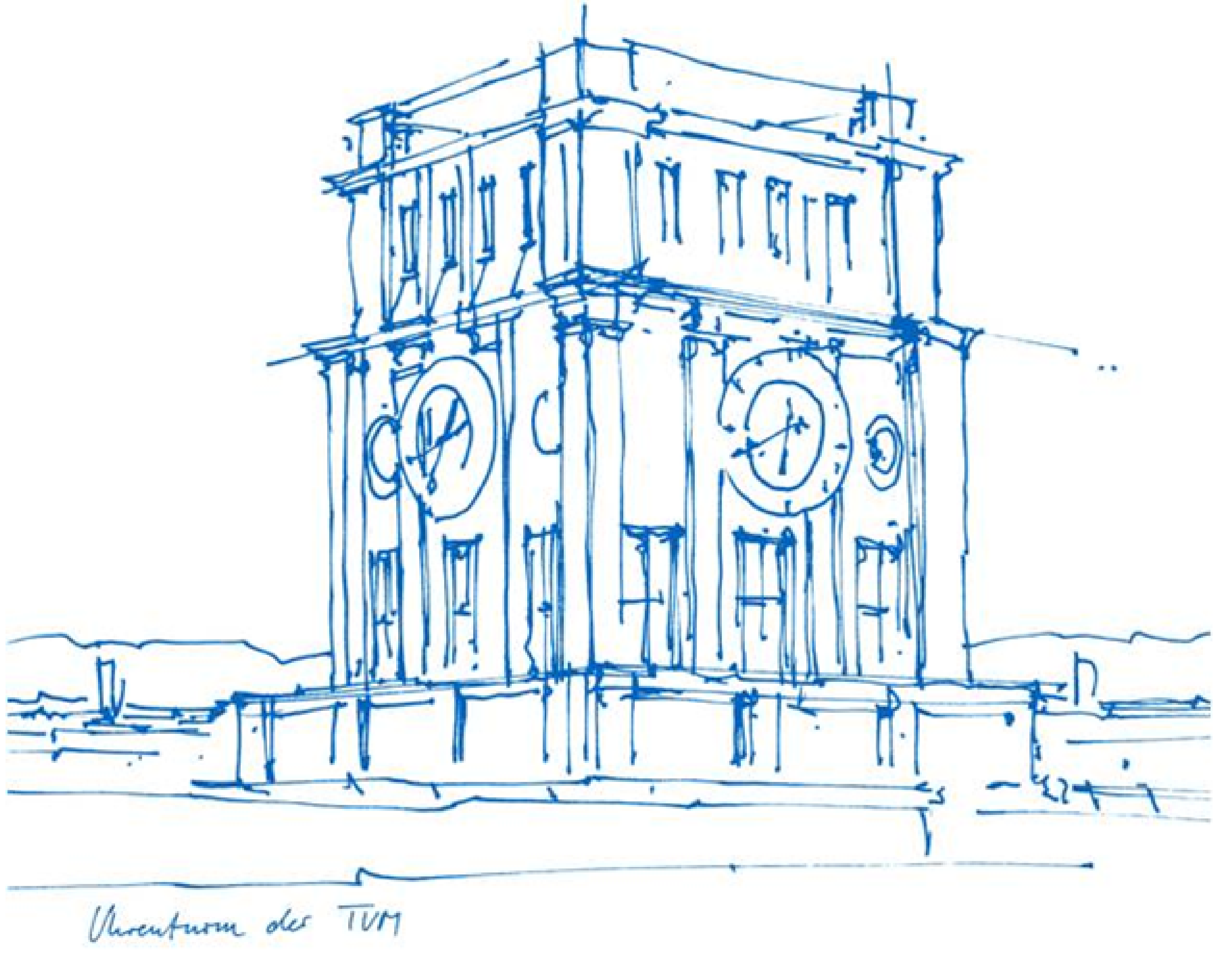
Enhancing the Degradation Patterns Using Advanced Machine Learning Techniques

Presented by: Aziz Ben Ali

Referents: Alessio Gagliardi
Ioannis Kouroudis

Technical University of Munich
TUM School of Computation Information and Technology



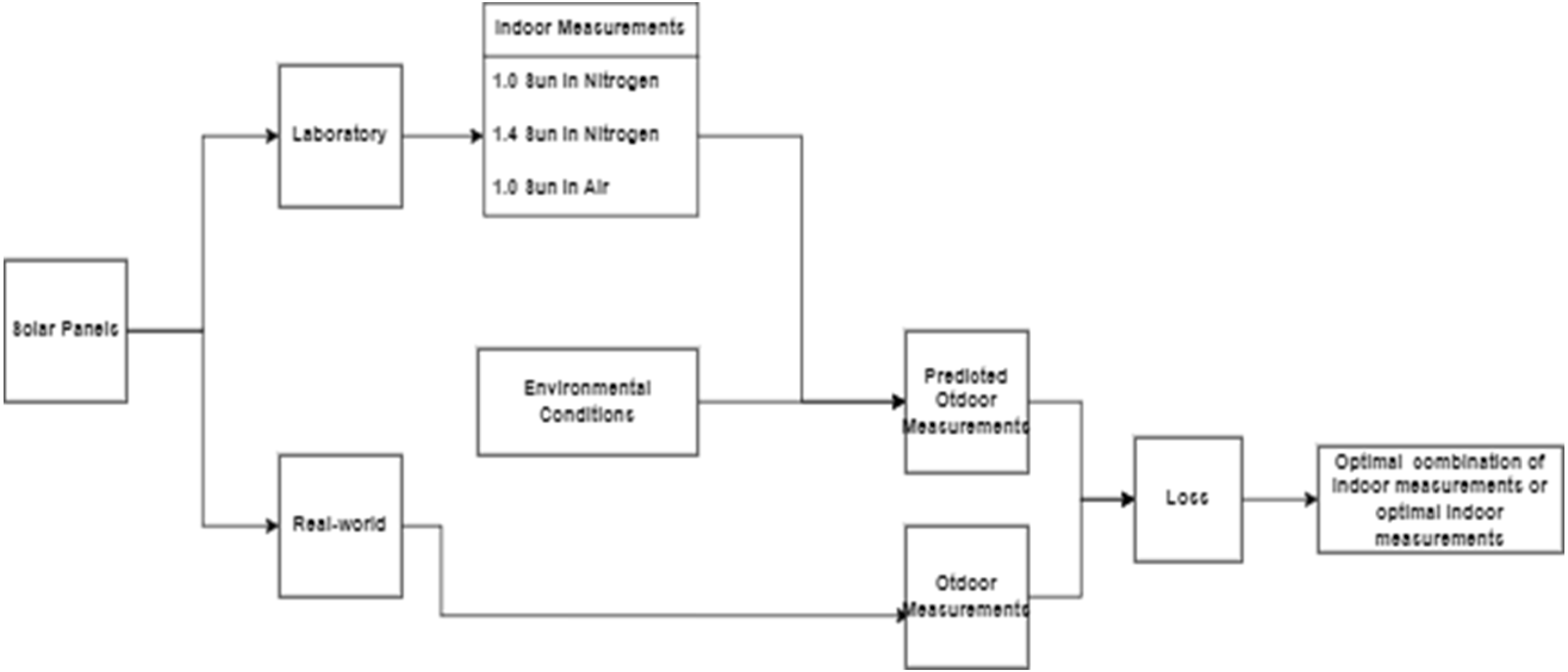
Munich, October 10th, 2023

Motivation

- Measuring the output power of solar panels in real-world conditions can be costly and time-consuming.

- Understanding the mechanisms behind the degradation patterns of solar panels is a highly complex task

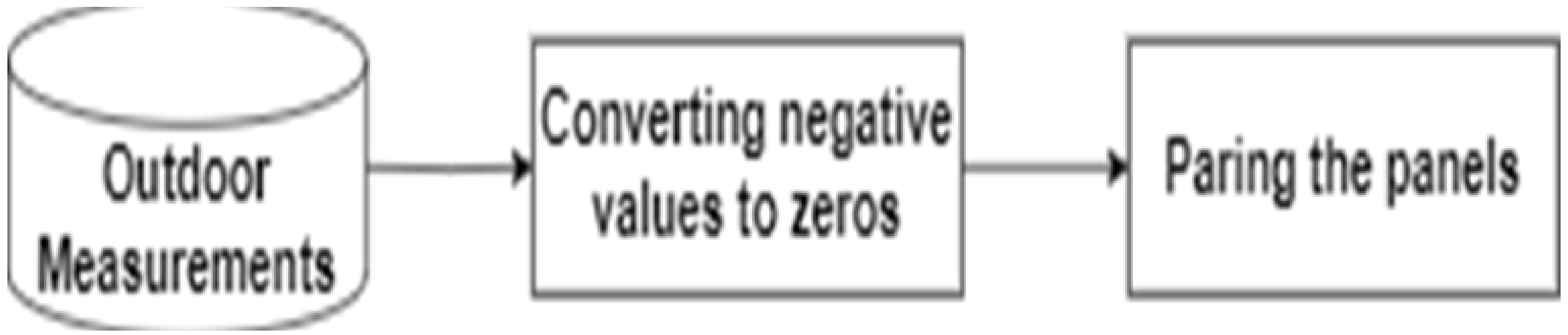
Approach

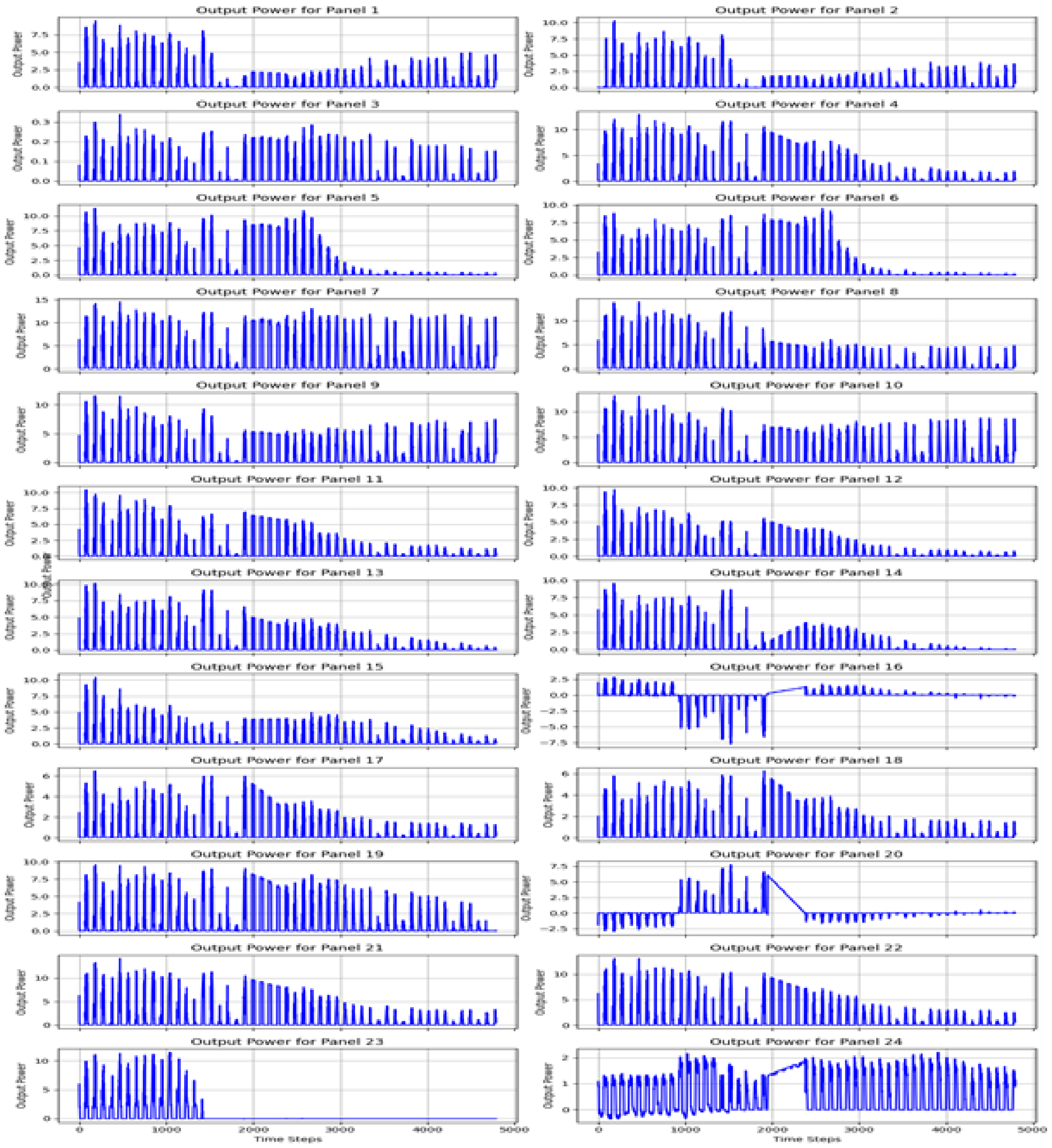


Solar Panels
Laboratory
Indoor Measurements
1.0 Sun in Nitrogen
1.4 Sun in Nitrogen
1.0 Sun in Air
Environmental Conditions
Real-world
Predicted Otdoor Measurements
Otdoor Measurements
Loss
Optimal combination of indoor measurements or optimal indoor measurements

