Using Lasso regression as a regularized regression method to identify wildfire predictors and estimate future burn area in western US ecoregions: California example

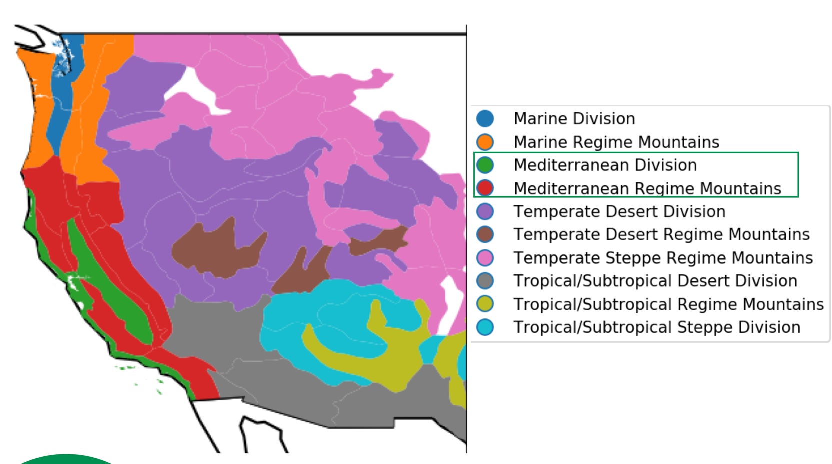
\# 381597

Steven Brey[1]* | sjbrey@rams.colostate.edu*
Libby Barnes[1], Jeff Pierce[1], and Emily Fischer[1]
[1]Colorado State University Department of Atmospheric Science
Fort Collins, Colorado

1 Introduction to the Lasso and western US fires

- This work aims to identify the environmental variables that offer the most skillful forecast of future western US burn area.
- These variables are identified using Lasso regression and nested leave-one-out cross validation. This allows us to identify the variables that make the most skillful forecast on unseen data.
- The model is trained to predict summer (JJA) burn area where historical burn area is obtained from the Monitoring Trends in burn Severity (MTBS) and environmental predicts are from ERA-Interim.

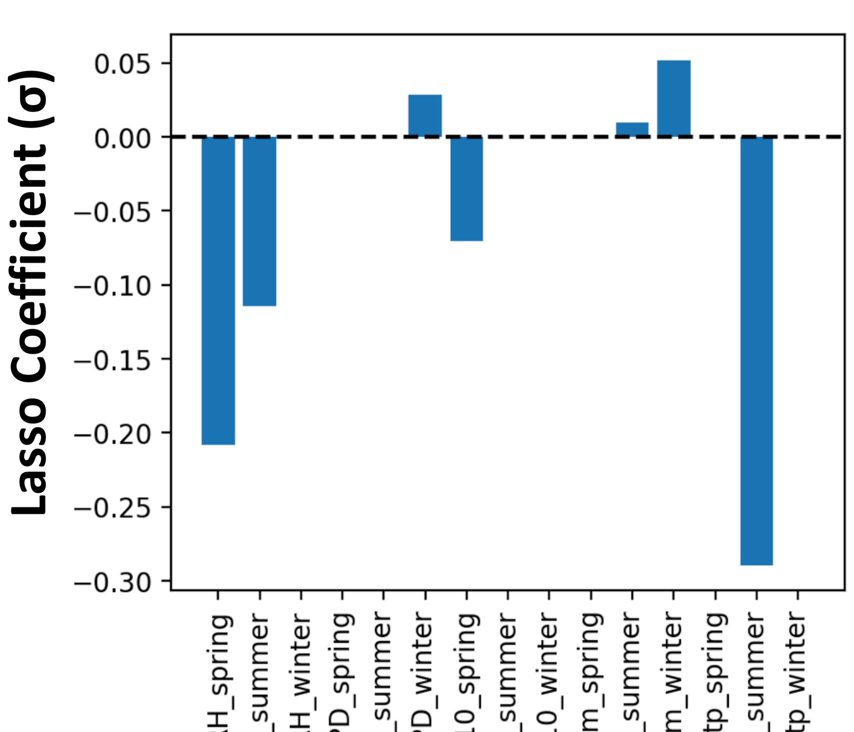
Lasso cost function : $\frac{1}{N}\sum(y_i - \mathbf{w}^T\mathbf{X})^2 + \lambda||\mathbf{w}||_1$



- Results in sparse number of predictors.
- Map shows ecoregions used in this analysis, which outline areas of similar ecosystems and climate.
- Poster shows California as an example, final work will have results for all.

