ADVANCES IN FOREST FIRE RESEARCH

2022

Edited by
DOMINGOS XAVIER VIEGAS
LUÍS MÁRIO RIBEIRO

Spatial predictions of human and natural-caused wildfire likelihood across Montana (USA)

Adrián Jiménez-Ruano*[1]; W.Matt Jolly[2]; Patrick Freeborn[2]; Daniel José Vega-Nieva[3]; Norma Angélica Monjarás-Vega[3]; Carlos Iván Briones-Herrera[3]; Marcos Rodrigues[1]

[1]Department of Geography and Land Management, University of Zaragoza. Pedro Cerbuna, 12, 50009, Zaragoza, Spain, {jimenez, rmarcos}@unizar.es
[2] Missoula Fire Sciences Laboratory, Rocky Mountain Research Station, USDA Forest Service. 5775 Hwy 10W, Missoula, MT 59808, US {matt.jolly, patrick.h.freeborn}@usda.gov
[3] Faculty of Forestry Sciences, University of Juarez of the State of Durango. Río Papaloapan y Blvd, Durango S/N Col. Valle del Sur, 34120 Durango, Mexico, {danieljvn, normonjaras, briones.ipi}@gmail.com

**Corresponding author*

Keywords

Wildfire occurrence; fire danger; Spatial GAM; BAM; Montana

Abstract

Spatial wildfire ignition predictions are needed to ensure efficient and effective wildfire response but robust methods for modelling wildfire occurrence have not been fully evaluated. Here we leverage high resolution, static spatial data to predict the ignition locations of human and naturally-caused wildfires across the state of Montana (USA). We leveraged a 25-year historical wildfire dataset (1992-2017) and four high resolution spatial variables that capture fuel availability, topography, geographic location and human transport infrastructure. We combined these data to train spatial logistic regression Generalized Additive Models (GAM) designed for big datasets (BAM) for both human and natural ignitions and we tested the efficacy of incremental changes in model complexity. Results showed that the best human and natural-caused ignition models were highly capable of distinguishing locations with and without new wildfire occurrences statewide (AUC = 0.89 and 0.84 respectively). Natural-caused ignitions were strongly influenced by slope and location, while human-caused fires were best predicted by distance to roads as well as terrain and fuel availability. Finally, these spatial fire occurrence models can be combined with temporally-variant data to predict the spatial and temporal distribution of wildfires across the state with the view that these methods can be used to develop predictive models at larger scales.

1. Introduction

Wildfires are present in all vegetated environments worldwide (Krawchuk et al. 2009; 2011; Archibald et al. 2013), where they would otherwise be beneficial, are suppressed to limit or prevent immediate human-related loss (Riley et al. 2018). Although effective wildfire planning requires early information about where new ignitions are likely to start (Belval, Stonesifer, and Calkin 2020), this information is generally not available at high spatial resolution.

New wildfires start when there is an ignition source, sufficient fuel and suitable weather conditions. Natural ignitions are mainly caused by cloud-to-ground lightning strikes and depend on the timing and lightning characteristics in relation to fuel condition (Mitchener and Parker 2005; Schultz et al. 2019). Negligence causes are the most common ones (Prestemon et al. 2013) being favored by road networks (Faivre et al. 2014; Benefield and Chen 2022). Therefore, differentiating the spatial distributions of wildfire ignitions is crucial to designing and implementing source-specific prevention and mitigation strategies.

Wildfire ignition likelihood may be modelled if proxies are included to represent the fuels, weather, topography and human presence that control their spatial distribution (Barreto and Armenteras 2020). Although fuels are seasonally dynamic, static factors such as loading and continuity may limit the wildfire location where insufficient fuels prevent propagation (Briones-Herrera et al. 2019). Road networks affect fire activity in two

ways: first, making wildlands accessible to humans, thus promoting anthropogenic wildfire ignitions (A. Syphard et al. 2007; Narayanaraj and Wimberly 2012); and second, roads enable ground-based firefighting resources to quickly respond to new events (Thompson, Gannon, and Caggiano 2021). Thus, if ignition sources are differentiated and adequate spatial biophysical-demographic drivers are identified, wildland fire ignition likelihoods could be modeled, mapped and predicted across large areas.

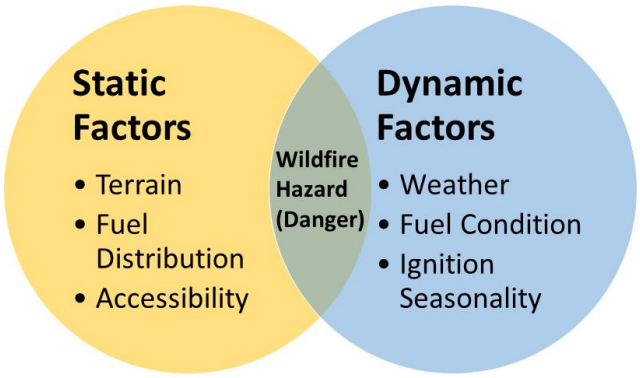


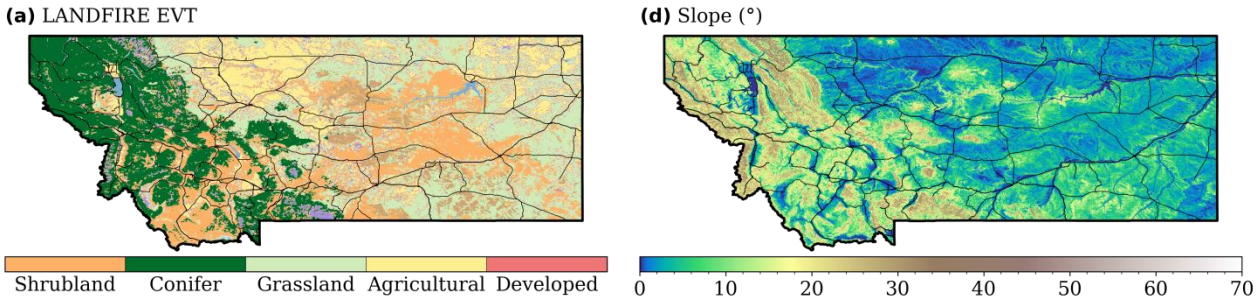
Figure 1. Conceptual scheme of the dual component of wildfire hazard or danger: the static factors and the dynamic factors.

