Automatization of the features selection process to improve temperature-based solar radiation models in Southern Spain

Bellido-Jiménez, J.A.; Estévez, J.; García-Marín, A.P.

Engineering Projects Area, Department of Rural Engineering, Civil Constructions and Engineering Projects. University of Córdoba, 14071 Córdoba, Spain.

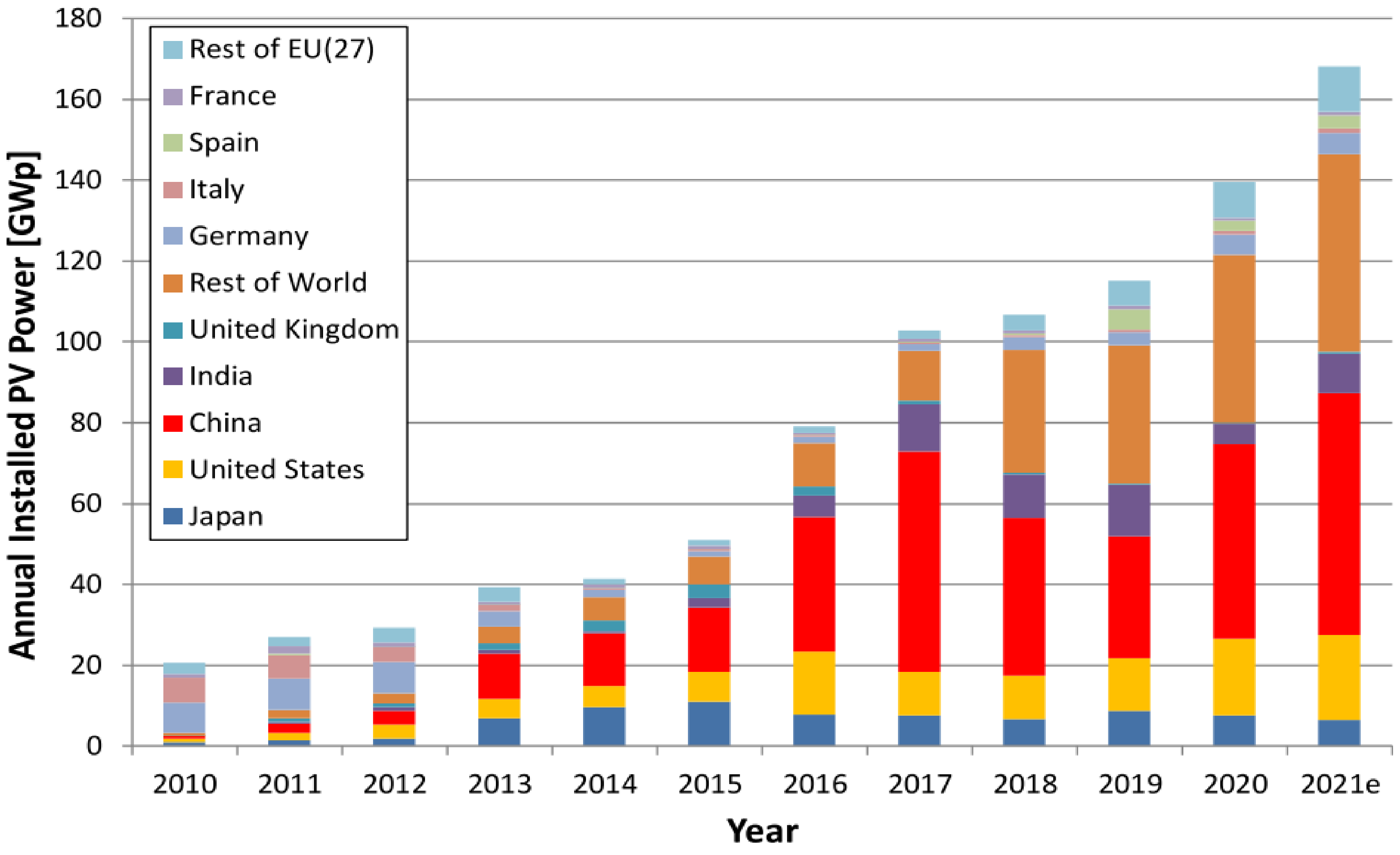
ITISE 2022. September 27th – 30th June, Gran Canaria, Spain

