Machine Learning for Exploring Wildfire Smoke Emissions: A Data-Driven Approach

Rahele B. Vaezi[1,2], Michael Nia[1,2], Farnaz Hosseinpour[1,2]

1. Desert Research Institute 2. University of Nevada, Reno

Correspondence: Rahele.BarjesteVaezi@dri.edu

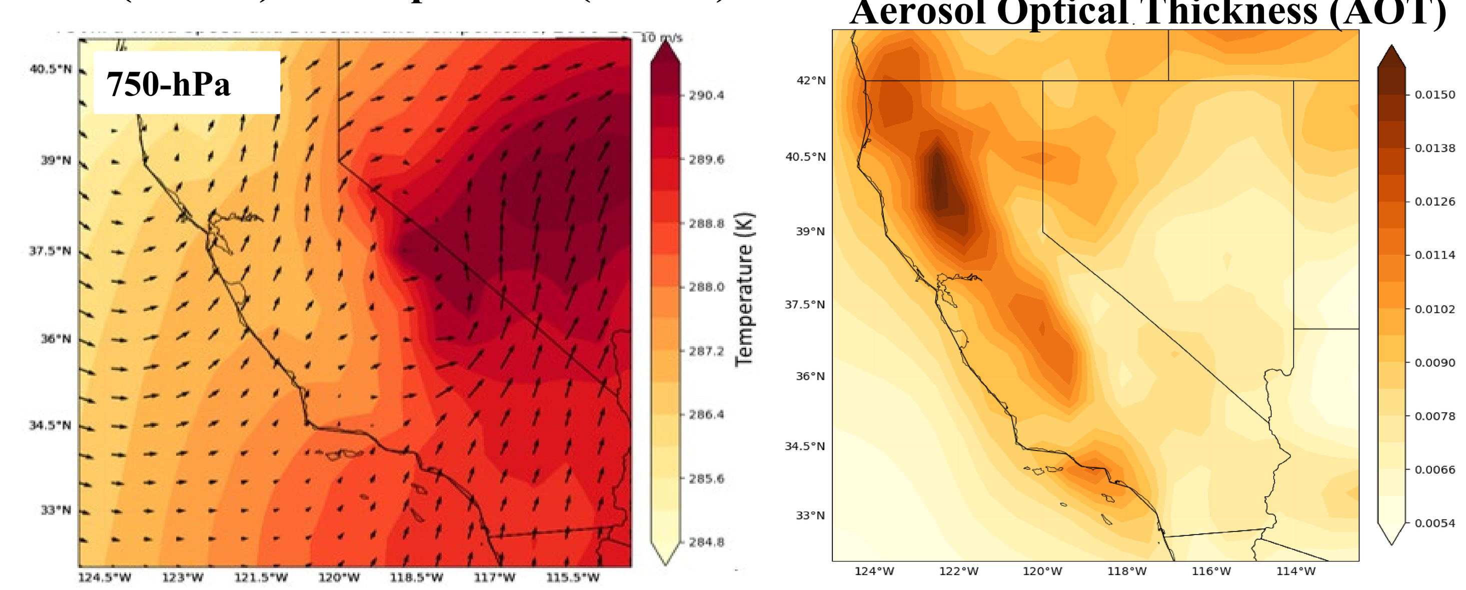
Introduction

Wildfires are increasing in frequency and intensity, releasing significant pollutants that impact regional and global climate. This study investigates the relationship between atmospheric conditions and wildfire emissions in California using machine learning (ML). We leverage an ensemble of remotely sensed observations and historical reanalysis data (MERRA-2) from 2000 to 2021. Our ML models incorporate fire-relevant atmospheric variables such as temperature, wind speed, potential vorticity, and pressure velocity across different vertical levels. By identifying patterns in these complex datasets, ML can provide insights that are difficult or impossible with traditional methods. This research will improve our understanding of how atmospheric conditions influence wildfire emissions, potentially aiding in wildfire prediction and mitigation efforts. (e.g., Hosseinpour et al., 2023).