In this study, we develop and evaluate several spatial models of wildfire occurrences to predict ignition likelihood across the state of Montana (US). To do this, we focus solely on constructing static ignition models by combining fire records with a suite of temporally invariant factors (Figure 1). We evaluate the efficacy of increasingly more complex models to map human and natural ignition probabilities. Our main objectives are: a.) to evaluate the contribution of different environmental-human factors considered crucial to wildfire ignition, b.) select the best predictive models' parameters configuration, and c.) map static human-natural wildfire ignition likelihood.

2. Material and Methods

2.1. Study area

Montana covers an area of 376,962 km^2 and it is characterized by west-to-east gradients in climate, topography, vegetation cover and wildfire causality (Figure 2).



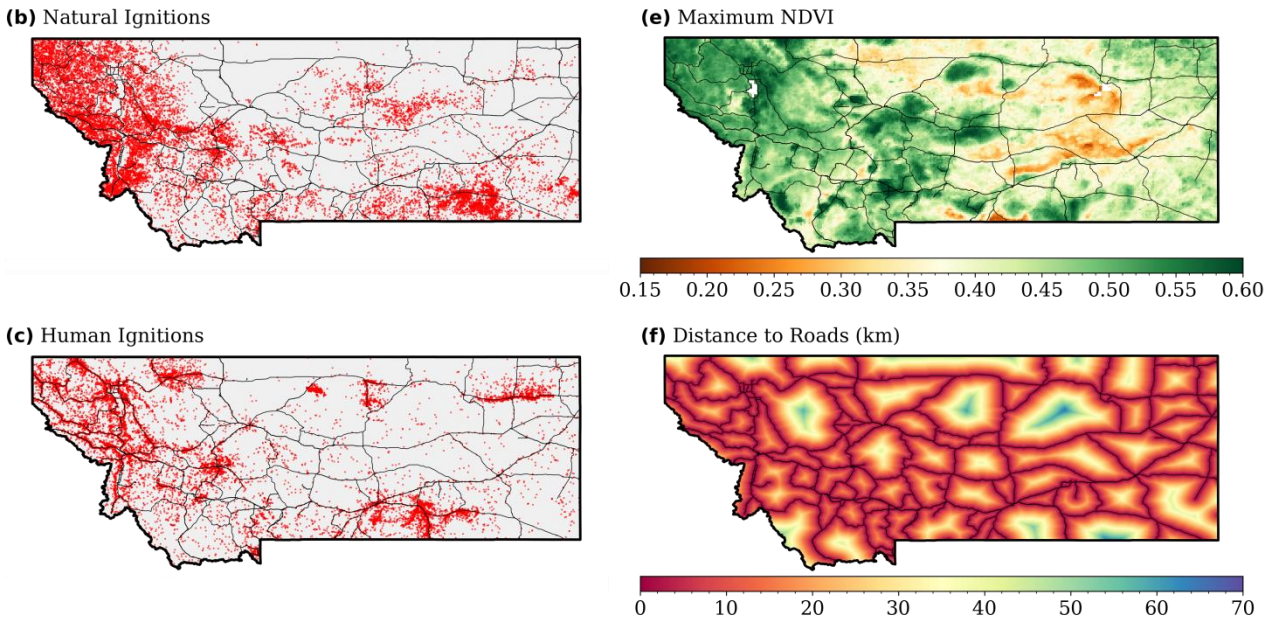


Figure 2. Maps of Montana shown with the road network. (a) Existing Vegetation Type (EVT). (b) and (c) Locations of natural and human ignitions (1992-2017). (d) Slope. (e) Maximum NDVI (1996-2017). (f) Distance to roads.

2.2. Data preparation

Wildfire point records from 1992-2017 were obtained from the Fire Program Analysis Fire-Occurrence Database (Short 2014), containing geographic coordinates of the origin, discovery date, ignition cause and final fire size. Wildfires were partitioned according to their causality (natural or human, excluding the unknown). The dependent variable (presence-absence) was obtained from the 30m binary masks of new wildfire reports. Several static potential covariates were explored (Figures 2d - 2f). To represent fuel availability, smoothed rasters of NDVI were retrieved from the Blended Vegetation Health Product (Kogan 2001; Yang, Kogan, and Guo 2020). Maximum-NDVI was chosen to capture the peak amount of above ground live biomass (Xu et al. 2012). Topography was represented by slope, obtained from the 30m resolution Digital Elevation Model (Rollins 2009). Distance from roads was used to represent accessibility, by downloading a road shapefile from the United States Census Bureau (US Census Bureau 2015), and creating a raster of distance to roads at 30m resolution.

