Machine Learning Based Analysis and Reconstruction of Soiling Loss Profiles on Photovoltaic Panels

Master's Thesis Presentation by Ludovica Cartolano

Advisor: Prof. Leonardo Micheli

Co-Advisor: Dr. Matthew Muller

Sapienza University of Rome, July 22nd, 2024

SAPIENZA
UNIVERSITÀ DI ROMA

Introduction

Agenda:

- **Motivation**
- **Methodology**
- **Key results:**
  - **Analysis**
  - **Reconstruction**
- **Conclusions and Future Research**

Motivation

The problem area, the importance of solar energy, and the specific challenge on photovoltaic panels.

- **Global climatic action:**
  - “The beginning of the end” of the fossil fuel era.
- **Renewable energy:**
  - What is renewable energy?
  - Solar energy as clean energy and free from carbon emission source.
- **Global energy needs:**
  - EU greenhouse gas emission reduction target 55% by 2030.
  - PV estimated growth to 455-605 GW of cumulative PV capacity by 2030.
- **Specific Challenges:**
  - Soiling on Photovoltaic (PV) panels.

Motivation

Soiling on photovoltaic panels.

- **What is soiling?**
- **Impact of soiling.**
- **Effects of soiling.**
- **Solutions and innovations.**

Research Objective

Reconstruct soiling loss profiles on PV panels via atmospheric data to enhance efficiency.

- **Primary Objective:**
  - Accurately reconstruct soiling loss profiles on PV panels.
- **Possible applications:**
  - Optimize cleaning schedules, enhance overall energy yield.
- **Focus Area:**
  - Atmospheric and meteorological parameter, predictive models.
- **How?**
  1. Understanding the relationship between atmospheric variables and soiling losses metrics.
  2. Make predictions of soiling losses and develop a robust reconstructing model for soiling profile.

Methodology

Data collection and model development to predict and reconstruct soiling loss profiles on PV panels.

- **Data collection:** December 30, 2014 – January 1, 2016.
  - Soiling Station: Calipatria, Imperial County, California, USA.
  - Soiling Metrics:

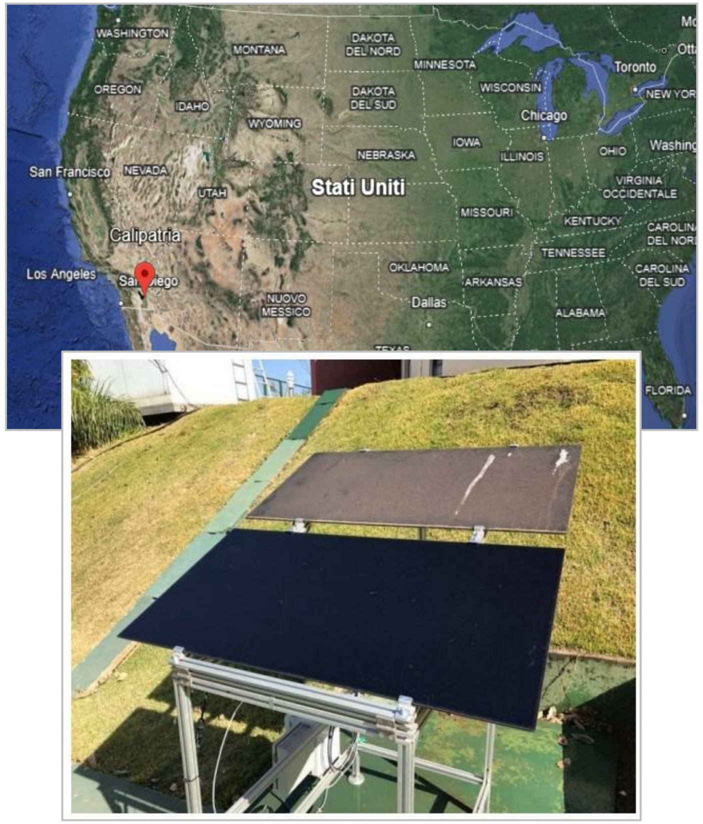
Soiling Ratio (SR): $SR_I = \frac{I_{soiled}}{I_{cleaned}}$

with $SR = 1$ clean condition and $SR < 1$ soiled conditions.

Soiling Rate (SRate %/day):

$$SRate_n = SR_n - SR_{n-1}$$

with $SRate \geq 0\%/day$ cleaning occurring (no soiling accumulation) and $SRate < 0\%/day$ faster soiling accumulation pace.



Methodology

Step 1: analysis of the data and use of machine learning models to predict Soiling Rate (%/day).

Exploratory data analysis: Understand the data, correlations between atmospheric variables and SRate (%/day) and physical meaning.

Feature engineering: Interaction terms, lagged variables, time and seasonal variables to capture complex interactions.

Clustering: Employment of K-means and agglomerative clustering to uncover patterns within the data.

Machine learning models: SVRs and GBMs to predict SRate (%/day).

Feature Importance: Which factors influence the prediction of SRate?

Methodology

Step 2: Soiling Ratio (SR) reconstruction

- **Model Architecture:** Definition of the neural network architecture and hyper-parameters.

- **Training:** Train and play with Random Search and Bayesian Optimizer hyper-parameter tuning.

- **Fine Tuning:** Find the best model and rigorously calibrate the predictions.

