TEXAS A&M UNIVERSITY
Landscape Architecture
& Urban Planning

Non-linearity and Spatial Heterogeneity of the Natural and Built Environment Factors to Wildfire Duration in Texas: An Interpretable Machine Learning Approach Using Random Forest and Geographically Weighted Random Forest

Team: UrbanDS
Member: Chunwu Zhu, Weishan Bai, Tianchen Huang, Jiaxin Du

Background

Wildfires can cause widespread damage and have long-lasting impacts on both the natural environment and human communities.

To mitigate the influence of wildfires, we need to **explore the contributing factors and their relationship** to wildfires from both natural environment and built environment that bores the disaster.

Natural environment: Water surface near wildfire[1]

Built environment: Road network near wildfire[2]

1.https://www.ocregister.com/2018/09/19/mountain-lions-can-usually-escape-wildfires-but-the-blazes-can-reduce-their-habitat/

2.https://www.cnn.com/2022/07/30/weather/mckinney-wildfire-northern-california/index.html

Problem statement

1.How to **measure wildfire** and its distribution?

How is the wildfire duration distributed in Texas?

Is the wildfire duration concentrated or randomly distributed?

If it is concentrated, where is **the hotspot** of the wildfire in Texas?

2.What **factors from the natural and built environment** contribute more to the duration of wildfire?

What is their **relationships** with the duration of wildfire in Texas?

Does the variable importance of natural and environmental factors vary across spatial scale?

Where do these factors demonstrate greater significance?

3. What are some **effective wildfire risk communication** might be?

Traditional tables and figures

Dashboard

Stable Diffusion Model Visualization

Framework

Aim:

Non-linear relationship and spatial heterogeneity of influential factors to the duration of wildfire

Measurement:

Outcome variable: Wildfire Duration Days (WDD).

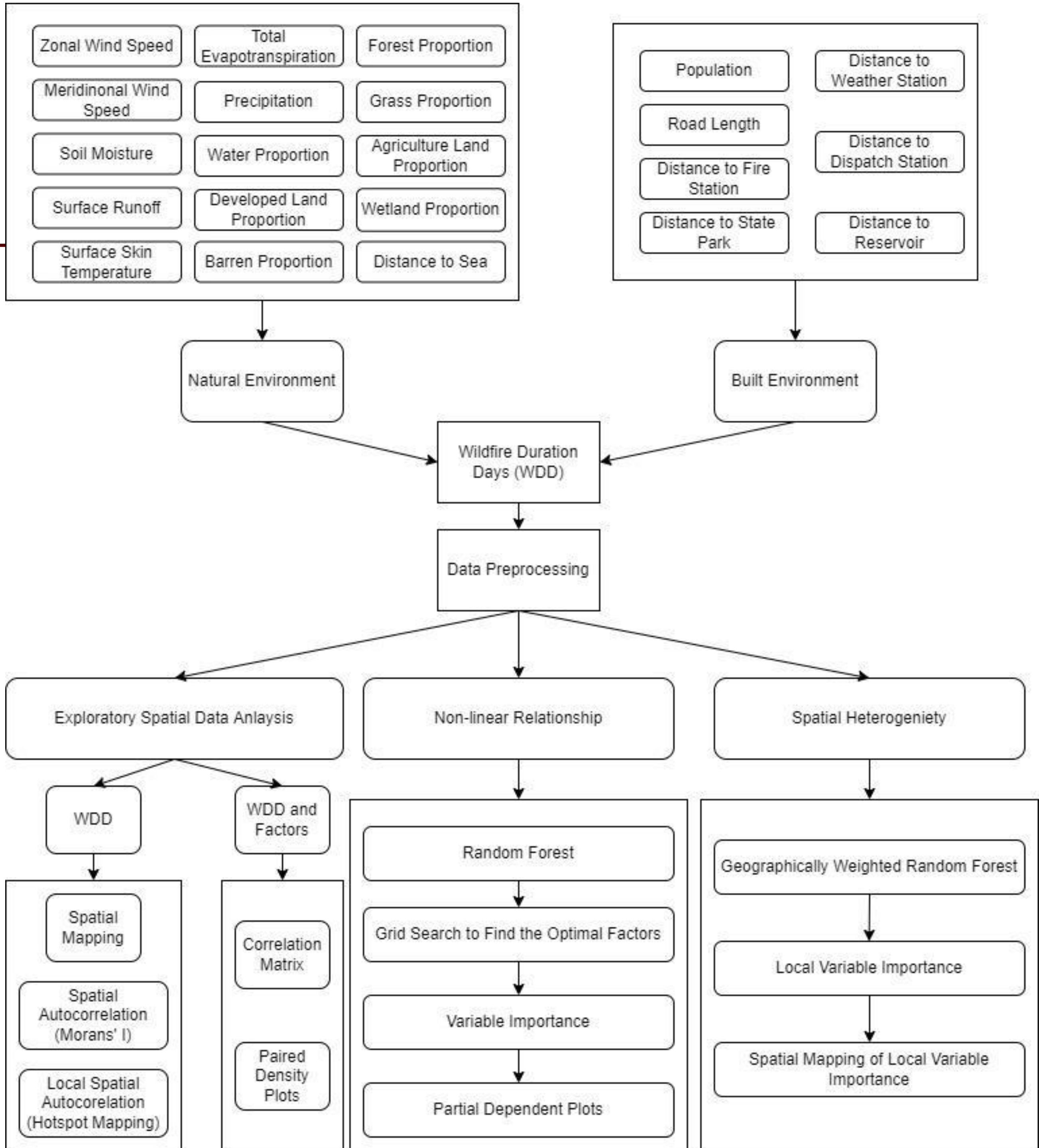
Input variables: Factors from the natural environment and built environment.