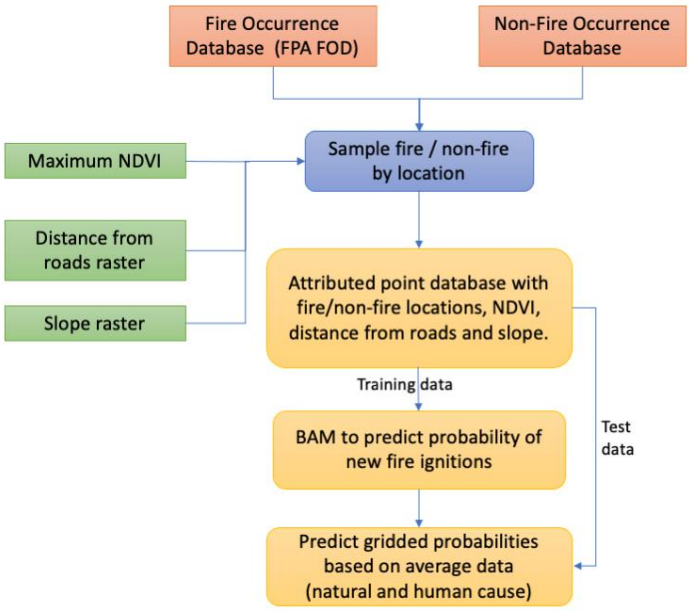


Figure 3. General workflow for model building and evaluation including inputs, sampling of wildfire presence-absence, extraction of covariates, creation of a point wildfire database, model fitting, and the final spatial predictions of wildfire ignition probabilities.

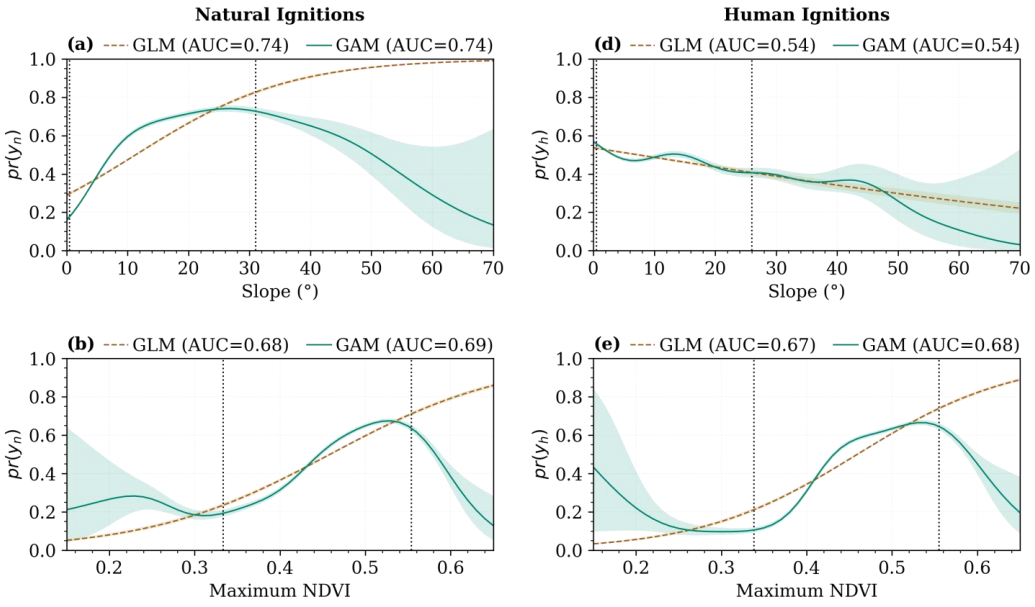
The datasets used were derived from the presence/absence masks of wildfires and the covariables' rasters (Figure 3). First, presence only datasets were created by using the grid cells containing a new wildfire report. This produced two presence only datasets (one for cause), together with their associated value for slope, maximum-NDVI, distance to roads and longitude-latitude. The presence only datasets were then split into training (70%) and testing (30%) datasets. Absence only datasets were created by randomly selecting grid cells from the binary masks at locations without any wildfire recorded. To provide balanced datasets, the number of absence grid cells was equal to the number of presence ones. After using the absence grid cells to extract the covariates, the absence only datasets were themselves spit again into training and testing datasets. Combining the presence and absence datasets for each ignition source yielded four datasets: the natural-caused training and test datasets, and the human-caused training and test datasets.

2.3. Model construction and evaluation

We included Generalized Additive Models (GAMs) (A.D. Syphard et al. 2008; Bar Massada et al. 2012) apart from Generalised Linear Model (GLMs), since the former overcomes the apriori assumption of linearity by replacing the linear terms with smoothed (Vilar et al. 2010). In this work, a suite of logistic-GLMs/GAMs were built in R (R Core Team 2021), starting with simple logistic-GLMs/GAMs to examine the isolated effects, using slope, maximum-NDVI and distance-to-roads as the sole explanatory variables. Then covariates were procedurally added using the default basis dimension (k = 10) to create increasingly more complex models with the goal of achieving the best fit with the fewest explanatory variables necessary. Additional variables were retained if they were statistically significant and if the percent deviance explained increased by more than 0.1%. Once added, the basis dimension in the spatial-GAM were iteratively increased to ensure that the smoothed terms had sufficient degrees of freedom without being overly complicated or computationally intractable. After fitting the models, each one was applied to the held back testing datasets to predict the natural and human-caused wildfire probabilities. Receiver operating characteristic (ROC) curves (Bradley 1997) and the area under the ROC Curve (AUC) were used to assess the models accuracy.

3. Results

Simple logistic-GLMs/GAMs built with one covariate illustrate the isolated effects of slope, maximum-NDVI and distance-to-roads on the probability of a new wildfire ignition (Figure 4). In contrast to the GLMs, the GAMs capture nonlinearities in the response, particularly at the lower and upper limits of slope and maximum-NDVI. Although there are strong differences between the GLM and GAM fits at the extremes, only a small fraction (5%) of the natural and human-caused ignitions in the training data had a slope >30º or a maximum-NDVI > 0.56. Moreover, both (GLM and GAM) fits exhibit monotonic behaviour over most of the range of the covariates, thus obtaining similar AUCs.