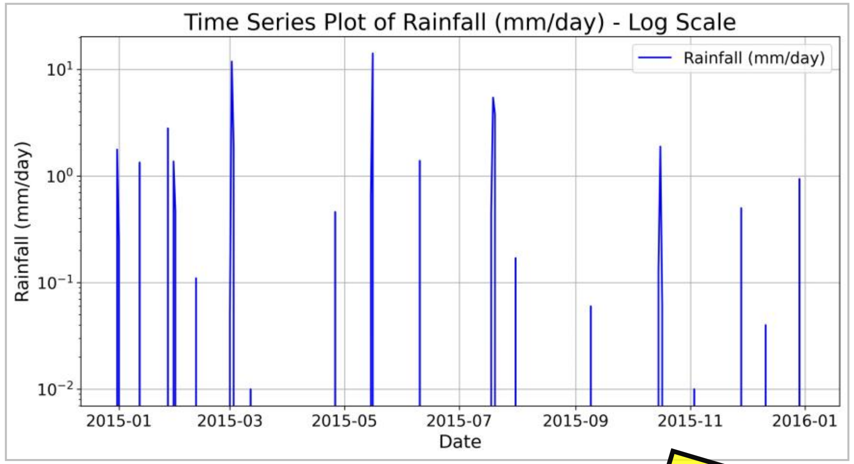
Methodology

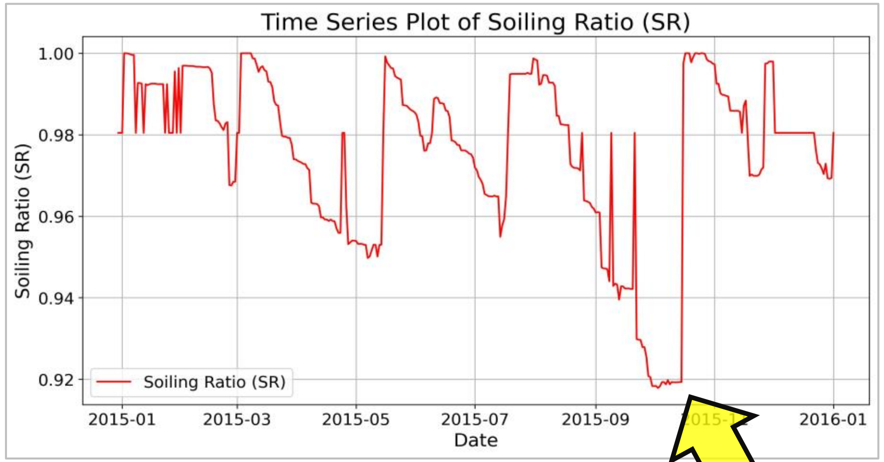
Meteorological Parameters.

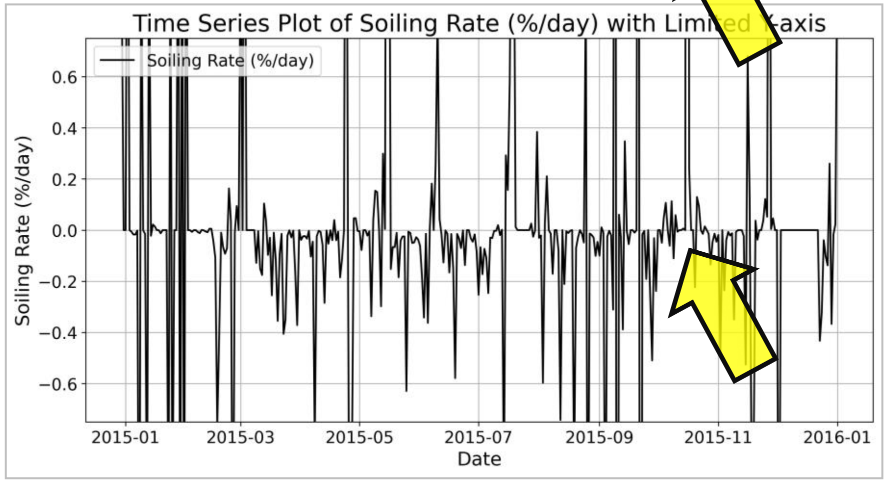
- Rainfall (mm/day).
- Temperature (°C).
- Dew Point (°C).
- Relative Humidity (%).
- Wind Speed (m/s).
- Particulate Matter: $PM_{10}(\mu g/m^3)$, $PM_{2.5}(\mu g/m^3)$ and $PM_1(\mu g/m^3)$.

Results - Analysis

Discuss the finds of the research during the analysis step.





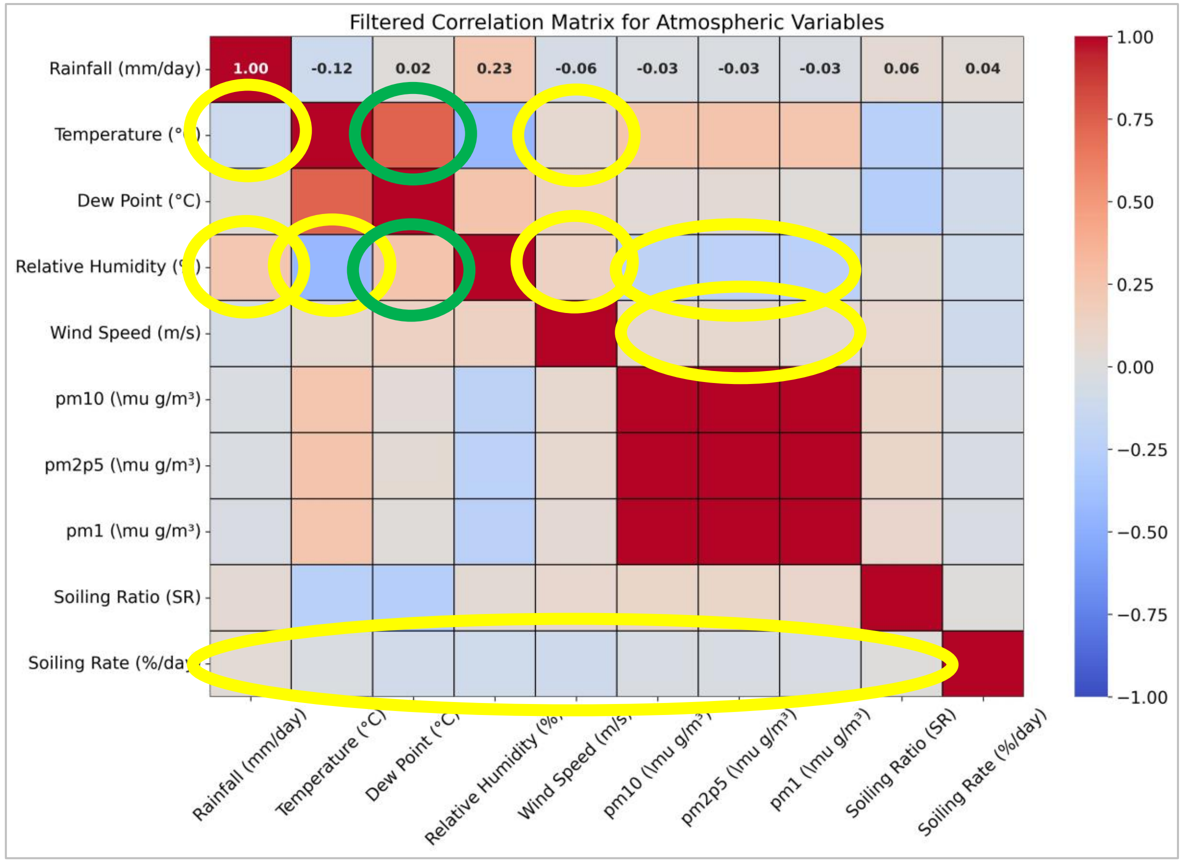


- **The effect of rainfalls.**
- **“Big Rain Events”:**
  - $SR > 0.995$
  - $SRate > 0.5\%/day$ and $SRate < -0.5\%/day$

