Biotic and physical drivers of fire in northwestern Patagonia

Supplementary information

Iván Barberá, Ana María Cingolani, Florencia Tiribelli, Mónica Alicia Mermoz, Juan Manuel Morales and Thomas Kitzberger

Appendix 1: vegetation type as a function of physical and human factors

To better understand how fire is affected by the interplay between vegetation type and physical and human factors we performed a quantitative analysis of how vegetation type responds to topography, precipitation and distance from human settlements and roads (hereafter, environmental variables). We based this analysis on the vegetation map from WWF Valdivian Ecoregion Vegetation assessment (Lara et al. 1999), reclassified in 8 categories following Kitzberger et al. (2022), which are shown in figure 1B: wet forest (original classes 01, 03, 05, 06, 07 and 11), subalpine forest (class 08 and 02), dry forest (class 04), shrubland (class 09), grassland (classes 13 and 14), plantations (class 17) and anthropogenic prairies (classes 16 and 18) and the non-burnable class (15, 19, 20, 21 and 22), which was excluded from our analysis. The class 02 in the WWF map comprised mostly subalpine forests dominated by *N. pumilio*, with *Araucaria araucana* presence, but Lara et al. (1999) labelled this community as *A. araucana* forest. We classified it as subalpine forests because the dominant vegetation is the most important for fire activity. We used the same database as in the analysis of spatial patterns of fire, which consisted of environmental data sampled at 8278 points in the burnable area (see the *Data analysis* section in the main text). We fitted Generalized Additive Models assuming a categorical distribution of the vegetation type, with topographic variables, precipitation, and distance from human settlements and roads as predictors. We used both univariate and multiple models (i.e., including only one or all predictors at once). In all cases we modelled the effect of predictors with cubic splines of three basis functions (k = 4 in mgcv), but for aspect we specified cyclic cubic splines. We did not include the NDVI as predictor because it is not a cause of vegetation type, but rather a consequence. As with burn probability models, we did not analyse temperature because it was highly correlated with elevation (Pearson's r = -0.77), and the latter had a much finer resolution (~ 930 m vs. 30 m). In these models, the vegetation classes plantation and anthropogenic prairie were not included due to their low abundance in the landscape.

Univariate models allow to assess the global effect of a predictor on vegetation at the regional scale; while the multivariate model allows a finer-scale description of how predictors affect vegetation, given that the most important regional-scale predictors are included. This is because in the multivariate model the partial effects may be assessed at the values of the most important drivers where each vegetation type occurs. For example, we can evaluate the topographic preferences of grasslands at the low elevation and precipitation range, where they are most abundant. In addition, the multiple model allows to identify which are the most relevant predictors when they are correlated (given that correlations are not too high).

Univariate models showed that vegetation type varied mostly as a function of elevation, precipitation and slope (Fig. S1, Table S1). Along the elevation gradient dry forests occupied the lowest portion together with grasslands (< 500 m a.s.l.), which extended to the mid portion of the elevation gradient. Shrublands and wet forests were more abundant between 500 and 1200 m a.s.l., and subalpine forests dominated above 1200 m a.s.l. and up to the timberline (Fig. S1A). Areas with slope below 10 ° were dominated by shrublands and grasslands; at steeper slopes, subalpine forests were the dominant type, and wet forests were

more abundant at intermediate slope areas (10 to 40 °; Fig. S1B). Along the precipitation gradient wet forests occupied the wettest extreme, followed by shrublands and subalpine forests, and then by dry forests. Finally, steppes occupied the driest portion of the gradient (Fig. S1E). Aspect showed influence only on subalpine forests, which tended to be located at south-eastern slopes, where grasslands and shrublands were less abundant (Fig. S1C). Shrublands, wet and dry forests were more abundant at low topographic position, while subalpine forests dominated the mid to high slopes, with grasslands being associated to high topographic position (Fig. S1D). In areas near human settlements and roads the most abundant types were grasslands, shrublands and dry forests; subalpine forests showed highest expected cover at intermediate distance, and wet forests, at large distance (Fig. S1F and S1G).

In the multiple regression model, where all predictors were included, the effects of elevation and precipitation remained unchanged, the effects of aspect were clearer and the remaining predictors showed milder effects (Fig. S2). Once elevation is considered, slope and topographic position lose importance, as they are positively correlated with elevation (Fig. S4). Dry forests and shrublands tend to prefer high topographic position, and subalpine forests show a mild preference for low topographic position (Fig. S2D). Moreover, dry forests are clearly associated to steep slopes (Fig. S2B). Regarding aspect, subalpine forests prefer south slopes, while wet forests, shrublands and grasslands prefer northern ones (Fig. S2C). Regarding distance from human settlements and roads, wet forests show their highest expected cover at short distance, the opposite from the univariate model (Fig. S2F and S2G). This happens because the areas with highest precipitation are also the furthest away from human settlements and roads (Fig. S4).