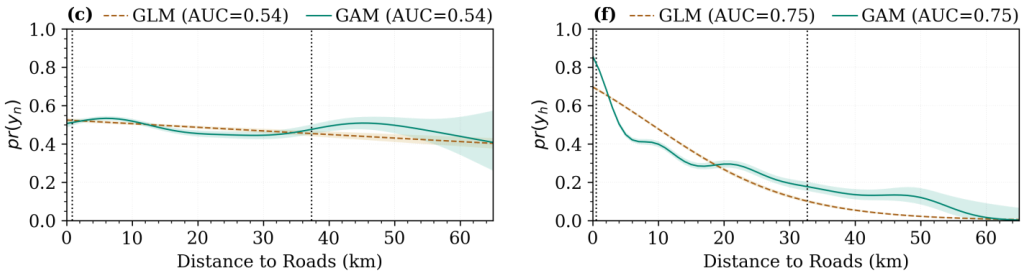


Figure 4. Simple General Linear Models (GLMs) and simple General Additive Models (GAMs) built with one covariate to predict the probability of natural ignitions (left column), and human ignitions (right column). The dotted vertical lines indicate the 5th and 95th percentiles of the covariates and capture 90% of the training data.

In the simplest version, all covariates were statistically significant ($p \leq 0.01$) and the percent deviance explained increased when spatial coordinates were added. However, whereas all variables were significant in the spatial-GAM of human-caused ignition probability, distance-to-roads was not significant in the spatial-GAM of natural ignition probability and was therefore excluded (Table 1). Initiating spatial-GAMs, the basis dimensions (k) were iteratively increased for all the smoothed terms. However, for slope, maximum-NDVI and distance to roads this had little effect on the percent deviance explained compared to the values of k for the geographic coordinates. Ultimately, the basis dimension for the smoothe geographic coordinates was increased until k = 50 and k = 35 for the natural and human-caused ignition models, respectively.

Table 1. Development from the General Linear Model (GLM) to the General Additive Model (GAM) and spatial-GAM. Formulas are presented, where α is the intercept, υ1= slope, υ2= maximum-NDVI, υ3 = distance to roads, x = longitude and y = latitude. The "s()" functions indicate the smoothed terms in the GAMs.

Version	*Cause*	*Formula*	*AUC*
GLM	Natural	$pr(y_n) = \alpha + \upsilon1 + \upsilon2 + \upsilon3$	0.75
	Human	$pr(y_h) = \alpha + \upsilon1 + \upsilon2 + \upsilon3$	0.77
GAM	Natural	$pr(y_n) = \alpha + s(\upsilon1) + s(\upsilon2) + s(\upsilon3)$	0.76
	Human	$pr(y_h) = \alpha + s(\upsilon1) + s(\upsilon2) + s(\upsilon3)$	0.80
Spatial GAM	Natural	$pr(y_n) = \alpha + s(\upsilon1) + s(\upsilon2) + s(x,y)$	0.84
	Human	$pr(y_n) = \alpha + s(\upsilon1) + s(\upsilon2) + s(\upsilon3) + s(x,y)$	0.89

According to the held-back testing (Figures 5 and 6), the logistic-GLMs exhibited the most uniform probability distributions and the greatest overlap between the presence-absence probabilities. In turn, the logistic-GAMs shifted the probability distributions towards higher values for the presence locations and towards lower values for the absence locations, indicating better differentiation. Overall, the spatial-GAMs (with geographic coordinates and optimized k values) yielded the greatest separation between the probability distributions. For the spatial-GAMs, half of the locations with a natural and human ignition had a predicted probability >76% and >80%, respectively; and half of the locations without a natural and human ignition had a predicted probability <25% and <18%, respectively. Consequently the spatial-GAMs offered the best performance, with AUC = 0.84 (natural-cause) and AUC = 0.89 (human-cause).