Results - Analysis

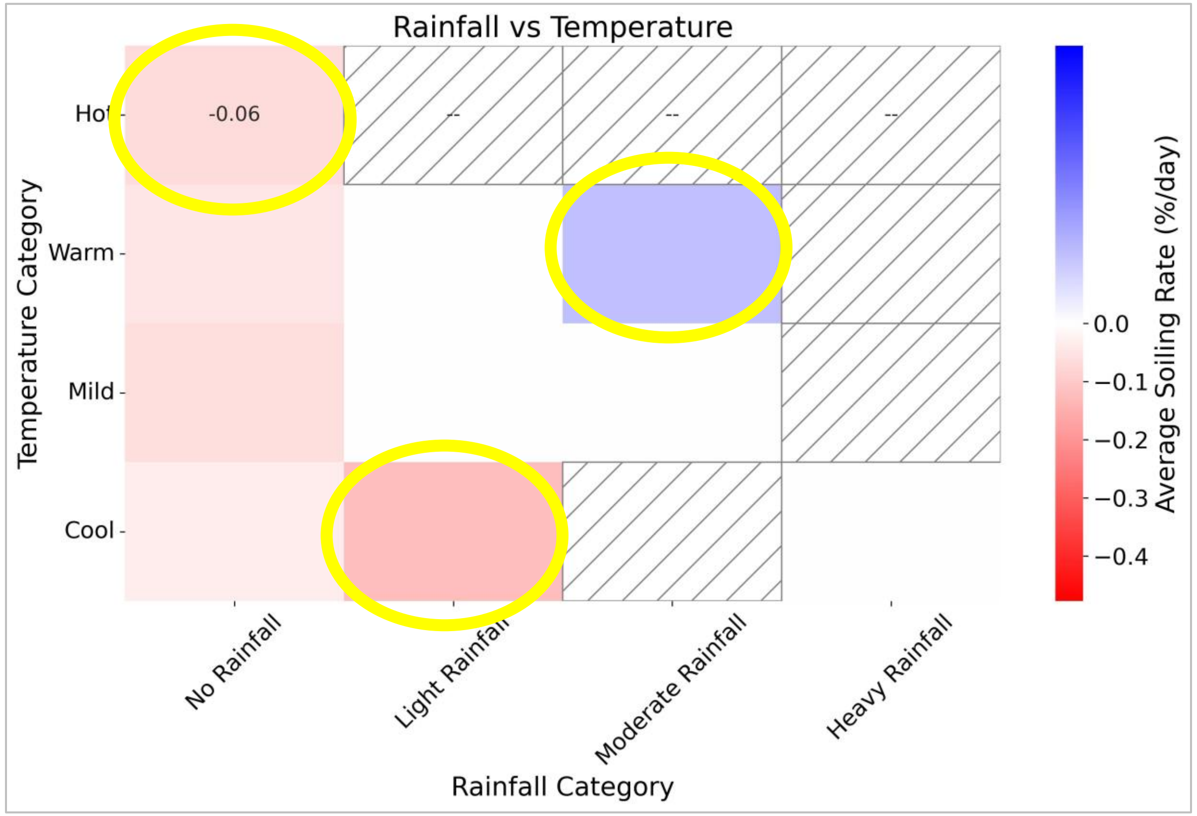
Identify the strongest correlations between atmospheric variables and Soiling Rate (%/day)

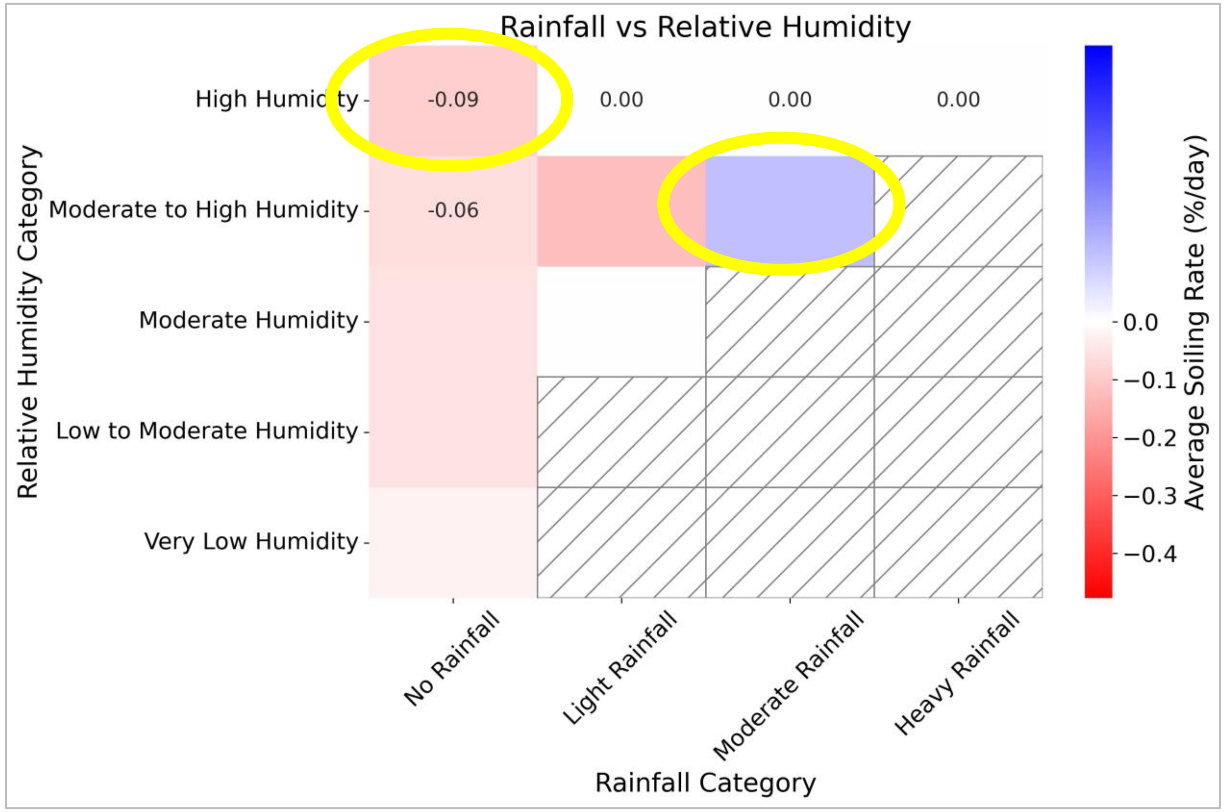
- **Significant relationships:**
  - Rainfall and Temperature
  - Rainfall and Relative Humidity
  - Temperature and Relative Humidity
  - Temperature and Wind Speed
  - Relative Humidity and Wind Speed
  - Relative Humidity with PM
  - Wind Speed with PM
- **Observe:**
  - Dew Point: $T_{dp} = T - \frac{100-RH}{5}$
  - with RH: r = 0.256, P-value = 6.316e-05;
  - with T: r = 0.736, P-value = 6.157e-42



