Predicting Post-Fire Vegetation Recovery Time in California

Brian Y. Xie[1,2], Ke Liu[2]
[1]Santa Monica College, [2]California institute of Technology

Predicting Recovery after Wildfire

California wildfires are becoming **larger** and more **severe**, increasing concern about whether burned landscapes can **recover** after fire. While fire occurrence and burn severity are well documented, recovery time across California remains **less understood**. Estimating how quickly vegetation returns can **help guide** reforestation efforts, carbon management, and future fire planning. This study uses **MODIS**, **ERA5**-Land climate data, and **SRTM** topography from **2001 – 2019** to predict recovery time across burned areas in California. **Predictive recovery** models can **support** restoration planning, carbon monitoring, re-burn risk assessment, and early identification of landscapes vulnerable to long-term ecosystem change.

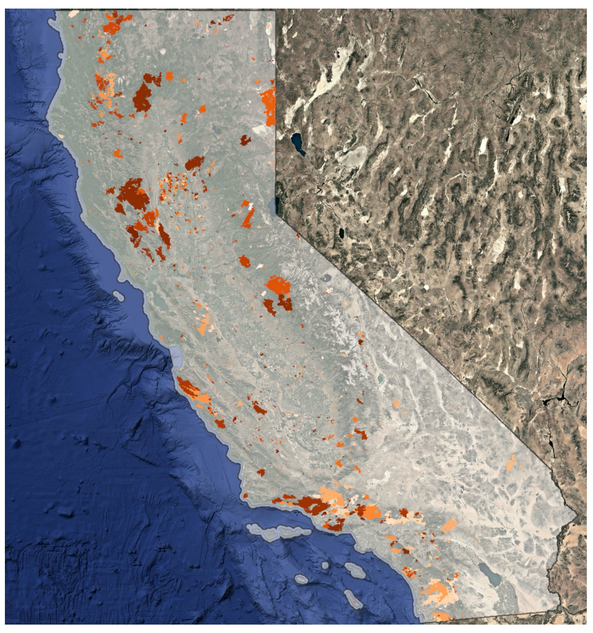


Figure 1: Generated Map of Wildfires from 2009 – 2019 from MODIS Burned Area, Colors progress from light red (2009) to dark red (2019)

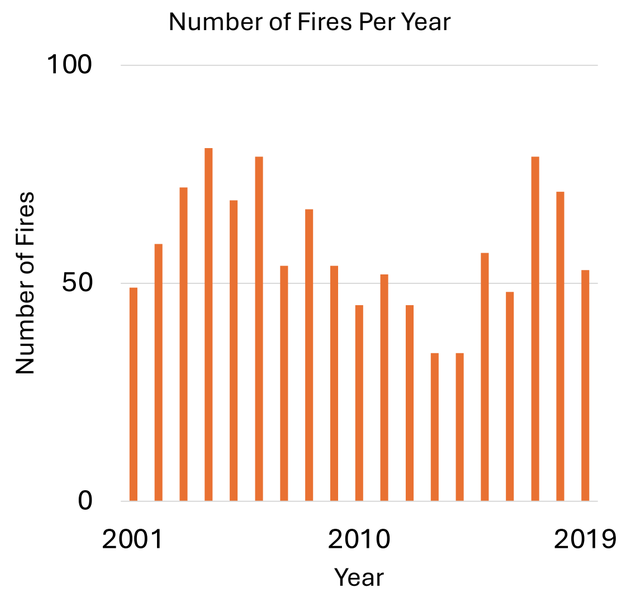


Figure 2: Number of fires per Year displayed in Figure 1

What can we learn about fires?

We used **NDVI**, or Normalized Difference Vegetation Index, to measure vegetation greenness before and after fire. As show in **Figure 3,** NDVI increases when vegetation is dense and healthy, so it provides a way to track how quickly a burned patch returns to its pre-fire condition. For each burn patch, we calculated pre-fire NDVI conditions, including peak NDVI, mean NDVI, minimum NDVI, NDVI variability, and NDVI amplitude. A patch was considered recovered when monthly NDVI returned to at least **90% of its pre-fire peak** for **three consecutive months**.