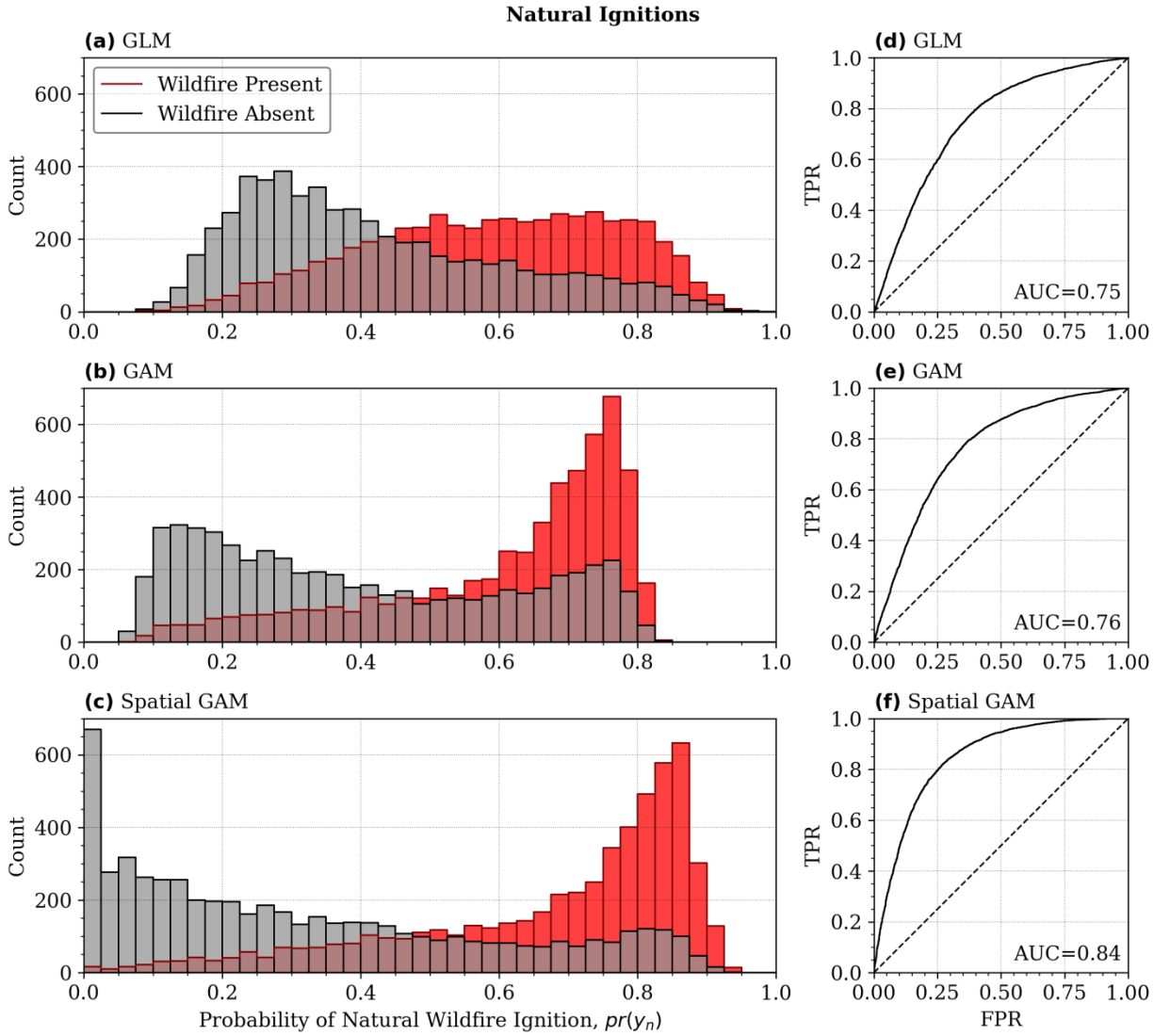


Figure 5. Evaluation of the General Linear Models (GLMs), General Additive Models (GAMs) and spatial-GAM used to model the probability of natural wildfire ignitions. Each model was applied to the held back testing datasets and the frequency distributions shown in (a)-(c) indicate the count of presence (red) and absence (black) grid cells according to their predicted probability.

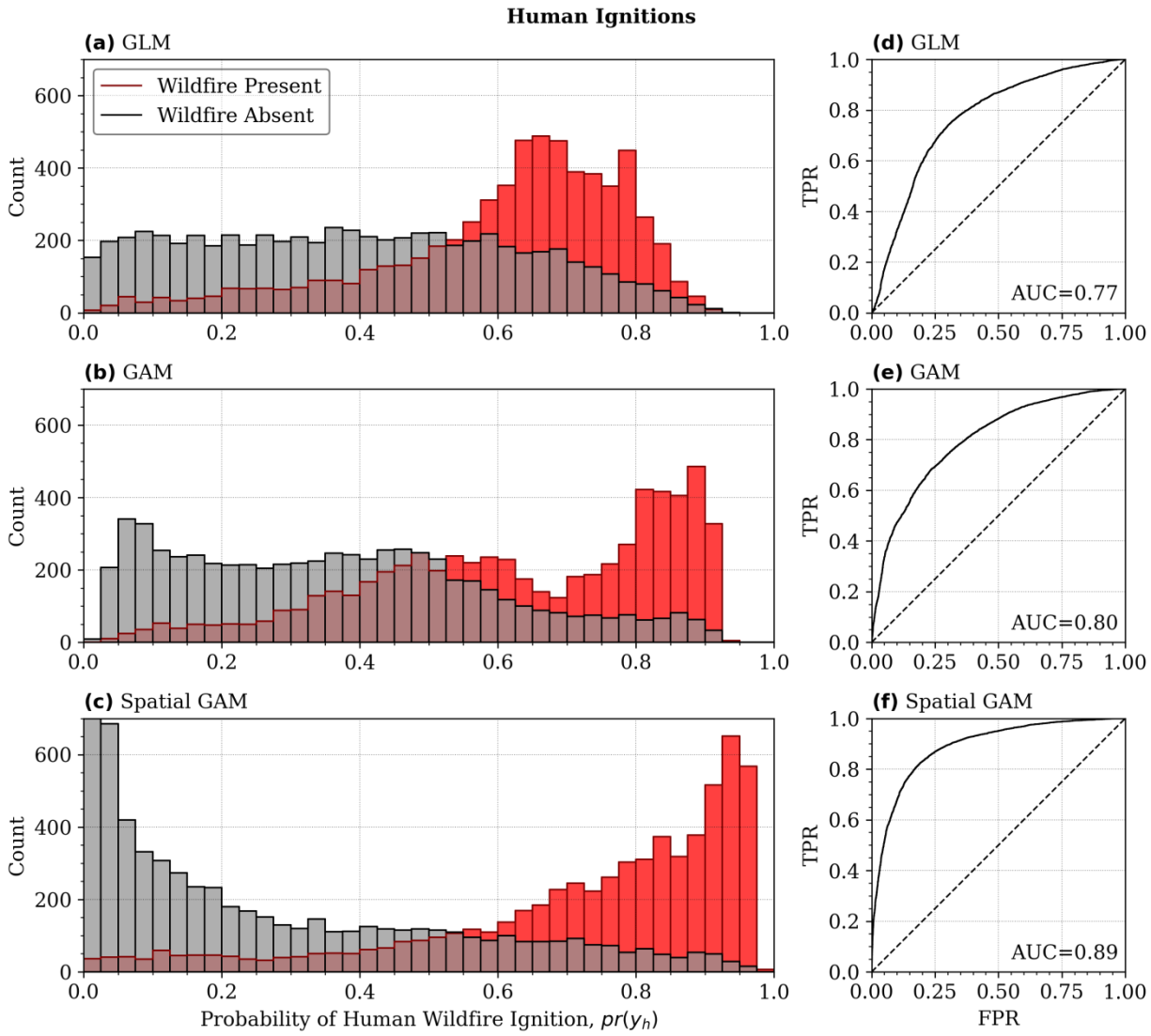


Figure 6. Same as Figure 5, but for human wildfire ignitions.

