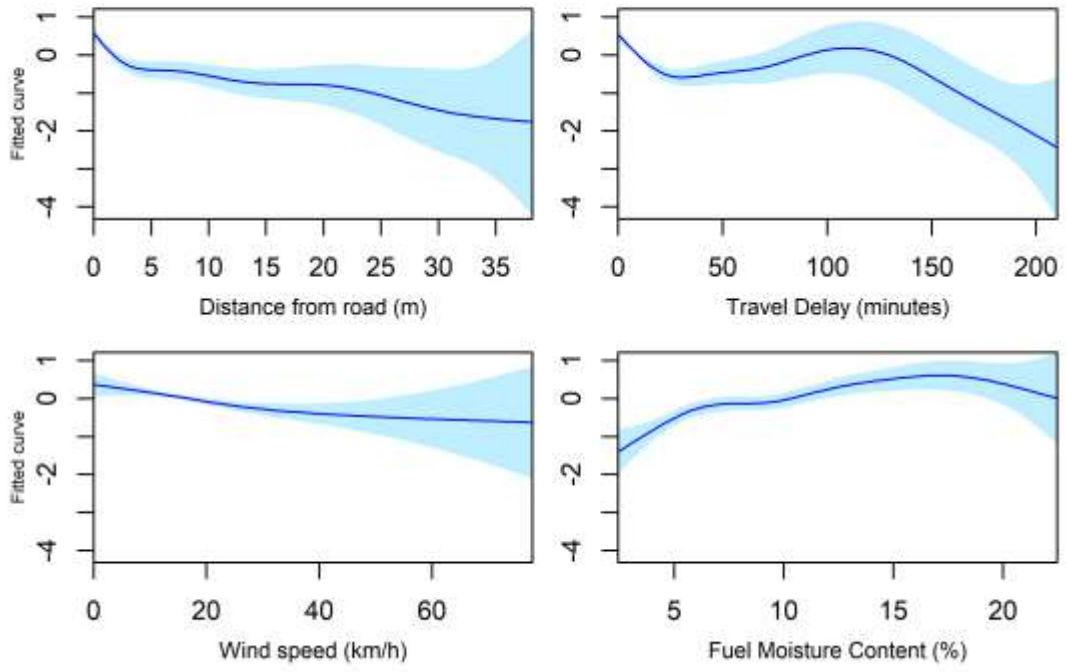

Figure S9. *The partial effects of selected explanatory variables in the forest Generalised Additive Model on the probability of initial attack success. The y-axis represents the partial effect of each variable. The shaded areas indicate the 95% confidence intervals.*

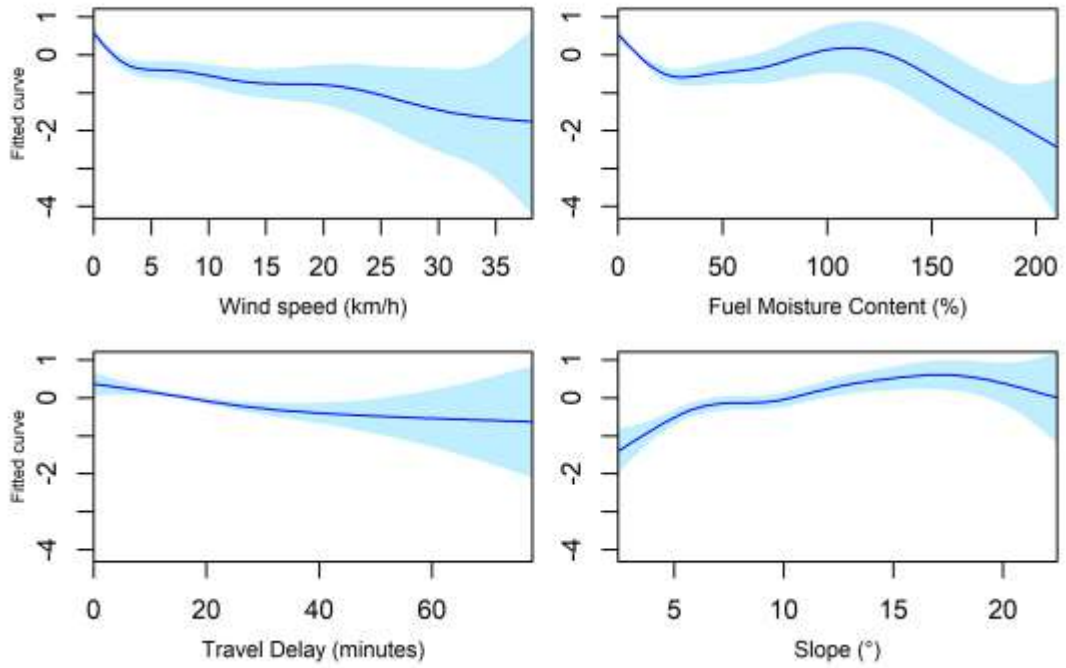

Figure S10. *The partial effects of selected explanatory variables in the shrubland Generalised Additive Model on the probability of initial attack success. The y-axis represents the partial effect of each variable. The shaded areas indicate the 95% confidence intervals.*

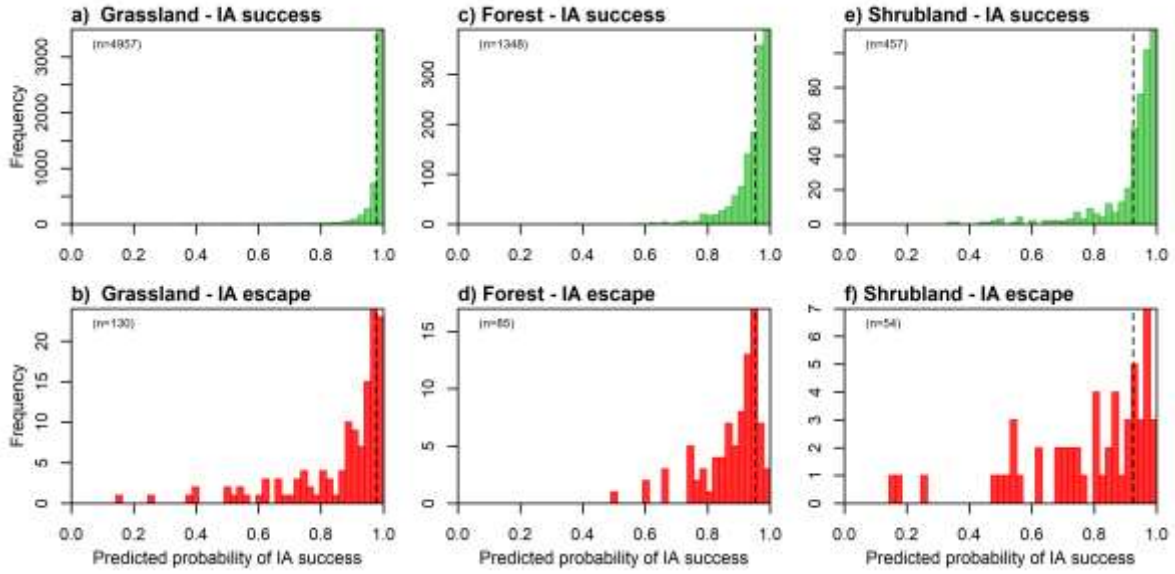

Figure S11. *Histograms of the distribution of Generalised Additive Model predictions for fires where IA has been successful (green) and fires that have escaped IA (red) in the validation datasets for (a, b) grassland, (c, d) forest and (e, f) shrubland vegetation types. The dashed vertical lines show the optimal threshold for each model. Note different y axis scales.*