Data Cleaning and Processing

1) Outdoor Measurements

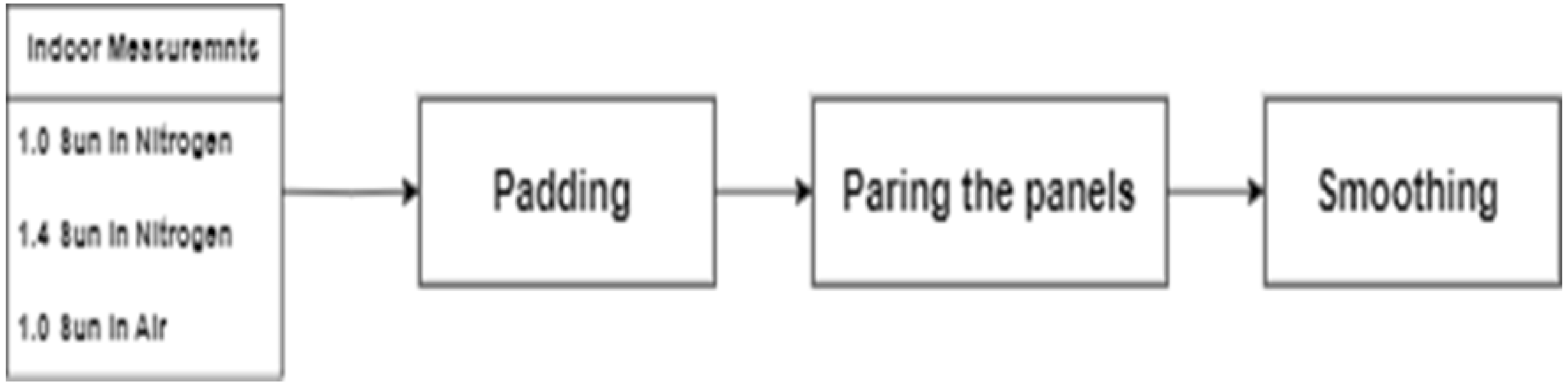


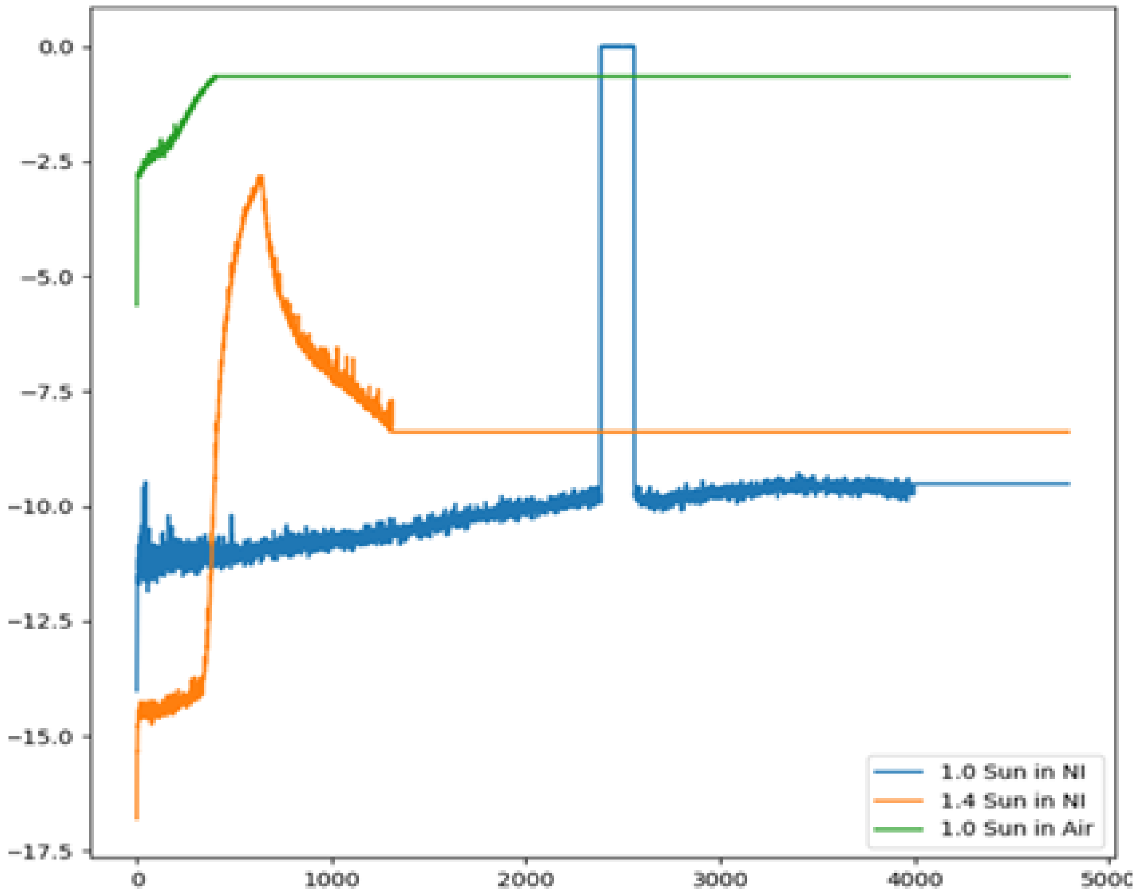


Actual Outdoor Measurements

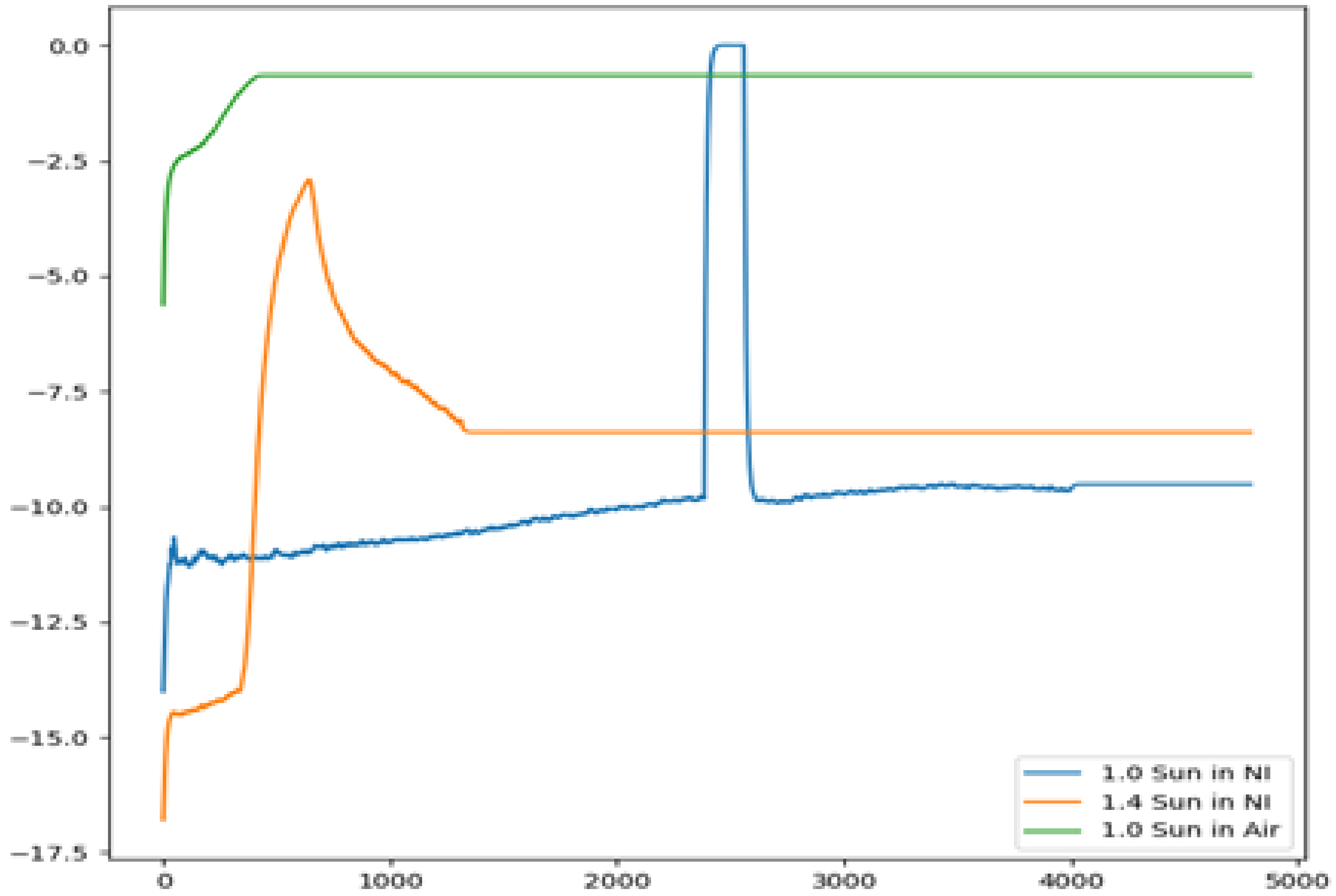
Data Cleaning and Processing

1) Indoor Measurements



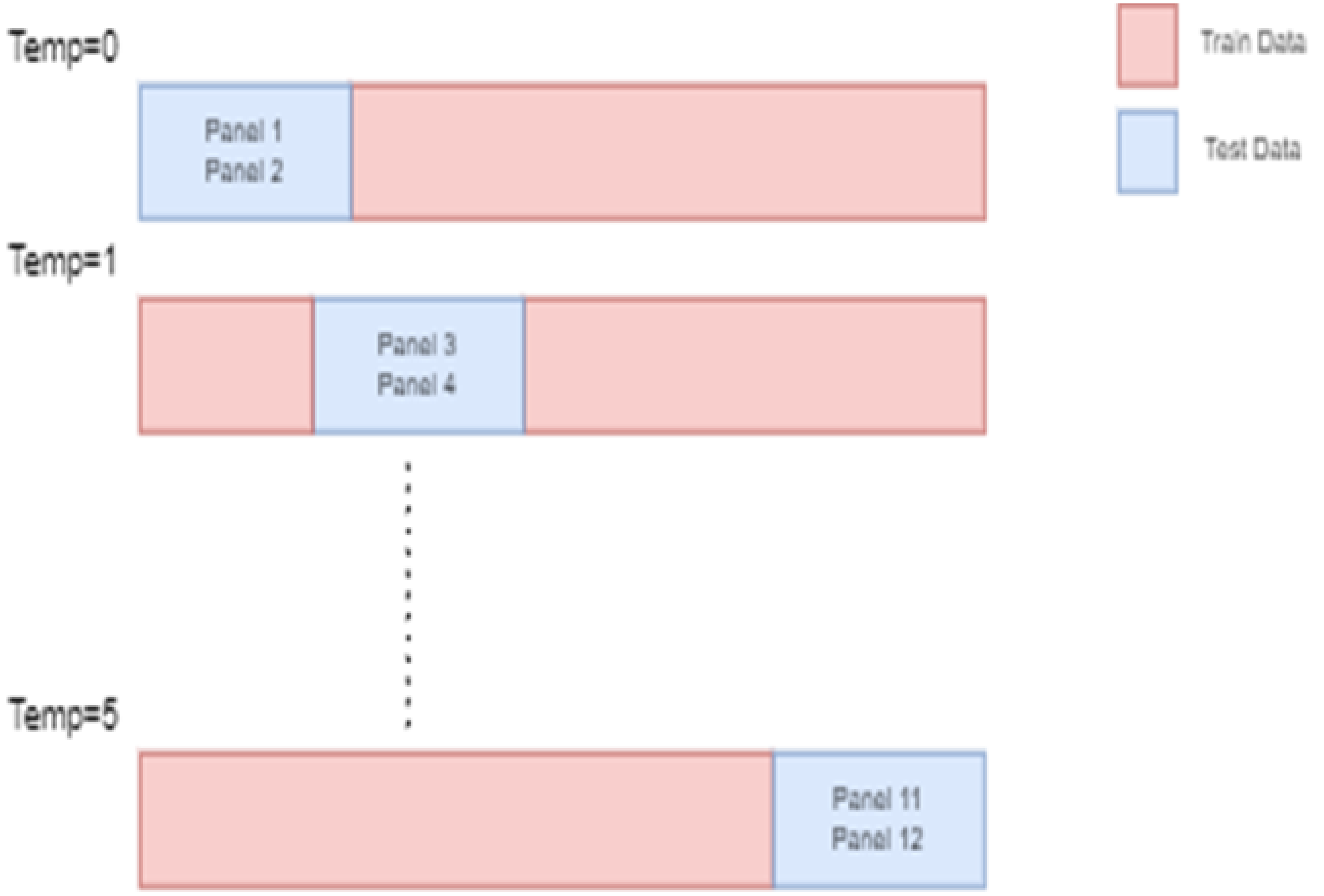


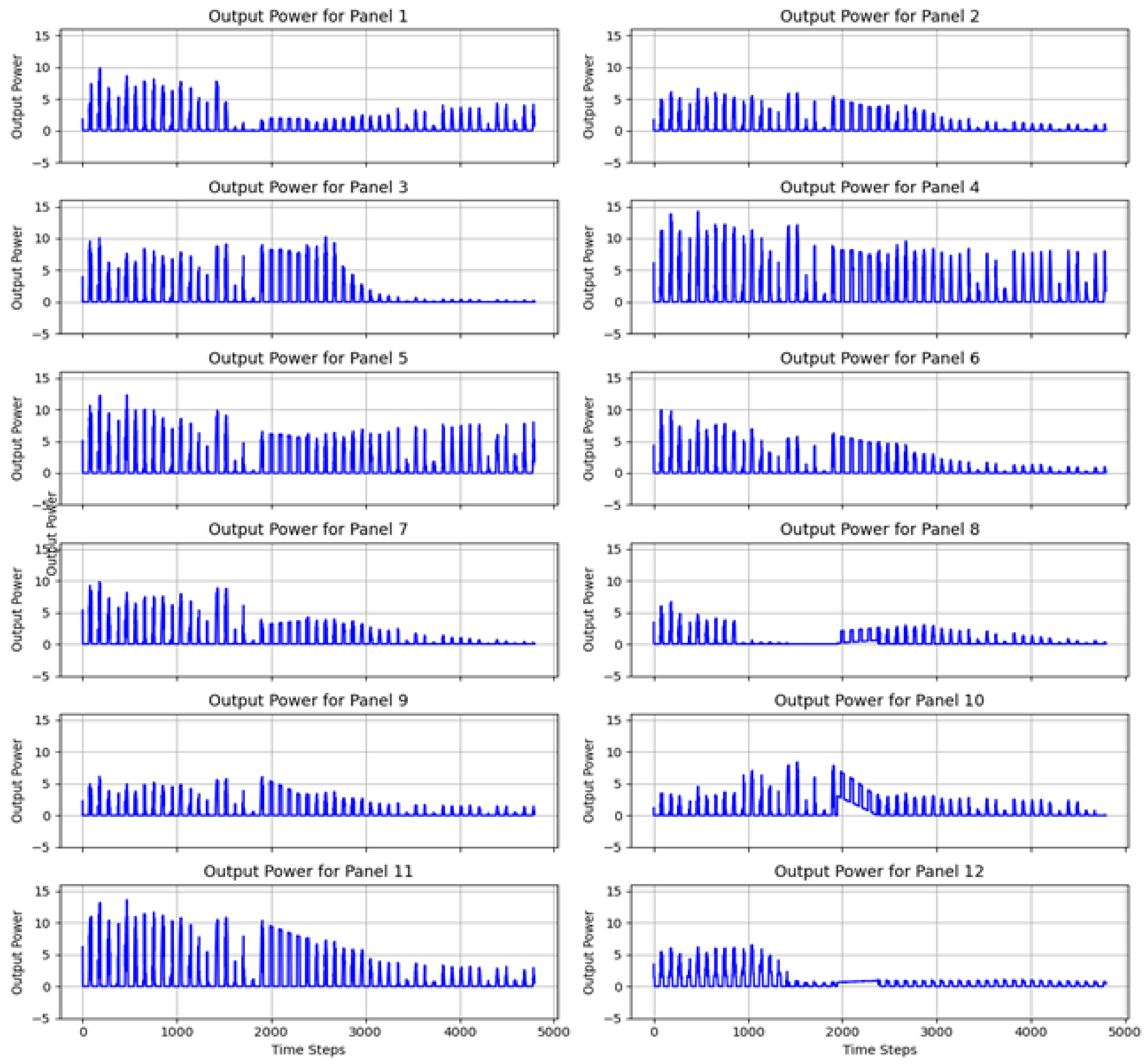
Unsmoothed Indoor Measurements



Smoothed Indoor Measurements

Training and Testing