Results - Analysis

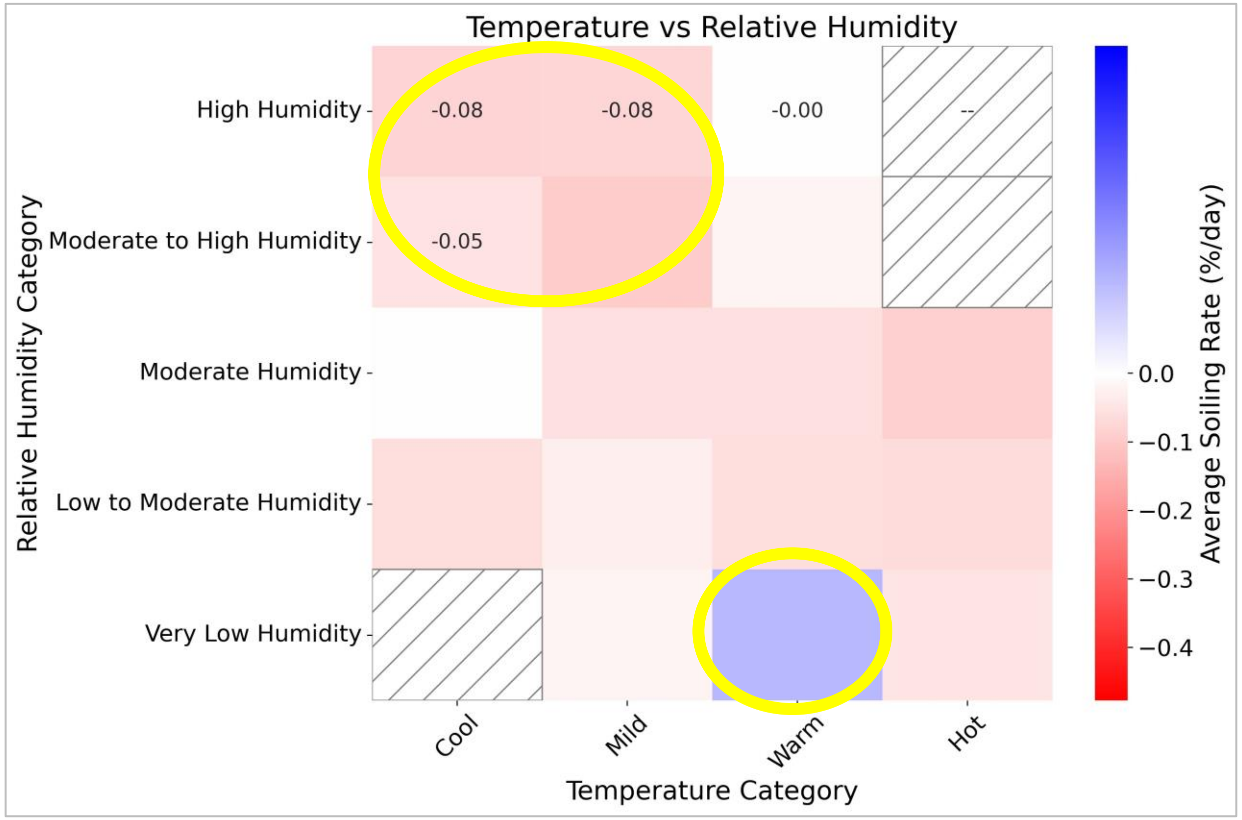
Understand the combined impact of atmospheric variables on the daily Soiling Rate (%/day) using heat-maps.