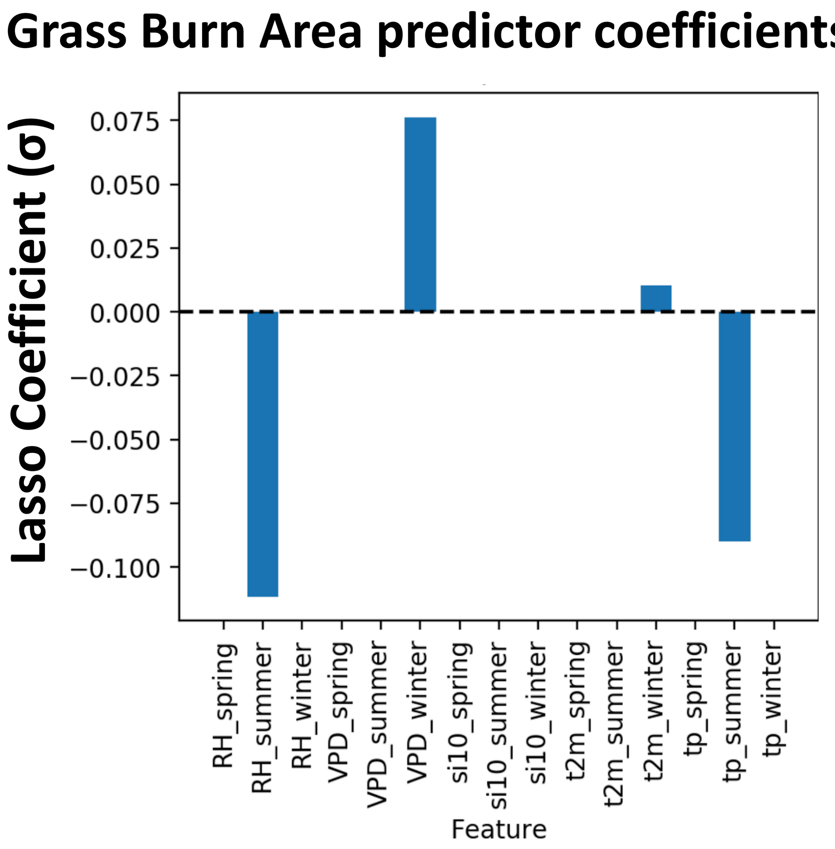
2 Wildfire drivers vary with land cover types

Forest JJA Burn Area predictor coefficients



- This figure shows the coefficients of the lasso regression that resulted in the lowest error on unseen data for western US forest burn area.
- Summer precipitation, spring RH%, and summer RH%, were identified as the most important predictors.

Grass Burn Area predictor coefficients

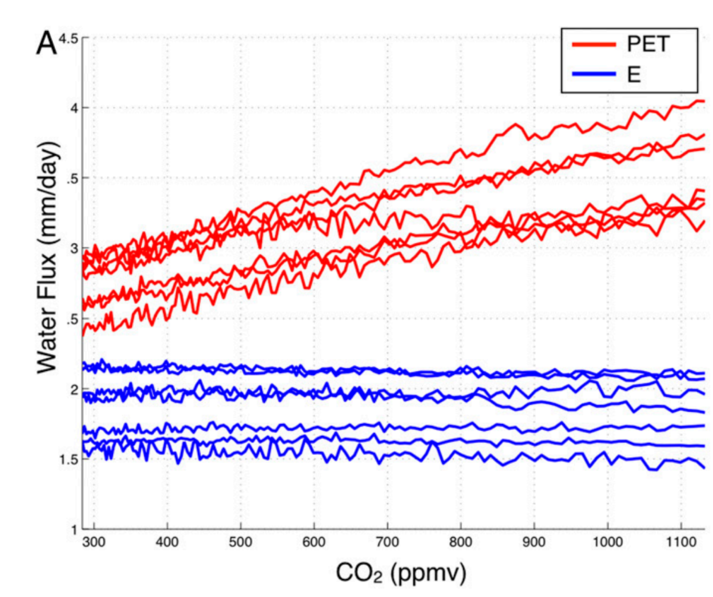


- This figure shows the coefficients of the lasso regression that resulted in the lowest error on unseen data for western US grass burn area.
- Summer RH%, Winter VPD, and summer precipitation result in the model with the most predictive power.