Cleaned Outdoor Measurements

Machine Learning Models
Multi-Layer Bidirectional LSTM-based Complex Neural Network

This Model involves three main components:

- Fully connected Layers (FC)
- Highway Layers (HL)
- Bidirectional LSTM Layers (BILSTML)

Machine Learning Models

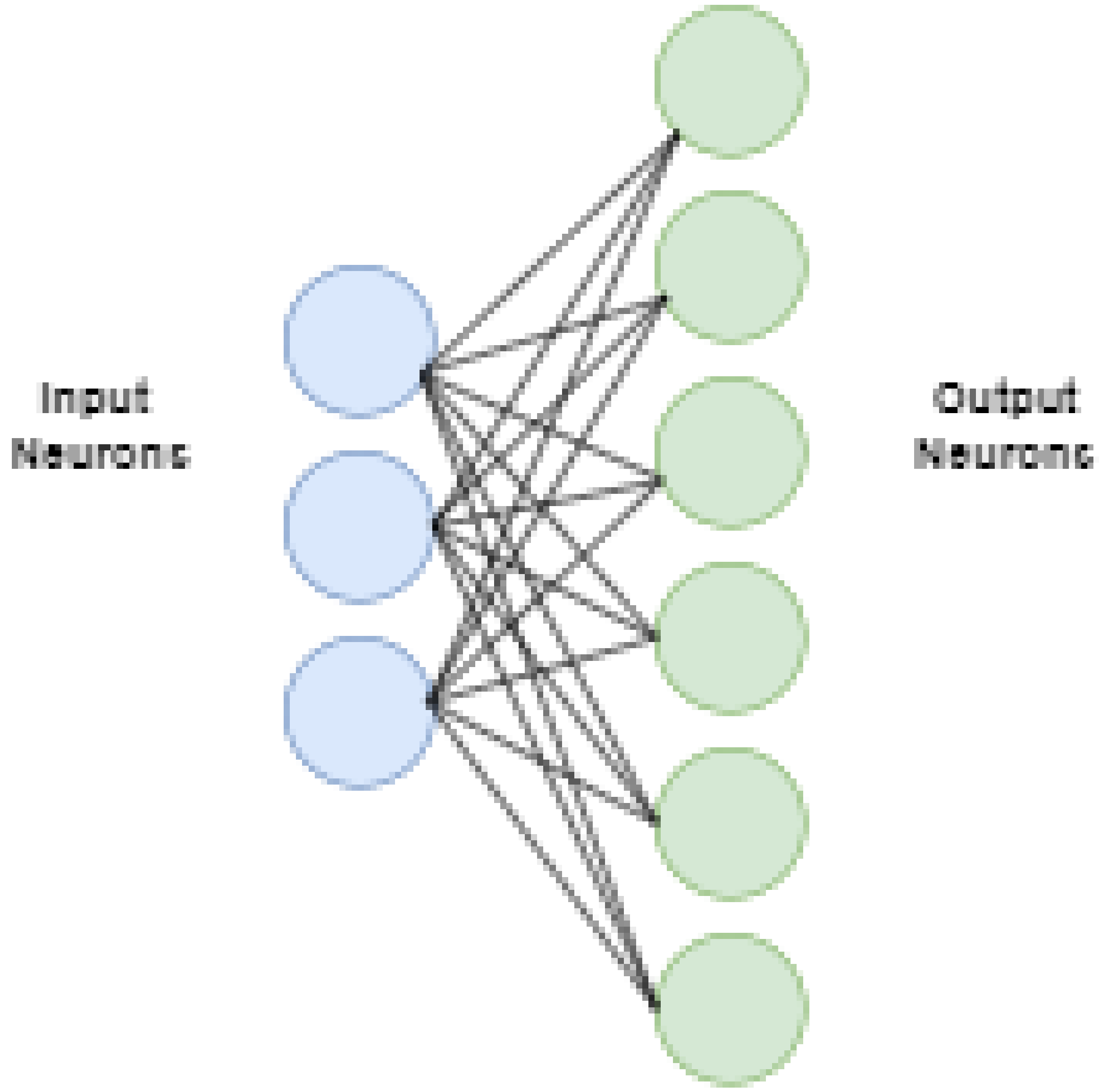
Multi-Layer Bidirectional LSTM-based Complex Neural Network

1) Exploring Fully connected Layers

Each input node is connected to each output node.

