Predicting Wildfire Risk Under Novel 21st-Century Climate Conditions

Matthew Cooper
Sust Global
595 Pacific Ave., Floor 4
San Francisco, California 94133

Abstract

Wildfires are one of the most impactful hazards associated with climate change, and in a hotter, drier world, wildfires will be much more common than they have historically been. However, the exact severity and frequency of future wildfires are difficult to estimate, because climate change will create novel combinations of vegetation and fire weather outside what has been historically observed. This provides a challenge for AI-based approaches to long-term fire risk modeling, as much future fire risk is outside of the available feature space provided by the historical record. Here, I give an overview of this problem that is inherent to many climate change impacts and propose a restricted model form that makes monotonic and interpretable predictions in novel fire weather environments. I then show how my model outperforms other neural networks and logistic regression models when making predictions on unseen data from a decade into the future.

Introduction

One way to describe the future effects of climate change is with the phrase *global weirding*. The 21st century will be increasingly uncanny, as we will see Caribbean beach weather in Iceland; deserts that become soggy and green; and an Arctic Ocean that is entirely free of ice, potentially by 2035 (Guarino et al. 2020). Novel assemblages of temperature, precipitation, land cover, and vegetation will emerge that are unlike anything in human history, giving rise to hazards unprecedented in severity and posing major challenges to adaptation. Additionally, these weird conditions are a challenge to any form of modeling that depends on rich training data, as much of the future will be entirely outside of the feature space of available observational data.

This is especially true in the case of wildfire, because fire depends on two things: burnable vegetation and dry enough conditions to ignite that vegetation. Under stable climate conditions, weather and vegetation reach an equilibrium, where the amount of burnable vegetation is proportional to the amount of rainfall (See Fig. 1). However, under climate change, we are seeing increasingly novel pairings of precipitation and vegetation (See Fig. 2). For example, California has historically had dry summers and wet winters, leading to chaparral and spare forest vegetation communities. However, in the past decade, California had weather conditions more characteristic of a desert climate. This extremely dry weather, coupled with high levels of vegetation, is what has caused the unprecedented fire crisis in California (Abatzoglou and Williams 2016). A similar situation is occurring in the Amazon, where tropical rainforest vegetation is experiencing increasingly long dry seasons and is converting into a tropical savanna, with fire consuming the excess biomass (Le Roux et al. 2022).

These emerging conditions are causing significant problems for sectors like the insurance industry, which has traditionally used historic risk to estimate future risk and appropriately price premiums. Unable to accurately estimate fire risk under unprecedented conditions, many home insurance companies are withdrawing from fire-prone areas, leaving homeowners without coverage (Poizner 2022; Singh 2022). Given that a typical home mortgage can last up to 30 years, a period over which climatological and ecological systems will continue to disequilibrate, it is imperative that we develop better methods for estimating fire risk that can make reasonable predictions outside of the existing feature space provided by historic data.

Data

For this analysis, I use data on fire occurrence provided globally and at a 500 meter resolution derived from NASA's MODIS satellite program (Giglio et al. 2009). This dataset goes back to November 2000 and provides a binary indicator of whether a fire was observed at a given pixel at a daily timestep. From this dataset, I collected 240 million sample locations on a given day across the terrestrial world, oversampling fire occurrence to make up approximately 10% of the dataset, but otherwise sampling completely at random.

For each sample point, I calculate a daily fire weather index known as the Keetch-Byram Drought Index, or KBDI (Brown, Wang, and Feng 2021; Gannon and Steinberg 2021). KBDI is an index updated on a daily time step and is indicative of the amount of water in the top 203 millimeters of soil. A KBDI score of 0 corresponds to saturated soil and very little fire risk, while a KBDI score of 203 indicates that soil is dry up to 203 millimeters deep and that fire risk is very high. To calculate historic values of this index, I use daily

Figure 1: Historically, precipitation and biomass have been in equilibrium.