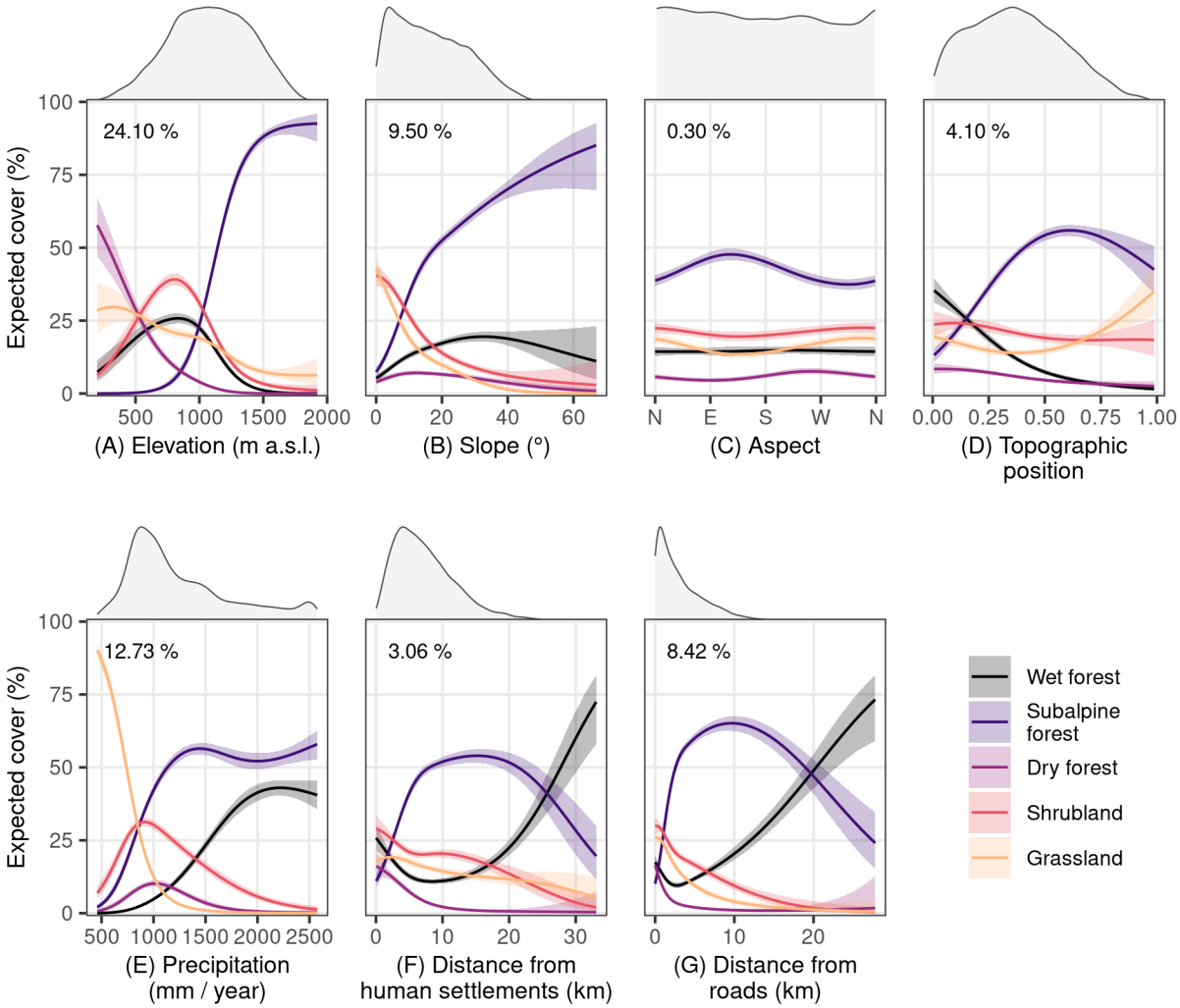

Figure S1. Expected cover (%) of each vegetation type as a function of physical and human factors. The lines show the estimated probability from univariate categorical GAMs with their 95 % confidence intervals. Note that for every value of the predictors all types sum to 100 %. Numbers inside panels show each model's Bayesian R^2, computed as the abundance-weighted average of the R^2 of each vegetation type treated as a Bernoulli variable. Gray shades above panels show the distribution of each predictor variable.

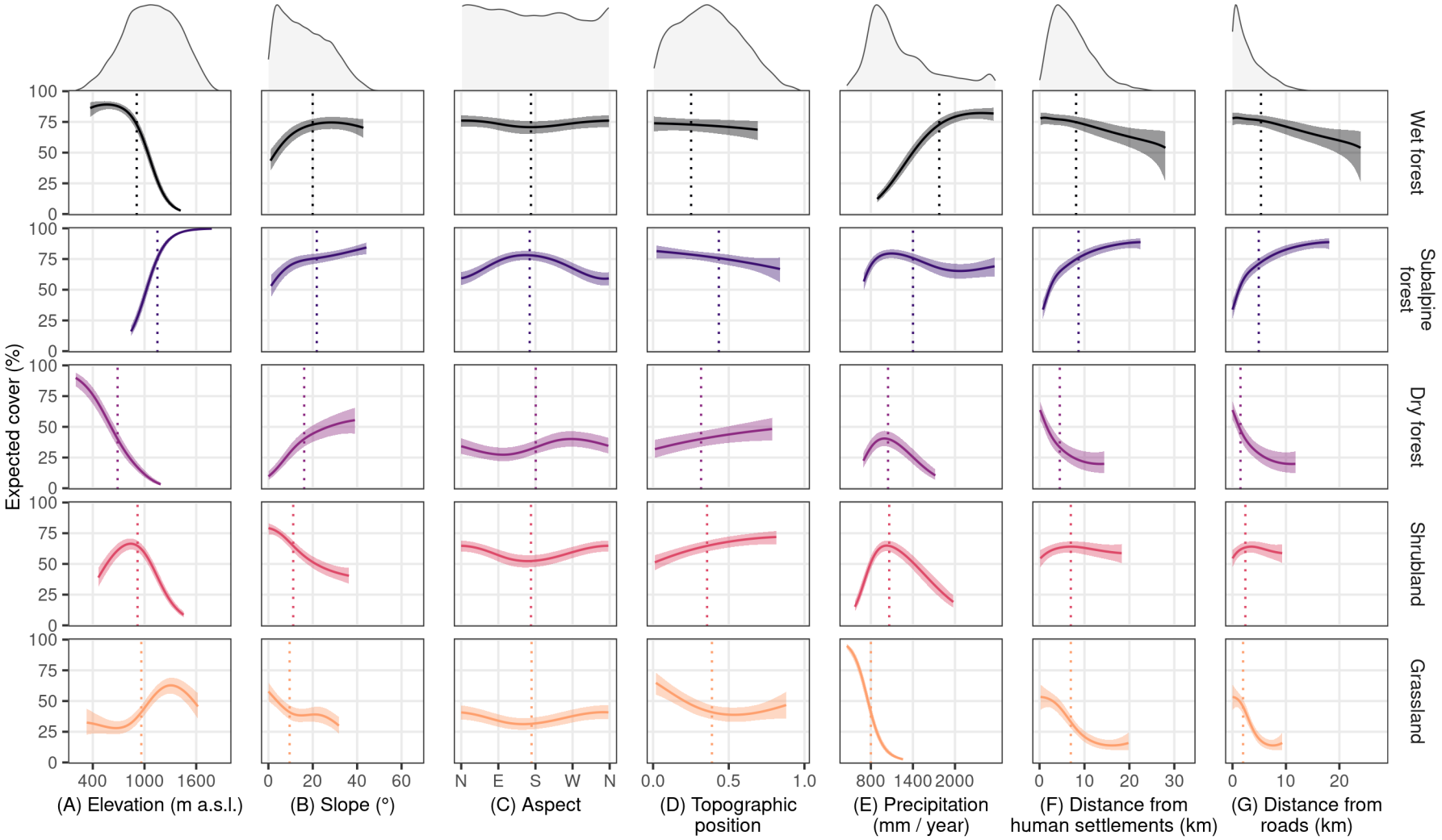

Figure S2. Expected cover of each vegetation type as a function of spatial fire drivers, computed from the multiple regression categorical GAM. In each panel the predictor is varied in the range of each vegetation type, with the remaining predictors fixed at the mean corresponding to each vegetation type (denoted by vertical lines). In the case of subalpine forest, elevation was fixed at 1150 instead of 1322 m a.s.l. to avoid the expected cover being nearly 100 %, as at this value we could not detect effects from other predictors. Gray shades above the panels show the distribution of each predictor considering all vegetation types. Predictions span the 98 % highest density interval of each predictor at each vegetation type.

Table S1. R^2 (%) for vegetation type models, both univariate and multivariate (bottom row). The R^2 was computed for every type separately as the Bayesian R^2 for a Bernoulli distribution. We highlight in bold the highest R^2 in each vegetation type from univariate models. The average R^2 weighted by abundance for each univariate model is shown in figure 4. Predictions from the multiple regression model are shown in figure S4.

