DUST RISK MAPPING ACROSS INDIA AND SRI LANKA USING MACHINE LEARNING AND EARTH OBSERVATION DATA

Introduction

Dust storms pose major environmental and health risks across South Asia, influencing air quality, visibility, and land degradation. This study applies satellite observations and simple machine-learning techniques to map dust occurrence and estimate dust risk across India and Sri Lanka. By combining MODIS Aerosol Optical Depth (AOD), vegetation indices, and climate variables from TerraClimate, a logistic regression model was developed to classify dust events and generate a monthly and annual dust-risk map for the region.

Study Area

The study area covers India and Sri Lanka, located in South Asia and exposed to frequent dust transport from arid regions such as the Thar Desert. India's diverse climate zones, ranging from desert regions to tropical forests, influence dust movement patterns, while Sri Lanka, located directly south of India, experiences seasonal dust episodes transported across the Indian subcontinent.

Figure 1: Map of the Study Area

Materials

Datasets: Aerosol Optical Depth (AOD), NDVI, TerraClimate (Rainfall, Soil Moisture, Wind Speed, Minimum Temperature, Maximum Temperature, Land cover (for water masking)

Temporal Coverage: India and Sri Lanka, 2020–2021 period

Platforms & Tools: Google Earth Engine (GEE), Google Colab Notebook, Python libraries (geemap, xee, xarray, numpy, pandas, scikit-learn, matplotlib)

APIs / Processing Engines: GEE High-Volume API Endpoint, Xarray Earth Engine plugin (xee)

Spatial Resolution: 0.1° spatial resolution (approx.)

Workflow Summary

Select Region & Time Period: India and Sri Lanka, monthly data from January 2020 to December 2021.

Collect Satellite Data: Load AOD, NDVI, rainfall, soil moisture, wind speed, and temperature from MODIS and TerraClimate.

Monthly Aggregation: Convert all datasets into monthly averages to ensure consistency.

Mask Out Water Bodies: Use MODIS land-cover data to remove water pixels from the analysis.

Build a Unified Dataset: Combine all variables into a single spatio-temporal dataset using Xarray.

Label Dust Events: Define "dust" when AOD ≥ 0.5 and create a dust/no-dust classification layer.

Train Machine Learning Model: Use logistic regression to predict dust occurrence using NDVI and climate variables.

Predict Dust Risk: Generate monthly dust-risk maps and a final mean dust-risk map for the region.

Visualize & Interpret Results: Produce maps for AOD, climate variables, dust classification, monthly risk, and final risk.

Results

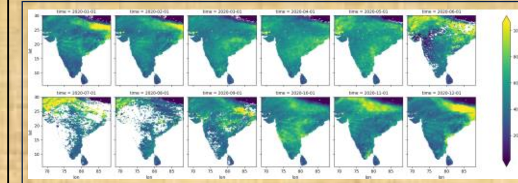

Figure 2: Monthly Aerosol Optical Depth (AOD)