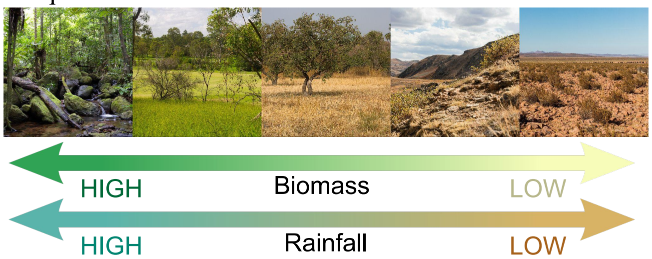
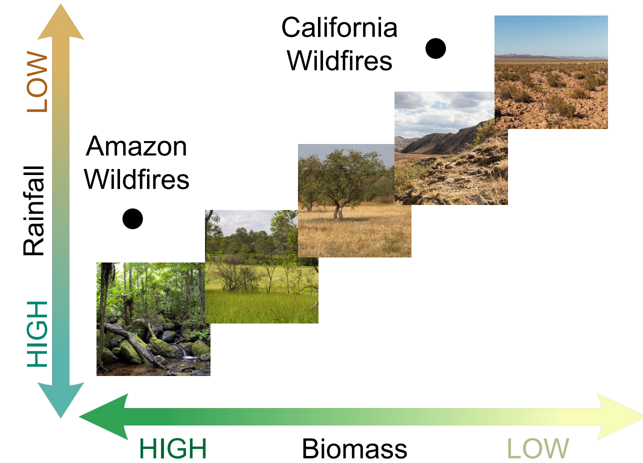


Figure 2: Under climate change, precipitation and biomass are decoupled, leading to unprecedented fire severity in California and the Amazon.



historic data on temperature and precipitation from the 10 kilometer ERA5-Land reanalysis dataset (Muñoz-Sabater et al. 2021). Additionally, to better determine the fire risk context I determine the local climate zone for each point using the Koppen-Geiger methodology (Köppen 2011), as well as the local land cover type using the 300 meter ESA land cover dataset (ESA 2017).

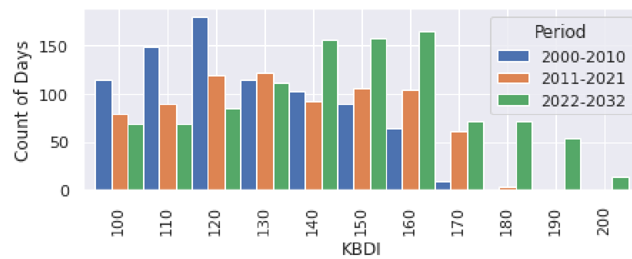
For my analysis, I use observed data from November 2000 to October 2011 as my training data (n = 135,559), and observed data from November 2011 to October 2021 as my validation data (n = 123,428). Testing my model on observations that occurred a decade beyond the end of the training data can give me an indication of how my model will perform over the course of the next decade. Additionally, I subset my analysis to eastern Oregon to constrain the discussion, although I have data processed and prepared for analyses at a global scale.

Finally, for future estimates of fire weather to use a features in model inference, I derive KBDI from ensembled and bias-corrected simulations of temperature and precipitation throughout the 21st century using Global Climate Models (GCMs) from the 6th Climate Model Intercomparison Project (CMIP6) (O'Neill et al. 2016).

The Problem

To better illustrate the modeling challenge presented by novel fire conditions, also referred to as domain shift, I show daily fire weather values (KBDI) in eastern Oregon for periods where observed KBDI scores were indicative of elevated fire risk (KBDI $>$ 100), typically in the summer (See Fig. 3). Eastern Oregon is an area without significant historic fire activity but is increasingly threatened by fire. There, KBDI values are increasing every decade, with the next decade modeled to have KBDI values at the maximum potential fire risk. This prevalence of increasingly out-of-sample and unprecedented fire weather is also associated with heightened fire risk, something models trained on only historic data will struggle to capture.

Figure 3: Shifting of fire weather towards unprecedented risk each decade complicates empirical AI modeling. Histogram of daily KBDI values in Eastern Oregon, by decade. Values for 2000-2010 and 2011-2021 are observed, values for 2022-2032 are taken from an ensemble of bias-corrected climate models.

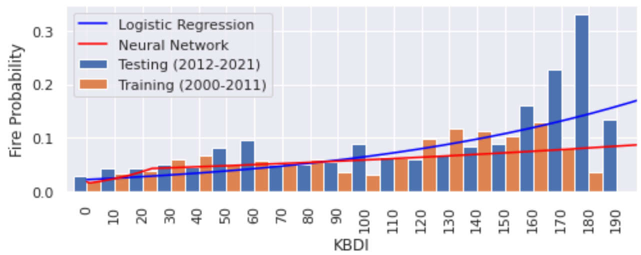


I further illustrate this domain shift modeling challenge by training a simple 3-layer feed-forward neural network to predict the probability of fire in eastern Oregon as a function of KBDI using sample data from 2000-2011 and validation data from 2012-2022. I compare that model against a logistic regression model using the same dataset. I find that the neural network under-estimated fire risk at high KBDI levels, while the logistic regression, due to its implicit monotonicity, better captured the trend of increasing fire risk with increasing KBDI levels (See Fig. 4).