During the training, FC experiences two main phases:

- **Forward pass**
- **Backward pass**



Fully Connected Layers

Machine Learning Models

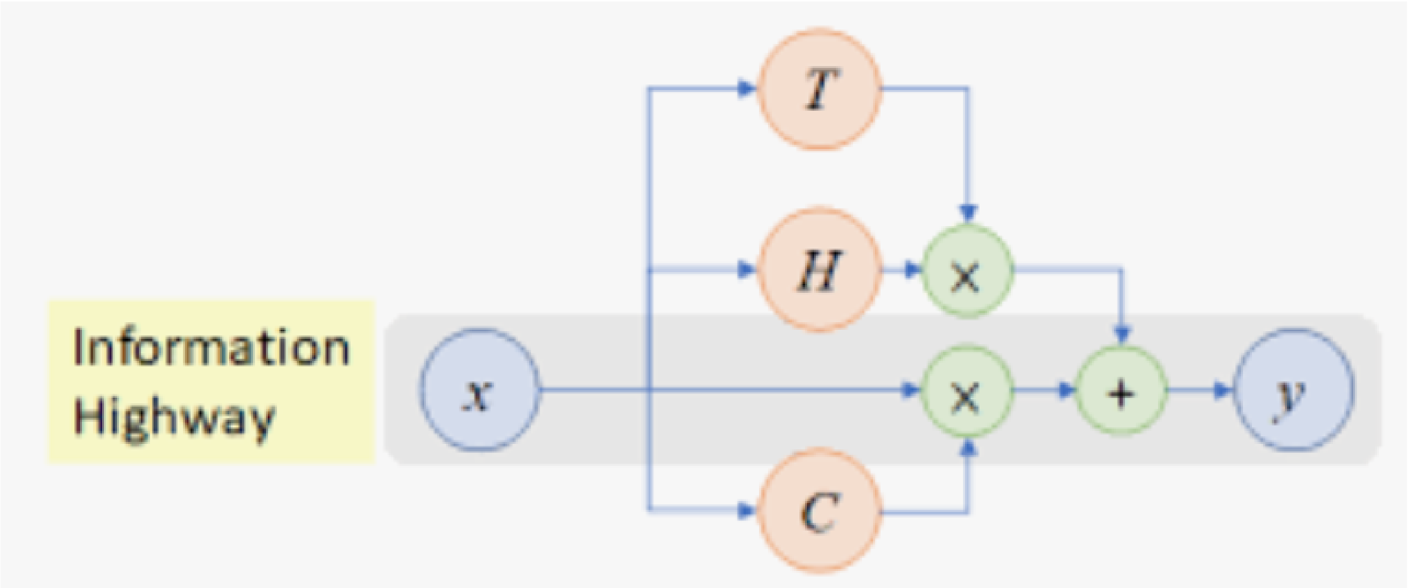
Multi-Layer Bidirectional LSTM-based Complex Neural Network

2) Exploring Highway Layers

A type of neural network.

The output is controlled by a gating mechanism.

→ allow the network to weather pass the input as it is or apply transformations.



Highway Layer -1-

This process involves three main components:

- Transformation Gate
- Nonlinear Transformation Function
- Carry Gate

-1- https://paperswithcode.com/method/highway-layer

Machine Learning Models

Multi-Layer Bidirectional LSTM-based Complex Neural Network

3) Long Short-Term Memory (LSTM) and Bidirectional LSTM Networks

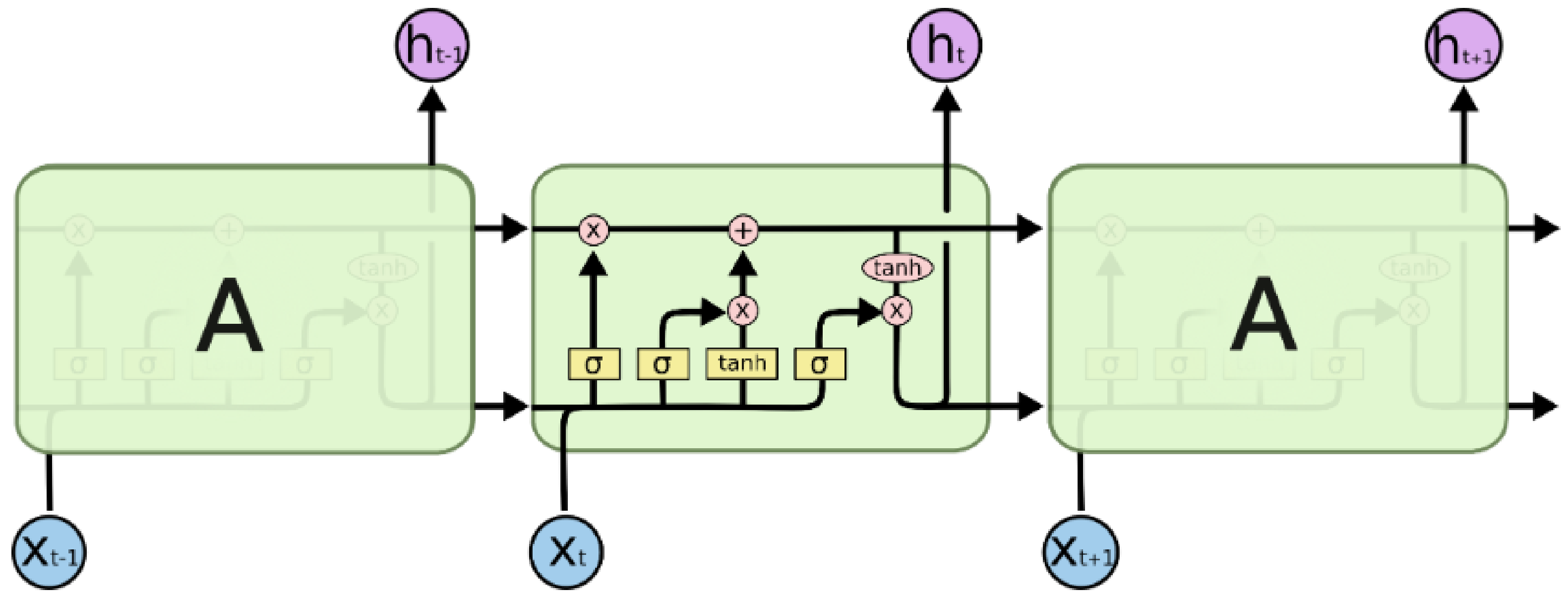
a) LSTM Layers

A type of recurrent neural networks (RNNs).

Capable of learning long-term dependencies.

Involves 3 main components:

- Forget Gate
- Input Gate
- Output Gate



LSTM layer -2-

-2- https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Machine Learning Models

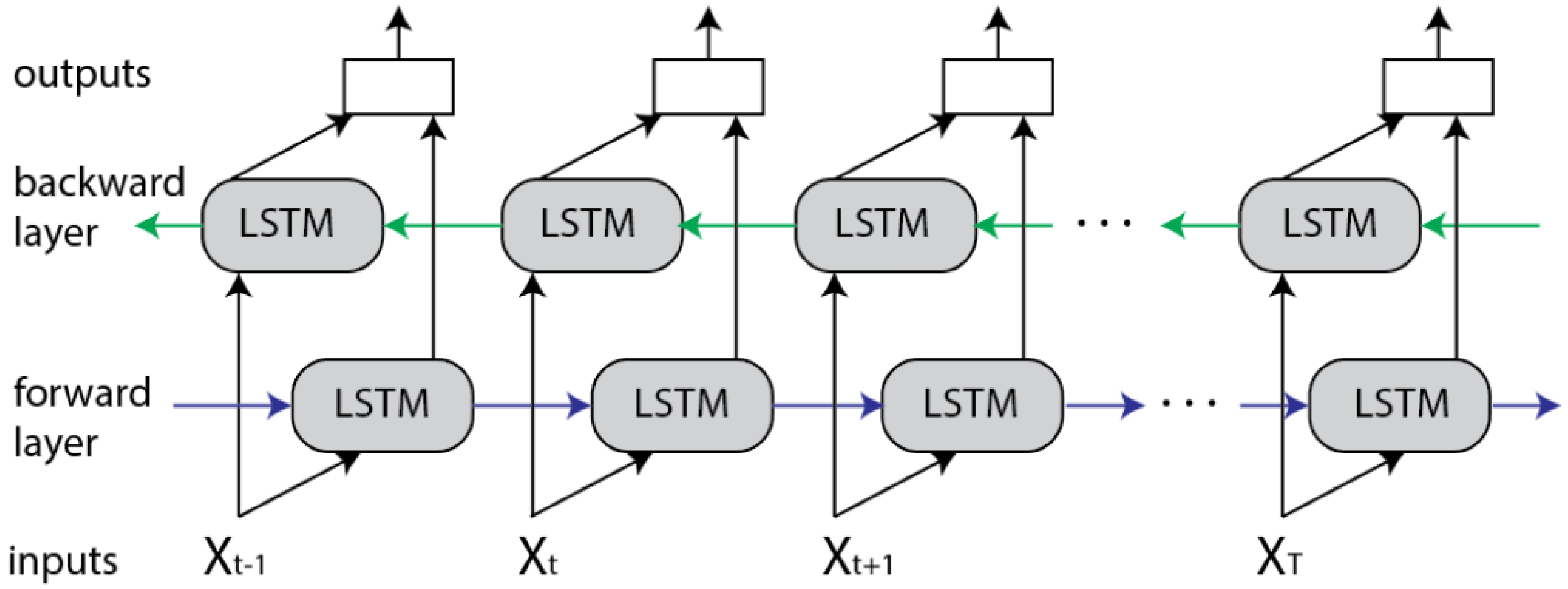
Multi-Layer Bidirectional LSTM-based Complex Neural Network

3) Long Short-Term Memory (LSTM) and Bidirectional LSTM Networks

b) BILSTM Layers

Same machanism as LSTM.

Able of capturing relevant information and relationships within the data in forward and backward directions.



BILSTM layer -3-

- 3- https://www.baeldung.com/cs/bidirectional-vs-unidirectional-lstm

Machine Learning Models

Multi-Layer Bidirectional LSTM-based Complex Neural Network

4) Model Architecture

- 4 BILSTML organized in a stack

  → capture of complex sequential pattern

  Problem: Exploding Gradient => Gradient Clipping

- Initialization of BILSTM Layers: The use of Kaiming–Normal Initialization

  → effective training convergence and less sensitivity to hyperparameters

  For each BILSTM Layer, a corresponding Highway layer is applied.

  → enhances information flow throughout the network.

  → mitigates the problem of Vanishing Gradient.