Methods:

Exploratory Spatial Data Analysis (ESDA)

Random Forest

Geographically Weighted Random Forest



Data

Generating dependent variable

Analysis unit: **10404 grids with 5-mile side length**

Wildfire duration days (WDD): **yearly average wildfire days within the grid (2014-2022)** from National Interagency Fire Center

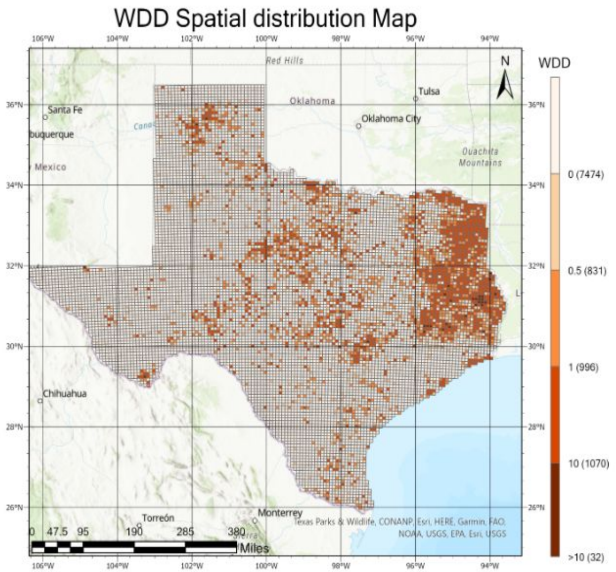


Figure 1. Spatial distribution of WDD

Generating independent variables

15 natural environment variables

1. Generating by calculating the sum value in each grid. For example:

WindSpdZ ->wind speed from zonal direction (m/s) in grid

SoilMoi ->Soil moisture (kg/m^2) in grid

2. Generating by proportion in each grid. For example:

PForest->Percent of forest in each grid (percentage)

PAgri->Percent of agricultural land in each grid (percentage)

7 Built environment variable

1. Generating by density in each grid. For example:

Pop-> population density in each grid(people)

RoadLen-> Road network density in each grid(miles)

2. Generating by distance to centroid of each grid. For example:

DFrieStn->The nearest distance to fire station from each grid(miles)

DDispatch ->The nearest distance to national dispatch office(miles)

Data

Natural environment

Factors	Variables	Source
Climate	Soil moisture	North American Land Data Assimilation System (NLDAS)
	Surface runoff	
	Surface temperature	
	Evapotranspiration	
	Precipitation	
	Wind speed (2 directions)	
Land Use	Developed land	USGS National Land Cover Database
	Barren land	
	Forest land	
	Grass land	
	Agricultural land	
	Wetland	
Location	Texas coastline	United States Census Bureau; USGS DEM data
	Elevation	

Built environment

Factors	Variables	Source
Population	Population density	LandScan Global Population Database
Road	Road Length	TxDOT Roadway Inventory
Fire station	Fire station	Google Map Place API
State Park	Texas State Park	TPWD SPATIAL DATA
Weather station and dispatch office	National Dispatch Office	National Interagency Fire Center
	Interagency Remote Automatic Weather Stations	
Reservoir	Existing Reservoirs	Texas Water Development

Methods

Random Forest

Random Forest is an ensemble machine learning method that uses decision trees to predict outcomes.

We run random forest in Python using sklearn package.

We also calculate variable importance and draw partial dependence plots (PDP) to interpretation the **non-linear relationship** of variables.