Predictor variable	Wet forest	Subalpine forest	Dry forest	Shrubland	Grassland
Elevation	7.16	**46.11**	**13.98**	**10.79**	2.75
Slope	1.46	14.34	0.19	7.79	9.72
Aspect	0	0.53	0.21	0.07	0.28
Topographic position	6.76	6.95	0.53	0.28	0.55
Precipitation	**17.61**	8.84	2.3	5.86	**31.31**
Distance from human settlements	2.44	5.41	3.41	0.69	0.44
Distance from roads	4.2	14.44	4.18	3.28	4.73
Multiple regression	43.04	56.77	27.72	24.87	40.67

These patterns indicate that despite environmental variables are associated to vegetation type, the relationship is not too tight, which allows to approximately separate the effects of vegetation type and environmental variables on fire. This is possible for two reasons. The first is that with considerable overlap among vegetation types along environmental gradients we can compare the estimated burn probability among vegetation types at the same environmental condition, depicting more pure effects of vegetation on fire. This allows to compare how the pattern of burn probability among vegetation types changes if environmental variables are controlled for or not (Fig. 6C vs. 6A, respectively). The second reason is the large environmental variation within vegetation types (see pink densities in Fig. 4 and 5, rows 2 to 6), which allow to estimate the effect of environmental variables within vegetation types, to compare it with their overall effect, which ignores vegetation type. These aspects of the data are depicted by the models predicting vegetation type (Fig. S1 and S2) and by the distribution of environmental variables within vegetation types (Fig. 4 and 5, rows 2 to 6). In addition, this can also be noticed in a more succinct way by identifying vegetation types in an ordination of data points based on the environmental variables (Fig. S3). The highest density regions for the Principal Component Analysis scores by vegetation type show considerable overlap, and also reflect high environmental variation within each type.

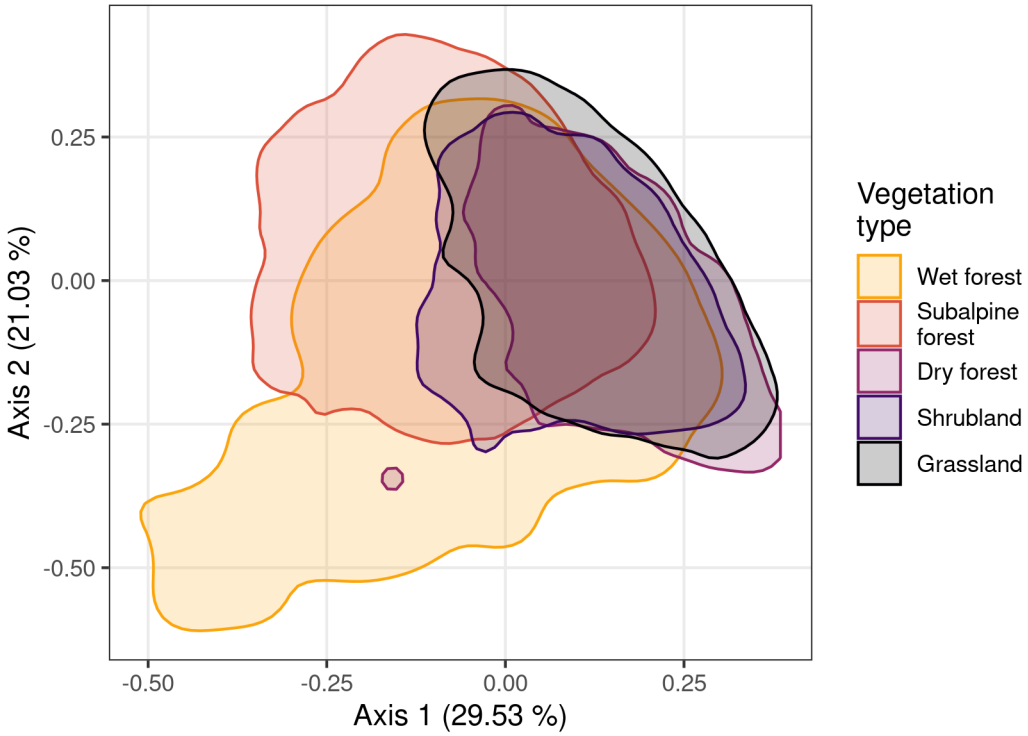

Figure S3. First two axes of a Principal Components Analysis (PCA) performed with the 7 environmental predictors: elevation, slope, aspect (transformed to northing as cos(aspect)), topographic position, precipitation, distance from human settlements and distance from roads. As we had too many data points to be visualized properly (N = 8018), we show only the 90 % highest density regions for the PCA scores of each vegetation type. Data points were grouped by vegetation type after computing the ordination. Percentages indicate the portion of variance explained by each axis.