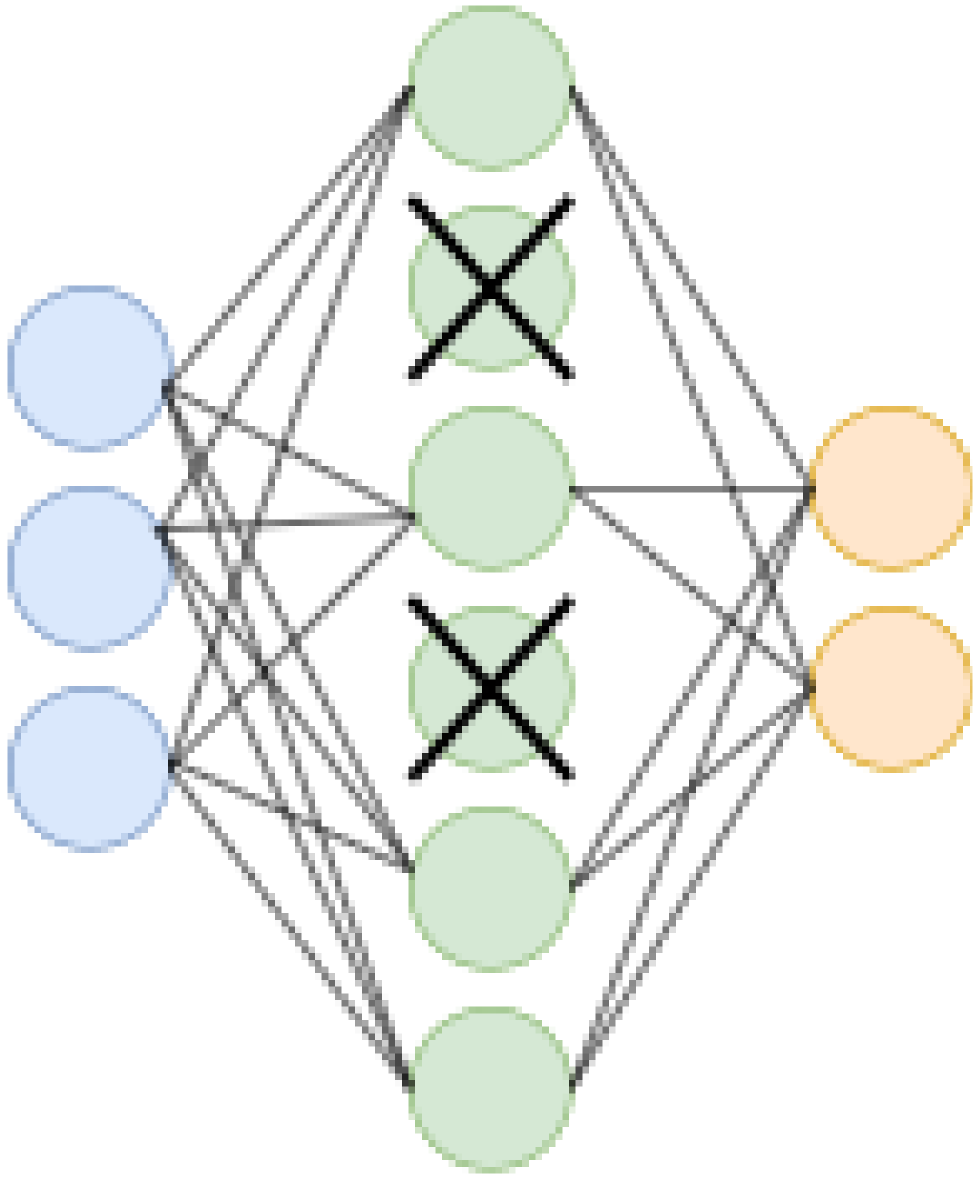
Machine Learning Models

Multi-Layer Bidirectional LSTM-based Complex Neural Network

4) Model Architecture

3 FCs → reduces the dimensionality while extracting the relevant features.

We employ Dropout technique → mitigates overfitting problem

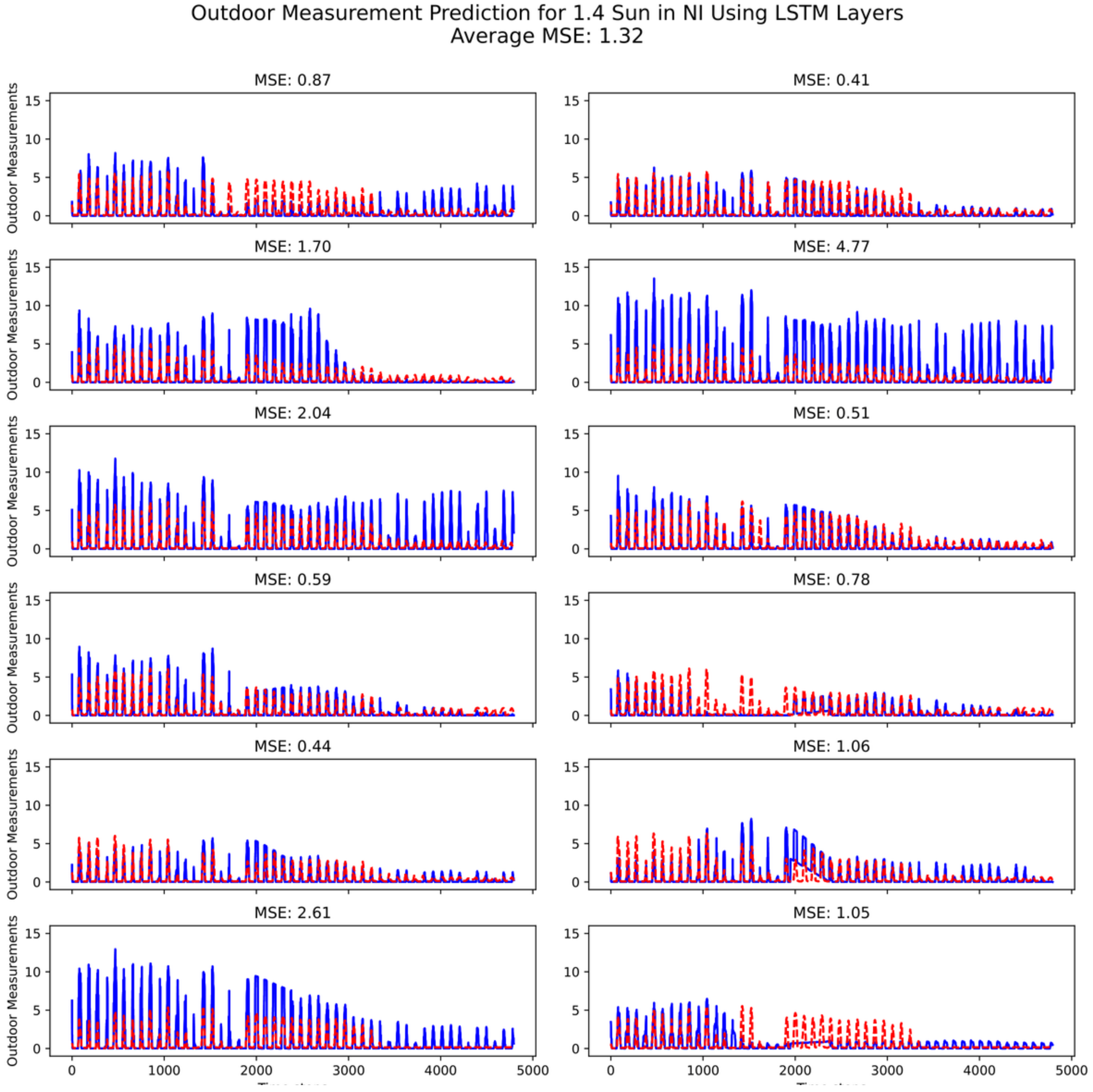


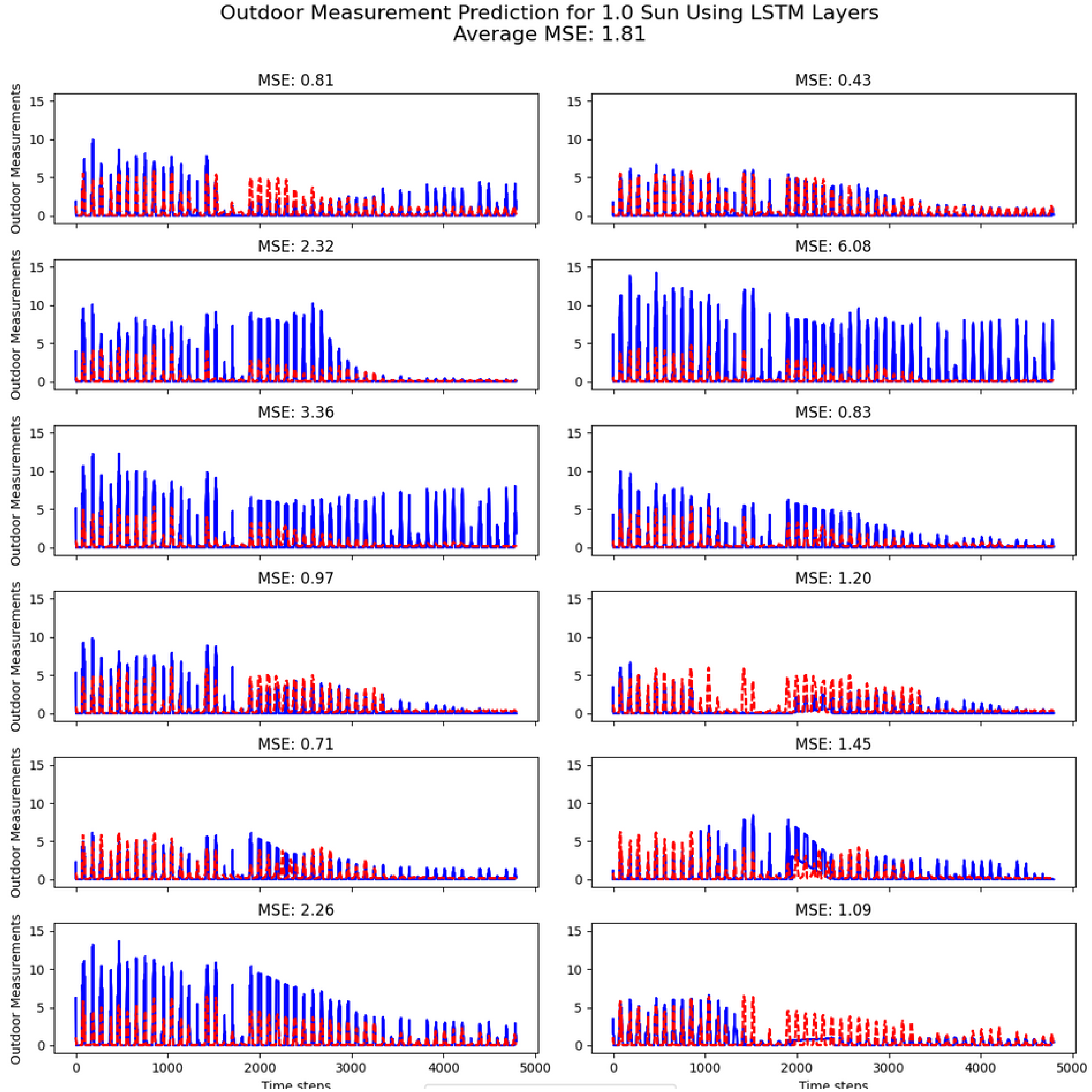
Dropout technique

Machine Learning Models

Multi-Layer Bidirectional LSTM-based Complex Neural Network

5) Results



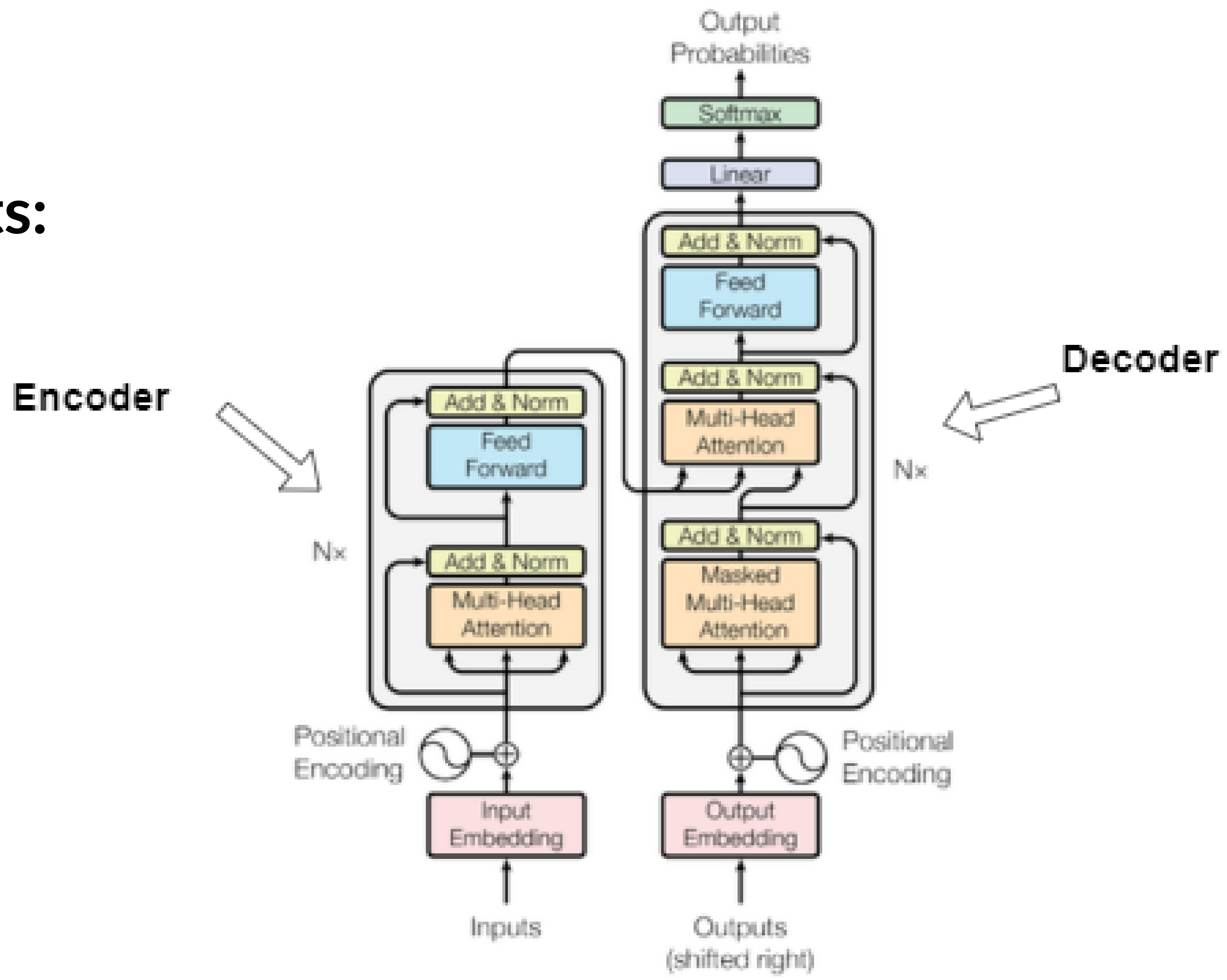


Machine Learning Models
Time Series Transformer (TST)

