

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 wildfire¹



Built environment: Road network near wildfire²



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



Generating independent variables

15 natural environment variables

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

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	National Dispatch Office	National Interacency					
and dispatch office	Interagency Remote Automatic Weather Stations	Fire Center					
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.

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

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

Correlation Matrix

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

Рор	1	0.41	-0.11	-0.12	-0.11	-0.052	0.013	-0.11	0.16	0.11	0.14	0.16	0.23	0.0073	0.87	0.013	-0.022	-0.3	-0.063	-0.0087	-0.12	-0.14	-0.12
RoadLen	0.41	1	-0.23	-0.14	-0.12	-0.024	-0.073	-0.044	0.12	0.11	0.12	0.12	0.14	-0.038	0.48	0.0036	0.0073	-0.19	0.012	-0.015	-0.12	-0.096	-0.068
DFireStn	-0.11	-0.23	1	0.28	0.3	0.25	0.16	0.12	-0.51	-0.51	-0.5	-0.49	-0.21	-0.09	-0.25	0.05	-0.32	0.45	-0.14	-0.17	0.43	0.39	0.28
DPark	-0.12	-0.14	0.28	1	0.16	0.47	-0.091	0.46					-0.28	-0.11	-0.19	-0.029	-0.37	0.15	0.29	-0.18	0.45	0.5	0.52
DDispatch	-0.11	-0.12	0.3	0.16	1	0.17	0.36	0.074					-0.14	-0.064	-0.2	0.067	-0.51	0.4	0.12		0.17	0.14	0.11
DWeather	-0.052	-0.024	0.25	0.47	0.17	1	0.057	0.2					-0.24	-0.092	-0.1	-0.033	-0.38	0.18	0.22	-0.22	0.35	0.36	0.37
WindSpdZ	0.013	-0.073	0.16	-0.091	0.36	0.057	1	-0.41	-0.083	-0.024	-0.016	0.021	-0.21	0.13	-0.016	0.14	-0.12	0.097	-0.12	0.056	0.14	-0.36	-0.53
WindSpdM	-0.11	-0.044	0.12	0.46	0.074	0.2	-0.41	1	-0.44	-0.43	-0.45	-0.51	-0.12	-0.15	-0.16	-0.05	-0.26	0.13	0.28	-0.27	0.24		
Preciptation	0.16	0.12	-0.51	-0.46	-0.58	-0.45	-0.083	-0.44	1	0.96	0.96	0.98	0.45	0.15	0.29	-0.034	0.62	-0.55	-0.14	0.51	-0.51	-0.75	-0.65
Evapotrans	0.11	0.11					-0.024	-0.43	0.96			0.96	0.5	0.098	0.24	-0.034			-0.1	0.45		-0.77	-0.68
SurRunoff	0.14	0.12					-0.016	-0.45	0.96			0.97		0.1	0.27	-0.031			-0.11	0.46		-0.77	-0.69
SurSkiTem	0.16	0.12					0.021	-0.51	0.98	0.96	0.97		0.42	0.11	0.29	-0.019			-0.16	0.49		-0.81	-0.73
SoilMoi	0.23	0.14	-0.21	-0.28	-0.14	-0.24	-0.21	-0.12	0.45	0.5	0.53	0.42	1	0.062	0.28	-0.017	0.062	-0.27	0.067	0.18		-0.28	-0.19
PWater	0.0073	-0.038	-0.09	-0.11	-0.064	-0.092	0.13	-0.15	0.15	0.098	0.1	0.11	0.062	1	0.03	0.15	-0.017	-0.3	-0.081	0.18	-0.15	-0.2	-0.23
PDvlpd	0.87	0.48	-0.25	-0.19	-0.2	-0.1	-0.016	-0.16	0.29	0.24	0.27	0.29	0.28	0.03		0.019	0.044		-0.041	0.039	-0.21	-0.24	-0.2
PBarren	0.013	0.0036	0.05	-0.029	0.067	-0.033	0.14	-0.05	-0.034	-0.034	-0.031	-0.019	-0.017	0.15	0.019	1	-0.042	-0.074	-0.058	0.14	0.0038	-0.075	-0.076
PForest	-0.022	0.0073	-0.32	-0.37	-0.51		-0.12	-0.26	0.62	0.62	0.59	0.62	0.062	-0.017	0.044	-0.042	1		-0.28	0.25	-0.28		
PGrass	-0.3	-0.19	0.45	0.15	0.4	0.18	0.097	0.13					-0.27	-0.3	-0.42	-0.074	-0.43	1	-0.51	-0.42	0.39	0.36	0.29
PAgri	-0.063	0.012	-0.14	0.29	0.12	0.22	-0.12	0.28	-0.14	-0.1	-0.11	-0.16	0.067	-0.081	-0.041	-0.058	-0.28	-0.51	1	-0.13	-0.07	0.19	0.26
PWetland	-0.0087	-0.015	-0.17	-0.18	-0.37	-0.22	0.056	-0.27	0.51	0.45	0.46	0.49	0.18	0.18	0.039	0.14	0.25		-0.13	1	-0.18	-0.38	-0.39
DReservoir	-0.12	-0.12	0.43	0.45	0.17	0.35	0.14	0.24	-0.51	-0.54	-0.53	-0.5	-0.36	-0.15	-0.21	0.0038	-0.28	0.39	-0.07	-0.18	1	0.49	0.32
Elevation	-0.14	-0.096	0.39	0.5	0.14	0.36	-0.36	0.67	-0.75	-0.77	-0.77	-0.81	-0.28	-0.2	-0.24	-0.075	-0.41	0.36	0.19	-0.38	0.49	1	0.88
DSea	-0.12	-0.068	0.28		0.11	0.37	-0.53		-0.65	-0.68	-0.69	-0.73	-0.19	-0.23	-0.2	-0.076		0.29	0.26		0.32	0.88	
	Pop	RoadLen	DFireStn	DPark	DDispatch	DWeather	WindSpdZ	WindSpdM	Preciptation	Evapotrans	SurRunoff	SurSkiTem	SoilMoi	PWater	PDvlpd	PBarren	PForest	PGrass	PAgri	PWetland	DReservoir	Elevation	DSea

Figure 3. Variable Correlation Matrix

-0.50

-0.75

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				

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



Figure 5. Partial Dependence Plots for All Variables

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

Dashboard

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

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



Stable Diffusion model

Simulate what the fire would look like in your familiar places



Visualization of Wildfire in Sam Houston National Forest

Team member: Jiaxin du, Zhunwu Zhu, Tianchen Huang, Weishan Bai



These pictures visualize an assumed wild in Sam Houston National Forest, Texas. Th mages include an existing situa fire situation for each scer we use to generate these on fire situation ned on text descriptions th iso be applied to other tasks such as i nling, outpainting, and generating in



On fire situation - Bird' eve view 1





isting situation - Bird' eye view 2

On fire situation - Bird' eye view 2





On fire situation - Bird' eve view 3





On fire situation - Perspective 1

On fire situation - Persper



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



Thank you all for your listening !