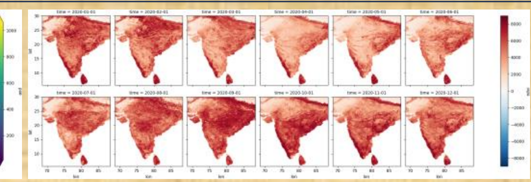

Figure 3: Monthly NDVI distribution

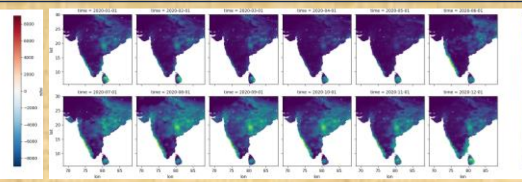

Figure 4: Rainfall patterns across the region

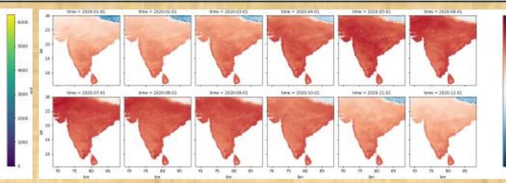

Figure 5: Monthly minimum temperature pattern

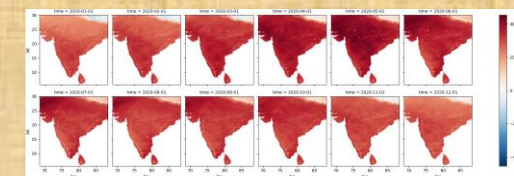

Figure 6: Monthly maximum temperature pattern

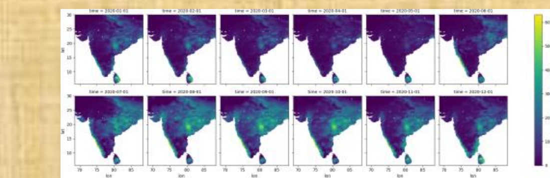

Figure 7: Soil moisture variability

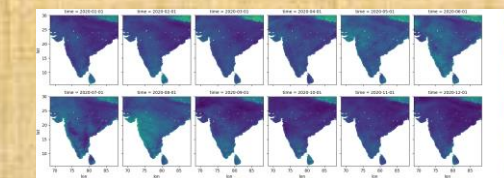

Figure 8: Spatial distribution of wind speed

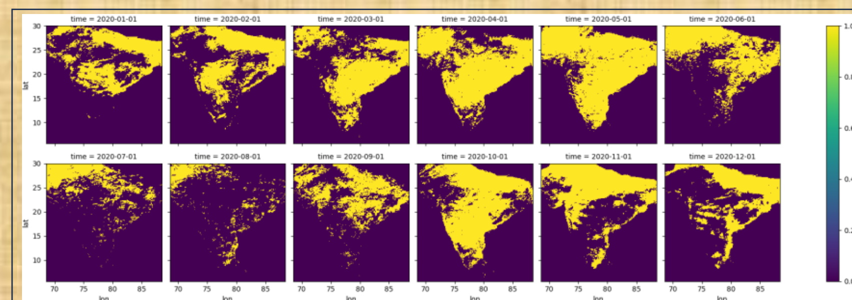

Figure 9: Binary dust classification (Dusty/Non-dusty Classification)

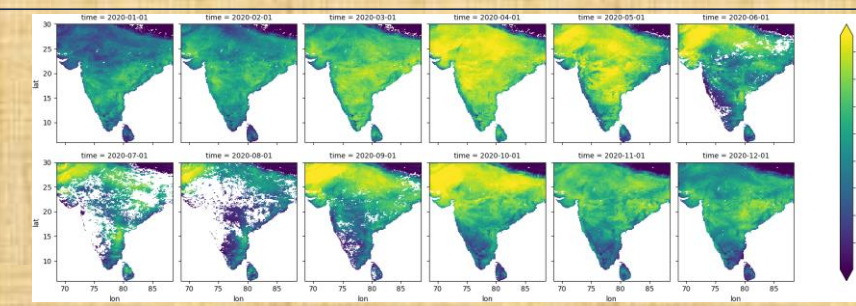

Figure 10: Predicted monthly dust-risk probabilities

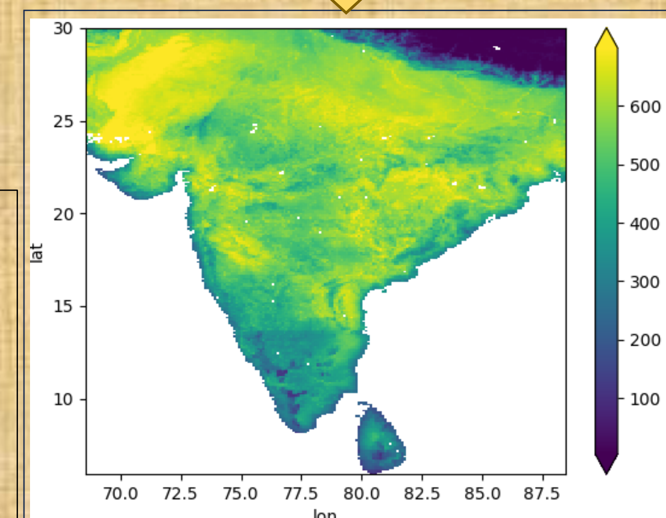

Figure 11: Annual mean dust-risk map

The environmental variable maps (AOD, NDVI, rainfall, wind speed, temperature, and soil moisture) show clear spatial differences across India and Sri Lanka, reflecting the region's climatic variability. High AOD values appear mainly in northwestern India, especially around the Thar Desert, while NDVI and rainfall maps highlight greener and wetter southern regions. These patterns align with the dust/no-dust maps, where dusty conditions are concentrated in arid and semi-arid zones, and rare in Sri Lanka and southern India. The monthly dust-risk maps reveal strong seasonal signals, with elevated risk during the pre-monsoon months and reduced risk during the monsoon. The final annual dust-risk map shows persistent high-risk corridors in northern and central India and low risk in the south.

Link to the Code

Scan Me

The logistic regression model performed reasonably well, achieving 69% accuracy, indicating good ability to distinguish dust from non-dust conditions, despite natural environmental noise. The model detected dust events effectively, with a recall of 74% and a precision of 67%, meaning it successfully identified most true dust cases while maintaining reasonable correctness in its predictions. Performance for non-dust conditions was slightly lower, with a recall of 63% and precision of 71%, but still consistent with expected variability across diverse landscapes. The F1-scores (0.71 for dust and 0.66 for no dust) further demonstrate balanced predictive ability. Importantly, the ROC-AUC of 0.75 shows good discriminatory power, confirming that the model distinguishes well between dusty and non-dusty conditions

Conclusion

This study demonstrates that combining satellite data with a simple machine-learning model can effectively map dust events and predict dust risk. The results highlight key high-risk regions and seasonal dust patterns, supporting environmental monitoring and early-warning systems. Future improvements may include using more advanced models and additional environmental predictors.

References

- MCD19A2: MODIS/Terra+Aqua Land Aerosol Optical Depth Daily L2G Global 1 km SIN Grid V061*. NASA EOSDIS Land Processes DAAC.
- MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V061*. NASA EOSDIS LP DAAC.
- MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V061*. NASA EOSDIS LP DAAC.
- Wu, Q. (2023). *geemap: A Python package for interactive mapping with Google Earth Engine*. Journal of Open-Source Software, 8(82), 5321
- Raspaud, M., et al. (2023). *Xarray Earth Engine (xee)*. GitHub Repository. (https://github.com/cheginit/xee)

Nelson Christopher
2025

nelsonchristophh@gmail.com