We also used **dNBR**, or differenced Normalized Burn Ratio, as a burn severity indicator. NBR compares near-infrared and shortwave-infrared reflectance, which are sensitive to healthy vegetation and burned surfaces. dNBR is calculated as:

dNBR = pre-fire NBR – post-fire NBR

Higher dNBR values generally indicate greater burn severity and more vegetation loss after fire. Additional predictor variables included **climate**, **topography**, **land cover**, **ecoregion**, **patch area**, and **fire year**. Climate variables from **ERA5-Land** included pre- and post-fire temperature and precipitation, as well as climate anomalies after fire. **Topographic** variables from **SRTM** included **elevation**, **slope**, and **aspect**. Aspect was converted into sine and cosine components so the model could better interpret slope direction. Land cover and ecoregion features helped the model account for differences between forests, shrublands, grasslands, and other vegetation types. Together, these features describe the condition of each burn patch before fire, the severity of vegetation loss, the environment where recovery occurred, and the post-fire climate conditions that may influence regrowth. This feature set was then used to train a Random Forest model to predict **months to vegetation recovery**.

Feature importance ranks the predictor variables based on how much they contributed to the Random Forest model's recovery-time predictions. Variables with higher importance scores had a stronger influence on the model's decisions.

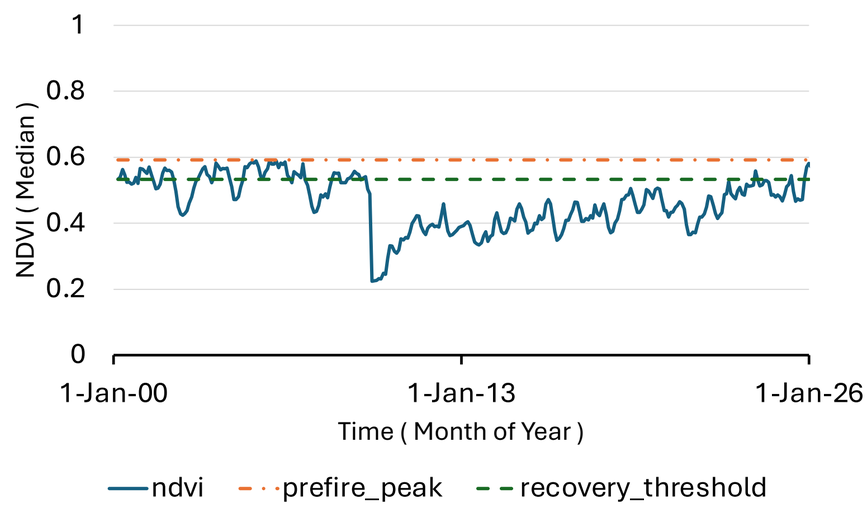


Figure 3: complete monthly NDVI record for one example burn patch from the beginning to the end of the MODIS time series.