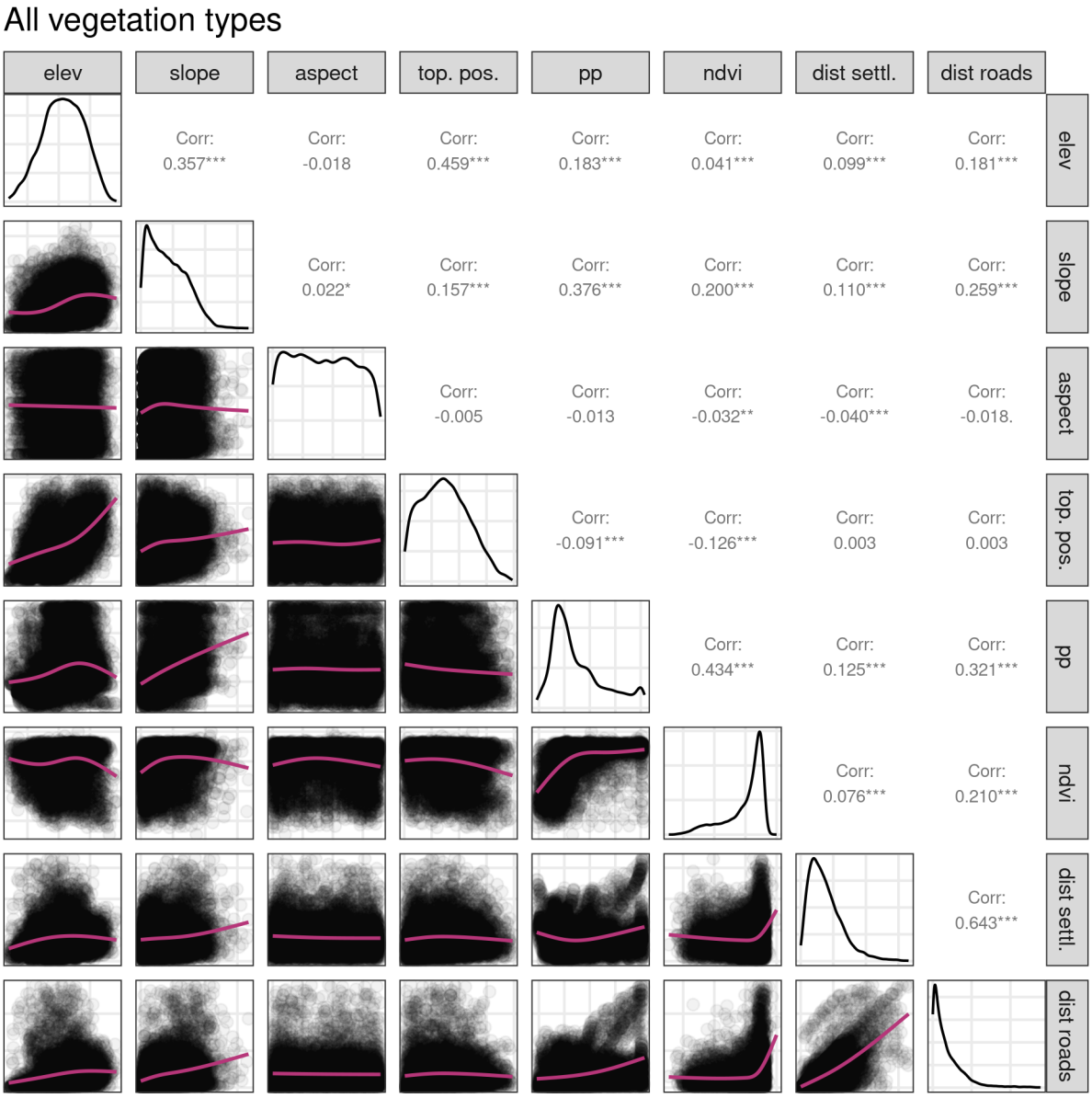

Figure S4. Scatter plots of all predictors of burn probability, using the 8228 points sampled at random from the burnable area. Lines show the fit of a GAM with three cubic spline basis functions (k = 4 in mgcv), assuming a normal distribution of the response. In the diagonal panels the distribution of each variable is shown, and numbers in the upper triangular panels show the Pearson's correlation coefficient. Abbreviations: elev = elevation, top. pos. = topographic position, pp = precipitation, dist settl. = distance from human settlements, dist roads = distance from roads.

Appendix 2: fire mapping methods

The mapping procedure was semi-automatic: first, we produced burned area scars with an automatic approach in Google Earth Engine (Gorelick et al. 2017) using burn and vegetation indices (Normalized Difference Vegetation Index -NDVI- and Normalized Burn Ratio -NBR-), and then we manually edited the fire polygons to amend errors. The automatic approach was reliable to locate where fires occurred, but because the burn signal (NBR and NDVI reductions) remains at least one year after the fire, many fires or parts of them were assigned incorrectly to adjacent years. In addition, burned areas show spectral properties similar to areas affected by other disturbances such as clearcuts and insect defoliations, causing the automatic procedure to generate some degree of commission error. The manual check allowed to overcome these problems. The database is publicly available at GitHub (https://github.com/barberaivan/patagonian_fires.git).

We mapped fires at annual steps, but as most fire activity in our study area occurs between December and March we consider years running from July to June, labelling them with the year of the ending month for simplicity. For example, a fire occurred in October 1998 is assigned to 1999.

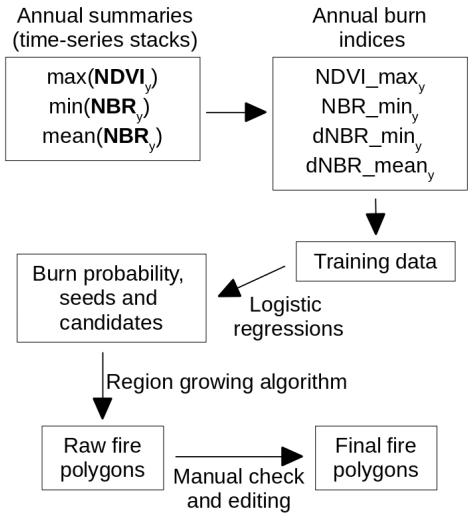

Figure S5. Fire mapping algorithm scheme. Most of the procedure was carried out in Google Earth Engine (Gorelick et al., 2017), while the logistic regressions were fitted in R (R Core Team, 2022) and for the manual check and editing we used both software combined with QGIS (QGIS Development Team 2022). The raw vegetation indexes used were the Normalized Difference Vegetation Index (NDVI) and the Normalized Burn Ratio (NBR). The suffix "y" indicates each year, including images only from December to April and taking the label of the ending year.

2.1 Automatic phase

The automatic mapping stage was similar as described by Long et al. (2019), based on a region-growing algorithm: for each year since 1999 to 2022 (included) separately we defined pixels with high probability of being burned, called seeds, and pixels with low or medium probability, called candidates. The region-growing algorithm, generates burned area clusters starting from seeds and extending the clusters across all the neighbouring (8 pixels) candidate

pixels (SNIC algorithm, Achanta and Susstrunk 2017). The pixel-level burn probability used to define seeds and candidates was defined by three logistic regression models, using NDVI- and NBR-based indices as predictors (Fig. S6A). For every year, we computed the NDVI and NBR from every Landsat image available (Surface Reflectance product, Collections 1 and 2) intersecting the study area, taken between December and April (included), as this period has low cloud cover. From these variables, we computed four derived indices by year (y):

$$NDVI\_max_y = max[max(\mathbf{NDVI}_y), max(\mathbf{NDVI}_{y-1})]$$
$$dNBR\_min_y = min(\mathbf{NBR}_{y-1}) - min(\mathbf{NBR}_y)$$
$$dNBR\_mean_y = mean(\mathbf{NBR}_{y-1}) - mean(\mathbf{NBR}_{y+1})$$
$$NBR\_min_y = min(\mathbf{NBR}_y),$$

and the maximum average NDVI across all years:

$$NDVI\_mean\_max = max[mean(\mathbf{NDVI}_{1999}), mean(\mathbf{NDVI}_{2000}), \ldots, mean(\mathbf{NDVI}_{2022})].$$

Bold variables represent vectors with all the available values of the corresponding spectral indices between December and April of year y, for every pixel (example: for 1999 it uses images from December 1998 to April 1999). From the four dynamic indices (with "y" suffix) we fitted three logistic regressions to model the pixel-level burn probability:
(R-like syntax)

$$logit(p\_dNBR\_min_y) = dNBR\_min_y + NDVI\_max_y + dNBR\_min_y \times NDVI\_max_y$$
$$logit(p\_dNBR\_mean_y) = dNBR\_mean_y + NDVI\_max_y + dNBR\_mean_y \times NDVI\_max_y$$
$$logit(p\_NBR\_min_y) = NBR\_min_y + NDVI\_max_y + NBR\_min_y \times NDVI\_max_y$$

We fitted the models using point data taken from in and around known fires. We chose five well-documented large fire events and visually selected ~ 300 burned and ~ 300 unburned pixels by fire, based on the RGB composite of Landsat images shortly after the fires occurred. We learnt to distinguish between burned and unburned pixels based on the image colour, assessing whether the NBR time-series showed a drop and inspecting Google Earth images where fire scars were evident. The NDVI_max was included to allow the thresholds based on NBR-based indices to vary as a function of the pixel productivity, as low productivity areas such as grasslands show a smaller decrease in NBR when they burn than high productivity areas, as forests (Fig. S6A). This strategy is similar to using relativized burn indices, such as RdNBR (Miller and Thode 2007), but exploratory data analyses suggested our approach would be more accurate. We also tried a Random Forest and a Support Vector Machines classifier, including as predictors the raw Landsat bands and several vegetation indices. As we found no improvements in accuracy, we chose the simpler regression methods.