- We allow predictors of summer burn area to be lagged at varying timescales to capture antecedent conditions that matter for wildfires.
- The above results justify partitioning burn area by ecoregions, so for the remainder of this poster, burn area and lasso model regressions will be presented for individual ecoregions.
- Temperature does not show up as a predictor in the regularized model. Without requiring temperature to explain historical variance, trends in future projections tend to be smaller.

3 Plant centric predictors should be used

- There are many **meteorological** or **biophysical** variables that are correlated with wildfire burn area. From the plant's perspective, it is only a drought if the plant is stressed (Swann et al 2016). We include predictors relevant to plant physiology to be selected by the Lasso. These variables are chosen more often than temperature.
- A plant centric view is important, as plants water use efficiency is expected to increase with increasing CO_2 concentrations.

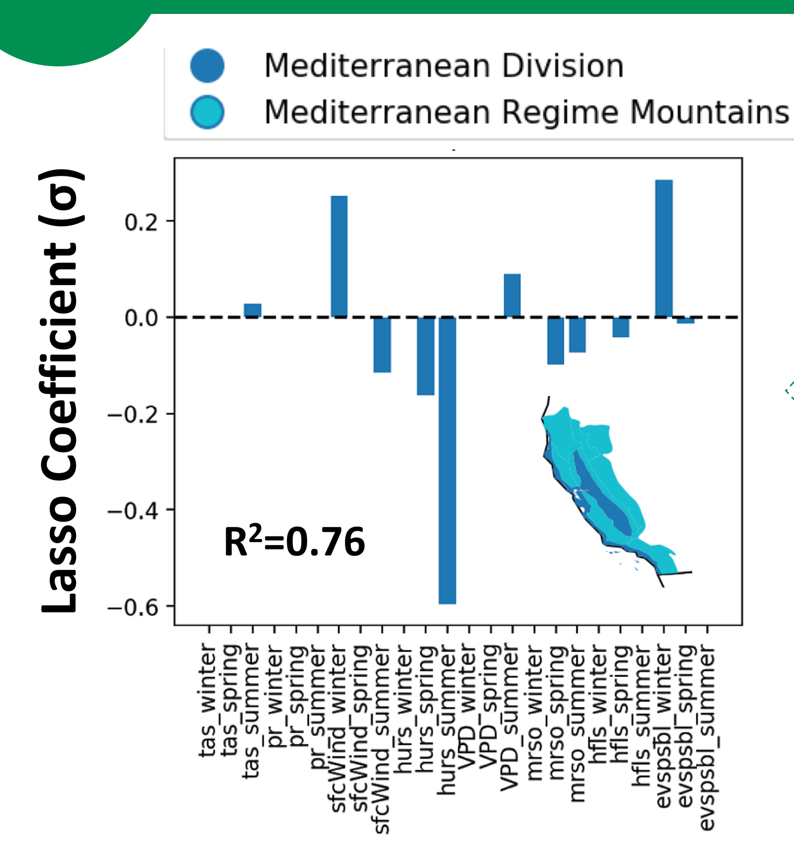


- Swann et al. 2016 show that there is a large divergence in CMIP5 models between atmospheric centric variables like Potential Evapotranspiration (PET) and actual evaporation and that plant physiology is responsible for the difference.
- Implication: Using atmospheric centric predictors of wildfire predictors (e.g. PET) may not capture actual trends live fuels vulnerability to burning.