The architecture of TST involves two main components:

Encoder:
- Positional Encodings
- Multi-Head Self-Attention:
- Position-wise Feedforward Network
- Residual Connections and Layer Normalization

Decoder:
- Positional Encodings
- Masked Multi-Head Self-Attention
- Position-wise Feedforward Network
- Residual Connections and Layer Normalization
- Output Layer



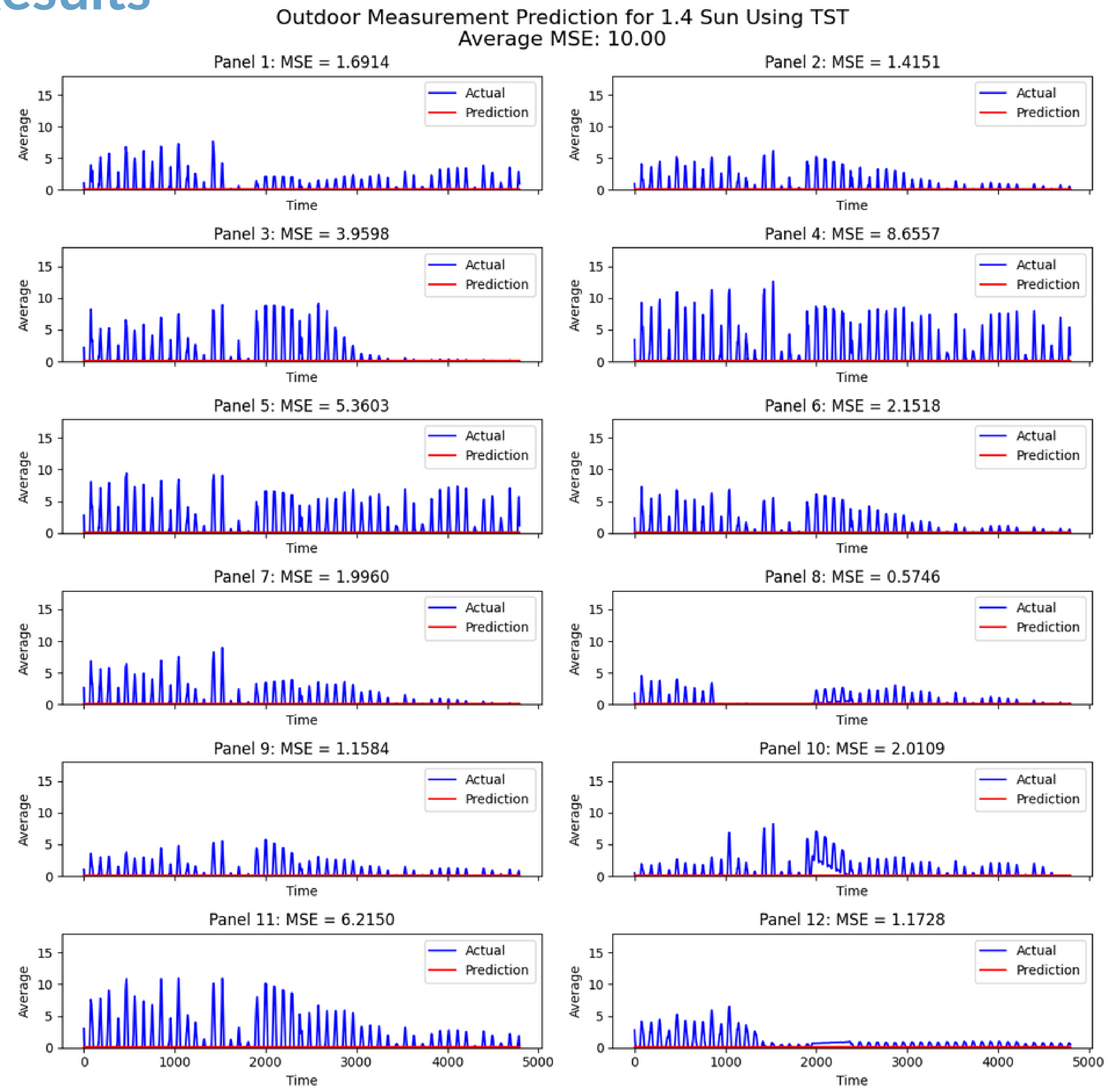
Time Series Transformer -4-

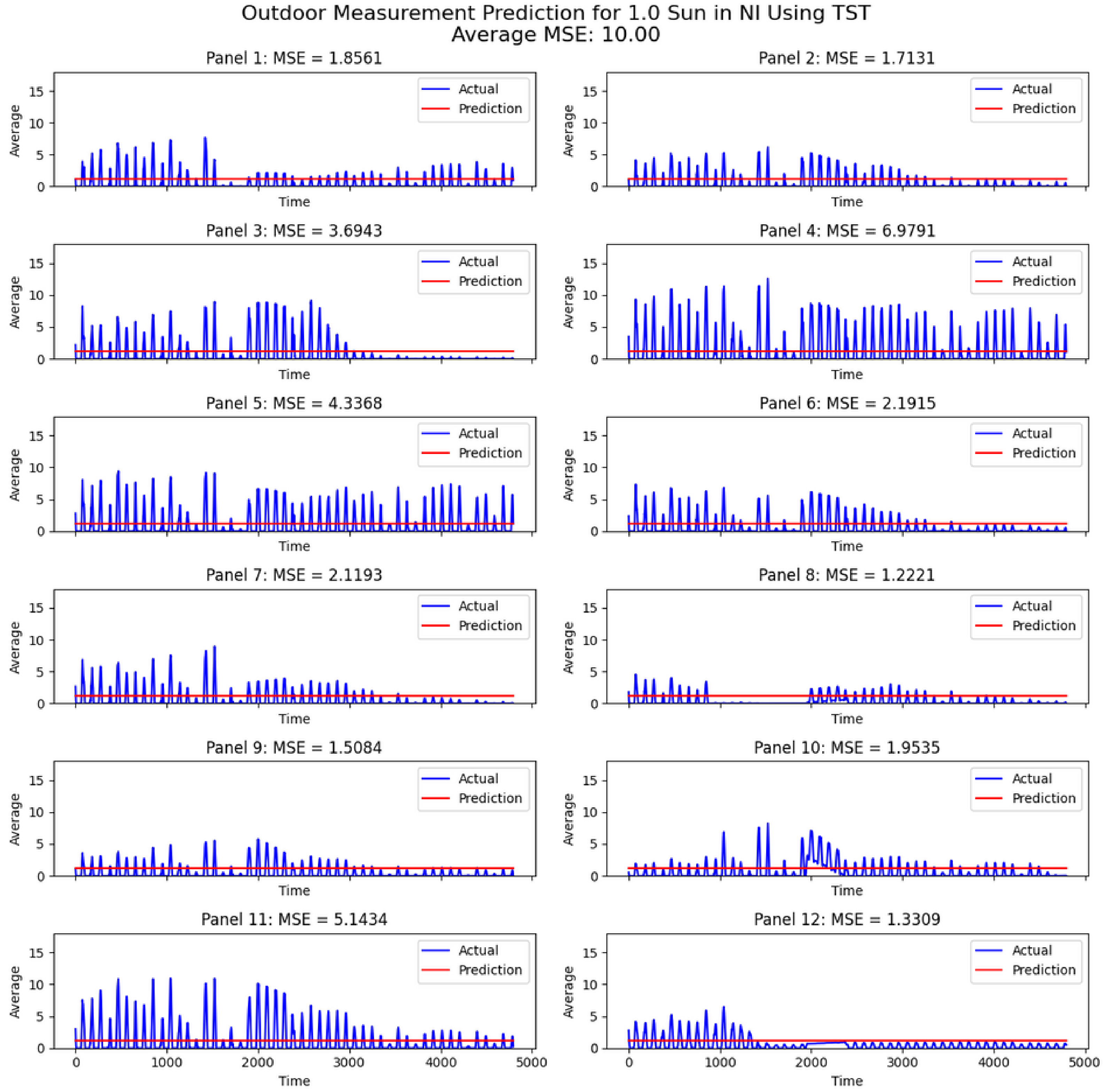
-4- https://machinelearningmastery.com/the-transformer-model/

Machine Learning Models

Time Series Transformer (TST)