We defined seeds as pixels with burn probability ≥ 0.98 for the three models also showing either high productivity (NDVI_mean_max ≥ 0.7) or low elevation (< 1300 m a.s.l.). This constraint was used to avoid false positives that appeared in high-Andean areas, frequently covered with snow. Candidates were defined separately according to the pixels' productivity:

(a) Low productivity: if $NDVI\_max_t \leq 0.7$, candidates where pixels where $(p\_dNBR\_min_t, p\_NBR\_min_t) \geq 0.25$,

(b) High productivity: if $NDVI\_max_t > 0.5$, candidates where pixels where $p\_dNBR\_mean_t \geq 0.20$.

In this way, the indices used to define burn probability for low productivity pixels (a) do not use the NBR in the following year (t+1), which is convenient because grasslands recover rapidly, so p_dNBR_mean does not detect correctly burned grasslands. On the contrary,

woody areas (b) show a prolonged decrease in NBR when burned, so the use of p_dNBR_mean is better to classify those pixels. Figure 3B shows the commission and omission error rates for each model.

Fire polygons from the same year were assigned the same fire ID if they were less than 85 m apart (closer than two diagonal 30-m pixels).

2.2 Manual phase

We checked all fire polygons ≥ 10 ha and removed those confounded with urbanization, cultivated areas, deforestation, snow or ice, flooded areas or residual fire scars corresponding to adjacent years. We edited or remapped overlapping adjacent fires occurred in consecutive years. In addition, we merged (split) nearby fire scars belonging to the same (different) fire event(s) based on local knowledge or on the burn dates.

In 1999 few images were available at the study area, so a few fires could not be mapped with desirable precision. To overcome this problem, we included in our database three fires from 1999 mapped by Administración de Parques Nacionales, Argentina (Mermoz 2002).

2.3 Fire date definition

Some fires were well documented, and the start date was obtained from fire management institutions or news. If the period when most of the area burned was known, the fire date was defined as the central date of the period. If the period was not available, the start date was set. For large fires, we complemented this information with the MODIS daily burned area product (MCD64A1) and MODIS hotspots (MCD14DL), which helped to define the period when most area burned. For fires with no information available that were not detected by MODIS, either because they occurred before 2001 or because they were too small, we defined the fire date based on the NBR time series computed from all available Landsat images. We visually inspected the time series within three years before and after the supposed burn year to detect a sharp NBR drop related to burning. The fire date was defined as the central date between the latest day when the NBR was high before the drop (plus one day) and the first day after the drop when the NBR was low.

2.4 Mapping error assessment

To quantify the error of the whole mapping procedure (automatic + manual) we assessed 22 fires that could be clearly identified in Google Earth. For each fire we visually selected burned and unburned points using Google Earth, and then evaluated whether they were within or outside their corresponding fire polygon. The number of points taken for each fire varied between 28 and 289 (considering separately burned and unburned ones), depending on fire size and the availability of clear images in Google Earth to identify them correctly (Table S2). To estimate the average commission and omission error across fires from the counts of incorrectly classified points over the total in each of the 22 test datasets, we fitted a Beta-Binomial model for each error type. This model estimates the error probability in each fire as a random effect, which follows a Beta distribution, which estimated mean represents the average error across fires. The average commission error was 0.64 % (95 % confidence interval, CI = [0.32 %; 1.30 %]), and the omission error was 3.36 % (95 % CI = [2.02 %; 5.52 %]). Table S2 shows the error rate and number of points by fire. The overall mapping accuracy was estimated in the same way, as 98.06 % (95 % CI = [96.92 %; 98.78 %]).

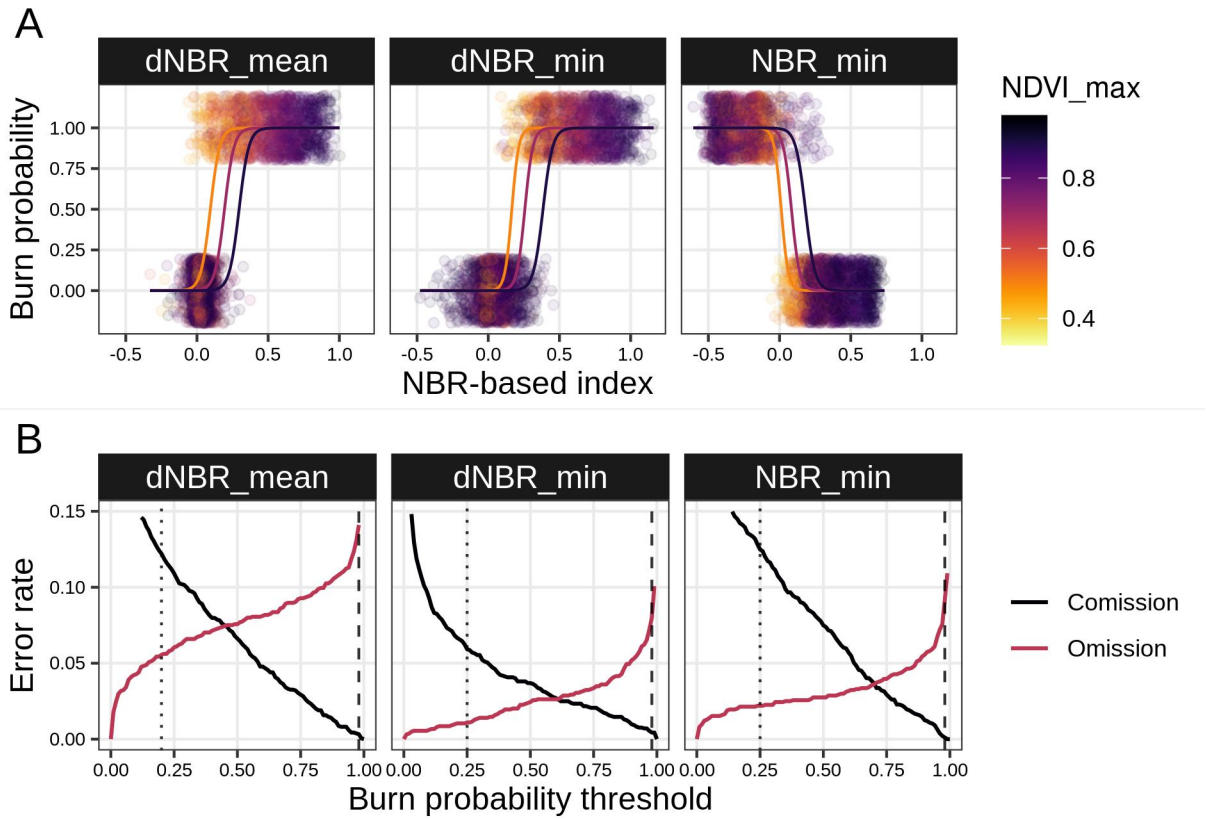

Figure S6. (A) Burn probability functions used to classify burned pixels, defined by spectral indices computed from Landsat imagery. Lines show the predicted burn probability for each model at different NDVI_max values (0.5, 0.7 and 0.9). Points show the training data (burned = 1, unburned = 0) with a jitter at the y axis for visibility. Models were trained with 3480 pixels belonging to five fire events. **(B)** Commission and omission error rates as a function of the burn probability threshold used to classify pixels as burned for each model, computed with a 5-fold cross-validation, where each fold was a separate fire. Commission error is the proportion of unburned pixels in the test set classified as burned, and omission error is the proportion of burned pixels classified as unburned. Vertical dotted lines show the burn probability thresholds used for candidates (0.20 or 0.25, depending on the model) and dashed lines show the threshold used for seeds (0.98) in the region growing algorithm. These thresholds achieved low omission error for candidates and low commission error for seeds.