Climatology and Correlation Studies

- **Climatology of Fire Season (JAS, 2000-2021)**

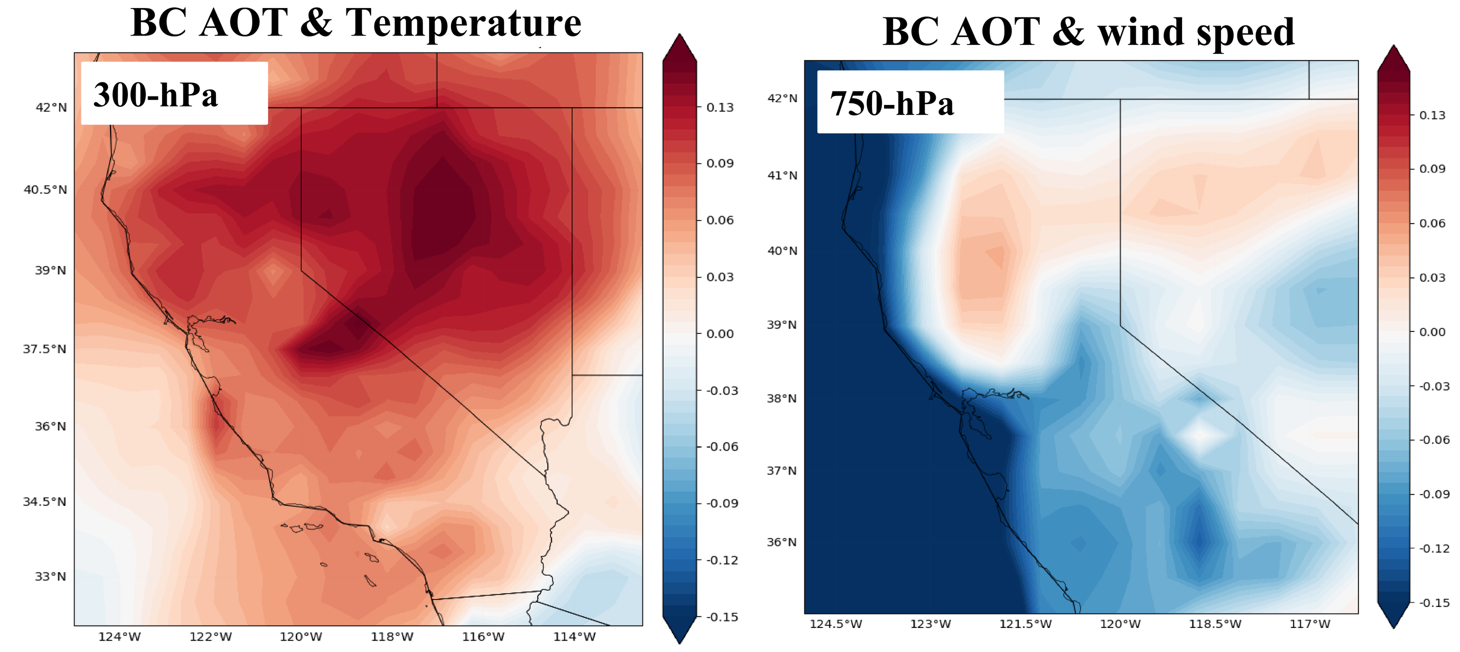
Wind (vectors) & Temperature (shaded) | **Black Carbon (BC) Aerosol Optical Thickness (AOT)**



- Southwesterlies ➔ smoke transport in the mid-troposphere (left graph)

- **Long-term time-correlation (JAS, 2000-2021)**

BC AOT & Temperature | **BC AOT & wind speed**



- High correlation between smoke AOT and temperature over northern Nevada (left graph)
- The enhanced AOD is associated with the increase of the mid-level westerlies (right graph)