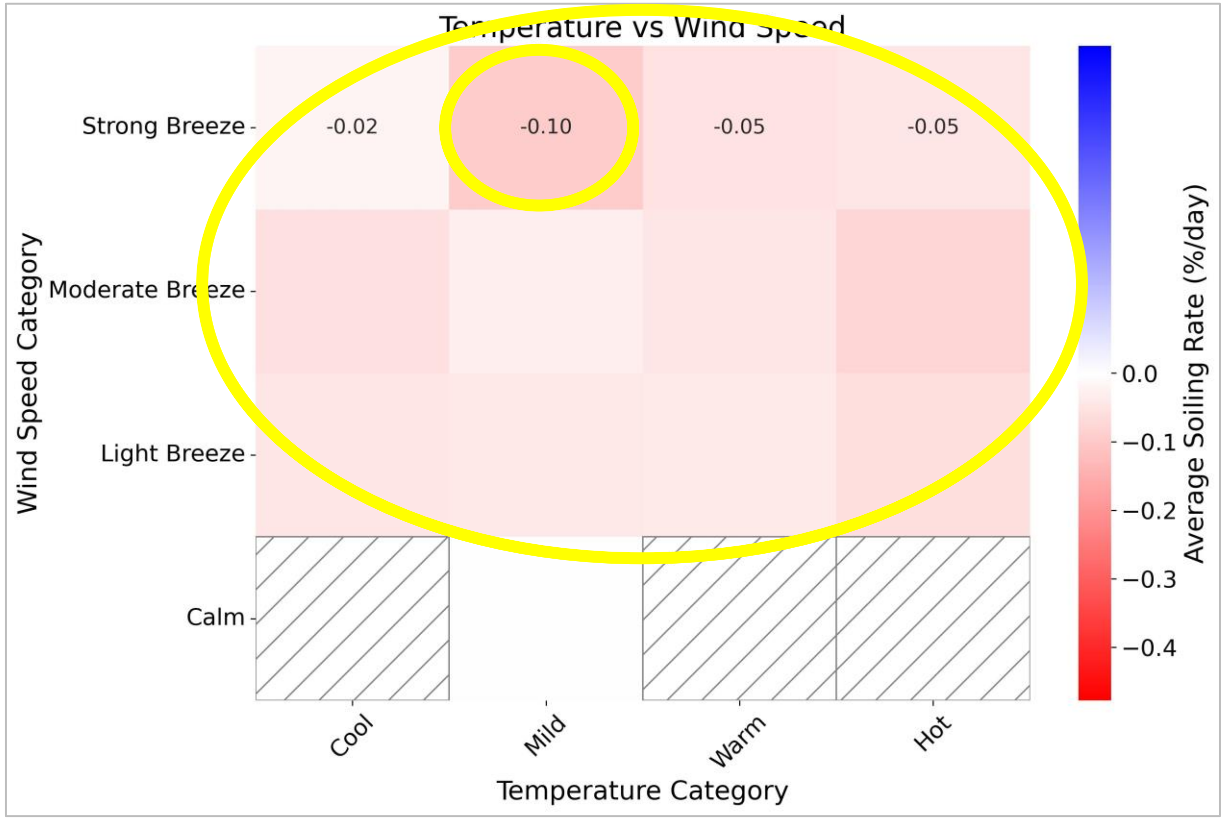




Results - Analysis

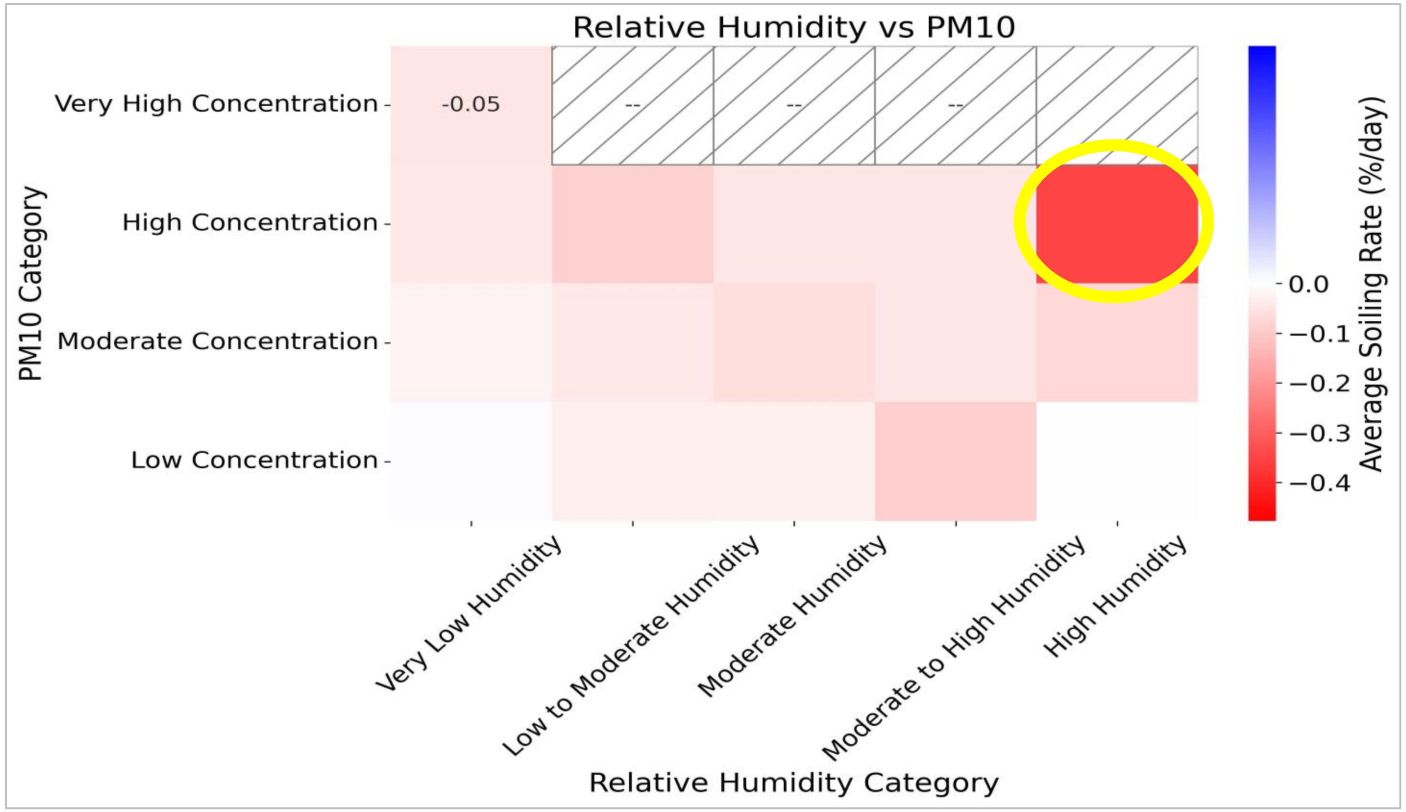
Understand the combined impact of atmospheric variables on the daily Soiling Rate (%/day) using heat-maps.

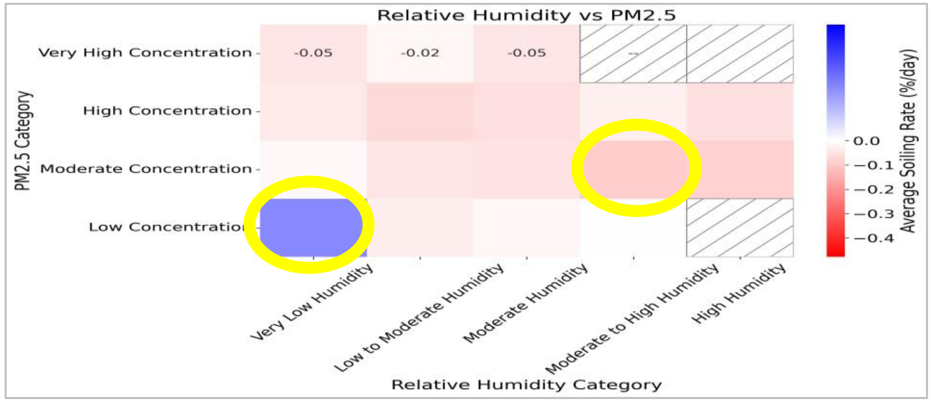


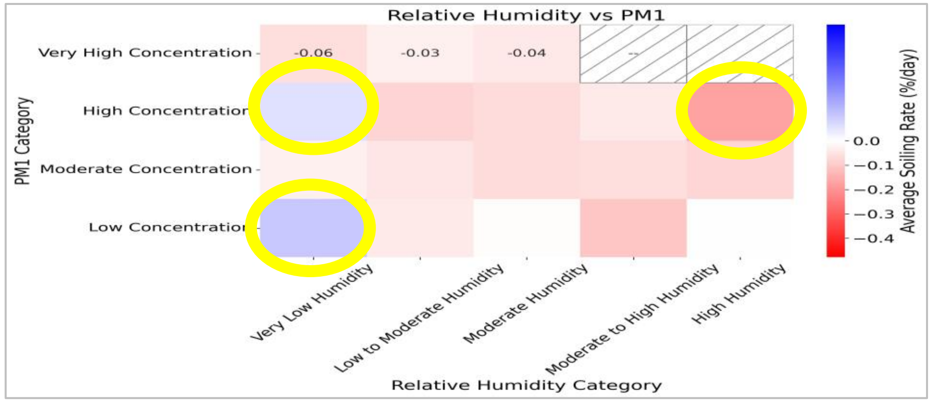


Results - Analysis

Understand the combined impact of atmospheric variables on the daily Soiling Rate (%/day) using heat-maps.