8th International conference on Time Series and Forecasting

- Introduction
- Materials and methods
- Results and discussion
- Conclusions

1. Introduction

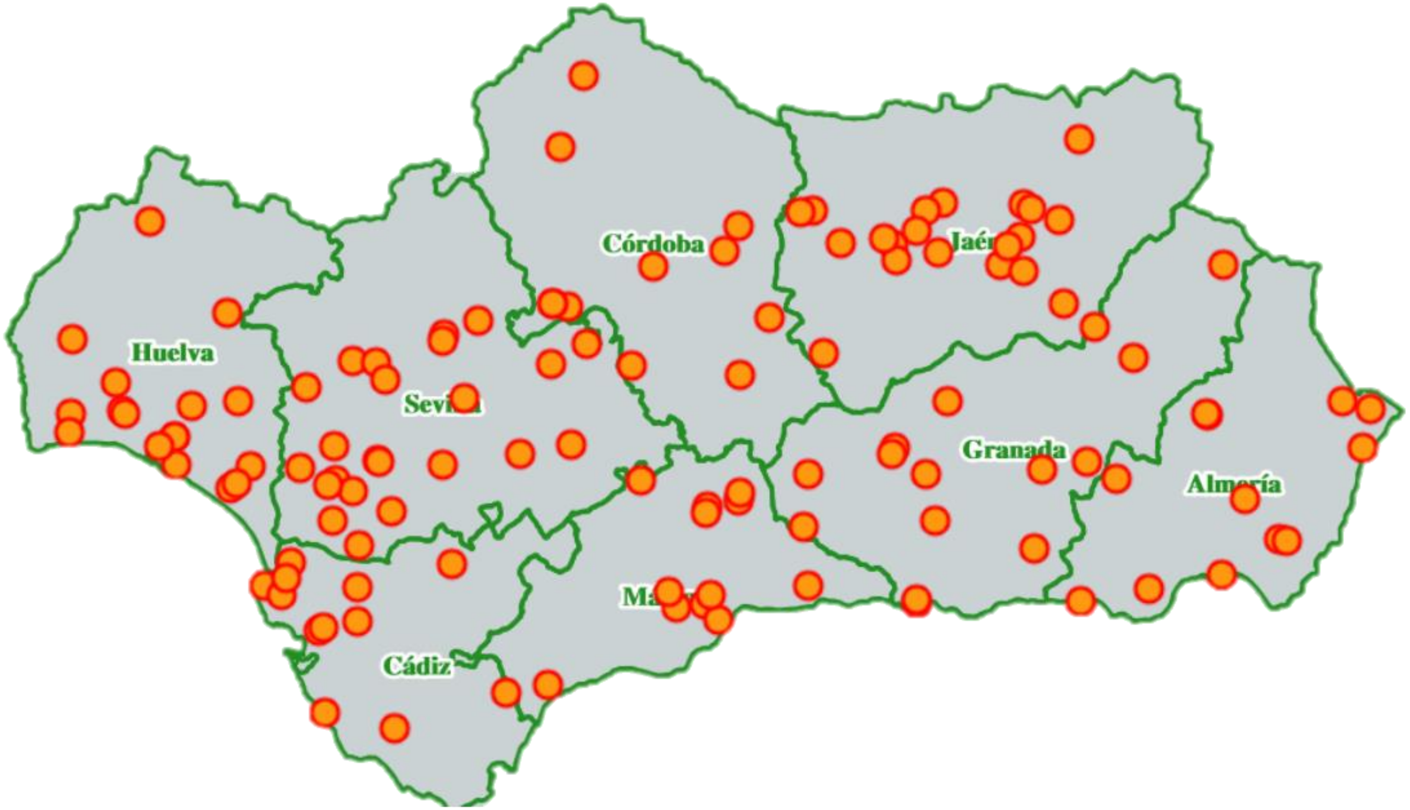
The importance of
Solar radiation estimates
and forecasts



Jäger-Waldau, A. (2021). Snapshot of photovoltaics− March 2021. EPJ Photovoltaics, 12, 2.