Geographically Weighted Random Forest (GWRF)

GWRF is a spatially explicit version of Random Forest that allows us to account for spatial heterogeneity in the relationship between environmental variables and wildfire duration.

We use the package ‘SpatialML’ in R to run GWRF.

We also plot the local variable importance graphs to demonstrate the **spatial heterogeneity** of the influential factors.

Result

Exploratory Spatial Data Analysis

Moran's I is a spatial statistical index that measures the degree of concentration.

The global Moran’s I of WDD is 0.32 ($p<0.01$), indicating **WDD is significantly concentrated** across whole Texas.

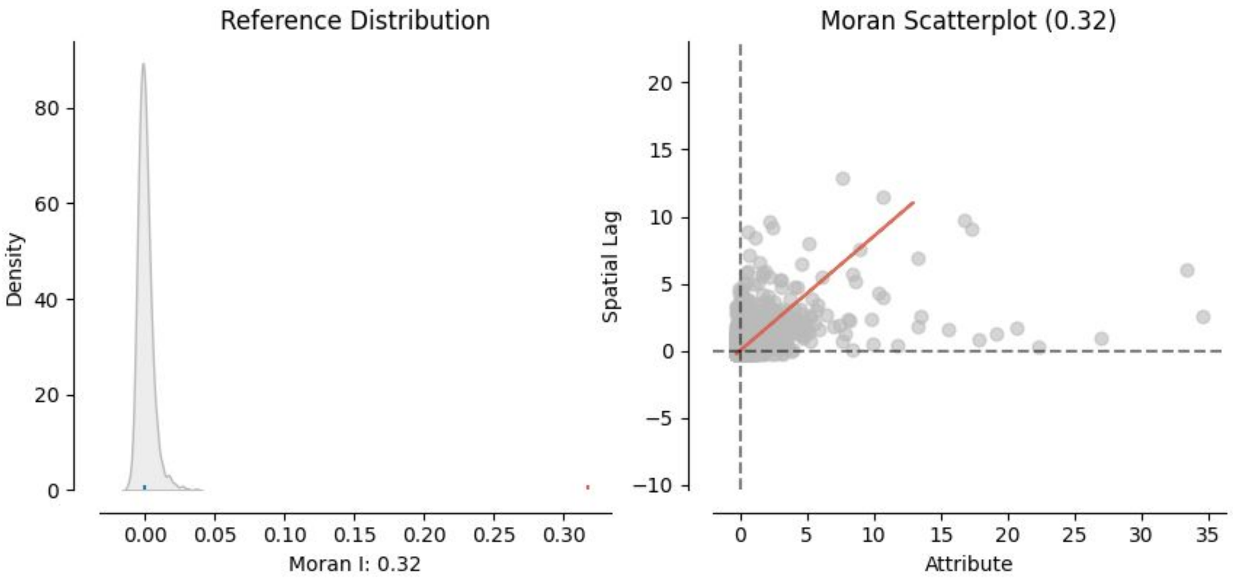


Figure 2. Moran’s I index of WDD

Result

Exploratory Spatial Data Analysis

We plot the hotspots of WDD using Getis-Ord Gi* Statistics.

Hotspots of WDD located at the **east region** of Texas where there are more forest.