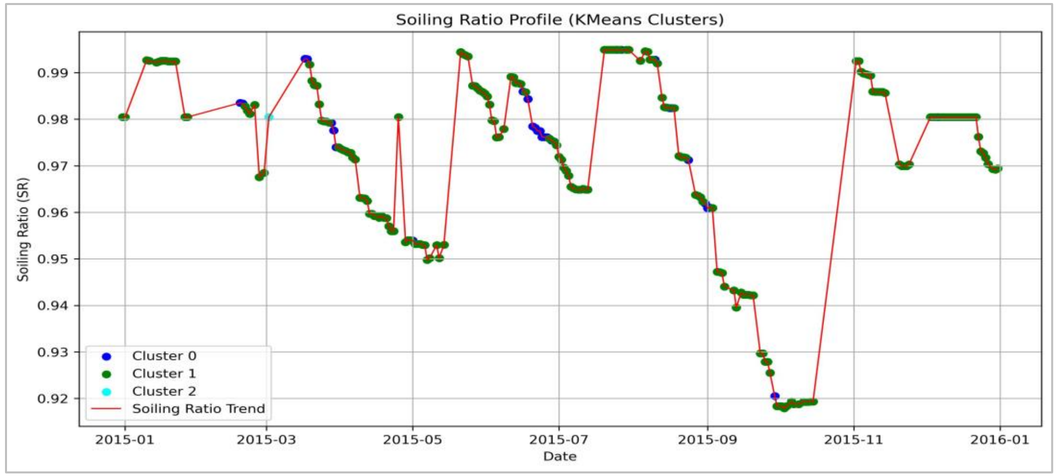
Results - Analysis

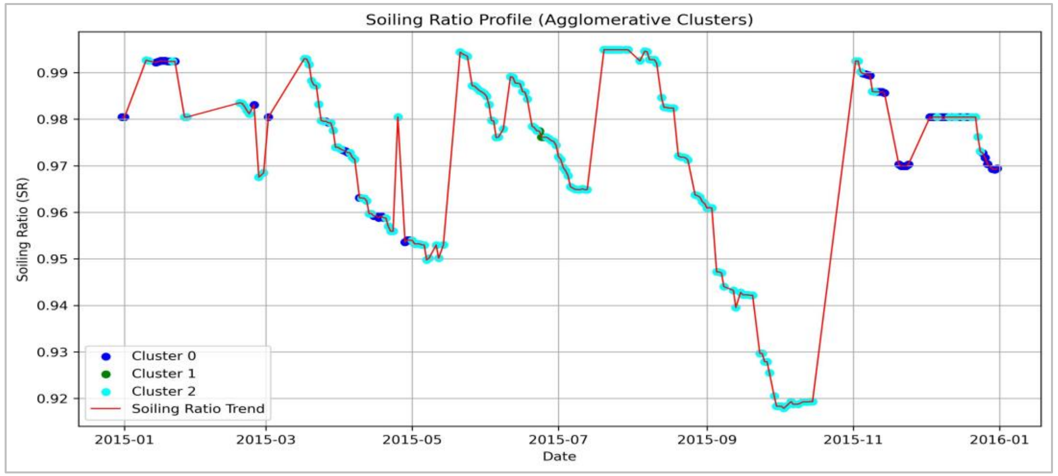
Feature engineering to capture additional nuances in the data.

- **Interaction features:**
  - Rainfall-Temperature, Relative Humidity- PM_{10}, Rainfall-Relative Humidity, Temperature-Relative Humidity, Temperature-Wind Speed, Wind Speed- PM_1.
- **Extraction features:**
  - Lagged SRate for 3 and 7 days and rolling averages and standard deviations for 7 and 30 days of Rainfall (mm/day).
- **Temporal data features:**
  - Day of the Year, Month, Quarter, Weekday, Spring, Summer, Fall and Winter.
- **Aggregation features:**
  - Days Since Last Rainfall and Dew Formation.

Results - Analysis

Clustering used to examine correlation between patterns in atmospheric variables and soiling rates.