For natural ignitions, the response to slope (Figure 7a) shows a likelihood increase up to ~30º and decreasing thereafter with a greater predictive uncertainty due to fewer training data. In turn, the likelihood of a new wildfire was relatively insensitive to maximum-NDVI until a value of ~0.5 and then decreased. The geographic coordinates in the logistic-GAM imparted a spatial smoother, with locations in the Northern Rocky Mountains and south-eastern MT, exhibiting the highest probabilities. For human ignitions, the response of the spatial-GAMs to slope and distance to roads was similar to GAMs with the exception that, again, of maximum-NDVI. Here, distance to roads played a prominent role in the spatial-GAM of human-caused ignitions (50% of the grid cells containing a human caused wildfire located < 2,500 meters of a road). The spatial smoother revealed similar though slightly different patterns, most noticeably in northern MT along the Canadian border.

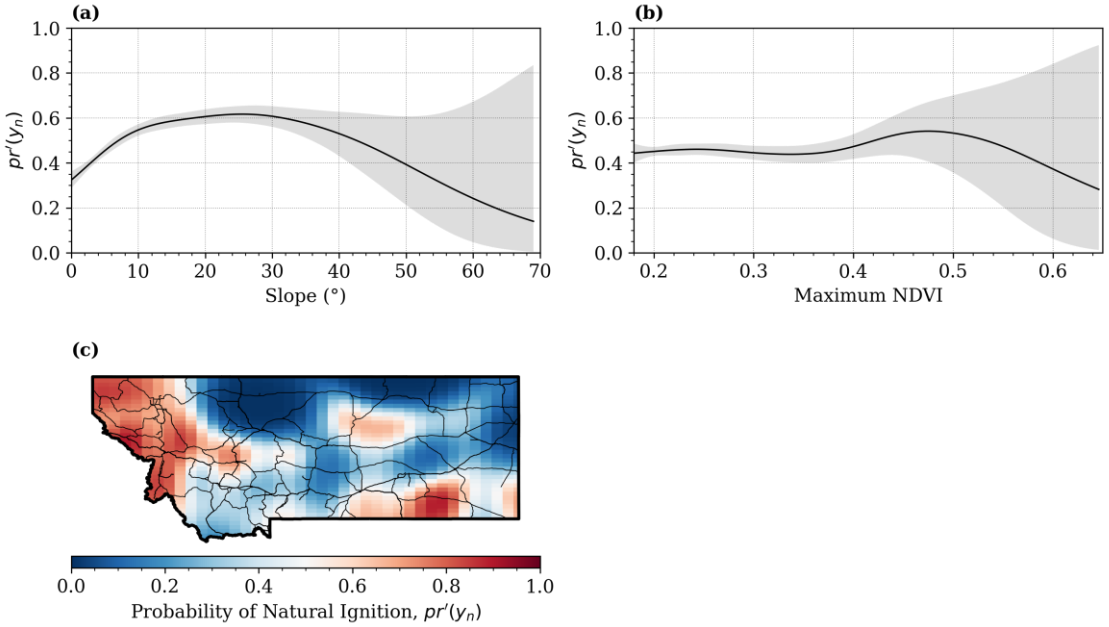


Figure 7. Partial effects plots for all smoothed variables in the spatial-GAM built to predict the probability of a natural wildfires.

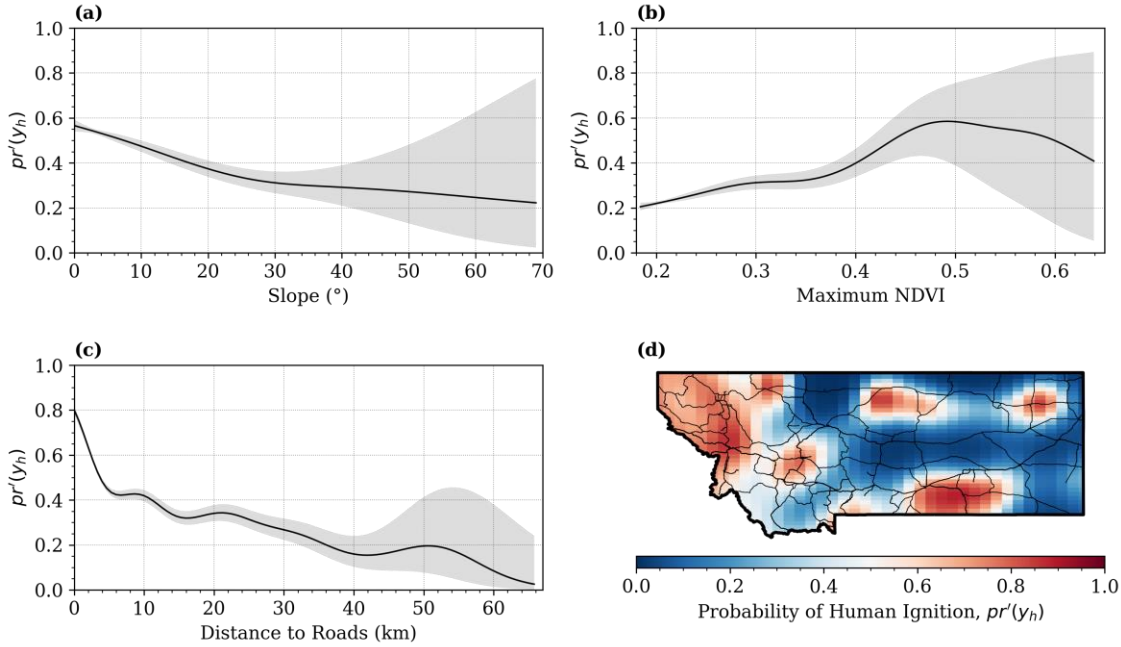


Figure 8. Same as Figure 7, but for human-caused wildfires.