Table S2. Omission and commission error and overall accuracy of the fire mapping procedure for 22 fires. The numbers between parentheses show the amount of points classified as unburned that were actually burned over the number of burned test points (omission), the amount of points classified as burned that were actually unburned over the number of unburned test points (commission), and the number of points -either burned or unburned- correctly classified over the total (accuracy).

Fire ID	Omission error (%)	Commission error (%)	Accuracy (%)
2008_1379405717	0 (0/133)	0 (0/116)	100 (249/249)
2008_3	0 (0/59)	1.32 (1/76)	99.26 (134/135)
2008_5	0 (0/62)	0 (0/179)	100 (241/241)
2012_57	1.96 (2/102)	0 (0/117)	99.09 (217/219)
2012_58	0 (0/91)	0 (0/120)	100 (211/211)
2014_3	2.75 (3/109)	0 (0/157)	98.87 (263/266)
2014_4	2.94 (2/68)	0 (0/159)	99.12 (225/227)
2015_16	0 (0/59)	0 (0/56)	100 (115/115)
2015_40	4 (6/150)	5 (7/140)	95.52 (277/290)
2015_53	0.86 (1/116)	0 (0/170)	99.65 (285/286)
2016_-665619614	0.84 (1/119)	2.02 (2/99)	98.62 (215/218)
2016_47j	0.63 (1/160)	0.63 (1/159)	99.37 (317/319)
2016_83N	3.92 (6/153)	0 (0/146)	97.99 (293/299)
2016_91	8.11 (3/37)	0 (0/90)	97.64 (124/127)
2018_44	0 (0/87)	0.8 (1/125)	99.53 (211/212)
2020_2089476162	3.13 (1/32)	0 (0/34)	98.48 (65/66)
2021_1229	1.31 (2/153)	1.52 (3/198)	98.58 (346/351)
2021_2146405150_E	4.24 (7/165)	0.79 (1/127)	97.26 (284/292)
2021_2146405150_W	10.55 (25/237)	0.4 (1/248)	94.64 (459/485)
2021_865	16.26 (47/289)	0.97 (2/207)	90.12 (447/496)
2021_936	7.14 (2/28)	0 (0/30)	96.55 (56/58)
2022_2125136700_r	3.81 (8/210)	0 (0/231)	98.19 (433/441)

Appendix 3: extended methods for spatial patterns analyses

3.1 Generalized Additive Models for burn probability

In the environmental and joint burn probability models we initially specified three basis functions by predictor (and by vegetation type, k = 4 in mgcv). However, elevation and precipitation showed wiggly curves in some vegetation types, which were not ecologically plausible, so we used only two basis functions for all of their associated terms (k = 3 in mgcv). Moreover, in the joint model, which included the vegetation type effect and its interactions, we forced the smooth terms to share the same smoothing parameter across vegetation types, allowing them to vary only across environmental predictors. This also helped to reduce the unreasonable wigglyness initially observed in some vegetation types.

3.2 Comparison of burn probability among vegetation types at equal conditions

The joint model (vegetation × environment) allows to compare the predicted burn probability among vegetation types while fixing the environmental predictors at a common value. This allows to explore differences in burn probability not related to environmental variation, which probably reflect contrasting fuel properties among vegetation types. The usual practice to make such comparison is to fix the the environmental predictors in a common value (usually their means) and compare the predicted burn probability across vegetation types conditional on that value; this approach is usually called partial prediction. But as elevation and precipitation have a large effect on vegetation type, their mean values are rare for some vegetation types, and the model predictions are not reliable because there is scarse data around those values. Specifically, dry and subalpine forests occur at the lowest and highest elevation range, respectively, and wet forests and grasslands occur at the moistest and driest precipitation ranges, respectively. To solve this problem, we compared burn proabability among vegetation types under four environmental conditions: low and high elevation combined with low and high precipitation.

Another aspect to consider is that we allowed the effect of environmental variables on burn probability to vary among vegetation types (because we included vegetation × environment interactions). This can make the prediction at a given point in the environmental space non-representative of the overall differences between vegetation types. Hence, we computed a partitioned version of a Partial Dependence Plot (PDP; Greenwell 2017) for the burn probability as a function of vegetation type. The PDP is conceptually similar to a partial prediction, but instead of fixing the non-focal covariates in a single value to compute a prediction, it averages over the predictions from all their available values in the dataset. In a classic PDP for vegetation type, the burn probability of each type is obtained as follows: compute the predicted burn probability over all data points, fixing the vegetation type at the focal level but leaving the remaining predictors at their observed value (unchanged); then, average the predicted burn probability across all data points. This procedure avoids the problem of fixing the remaining predictors at a single value, but still creates unrealistic combinations of vegetation type and environmental variables, assuming they are independent. Hence, we performed a partitioned PDP for vegetation type, where the PDP was computed separately for four subsets of the data, divided in four combinations of elevation and precipitation ranges:

	Low	High
Elevation (m a.s.l.)	[500; 900)	[900; 1300)
Precipitation (mm / year)	[750; 1000)	[1000; 1500)

To avoid unrealistic predictions where the models were not reliable, the PDP was computed in each subset considering only the vegetation types that were likely to occur in that condition (Fig. S1): dry forest missing at high elevation; subalpine forest missing at low elevation; wet forest missing at low precipitation, and grassland missing at high precipitation (note the missing bars in Fig. 6C).

A final consideration is that the NDVI should not be treated as an environmental variable. Despite it responds to the environment, it is tightly related to the vegetation type. To avoid using unrealistic combinations of NDVI and vegetation types in the partitioned PDP, within each partition of the data we predicted the NDVI for each data point corresponding to each vegetation type, by using a model where the NDVI is the response variable, as a function of environmental variables interacting with vegetation type (see details in the next section).

3.3 Dealing with correlated predictors

Results from multiple regression models may be delusive if predictors show considerable correlation, even when correlation coefficients are not too high (e.g., < 0.7). To overcome this issue, we considered the correlation between correlated predictors when assessing their effects, as ignoring correlations led to unrealistic patterns. Despite multiple regression models as the ones we used may identify separate effects of each predictor when the correlation among them is moderate, some separate effects may be less ecologically meaningful than the total effects, i.e., those considering the co-variation among predictors.