Machine Learning (ML) Approach

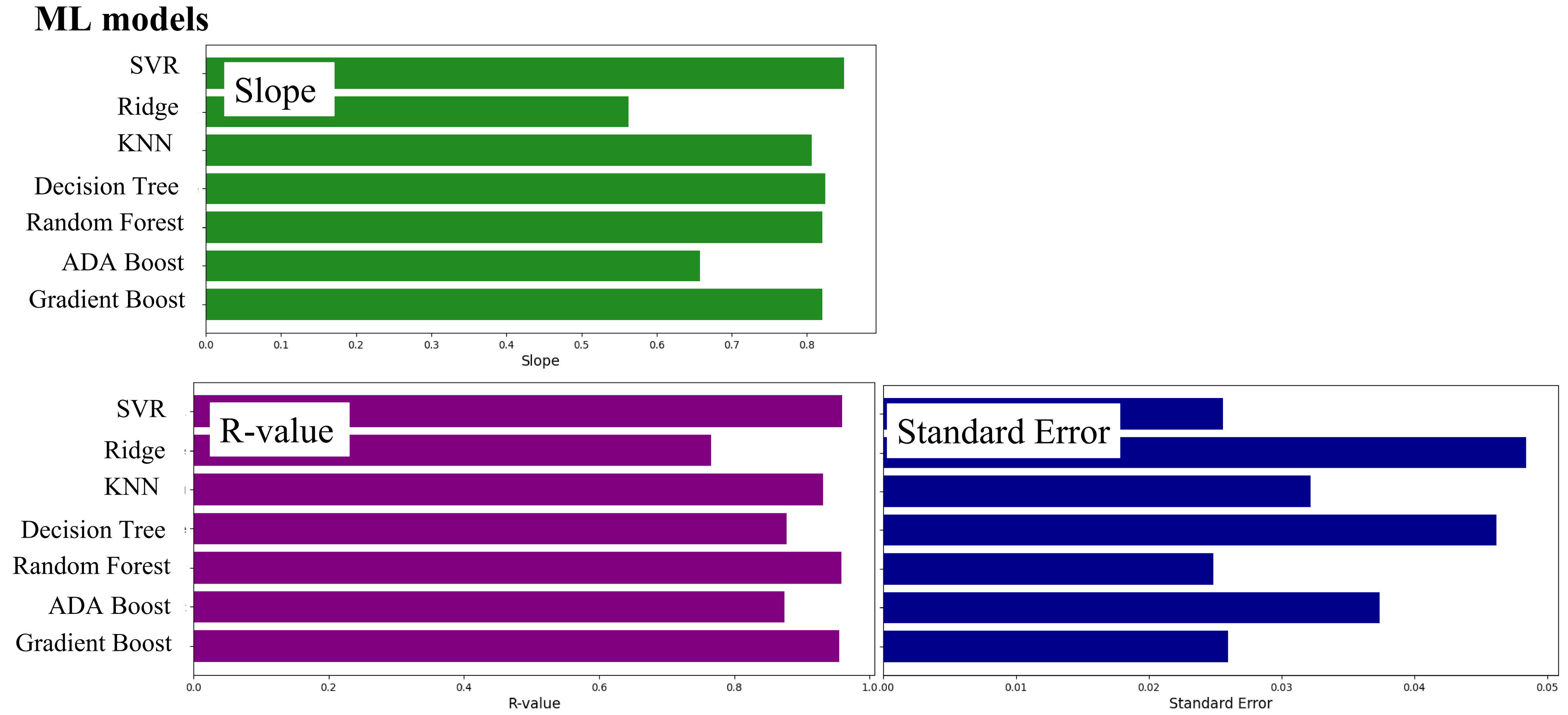
Developed various ML models

- ML model inputs: Temperature, wind speed, Ertel's potential vorticity, and vertical pressure velocity based on 3 vertical levels.
- ML model output: BC Extinction AOT
- Avoid over-fitting: partition data into separate training and test sets.
  - Training set (80% of total data)
  - Test set (20% of total data)
- ML is developed based on the training set and prediction is based on the test set.