Generally, the 30m static probability maps were broadly influenced by the spatial smoothers (Figures 9a and 9c versus Figures 7c and 8d) and locally modulated by slope, maximum-NDVI and distance to roads. This is most evident in the spatial-GAM of natural ignitions where probabilities are lower in the valley bottoms and nearby shallower slopes, and in the spatial-GAM of human-caused ignitions where probabilities are highest along the road network. In fact, both spatial-GAMs predicted large areas with low probabilities of wildfire ignition (50% of the territory have <23% and <15% chances of a natural and human-caused ignition, respectively). Conversely, locations with a high predicted probability of a new wildfire are rare (~2% of the state had >90% of ignition chance).

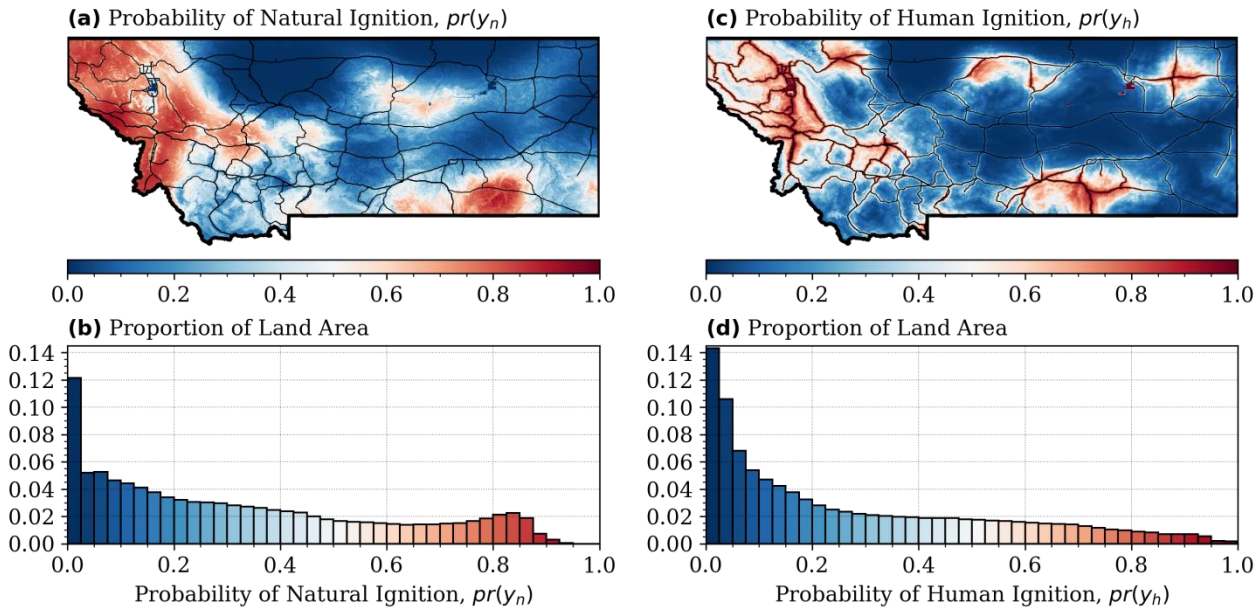


Figure 9. Maps of (a) the probability of a natural wildfire, and (b) the probability of a human wildfire obtained from the spatial-GAMs. Distributions of the land area proportion as functions of both causes are shown in (b) and (d), respectively.

4. Conclusions

In this study we have tested two different spatial-GAMs models using a daily wildfire database and drivers through a 25-year time span. Spatially, human model produced the highest wildfire probabilities, although the homogeneity of its greatest values were very clustered in particular areas. The inclusion of environment-human variables allowed us to forecast static wildfire probability in areas with a high wildfire activity and, thus, a future danger is expected. In addition to the advantage of easily updating the models as new available wildfire data, we find the selection of these basic variables to be practical.

5. References

Archibald, Sally, Caroline E R Lehmann, Jose L Gómez-dans, and Ross A Bradstock. 2013. "Defining Pyromes and Global Syndromes of Fire Regimes." Proceedings of the National Academy of Sciences of the United States of America 110 (16): 6445–47. https://doi.org/10.1073/pnas.1211466110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1211466110.

Balch, Jennifer K., Bethany A. Bradley, John T. Abatzoglou, R. Chelsea Nagy, Emily J. Fusco, and Adam L. Mahood. 2017. "Human-Started Wildfires Expand the Fire Niche across the United States." Proceedings of the National Academy of Sciences of the United States of America 114 (11): 2946–51. https://doi.org/10.1073/pnas.1617394114.

Bar Massada, A., A.D. Syphard, S. Stewart, and V.C. Radeloff. 2012. "Wildfire Ignition-Distribution Modelling : A Comparative Study in the Huron – Manistee National Forest , Michigan, USA." International Journal of Wildland Fire 22: 174–183. https://doi.org/https://doi.org/http://dx.doi.org/10.1071/WF11178.

Barreto, Joan Sebastian, and Dolors Armenteras. 2020. "Open Data and Machine Learning to Model the Occurrence of Fire in the Ecoregion of 'Llanos Colombo–Venezolanos.'" Remote Sensing 12 (3921): 1–18. https://doi.org/10.3390/rs12233921.

Belval, Erin J., Crystal S. Stonesifer, and David E. Calkin. 2020. "Fire Suppression Resource Scarcity: Current Metrics and Future Performance Indicators." Forests 11 (2). https://doi.org/10.3390/f11020217.

Benefield, A., and J. Chen. 2022. "Examining the Influence of Outdoor Recreation on Anthropogenic Wildfire Regime of the Southern Rocky Mountains." Natural Hazards 111: 523–45. https://doi.org/10.1007/s11069-021-05065-1.