While these test datasets illustrate the nature of the problem, both models used here were quite simple. In addition to fire weather, fire risk is heavily determined by other contextual factors, including biomass, land cover, long-term climate conditions, and elevation. I therefore construct more complex models based on 24 features derived from my sample dataset, one-hot encoding for land cover type and climate zone, as well as including terms for latitude and longitude, allowing the models to learn location-specific fire risk relationships. Additionally, I fit a hierarchical logistic regression using the same features as the multivariate neural network.

Overall, I find that multivariate models perform better than univariate models based only on KBDI when evaluated on a held out test dataset from the next decade (See Table 1). Additionally, I find that logistic regression models outperform neural networks on the test data, because they make

Figure 4: Observed probability of fire by KBDI value, in the training and testing datasets. Additionally, I show the predictions of a simple feed-forward neural network and a logistic regression. Note that the neural network under-estimates out-of-sample future fire risk.



predictions that are monotonic. This suggests that the neural networks struggle to capture extreme behavior.

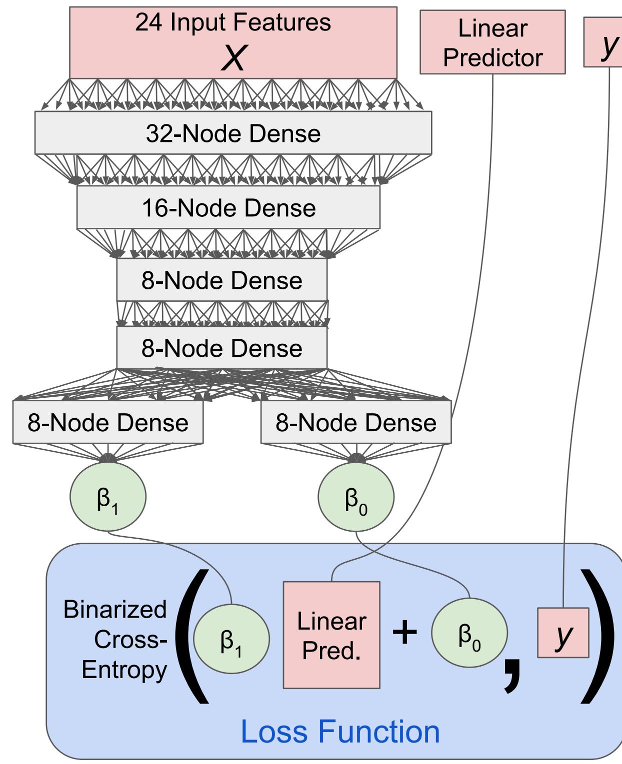
New Architecture

Because simple neural networks struggle to capture fire extremes under novel data domains, I propose a new neural network architecture, based on two premises. The first is that the relationship between KBDI and fire probability is monotonic, and as ongoing climate change leads to conditions drier than any previously observed in many locations, it will be necessary to use models that can extrapolate monotonically, such as logistic regression models. Secondly, the parameterization of the weather-fire relationship is complex and context dependent, with a large number of influencing variables that interact nonlinearly, requiring models like neural networks that can handle such estimation problems.

Drawing from both of these premises, I have implemented a neural network architecture that uses a large number of features describing the geographic context to estimate the parameters of a logistic model that describes the KBDI-fire relationship in that context. In this case, I use features for the spatial location, local land cover type, and historic climate zones indicative of prevailing vegetation communities; however, this architecture could be extended to incorporate other important features, such as topography, proximity to human settlements, or aboveground biomass. This approach has the advantage of drawing on complex interactions within the geophysical environment that influence the relationship between fire and weather conditions, while still being constrained to make predictions in line with my strong prior assumption that the relationship between dryness and fire risk is monotonic.

The model feeds a large number of features in four dense hidden layers that condense from 32 to 8 nodes with a ReLU activation function. The model then diverges into two separate hidden layers, each of which converges into a single-parameter output, which are treated as the two parameters in a logistic regression (β_0 and β_1). The model's loss function is therefore the performance of those two parameters in a logistic regression using observed KBDI, evaluated with binary cross-entropy (See Fig. 5).

Figure 5: Diagrammatic representation of fire neural network used to estimate logistic regression parameters.



Model	**R^2**	**MSE**
Univariate NN	0.0091	0.0442
Logistic Regression	0.0139	0.0440
Multivariate NN	0.0156	0.0439
Hierarchical Logistic Regression	0.0166	0.0438
NN-Estimated Logistic Regression	0.0202	0.0436

Table 1: Model performance by R^2 and mean squared error (MSE).