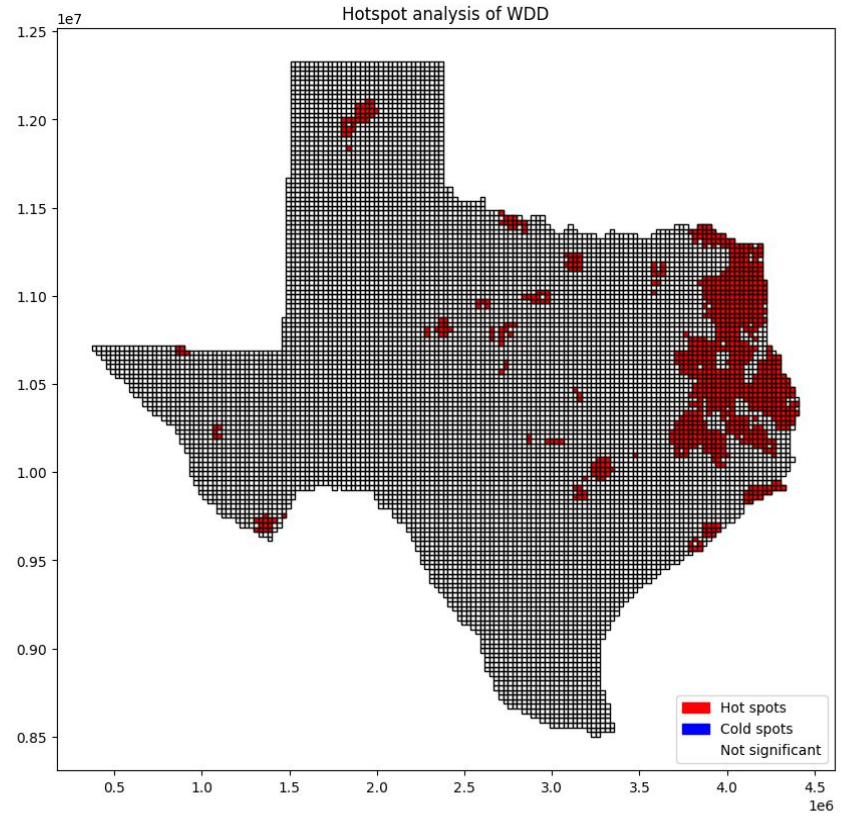


Figure 3. Hot spot analysis of WDD

Result

Correlation Matrix

Correlation matrix show that Perciptation, Evapotrans, SurRunOff, and SurSkiTem are highly correlated.