5) Results

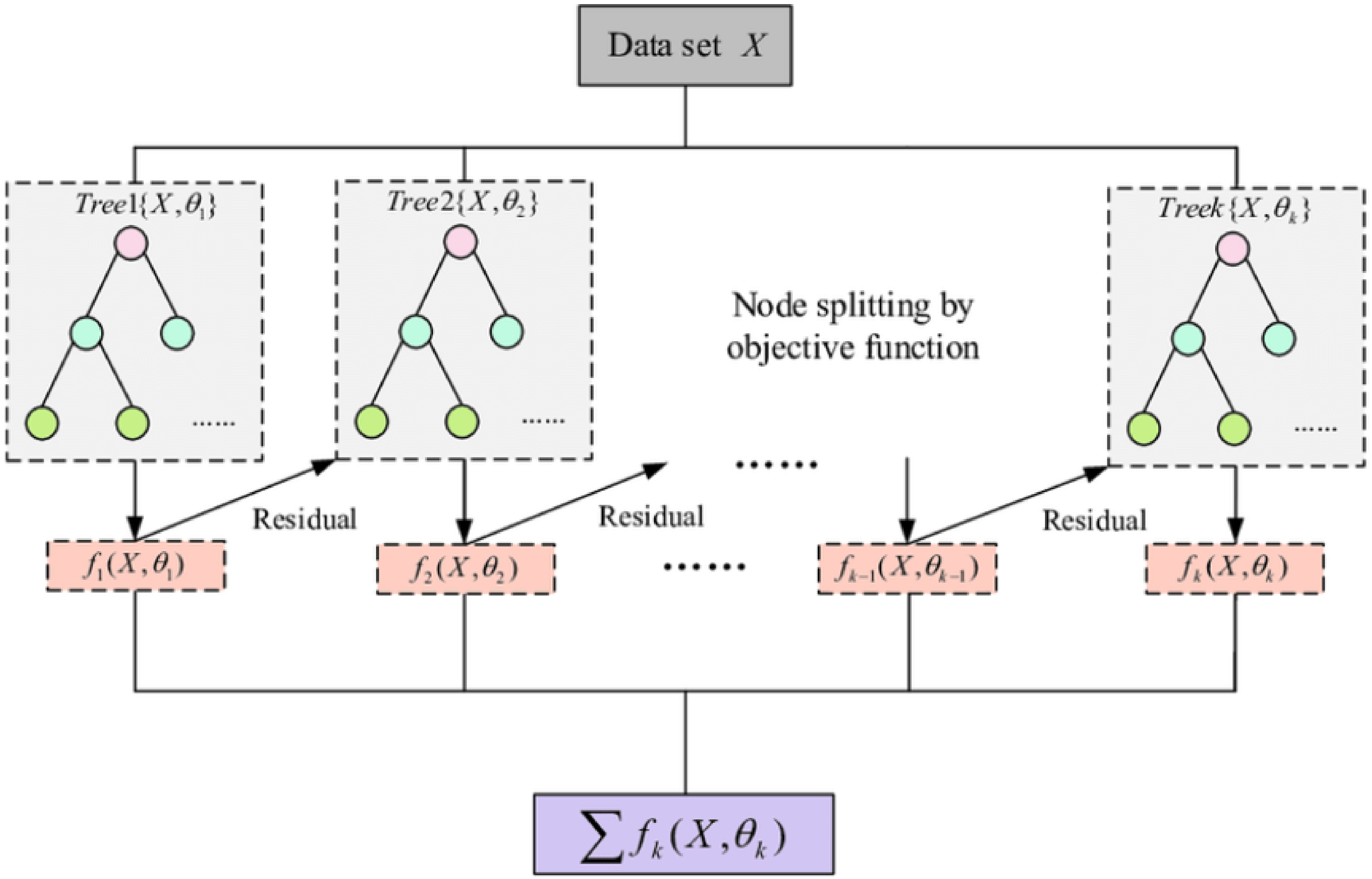




Machine Learning Models

XGBoost

- Belongs to the category of ensemble learning.
- Employs gradient boosting.
- Uses decision trees as its weak learners.
- Each new tree is trained to minimize the error made by the previous trees.



XGBoost

https://www.researchgate.net/figure/Flow-chart-of-XGBoost_fig3_345327934

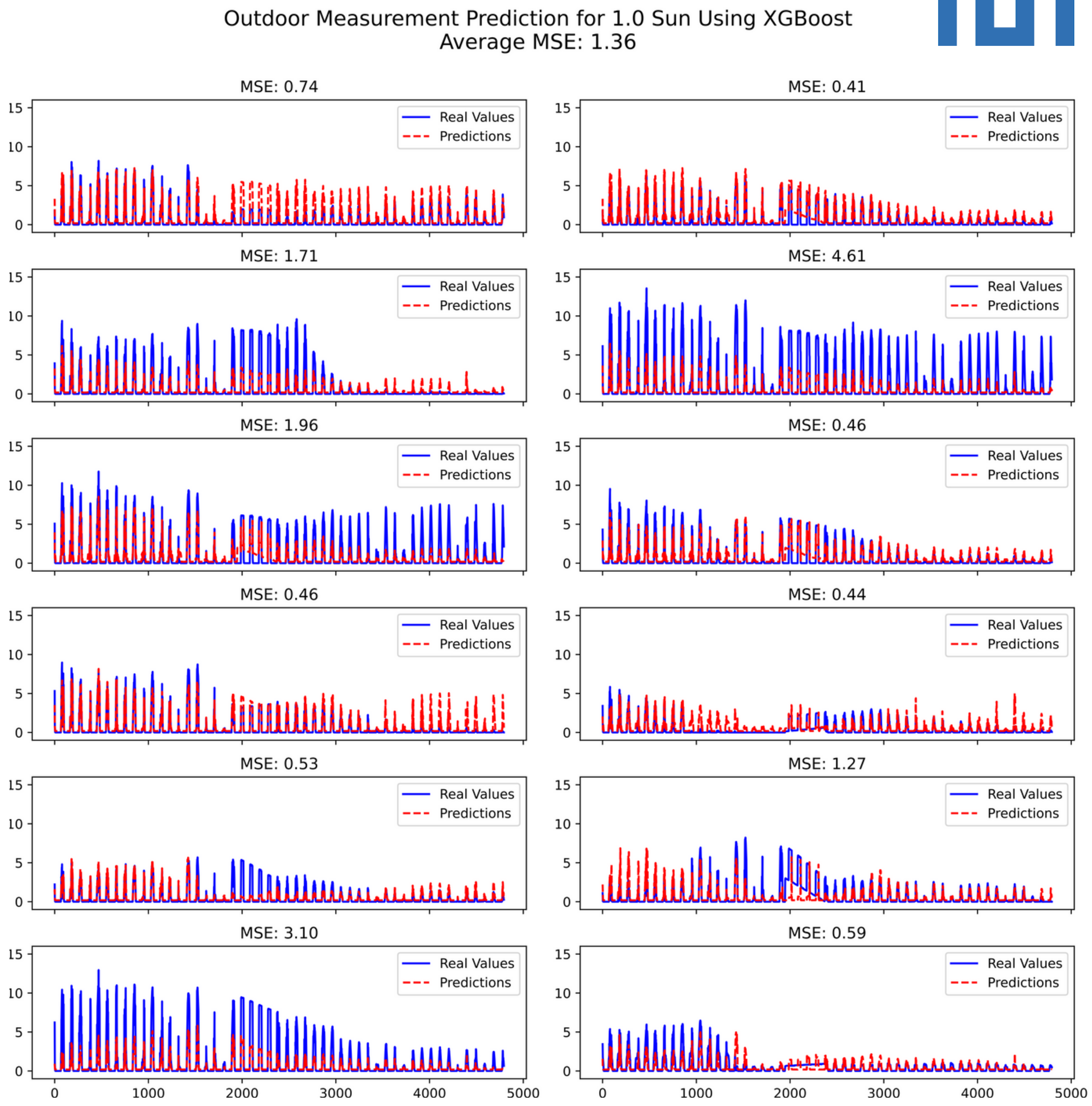
Machine Learning Models

XGBoost

5) Results

Degradation Accelerators	Mean Squared Error (MSE)
1.0 Sun in NI	1.36
1.4 Sun in NI	1.39
1.0 Sun in Air	1.49
1.0 Sun in NI + 1.4 Sun in NI	1.31
1.0 Sun in Air + 1.4 Sun in NI	1.34
1.0 Sun in Air + 1.0 Sun in NI	1.32
1.0 Sun in NI +1.4 Sun in NI +1.0 Sun in Air	1.37

Total Error Average	1.37



Machine Learning Models

Gaussian Process Regressor (GPR)

GPR is a non-parametric, probabilistic, and bayesian approach to regression, characterized by two main concepts:

- **The prior over function:** $f(x) \sim GP(m(x), k(x, x'))$
- **Likelihood:** $p(y|f(x) \sim N(f(x), \sigma^2 I)$

By applying Bayes' theorem, GPR combines the prior and likelihood to calculate the posterior distribution over functions p(f (x)|y)

$$p(f(x)|y) = (p(y|f(x) * p(f(x)) \div p(y)$$

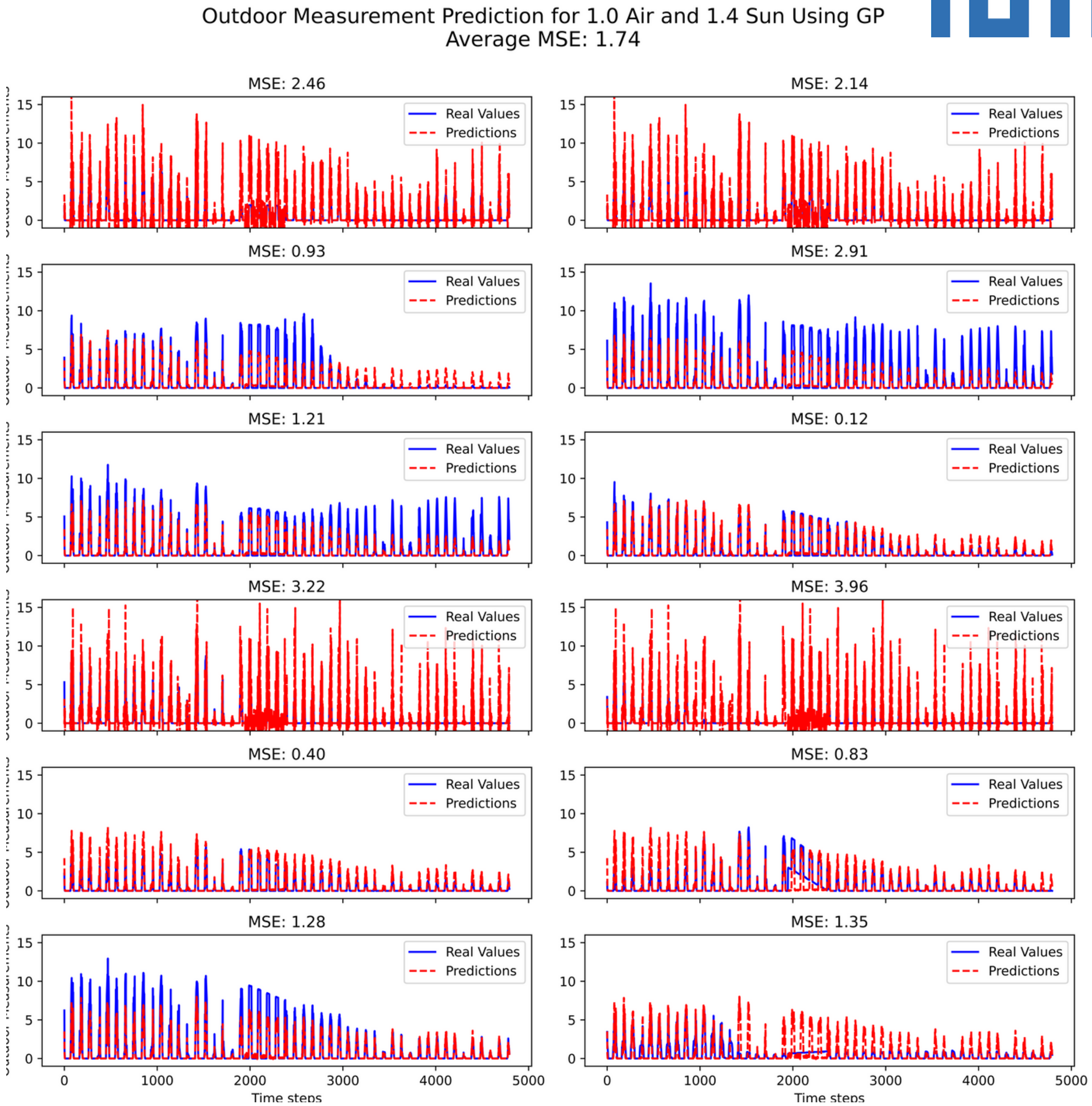
Machine Learning Models

Gaussian Process Regressor (GPR)