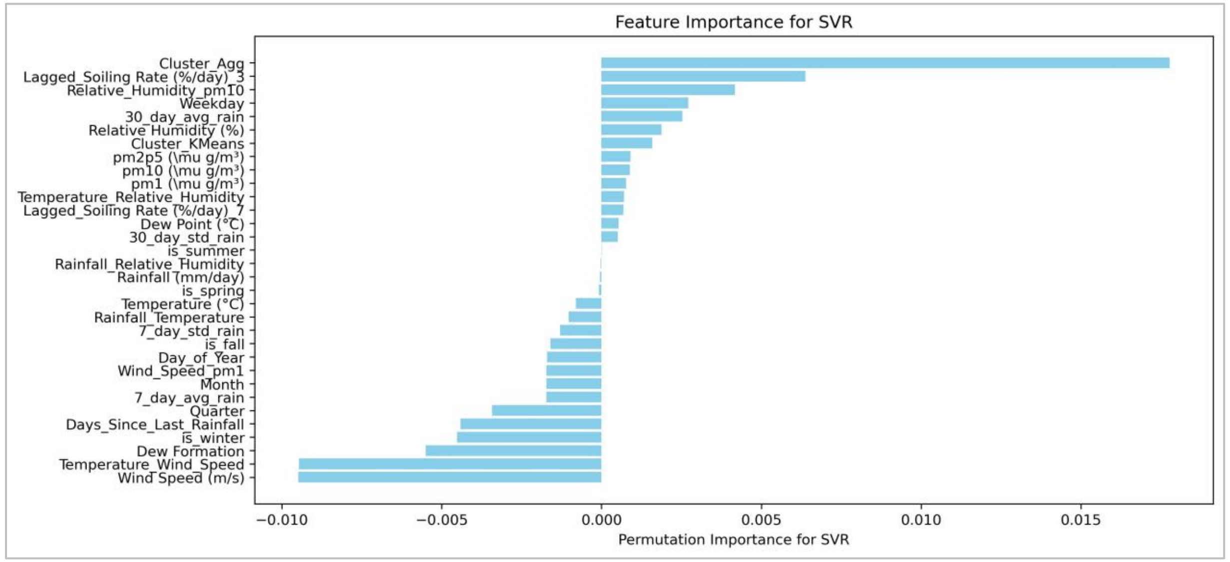


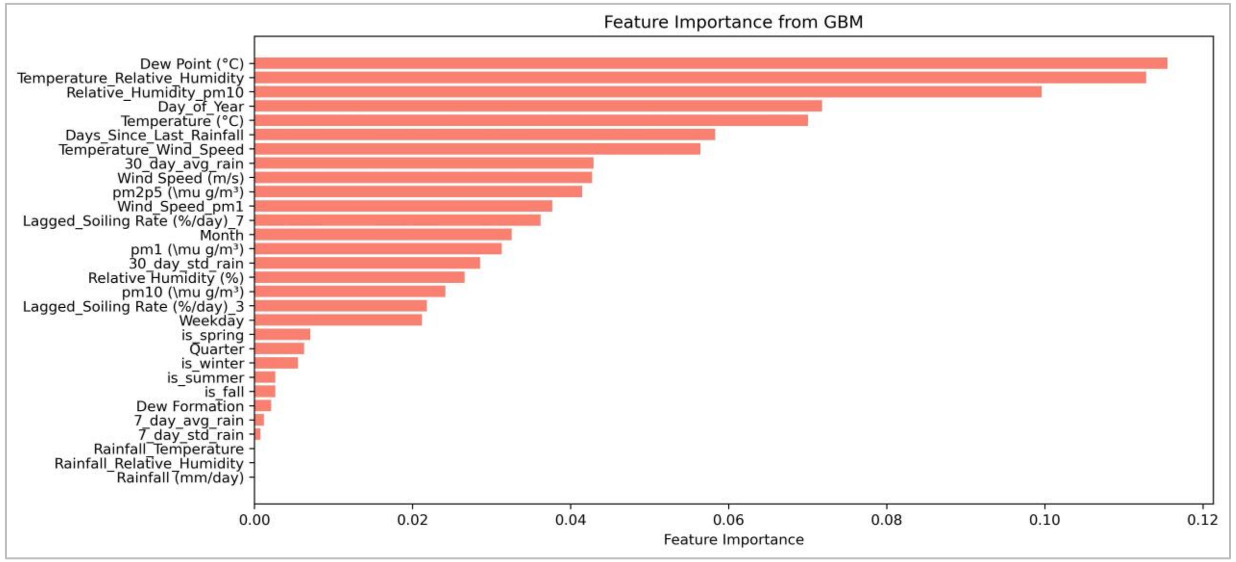
- **Cluster 0:** Less frequent, significant soiling changes, unusual weather.
- **Cluster 1:** Predominant, lower or stable soiling changes, usual weather.
- **Cluster 2:** Sporadic condition.

- **Cluster 0:** Frequent during certain periods, significant soiling changes.
- **Cluster 1:** Sporadic, more localized.
- **Cluster 2:** Predominant, lower or stable soiling changes, usual weather.

Results - Analysis

Discuss feature importance after SVR and GBM prediction of SRate (%/day).





- **Top:** Cluster Agg, Lagged SR 3 days, RH - PM_{10}.
- **Mid-level:** Weekday, RH, PMs.
- **Least important:** WS, T - WS, Dew Formation, Winter, Fall, 7-day avg/std rain.

- **Top:** Dew Point, T - RH, RH - PM_{10}.
- **Mid-level:** Days Since Last Rainfall.
- **Least important:** RF – T, 7-day avg/std rain, Dew Formation, seasons.

Results - Analysis

List the final features of the dataset.

- **Atmospheric variables:**
  - Rainfall (mm/day), Temperature (°C), Dew Point (°C), Relative Humidity (%), Wind Speed (m/s), Particulate Matter: $PM_{10}(\mu g/m^3)$, $PM_{2.5}(\mu g/m^3)$ and $PM_1(\mu g/m^3)$.
- **Feature engineered variables:**
  - Rainfall-Temperature, Rainfall-Relative Humidity, Temperature-Relative Humidity, Temperature-Wind Speed, Relative Humidity- PM_{10}, Wind Speed- PM_1, 7-day avg rain, 30-day avg rain, 7-day std rain, 30-day std rain, Lagged Soiling Rate (%/day) 3, Lagged Soiling Rate (%/day) 7, Month, Quarter, Day of Year, Weekday, Spring, Summer, Fall, Winter, Days Since Last Rainfall, Dew Formation.
- **Clusters:**
  - K-Means Clustering and Agglomerative Clustering.
- **Soiling metrics:**
  - Soiling Rate and Soiling Ratio.
- **Predictions:**
  - SVR Prediction SRate, GBM Prediction SRate, SVR Prediction SR, GBM Prediction SR.

Results - Reconstruction

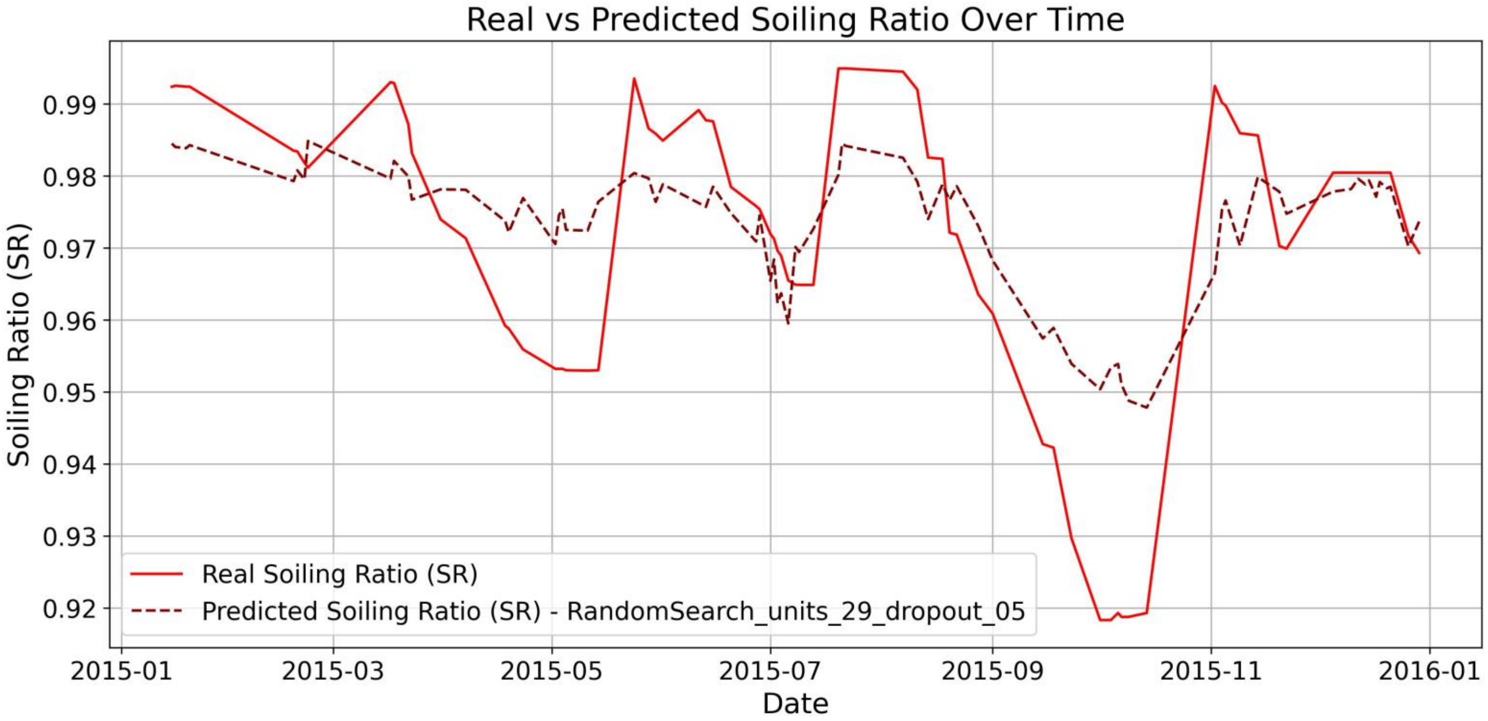
NN architecture, hyper-parameters tuning, play with intervals, feature selection.