In our study area slope steepness and topographic position increase with elevation. However, they showed the opposite effect on burn probability with respect to elevation. This led to an estimation of large effects of these three predictors, but they were compensated when all of them were considered at the same time. Hence, we preferred to show their joint effect, evaluating the change in burn probability as these predictors co-vary as expected in the landscape.

Analysing the effects of topography, precipitation and productivity on burn probability was also challenging, as the effects of topography and precipitation are expected to affect burn probability partly by affecting productivity (approximated here with the NDVI). Hence, when analysing the effect of topographic variables and precipitation we co-varied the NDVI. However, we did not co-vary topographic variables and precipitation to compute the partial prediction of NDVI, because NDVI is partly a consequence of these physical variables, not a cause.

Finally, distance from human settlements was highly correlated with distance from roads, so we co-varied their values when assessing their effects.

The effects of environmental variables and NDVI were analysed by computing partial prediction curves, by computing the expected burn probability in a sequence of each predictor fixing the remaining ones at their means. To consider correlations, we co-varied the correlated predictors when making these predictions (making partial predictions not-so-partial). To this end, we fitted three groups of models, where some predictors were the response variables, and others, the explanatory variables:

1. Topographic models: univariate models where elevation, slope and topographic position were either response or predictor.
2. Distance models: two models, to predict distance from roads using distance from human settlements and the other way around.

3. NDVI models, where NDVI was the response variable and all topographic variables and precipitation were the predictors. This model was also used to compute the partitioned PDP for vegetation type (see section 3.2 above).

All these models for predictor variables had two variants, equivalent to the environmental and joint models of burn probability. In the former, vegetation type was not included, but it was in the latter, including interaction terms. Thus, the two models for burn probability had their corresponding predictor-models. For elevation we assumed normal distribution, gamma for slope and distances, and beta for topographic position and NDVI. Topographic models were GAMs, were cubic splines of three basis functions were used. Models for distances were GLMs (just linear effects at the log scale), and the NDVI models had exactly the same structure as burn probability models (see section 3.1 of this appendix).

We illustrate the prediction method with a few examples. To compute the not-so-partial prediction curve for elevation, we initially created a sequence of elevation values. From this sequence we predict slope and topographic position using the models were elevation is the explanatory variable. Aspect, precipitation, and distance from human settlements and roads were fixed at their means, and then, the NDVI is predicted using precipitation and all the topographic variables (slope and topographic position co-varying with elevation and aspect fixed at its mean). After that, the burn probability was predicted. To compute the partial prediction curve of distance from human settlements, we start with a sequence, and simply predict the distance from roads. The remaining predictors are fixed at their means and then the burn probability is predicted. If burn probability was predicted ignoring vegetation types, the environmental predictor-models was used; otherwise, the joint predictor-models, which take vegetation into account, are used.

Figure S7 shows the predictions from the NDVI models, and Figure S8 shows the strictly partial prediction curves for burn probability (“fixed”), which ignore correlation between predictors, along with the not-so-partial prediction curves, considering correlation (“co-varying”). In the main text, figures 4 and 5 show the latter.

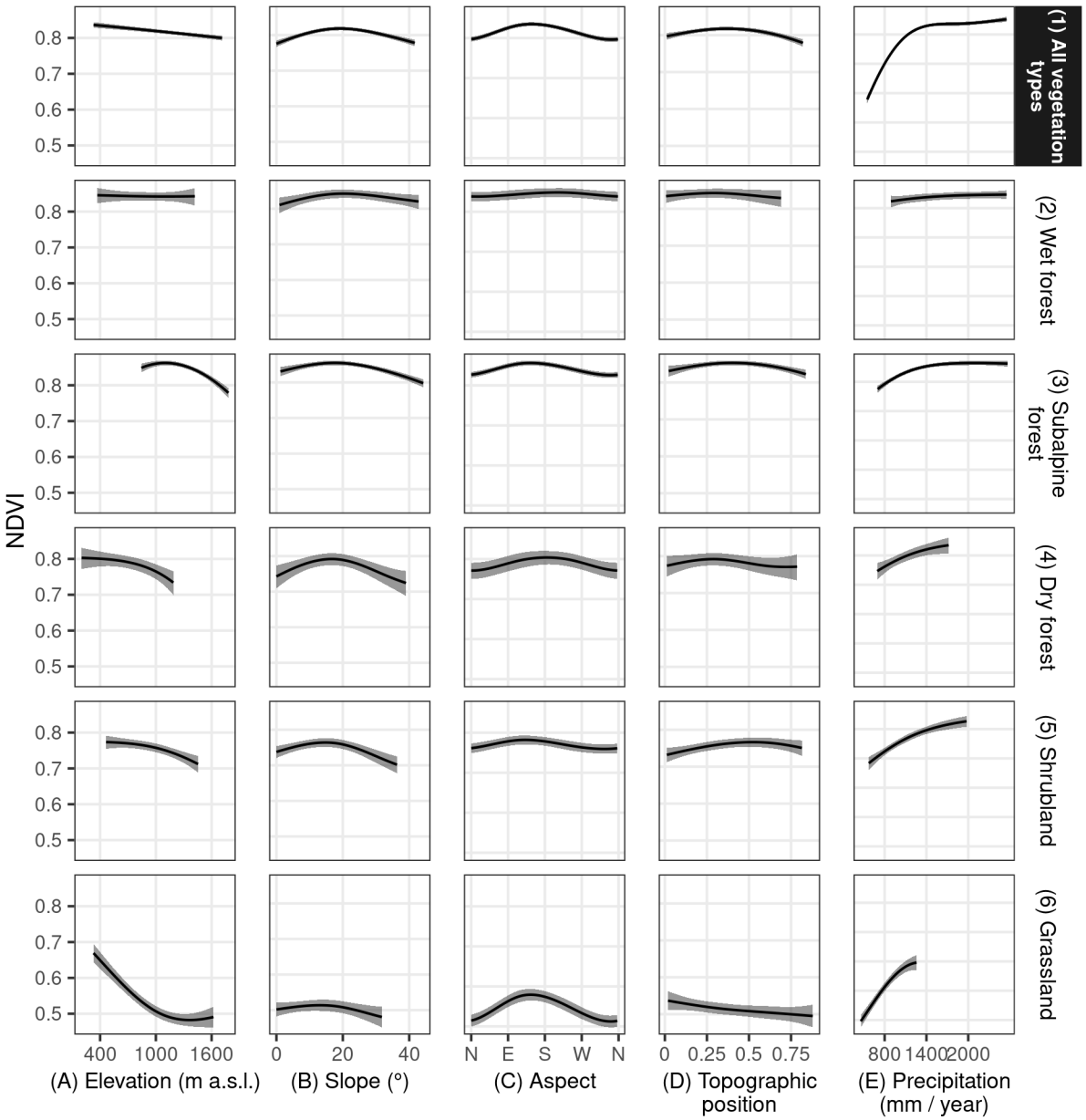

Figure S7. NDVI as a function of topographic variables and precipitation, computed along the 98 % highest density interval at each vegetation type. Lines show the prediction from the maximum likelihood estimate, and ribbons, the 95 % confidence interval.

