Fire-Climate Relationships in Continental Southeast Asia (SEA)

Alex Han, Qichen Huang, Vanessa Cheung

Supervised by Professors Keegan Korthauer, Melissa Lee, and Rodolfo Lourenzutti from UBC Department of Statistics

Abstract

Our project investigates relationships between climate factors (i.e. rainfall) and fire activity in continental SEA. We find that drought severity and percentage of vegetation have the greatest influence. Additionally, fire patterns do vary according to geographical patterns across SEA.

Introduction

Research Questions

1. How do fire patterns vary across continental SEA?
2. How do rainfall patterns influence the variations in fire activity?

Exploratory Data Analysis

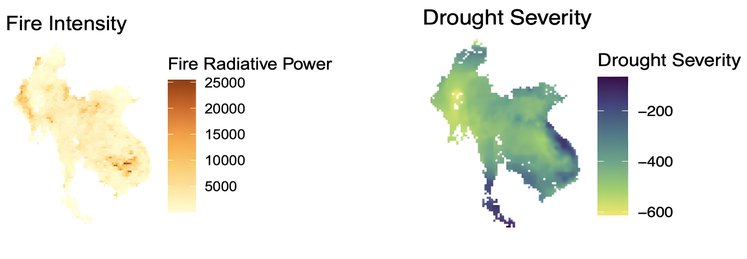

Figure 1: Visualizing data by raster graphics.

Regionalization

- Aim to cluster SEA into "similar" regions by fire patterns using Spatial 'K'luster Analysis by Tree Edge Removal (SKATER) algorithm.

Statistical Analysis

- Analyze model performance and variable importance by Generalized Additive Models (GAM) and Random Forest Regression (RF).

Results on Regionalization

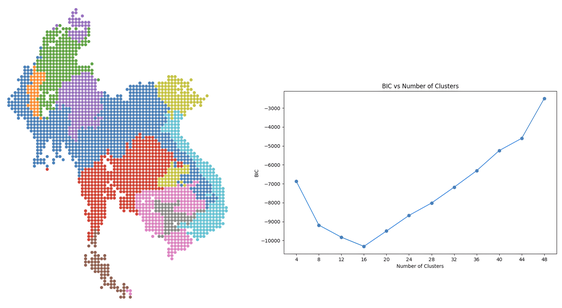

Figure 2: Based on the lowest BIC score, 16 clusters were identified in the region. The VRC and Gap statistics are 250 and 1.74 respectively. Our clusters demonstrate that there is variability in fire patterns across SEA.

Results on Fire-Rainfall Relationships

Generalized Additive Model

Model: $log(y) = s(MAP) + s(drought\ severity) + s(seasonality\ index) + s(PNV)$, where s is the smooth function.

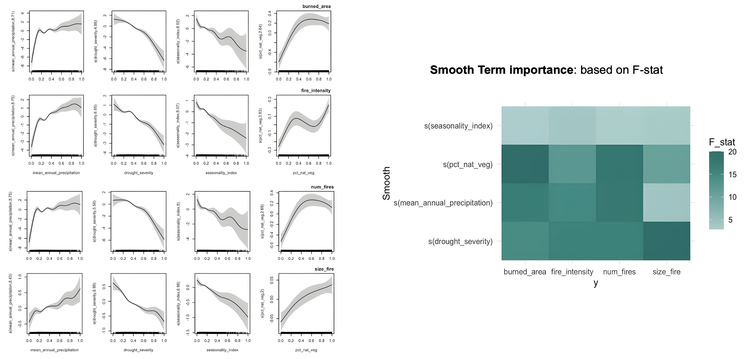

Figure 3: **Left**: Drought severity and Seasonality index appear to be negatively correlated with all fire variables. **Right**: Based on F-stat, drought severity seems to be most important in fire size model and percentage of natural vegetation seems to be most important in burned area and number of fires model. F-stat indicates whether the smooth function term for the variable is significant in a particular model.

Random Forest Regression

Model: $log(y) = MAP + drought\ severity + seasonality\ index + PNV$

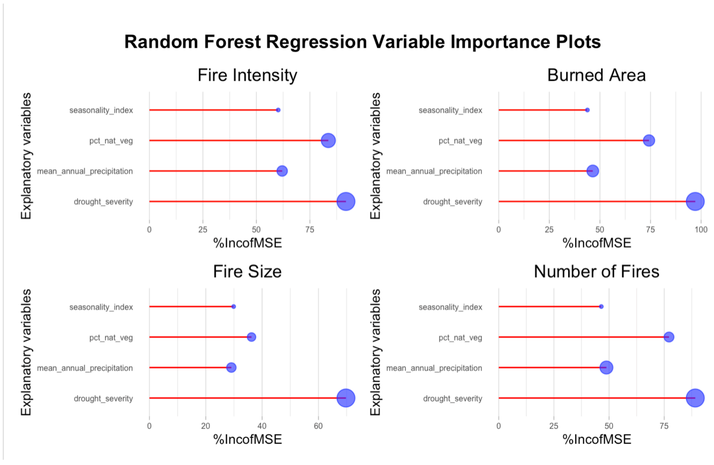

Figure 4: Variable Importance in Regression Model measured by percentage increase in mean square error (MSE). The metric percentage of increase in MSE indicates that how much MSE will the model increase after removing the corresponding explanatory variable. Drought Severity and Percentage of Natural Vegetation seem to be the most important variables in all four models. Seasonality index seems to be the least important here.

Statistical Model Performance

Generalized Additive Model performance

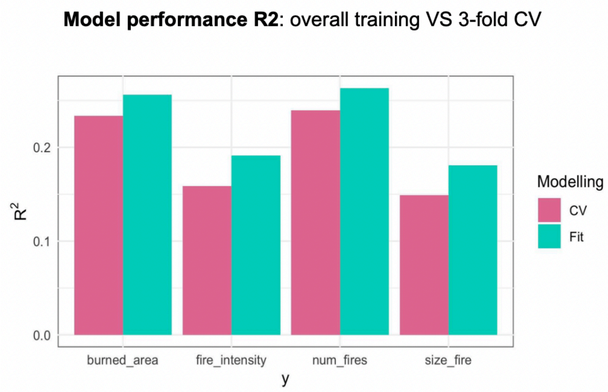

Figure 5: Cross-validation and fit R-squared scores are relatively comparable. The R-squared values range around 0.15 to 0.28.

Random Forest Regression performance

	Number of Trees	Number of variables tried each split	MSE	% of Variance Explained
fire_intensity	500	2	0.66	32.66
burned_area	500	3	2.90	33.09
size_fire	500	2	0.09	25.53
num_fires	500	2	2.36	36.64

Figure 6: Hyperparameters for each model obtained from 3-fold cross-validation repeated 2 times. Relatively low MSE for fire intensity and fire size models. Percentage of variance explained ranges from 26-37%.

Conclusions and Discussions

Regionalization

- Our cluster metrics show there is a significant variability in fire patterns across SEA.

Rainfall Pattern Analysis

- Drought severity seems to have the greatest explanatory significance to the fire pattern data.
- Both GAM and RF models agree on the direction of effect of mean annual precipitation, drought severity, and seasonality index.

Discussions

- Data is aggregated over 20-year average, hence, we have a rather low model performance/prediction score due to nature of data.
- Might also consider the correlation between rainfall patterns.
- Potentially capture non-linearity via polynomial regression while watching for overfitting.