Number of Features	Dropout Interval Input	Dropout Interval Hidden Layers	Tunin Method	RMSE	MAE
37	[0.0, 0.5]	[0.5, 0.8]	Random Search	0.01463	0.120698
37	[0.0, 0.5]	[0.5, 0.8]	Bayesian Optimization	0.018145	0.014381
37	[0.0, 0.2]	[0.5, 0.8]	Random Search	0.017837	0.015145
37	[0.0, 0.2]	[0.5, 0.8]	Bayesian Optimization	0.016448	0.013579
37	[0.0, 0.5]	[0.0, 0.5]	Random Search	0.015921	0.012897
37	[0.0, 0.5]	[0.0, 0.5]	Bayesian Optimization	0.013455	0.010996
37	[0.0, 0.2]	[0.0, 0.5]	Random Search	0.017228	0.014366
37	[0.0, 0.2]	[0.0, 0.5]	Bayesian Optimization	0.014383	0.011719
31	[0.0, 0.5]	[0.5, 0.8]	Random Search	0.015370	0.012802
31	[0.0, 0.5]	[0.5, 0.8]	Bayesian Optimization	0.018750	0.015003
31	[0.0, 0.2]	[0.5, 0.8]	Random Search	0.017804	0.015014
31	[0.0, 0.2]	[0.5, 0.8]	Bayesian Optimization	0.018311	0.014987
31	[0.0, 0.5]	[0.0, 0.5]	Random Search	0.020415	0.017203
31	[0.0, 0.5]	[0.0, 0.5]	Bayesian Optimization	0.014749	0.012500
31	[0.0, 0.2]	[0.0, 0.5]	Random Search	0.017755	0.0148275
31	[0.0, 0.2]	[0.0, 0.5]	Bayesian Optimization	0.013909	0.011212
29	[0.0, 0.5]	[0.5, 0.8]	Random Search	0.013455	0.010776
29	[0.0, 0.5]	[0.5, 0.8]	Bayesian Optimization	0.018484	0.015338
29	[0.0, 0.2]	[0.5, 0.8]	Random Search	0.018972	0.015844
29	[0.0, 0.2]	[0.5, 0.8]	Bayesian Optimization	0.017955	0.014162
29	[0.0, 0.5]	[0.0, 0.5]	Random Search	0.018684	0.016146
29	[0.0, 0.5]	[0.0, 0.5]	Bayesian Optimization	0.014370	0.011371
29	[0.0, 0.2]	[0.0, 0.5]	Random Search	0.017503	0.014627
29	[0.0, 0.2]	[0.0, 0.5]	Bayesian Optimization	0.017955	0.014162

Results - Reconstruction

Best model.

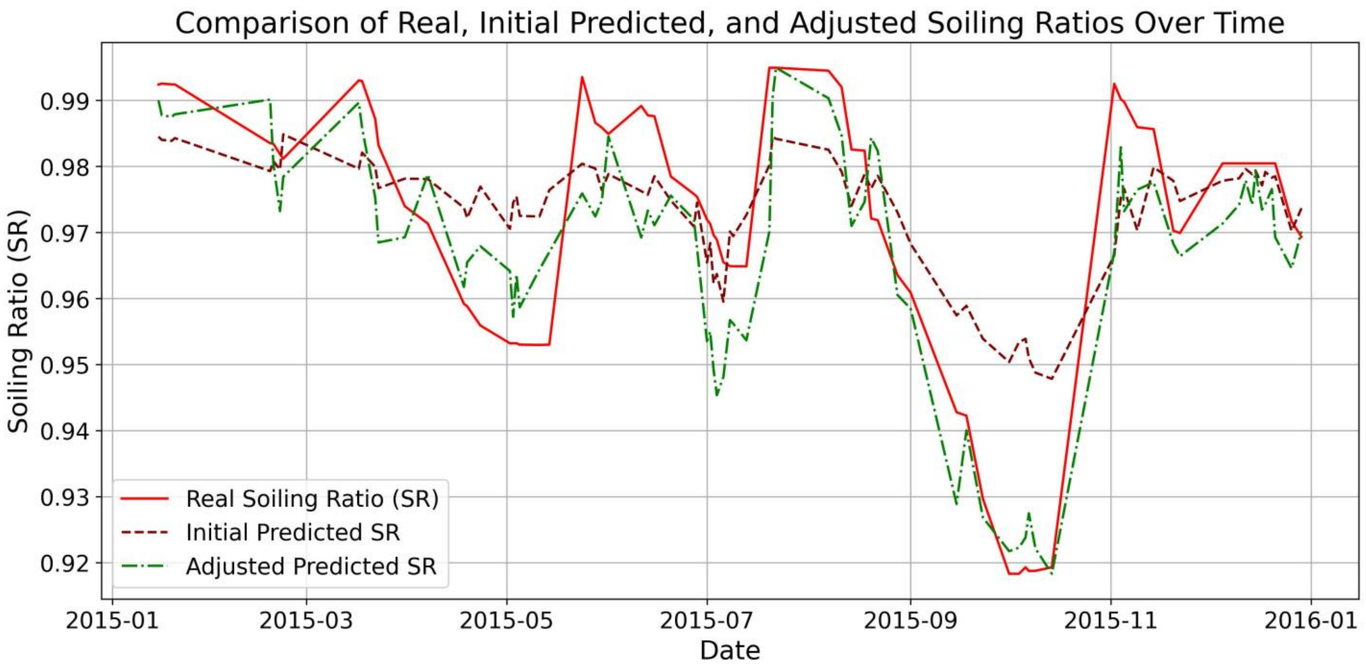
- **Model Architecture:**
  - Input units: 145
  - Input dropout: 0.3
  - Number of layers: 1
  - Layer 2 units: 116
  - Layer 2 dropout: 0.6
  - Optimizer: SGD
  - Learning rate: 0.0071



Results - Reconstruction

Fine tuning: transformation and more focused hyperparameter tuning

Model	RMSE	MAE
Javed et al.'s Study		
MLR	0.0904	0.684
ANN-10	0.0617	0.520
This Work		
Proposed Model	**0.0102**	**0.0083**



Conclusions

Summary of finds and future work.

- **Study Overview:**
  - Objective: Reconstruct soiling profiles using ML.
  - Data used: Extensive atmospheric and time data.
- **Key Findings:**
  - Coupled variables significantly impact on SRates.
  - Relative Humidity crucial for soil adhesion on PV surfaces.
- **The Idea:**
  - Phase 1: Data analysis.
  - Phase 2: SR reconstruction.
- **Results:**
  - Improved accuracy over previous studies.
- **Future directions:**
  - Collect more diverse datasets and on-site data.
  - Develop hybrid approaches and refine deep learning models.
  - Integrate economic models for cost-effective cleaning.
  - Localize models for specific climatic regions.

Thank You for the Kind Attention!

Any Question?