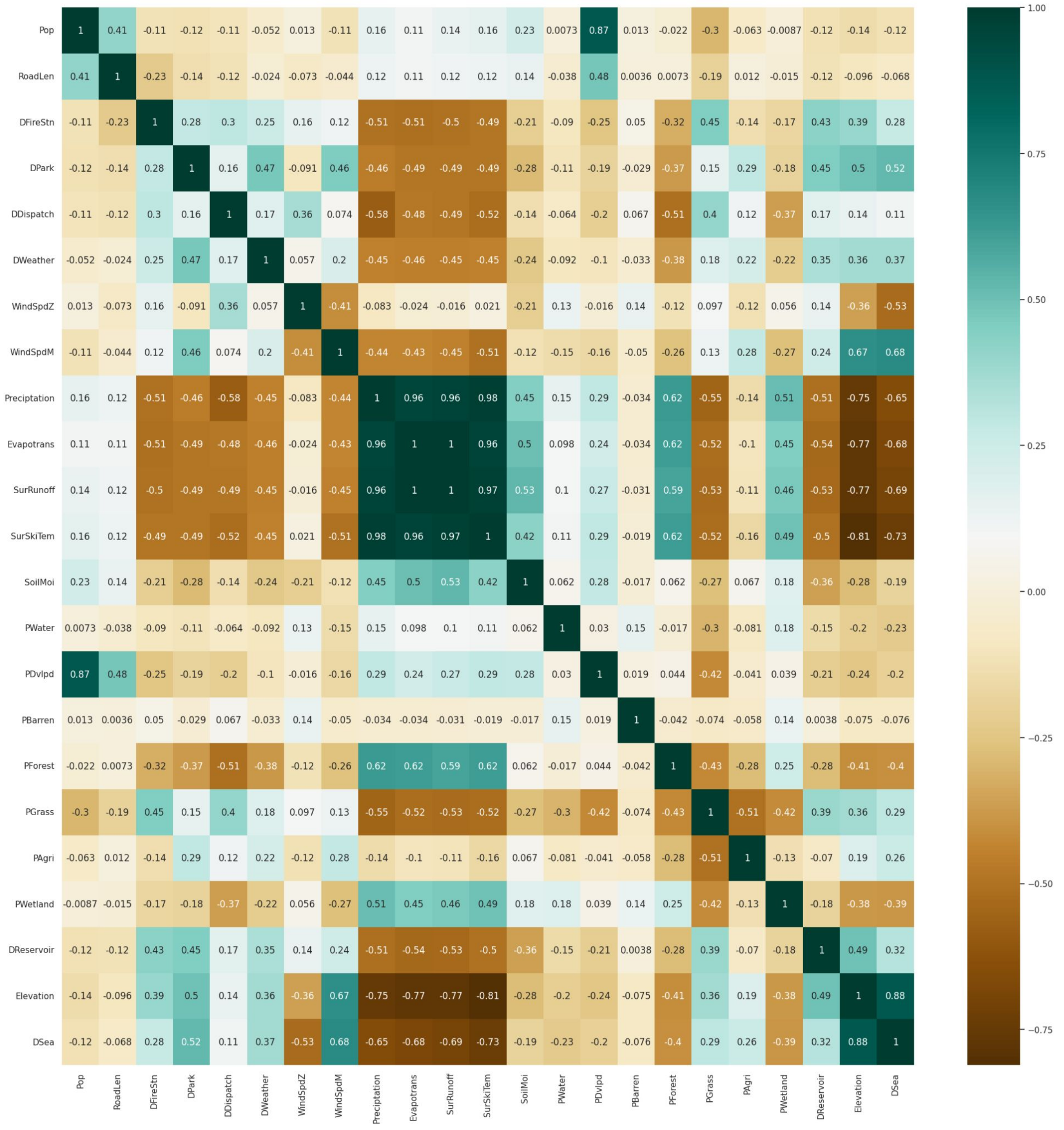


Figure 3. Variable Correlation Matrix

Result

Random forest

Calculate variable Precipitation-Evaporation Index (PEI) using Principle Component Analysis (PCA) for Precipitation, Evapotrans, SurRunOff, and SurSkiTem.

Compare model using PEI and using only one of the variable (We choose SurSkiTem), split the training set and test set to 8:2.

Table 1. Model Performance Comparison between Standardized and Unstandardized Model with PEI and SurSkiTem

Models	RMSE (Test Set)	R2 (Training Set)
Random Forest with PEI	**0.58**	**0.20**
Random Forest with PEI (standardized)	0.58	0.20
Random Forest with SurSkiTem	0.65	0.19
Random Forest with SurSkiTem(standardized)	0.65	0.19

Result

Variable importance

Proportion of Forest, Distance to Weather Station, Proportion of Grass land, and Distance to Dispatch are the top 4 influential factors.

Forest and grass are related to the happening of wildfire, while weather station and dispatch office are related to the discovery and suppression of wildfire.