I fit a model with this architecture using the same features as the aforementioned multivariate neural network and find that it improves performance on R^2 by 22%. This architecture is able to draw on the advantages of using gradient descent to explore complex relationships among features, while still making predictions that are interpretable and extrapolate well outside of the observed range of fire weather values.

Conclusion

While there would be many benefits of using this methodology, it would have the drawback of requiring a very large dataset, as is typical of neural network based approaches. This would evolve the state of the art of predicting wildfires by focusing specifically on making predictions outside of the feature space available for training. Having better long-term fire predictions would help state agencies and governments to eliminate risks, as they currently rely on projections that are more near-term, focusing on weekly to seasonal timescales.

Neural networks provide a number of advantages and can explore a hyper-dimensional and complex feature space efficiently. However, they are brittle outside of their training space. In such situations where it is necessary to make predictions in the absence of available training data, predictions must be guided by theory and model behavior must be interpretable. I therefore developed an architecture that flexibly draws on complex environmental variables while still making predictions that are aligned with my theoretical prior that drier weather leads to increased fire risk. I find that this model performs better than other approaches when used to make predictions a decade into the future. Given the theoretical support of this approach, it is likely to be especially useful for making estimates at even longer timescales of up to two or three decades. This approach has relevance for modeling many of the novel risks posed by climate change.

References

Abatzoglou, J. T., and Williams, A. P. 2016. Impact of anthropogenic climate change on wildfire across western US forests. *Proc. Natl. Acad. Sci. U.S.A.* 113(42):11770–11775.

Brown, E. K.; Wang, J.; and Feng, Y. 2021. US wildfire potential: a historical view and future projection using high-resolution climate data. *Environ. Res. Lett.* 16(3):034060.

ESA. 2017. Land cover cci product user guide. Technical report.

Gannon, C. S., and Steinberg, N. C. 2021. A global assessment of wildfire potential under climate change utilizing keetch-byram drought index and land cover classifications. *Environmental Research Communications* 3(3):035002.

Giglio, L.; Loboda, T.; Roy, D. P.; Quayle, B.; and Justice, C. O. 2009. An active-fire based burned area mapping algorithm for the MODIS sensor. *Remote Sens. Environ.* 113(2):408–420.

Guarino, M.-V.; Sime, L. C.; Schröeder, D.; Malmierca-Vallet, I.; Rosenblum, E.; Ringer, M.; Ridley, J.; Feltham, D.; Bitz, C.; Steig, E. J.; et al. 2020. Sea-ice-free arctic during the last interglacial supports fast future loss. *Nature Climate Change* 10(10):928–932.

Köppen, W. 2011. The thermal zones of the earth according to the duration of hot, moderate and cold periods and to the impact of heat on the organic world. *Meteorologische Zeitschrift* 20(3):351–360.

Le Roux, R.; Wagner, F.; Blanc, L.; Betbeder, J.; Gond, V.; Dessard, H.; Funatzu, B.; Bourgoin, C.; Cornu, G.; Herault, B.; Montfort, F.; Sist, P.; Begue, A.; Dubreuil, V.; Laurent, F.; Messner, F.; Hasan, A. F.; and Arvor, D. 2022. How wildfires increase sensitivity of Amazon forests to droughts. *Environ. Res. Lett.* 17(4):044031.

Muñoz-Sabater, J.; Dutra, E.; Agustí-Panareda, A.; Albergel, C.; Arduini, G.; Balsamo, G.; Boussetta, S.; Choulga, M.; Harrigan, S.; Hersbach, H.; Martens, B.; Miralles, D. G.; Piles, M.; Rodríguez-Fernández, N. J.; Zsoter, E.; Buontempo, C.; and Thépaut, J.-N. 2021. ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data* 13(9):4349–4383.

O'Neill, B. C.; Tebaldi, C.; Van Vuuren, D. P.; Eyring, V.; Friedlingstein, P.; Hurtt, G.; Knutti, R.; Kriegler, E.; Lamarque, J.-F.; Lowe, J.; et al. 2016. The scenario model intercomparison project (scenariomip) for cmip6. *Geoscientific Model Development* 9(9):3461–3482.

Poizner, S. 2022. Op-Ed: Wildfires never threatened my home. But my insurer said they do — and dumped me. *Los Angeles Times*.

Singh, A. G. 2022. The need to modernize california wildfire insurance regulation with climate science. *Journal of Science Policy and Governance* 20(1).