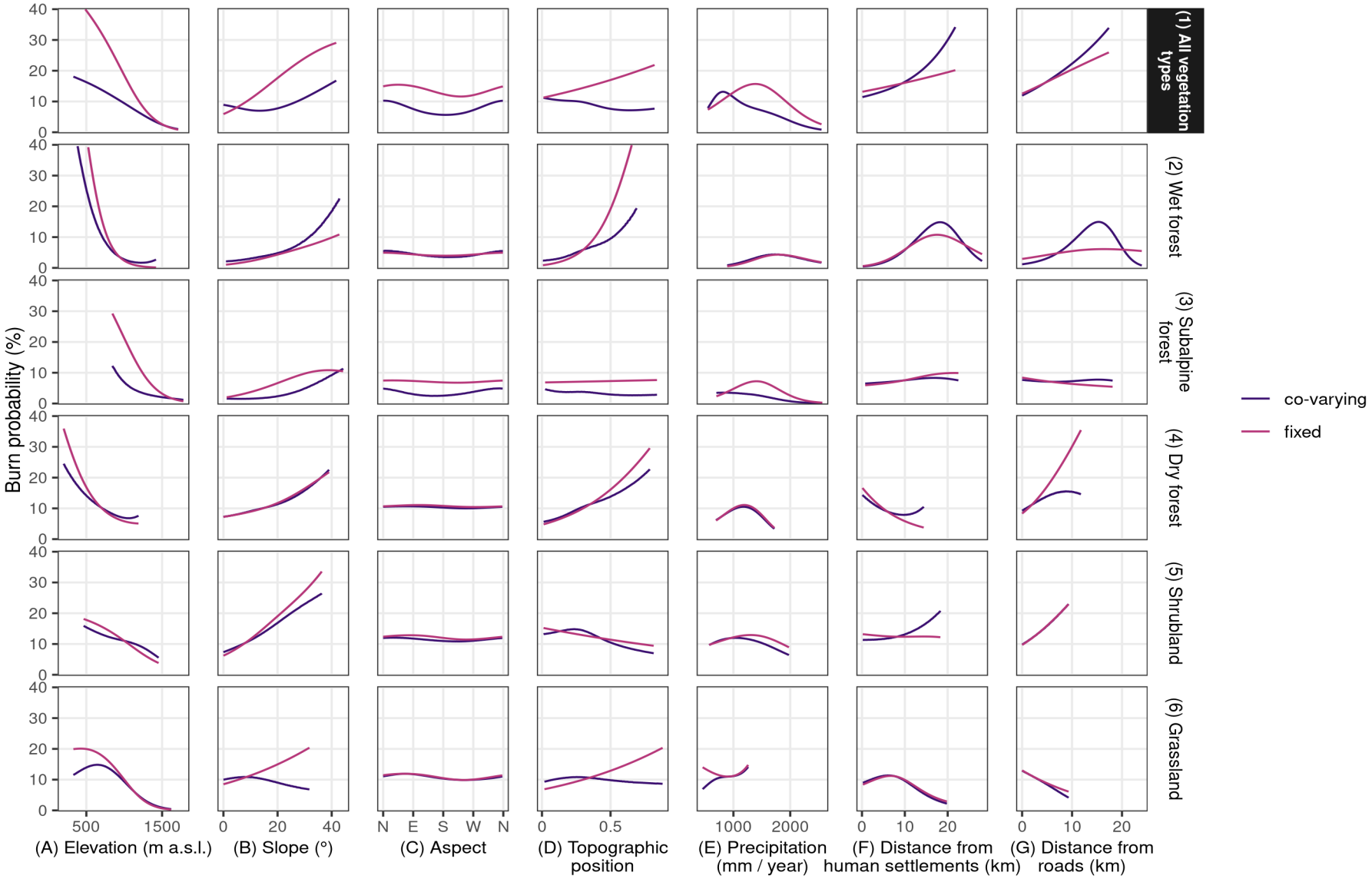

Figure S8. Burn probability as a function of topography, precipitation and distance from human settlements and roads, predicted considering the co-variation between predictors (co-varying) or not (fixed, strictly partial predictions). The darker lines show the same predictions as Fig. 5 and 6 in the main text.

3.4 Effect size metrics

For the environmental variables and NDVI (continuous predictors) the effect size was defined as the mean absolute deviation of the not-so-partial prediction curve, weighting each value of the predictor by its empirical density, both to compute the mean across the curve and to average the absolute difference with the mean. In the case of vegetation type, the computation was similar, but instead of using a prediction curve we used the predicted burn probability for all vegetation types, either from the vegetation model, or from the partitioned PDP using the joint model (section 3.2 in this appendix). For the vegetation model, weights were defined as the relative abundance of vegetation types in the burnable area. When vegetation types were compared at common environmental conditions with the joint model (partitioned PDP), the effect size of vegetation type was computed in each partition. In that case, the weights were defined by the relative abundance of the vegetation types compared in each combination of elevation and precipitation ranges.

3.5 Detecting patterns under spatial correlation

As spatial correlation leads to an underestimation of uncertainty, instead of computing confidence intervals, we assessed uncertainty by comparing our results to those from randomized fire datasets that exhibited similar correlation. We created 500 synthetic burn layers, by turning every observed fire into a circle of the same area and relocating their centroids at random within the burnable area, allowing fires to overlap. These burn layers were rasterized and sampled at the same 8228 points from where observed data were extracted. Then, all the analyses performed with observed data were repeated for each randomized layer. For predictors effects, we computed one-tailed P-values as $1 - \pi$, where π is the percentile corresponding to the observed effect in the distribution of effects from randomized fire datasets. For burn probabilities, we computed two-tailed P-values as follows:

$$\mathrm{P} = \begin{cases} \pi\ 2, & \text{if } \pi < 0.5 \\ (1 - \pi)\ 2, & \text{otherwise.} \end{cases}$$

Where π is the percentile of the observed burn probability in the distribution of randomized fires. In addition, we computed the 80 % and 95 % equal-tailed intervals of all randomized partial prediction curves (based on the maximum likelihood estimate) and visually assessed how far the curve obtained with observed data fell from those intervals.

References

Achanta, R., & Susstrunk, S. (2017). Superpixels and polygons using simple non-iterative clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4651-4660).

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote sensing of Environment, 202, 18-27.

Greenwell, B.M. (2017). pdp: An R Package for Constructing Partial Dependence Plots. The R Journal 9, 421-436.

Long, T., Zhang, Z., He, G., Jiao, W., Tang, C., Wu, B., Zhang, X., Wang, G., & Yin, R. (2019). 30 m resolution global annual burned area mapping based on Landsat Images and Google Earth Engine. Remote Sensing, 11(5), 489.

Mermoz. M. (2002). Detección y mapeo de incendios forestales en los parques nacionales de nor-patagonia, período 1985-1999. Delegación Regional Patagonia, Administración de Parques Nacionales.

Miller, J. D., & Thode, A. E. (2007). Quantifying burn severity in a heterogeneous landscape with a relative version of the delta Normalized Burn Ratio (dNBR). Remote Sensing of Environment, 109(1), 66-80.

QGIS Development Team, 2022. QGIS Geographic Information System. QGIS Association. http://www.qgis.org

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.