Conducted Statistical Analysis

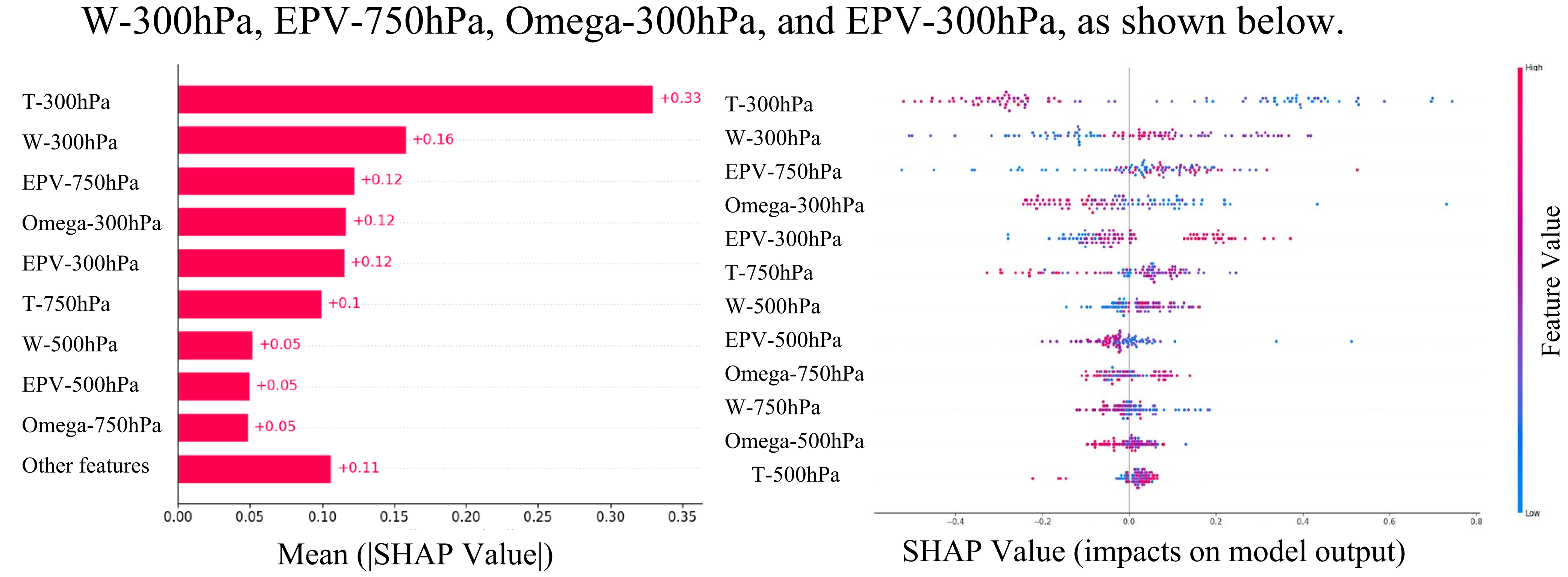
- The R-value, standard error, and slope analyses indicate that the most suitable ML-based model with the highest accuracy is the Support Vector Regression (SVR), Random Forest, and Gradient Boost algorithms.

ML models



Investigated feature importance based on SHAP analysis

- SHAP provides an understanding of how important each input variable (feature) is in predicting smoke emissions (i.e., the ML model's output).
- SHAP analysis highlights that the critical features of the RF model are T-300hPa, W-300hPa, EPV-750hPa, Omega-300hPa, and EPV-300hPa, as shown below.



Results

- We utilized a machine Learning approach to study the impacts of the climate and atmospheric variables on smoke emissions in the California and Nevada regions from 2000 to 2021.
- The presence of high AOT, associated with the vertical motions along with the strong south-westerlies over the region, suggests the impacts of mid-level circulations on the elevation of smoke concentration during fire seasons.
- The statistical metrics and their corresponding values suggested that Support Vector Regression (SVR), Random forest, and Gradient Boost algorithms perform better in predicting fire emissions, compared to the other ML algorithms used in this study.
- We applied SHAP analysis to investigate the most important climate variables that affect the prediction of smoke emissions from wildfires.

Acknowledgments

We gratefully acknowledge the support of the Nevada NASA EPSCoR Program for their generous grant # GR18437, which has made this project possible. We would like to thank Dr. Erfani for his invaluable guidance. We also extend our appreciation to the SMART Lab at the Desert Research Institute, Reno campus, for their advisory and support throughout this project.

References

- Hosseinpour, F., Kumar, N., Tran, T., & Knipping, E. (2024). Using machine learning to improve the estimate of US background ozone. Atmospheric Environment, 316, 120145.
- Hosseinpour, F., Brown, T. (2023). Developing sub-seasonal to seasonal vegetation greenness outlooks. American Meteorological Society, 14th Symposium on Forest and Fire Meteorology
- Hosseinpour, F., Brown, T. (2021). Developing a Smoke Concentration model based on machine learning, Final report, 10 pp, Nov. 2021.
- Marlon, J. R., Bartlein, P. J., Gavin, D. G., Long, C. J., Anderson, R. S., Briles, C. E., & Walsh, M. K. (2012). Long-term perspective on wildfires in the western USA. Proceedings of the National Academy of Sciences, 109(9), E535-E543.
- Reinhardt, T. E. (2000). Smoke exposure at western wildfires (Vol. 525). US Department of Agriculture, Forest Service, Pacific Northwest Research Station.
- Stevens-Rumann, C. S., & Morgan, P. (2019). Tree regeneration following wildfires in the western US: a review. Fire Ecology, 15(1), 1-17.