Atmospheric Centric	Plant Centric
Temperature	Soil Moisture
Total Precipitation	Runoff Deficit
Wind Speed	Evaporation
Vapor Pressure Deficit	Latent heat flux

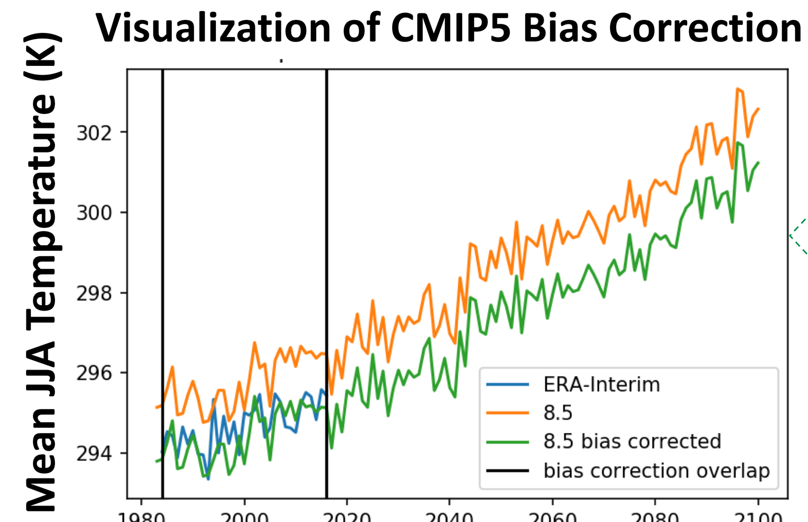
Not sure here, considering a diagram of a stomata? Or the table to the right

4 Future burn area for ecoregions are uncertain uncertain!

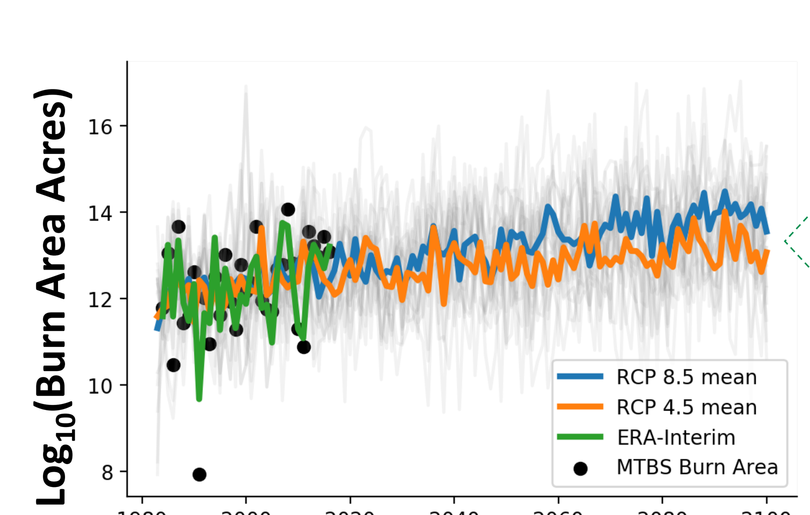


- This section presents California ecoregions as an example.
- Coefficients for California Specific ecoregions where additional plant centric predictors (e.g. soil moisture, evaporation) are included.
- Variance can be explained primarily by summer RH%, winter evaporation, and winter surface wind speeds.

Visualization of CMIP5 Bias Correction



- CMIP5 model output bias is accounted for using 1984-2016, where CMIP5 data overlap with ERA-Interim.
- Figure shows the upward trend in western US mean summer temperature in the.



- Using CMIP5 projections with the trained linear model results in increased California burn area, with a large range of possible outcomes.
- The gray lines represent individual model predictions and the uncertainty is large.
- Variance can be explained primarily by summer RH%, winter evaporation, and winter surface wind speeds.

- This figure shows the MODIS fire detection confidence is generally high in all regions for MODIS detections within 4 km of HMS fires.

Citations: Swann, Abigail L. et al., 2016. "Plant Responses to Increasing CO2 Reduce Estimates of Climate Impacts on Drought Severity." *Proceedings of the National Academy of Sciences of the United States of America* 113 (36): 10019–24..

EPA FUNDING NUMBER

Code and data available | https://github.com/stevenjoelbrey/metSpread