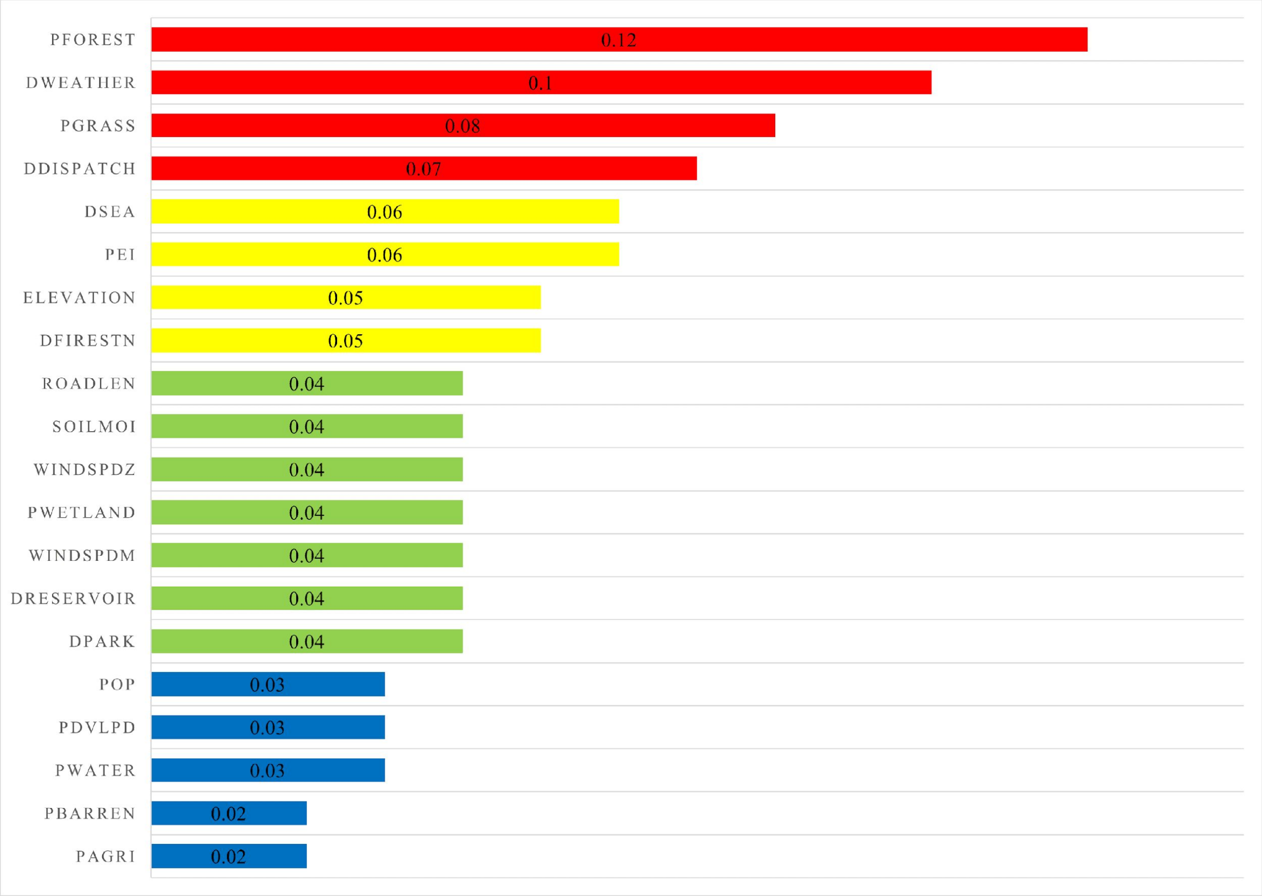


Figure 5. Variable Importance of the Random Forest Model

Result

Partial Dependence Plots

Non-linear relationship of, for example:

Proportion of forest: As proportion of forest increase, the WDD increase. When the proportion of forest is larger than 43%, the WDD will increase faster.

Distance to dispatch office: If the distance to dispatch office is less than 100 miles, the WDD will decrease as the distance to dispatch office increase. When the distance to dispatch office is greater than 100 miles, it does not have a significant effect.

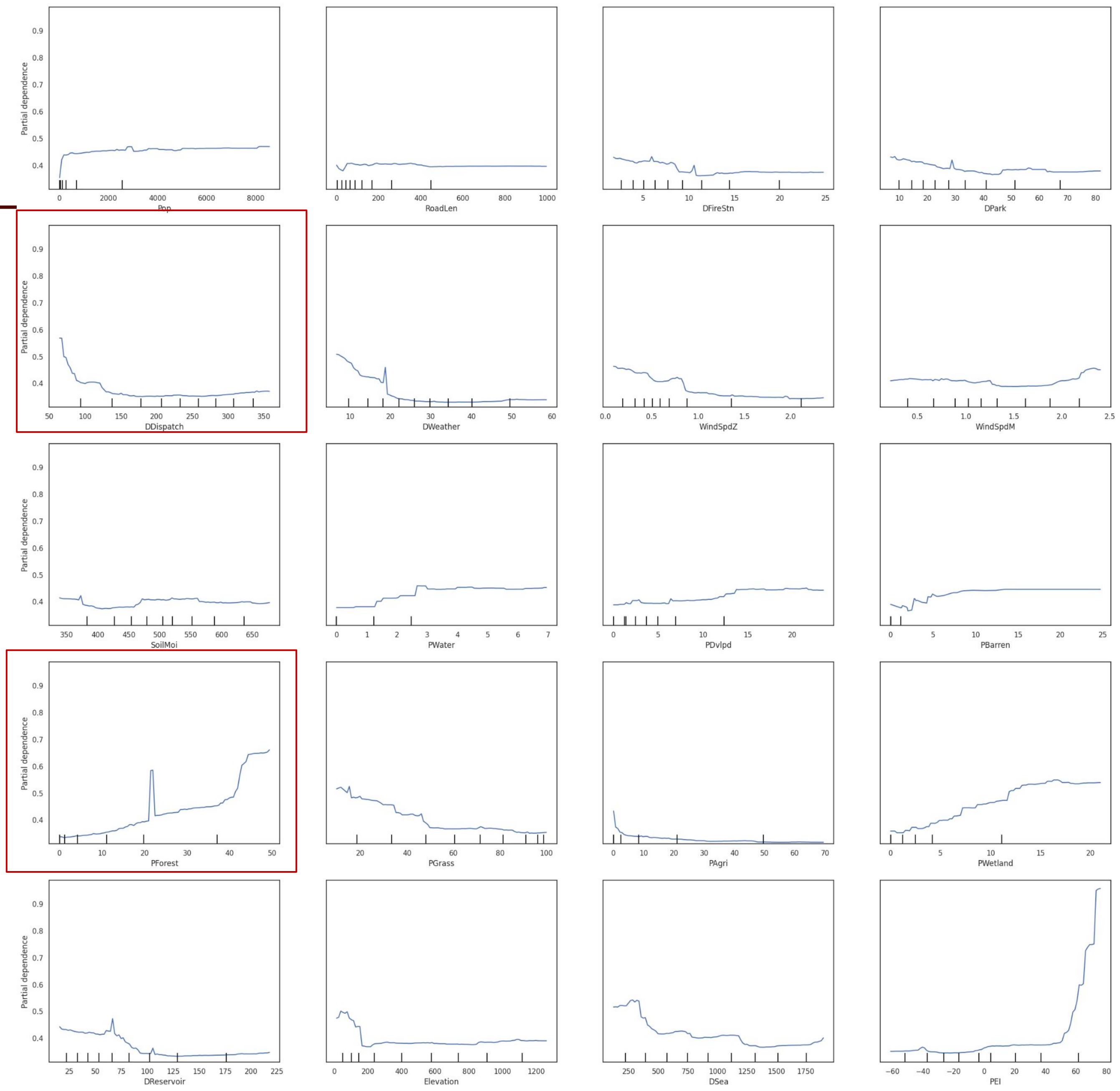


Figure 5. Partial Dependence Plots for All Variables

Result

Local variable importance(GWRF)