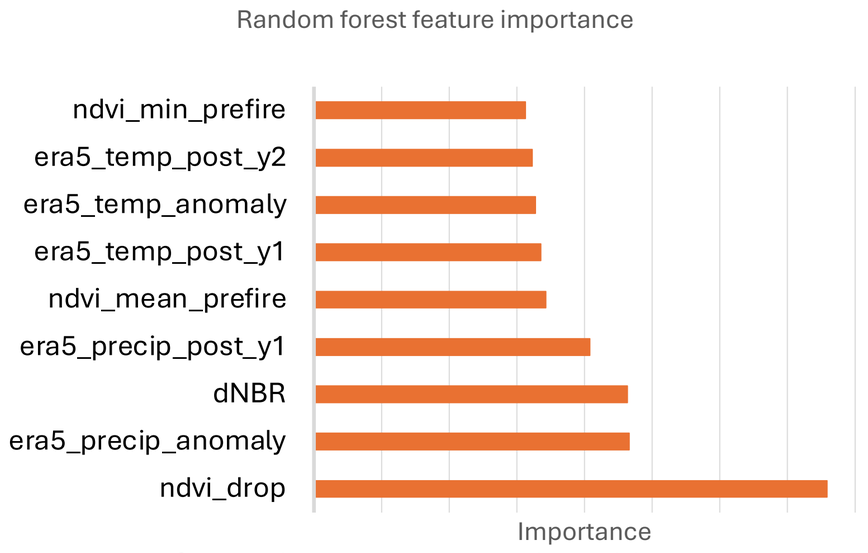
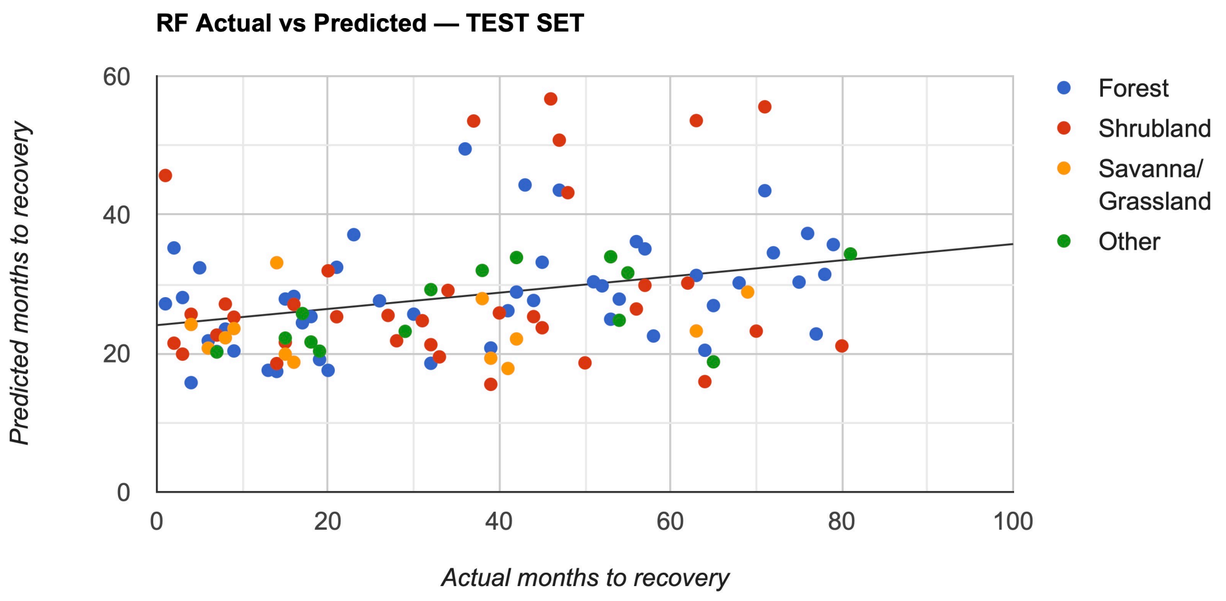


Figure 4: Top 9 features that were important to the model

What does the model say?



The model was built as a **Random Forest regression model**, to predict recovery time in months. Instead of relying on one tree, the Random Forest averages predictions from many trees, which helps reduce overfitting and makes the model more stable. In this analysis, the model used **200 trees**, tested a subset of predictor variables at each split, and required a minimum number of patches in each final leaf so that predictions were not based on single isolated examples. To evaluate the model more realistically, the data were split using **spatial blocks** rather than a fully random split. Each patch was assigned to a geographic block based on **latitude** and **longitude**, and whole blocks were assigned to either training or testing. This matters because nearby burn patches can be environmentally similar. A random split could place neighboring patches in both training and testing, making the model look better than it really is.

Conclusion

The Random Forest model showed **moderate predictive strength** for estimating post-fire vegetation recovery time. It performed better on the training set than the test set, but the test performance remained reasonable, with an **R^2 of 0.61**, meaning the model explained about **61% of the variation** in recovery time for spatially separated test patches. The test MAE of about **12.6 months** shows that predictions were typically within about one year of the observed recovery time. The small negative test bias, about **–1.8 months**, means the model slightly underestimated recovery time on average. Overall, the model captured meaningful recovery patterns across California burn patches, but the larger test errors show that recovery remains difficult to predict in new areas and may require additional local or ecological variables to improve accuracy. Thank you to Caltech, Ke Liu, and all my friends for this opportunity.