5) Results

Degradation Accelerators	Mean Squared Error (MSE)
1.0 Sun in NI	1.42
1.4 Sun in NI	2.57
1.0 Sun in Air	1.15
1.0 Sun in NI + 1.4 Sun in NI	2.10
1.0 Sun in Air + 1.4 Sun in NI	1.74
1.0 Sun in Air + 1.0 Sun in NI	1.47
1.0 Sun in NI +1.4 Sun in NI +1.0 Sun in Air	1.04

Total Error Average	1.65



Machine Learning Models

Gaussian Process Regressor (GPR)

Optimization

Discrete Fourier Transformation (DFT)

Decomposes the discrete time series into its constituent frequencies in the frequency domain.

$$X[K] = \sum_{n=0}^{N-1} x[n] e^{-j*2\Pi \frac{kn}{N}}$$

=> We add the magnitude in decibels of the transformed output as an independent feature.

$$DBMAG = 20 * log(MAG)$$

Machine Learning Models

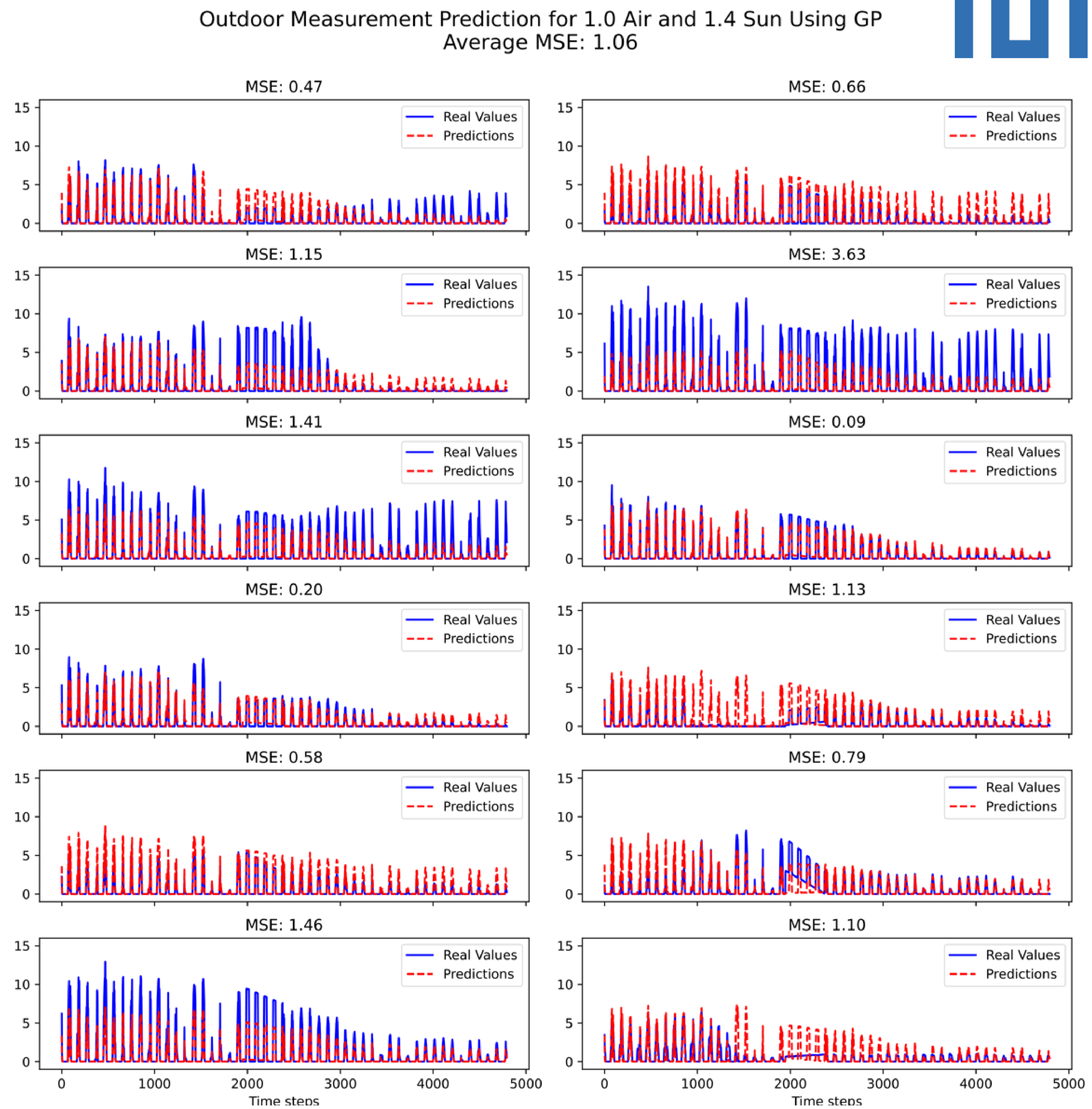
Gaussian Process Regressor (GPR)

Discrete Fourier Transformation (DFT)

Results

Degradation Accelerators	Mean Squared Error (MSE)
1.0 Sun in NI	1.06
1.4 Sun in NI	1.00
1.0 Sun in Air	1.06
1.0 Sun in NI + 1.4 Sun in NI	0.94
1.0 Sun in Air + 1.4 Sun in NI	1.06
1.0 Sun in Air + 1.0 Sun in NI	1.05
1.0 Sun in NI +1.4 Sun in NI +1.0 Sun in Air	1.04

Total Error Average	1.03



Machine Learning Models

Kernel Ridge Regression (KRR)

KRR is a powerful machine learning model designed for regression tasks.

Unlike traditional methods that map data to high-dimensional feature spaces (computationally expensive), KRR uses the "kernel trick”.

→ allows KRR to work with inner products of mapped data points directly using a kernel function

The central idea is to find optimal coefficients that minimize a dual objective function:

$$\alpha^* = \arg\min_{\alpha} \left(\frac{1}{2} \alpha^T k \alpha - \alpha^T y + N\lambda \alpha^T k \alpha \right)$$

by applying the kernel function and combining the results with the learned α coefficient, KRR can predict the target values of new inputs.

Machine Learning Models

Kernel Ridge Regression (KRR)

Results

Degradation Accelerators	**Mean Squared Error (MSE)**
1.0 Sun in NI	1.35
1.4 Sun in NI	1.11
1.0 Sun in Air	1.30
1.0 Sun in NI + 1.4 Sun in NI	1.34
1.0 Sun in Air + 1.4 Sun in NI	1.26
1.0 Sun in Air + 1.0 Sun in NI	1.39
1.0 Sun in NI +1.4 Sun in NI +1.0 Sun in Air	1.34

Total Error Average	1.29

Machine Learning Models
Kernel Ridge Regression (KRR)

Optimization

Grid Search
Widely recognized hyperparameter tuning technique

Works as following:

- Define the Search Space (Grid)
- Use Cross-Validation
- Select the Best Model
- Train and evaluate the best Model

Total Error Average: 1.01

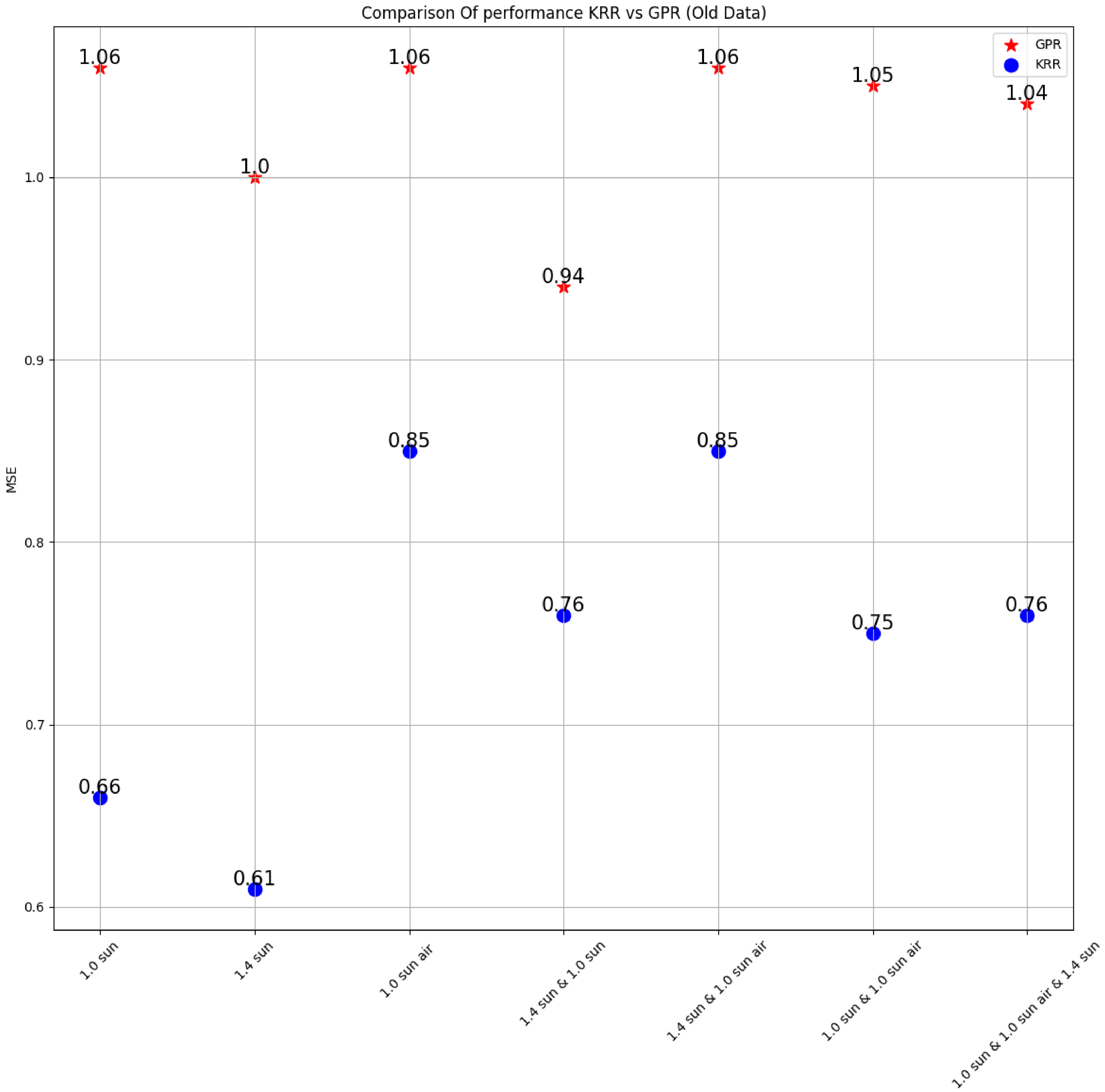
Bayesian Optimization
More powerful hyperparameter technique than Grid Search.

Involves three key concepts:

- Objective Function
- Surrogate Function (Gaussian Process)
- Acquisition Function

Total Error Average: 0.74

Discussion



Comparison Of performance KRR vs GPR (Old Data)
GPR
KRR
MSE
1.06
1.0
1.06
0.94
1.06
1.05
1.04
0.66
0.61
0.85
0.76
0.85
0.75
0.76
1.0
0.9
0.8
0.7
0.6
1.0 sun
1.4 sun
1.0 sun air
1.4 sun & 1.0 sun
1.4 sun & 1.0 sun air
1.0 sun & 1.0 sun air
1.0 sun & 1.0 sun air & 1.4 sun

Conclusion

Discussion

Wavelet

powerful method for analyzing time-series data across different scales or frequencies
=> No need to the use of padding.

Decomposition into Components:

Approximation Coefficients (Aj):

These coefficients represent the low-frequency components of the signal at scale j.

Detail Coefficients (Dj):

These coefficients represent the high-frequency components of the signal at scale j.

Extract and use the detail coefficients as features for indoor measurements.
Total error average: 1,11 > 1,03