Spatial heterogeneity of, for example:

Proportion of forest: Proportion of forest is more important in the west region, north region, and southwestern part.

Proportion of Agricultural land: Proportion of Agricultural land is more important in north region, but not influential in the remaining region.

Agricultural land involves massive human activities which might help to discover and suppress the wildfire faster

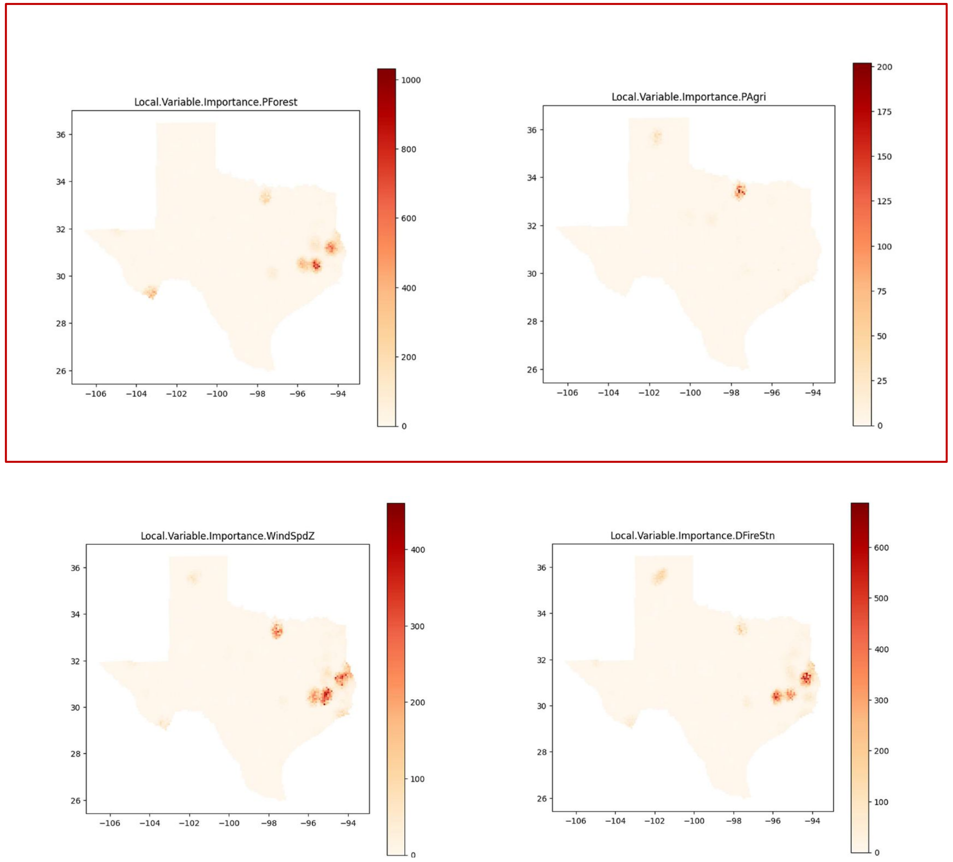


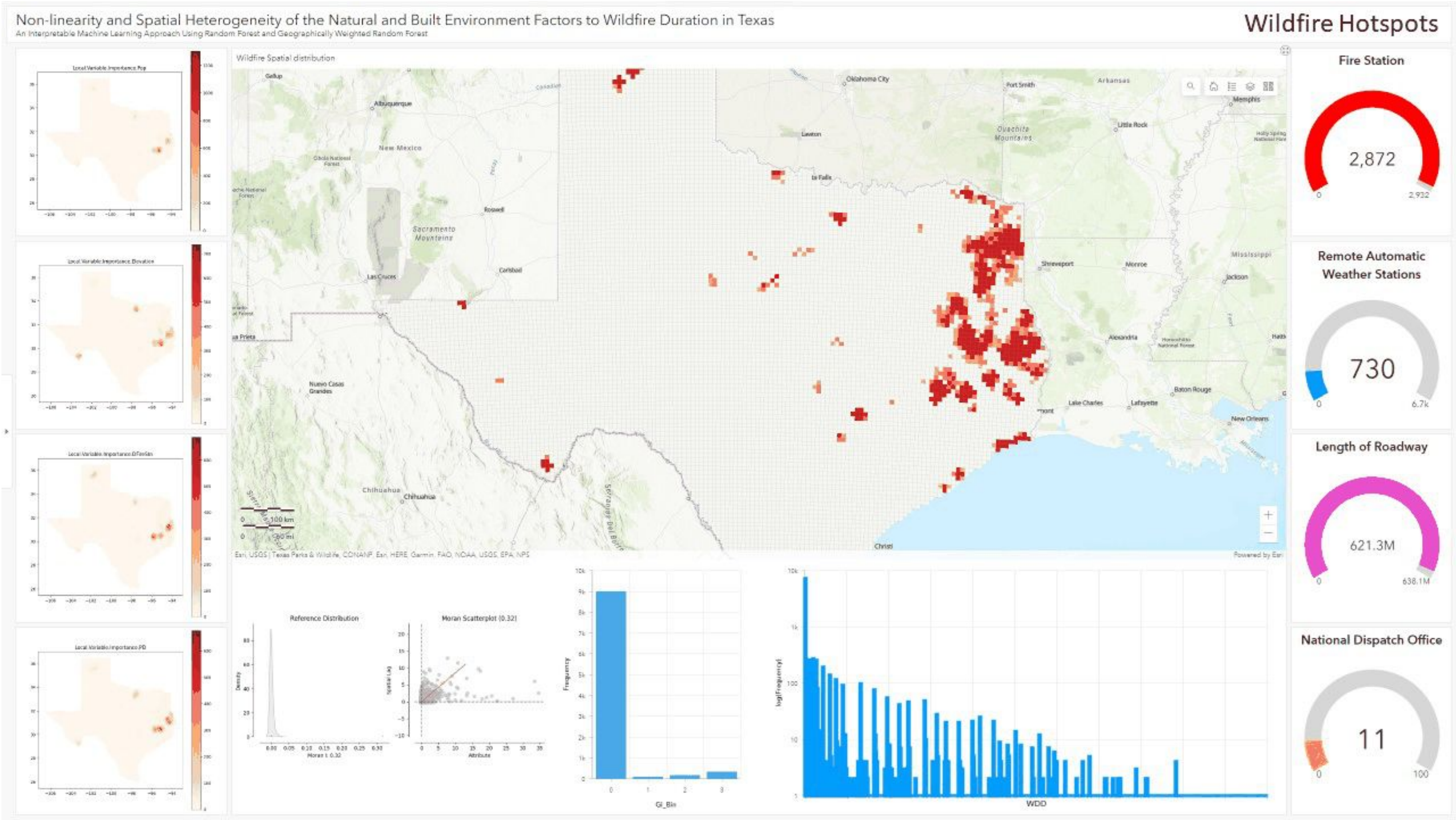
Figure 6. Local variable importance plots

Result

Dashboard