Bradley, Andrew P. 1997. "The Use of the Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms." Pattern Recogn. 30 (7): 1145–59. https://doi.org/10.1016/S0031-3203(96)00142-2.

Briones-Herrera, Carlos Ivan, Daniel José Vega-Nieva, Norma Angélica Monjarás-Vega, Favian Flores-Medina, Pablito Marcelo Lopez-Serrano, José Javier Corral-Rivas, Artemio Carrillo-Parra, et al. 2019. "Modeling and Mapping Forest Fire Occurrence from Aboveground Carbon Density in Mexico." Forests 10 (5): 1–19. https://doi.org/10.3390/f10050402.

Faivre, Nicolas, Yufang Jin, Michael L. Goulden, and James T. Randerson. 2014. "Controls on the Spatial Pattern of Wildfire Ignitions in Southern California." International Journal of Wildland Fire 23 (6): 799–811. https://doi.org/10.1071/WF13136.

Kogan, Felix N. 2001. "Operational Space Technology for Global Vegetation Assessment." Bulletin of the American Meteorological Society 82 (9): 1949–64. https://doi.org/10.1175/1520-0477(2001)082<1949:OSTFGV>2.3.CO;2.

Krawchuk, Meg A., Max A. Moritz, Marc André Parisien, Jeff Van Dorn, and Katharine Hayhoe. 2009. "Global Pyrogeography: The Current and Future Distribution of Wildfire." PLoS ONE 4 (4). https://doi.org/10.1371/journal.pone.0005102.

Krawchuk, Meg A, Max A Moritz, Meg A Krawchuk, and Max A Moritz. 2011. "Constraints on Global Fire Activity Vary across a Resource Gradient." Ecology 92 (1): 121–32. https://www.jstor.org/stable/29779580.

Mitchener, Lori Jean, and Albert J. Parker. 2005. "Climate, Lightning, and Wildfire in the National Forests of the Southeastern United States: 1989-1998." Physical Geography 26: 147–62. https://doi.org/https://doi.org/10.2747/0272-3646.26.2.147.

Narayanaraj, Ganapathy, and Michael C. Wimberly. 2012. "Influences of Forest Roads on the Spatial Patterns of Human- and Lightning-Caused Wildfire Ignitions." Applied Geography 32 (2): 878–88. https://doi.org/10.1016/j.apgeog.2011.09.004.

Prestemon, J.P., T.J. Hawbaker, M. Bowden, J. Carpenter, M.T. Brooks, K.L. Abt, R. Sutphen, and S. Scranton. 2013. "Wildfire Ignitions : A Review of the Science and Recommendations for Empirical Modeling," 20.

R Core Team. 2021. "R: A Language and Environment for Statistical Computing." Vienna: R Foundation for Statistical Computing. https://www.r-project.org/.

Riley, Karin L., Matthew P. Thompson, Joe H. Scott, and Julie W. Gilbertson-Day. 2018. "A Model-Based Framework to Evaluate Alternative Wildfire Suppression Strategies." Resources 7 (1): 1–26. https://doi.org/10.3390/resources7010004.

Rollins, Matthew G. 2009. "LANDFIRE: A Nationally Consistent Vegetation, Wildland Fire, and Fuel Assessment." International Journal of Wildland Fire 18 (3): 235–49. https://doi.org/10.1071/WF08088.

Schultz, Christopher J., Nicholas J. Nauslar, J. Brent Wachter, Christopher R. Hain, and Jordan R. Bell. 2019. "Spatial, Temporal and Electrical Characteristics of Lightning in Reported Lightning-Initiated Wildfire Events." Fire 2 (2): 1–15. https://doi.org/10.3390/fire2020018.

Short, K. C. 2014. "A Spatial Database of Wildfires in the United States, 1992-2011." Earth System Science Data 6 (1): 1–27. https://doi.org/10.5194/essd-6-1-2014.

Syphard, A.D., Volker C. Radeloff, Nicholas S. Keuler, Robert S. Taylor, Todd J. Hawbaker, Susan I. Stewart, and Murray K. Clayton. 2008. "Predicting Spatial Patterns of Fire on a Southern California Landscape." International Journal of Wildland Fire 17 (5): 602–13. https://doi.org/10.1071/WF07087.

Syphard, AD, VC Radeloff, JE Keeley, TJ Hawbaker, MK Clayton, SI Stewart, and RB Hammer. 2007. "Human Influences on California Fire Regimes." Ecological Applications 17 (5): 1388–1402. https://doi.org/10.1890/06-1128.1.

Thompson, Matthew P., Benjamin M. Gannon, and Michael D. Caggiano. 2021. "Forest Roads and Operational Wildfire Response Planning." Forests 12 (2): 1–11. https://doi.org/10.3390/f12020110.

US Census Bureau. 2015. "TIGER/ Line Shapefiles."

Vilar, Lara, Douglas G. Woolford, David L. Martell, and M. Pilar Martn. 2010. "A Model for Predicting Human-Caused Wildfire Occurrence in the Region of Madrid, Spain." International Journal of Wildland Fire 19 (3): 325–37. https://doi.org/10.1071/WF09030.

Xu, Chi, Yutong Li, Jian Hu, Xuejiao Yang, Sheng Sheng, and Maosong Liu. 2012. "Evaluating the Difference between the Normalized Difference Vegetation Index and Net Primary Productivity as the Indicators of Vegetation Vigor Assessment at Landscape Scale." Environmental Monitoring and Assessment 184 (3): 1275–86. https://doi.org/10.1007/s10661-011-2039-1.

Yang, Wenze, Felix Kogan, and Wei Guo. 2020. "An Ongoing Blended Long-Term Vegetation Health Product for Monitoring Global Food Security." Agronomy 10 (12). https://doi.org/10.3390/agronomy10121936.