https://tamu.maps.arcgis.com/apps/dashboards/83a779c859a84203addb316bcf30c012

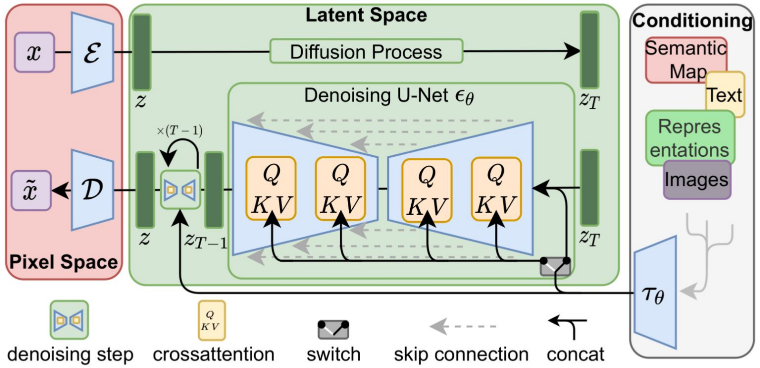
Github repo: https://github.com/UrbanDS/UrbanDS_Team



Result

Stable Diffusion model

Simulate what the fire would look like in your familiar places



TEXAS A&M UNIVERSITY
Landscape Architecture
& Urban Planning

Acknowledgement

Special thanks to Professor Xinyue.Ye for his guidance and to Shoujia Li for her support in the work.

TEXAS A&M UNIVERSITY
Landscape Architecture
& Urban Planning

Q&A

Thank you all for your listening !