1. Introduction

Measuring solar radiation



1. Introduction

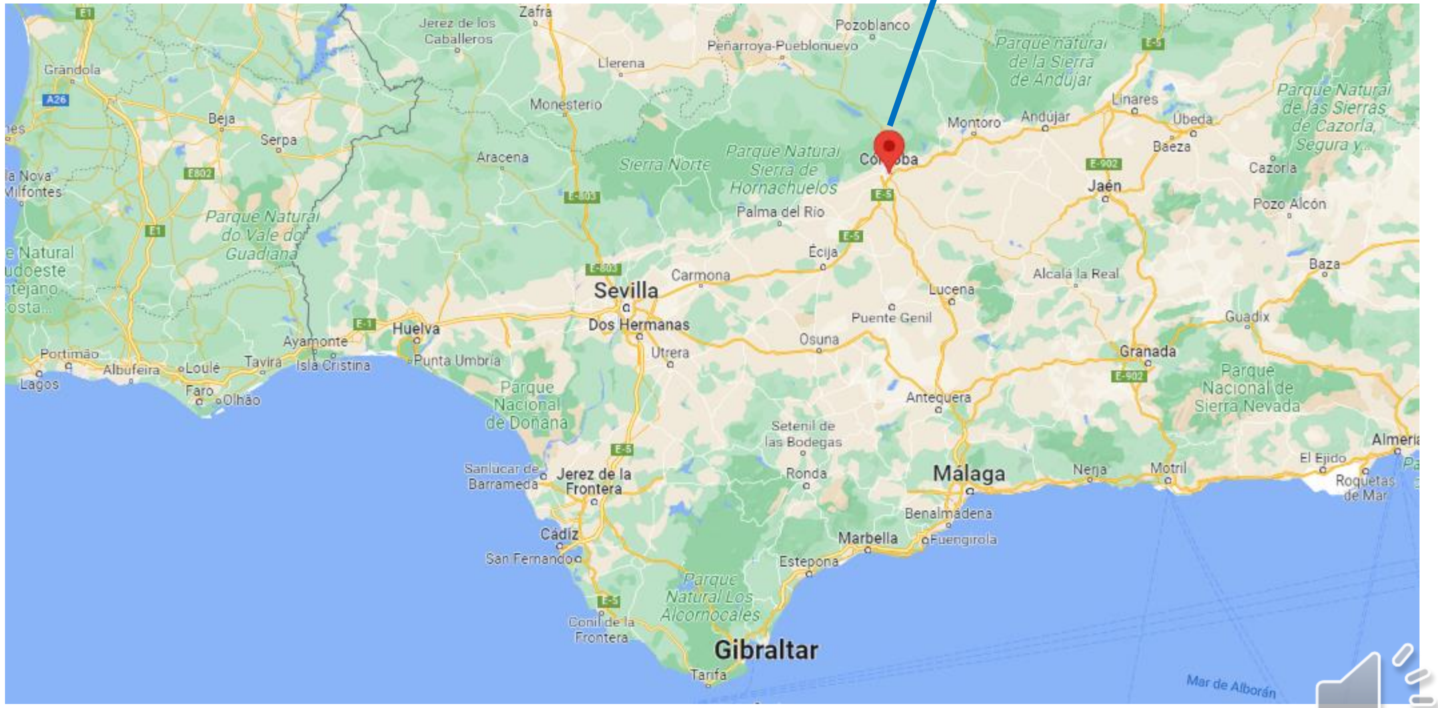
Machine learning models

MAIN GOAL

Solar radiation estimates in Córdoba station (Andalusia region)

2. Materials and methods

Study area



Cordoba station:

Latitude: 37º 51’ 25” N

Longitude: 04º 48’ 10” W

Elevation: 94 m

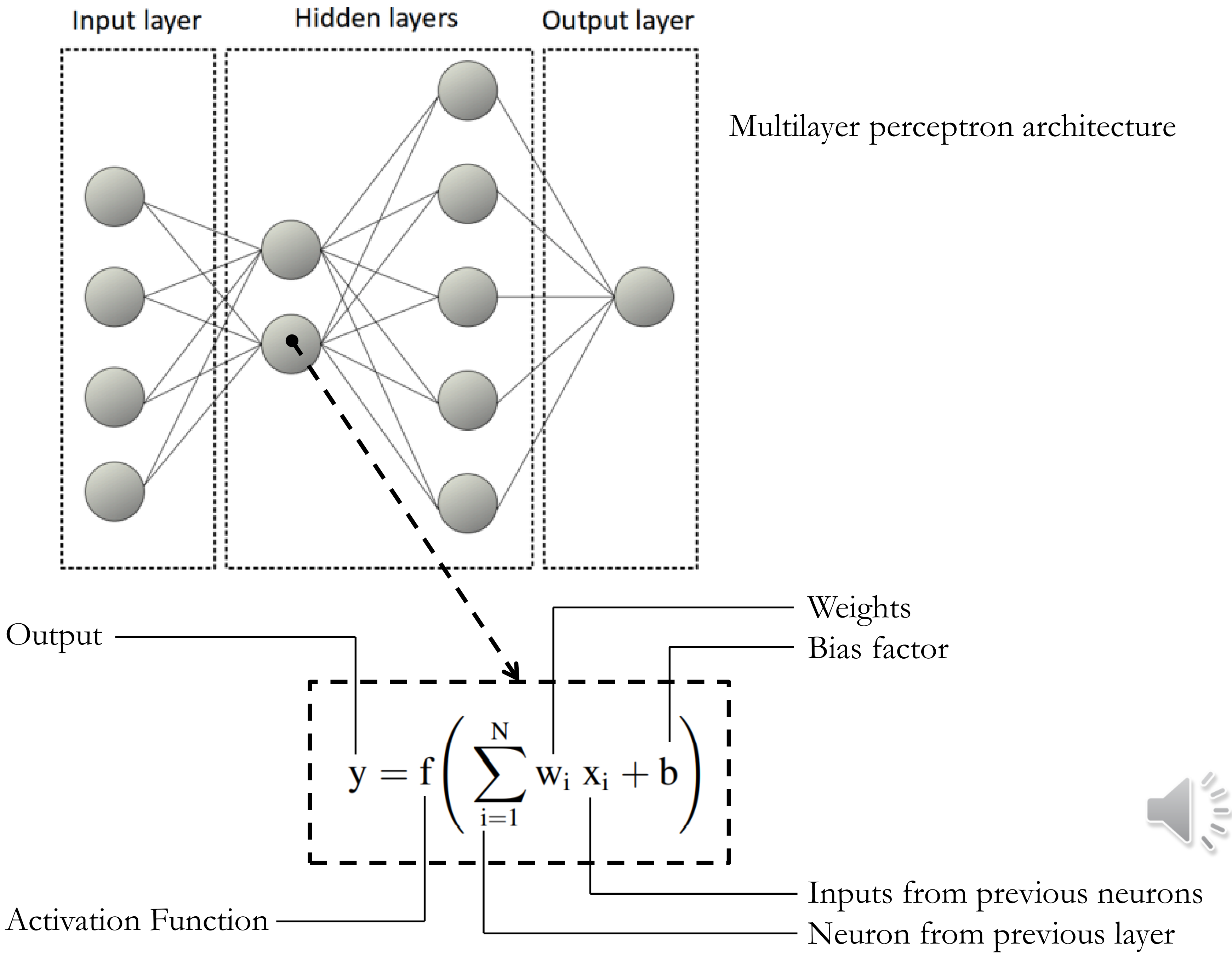
2. Materials and methods

Temperature-based features

Features	Definition
DOY	Day of the year. A value between 1 and 366
Tx	Maximum air daily temperature
Tm	Medium air daily temperature
Tn	Minimum air daily temperature
Ra	Extraterrestrial solar radiation
HourminTx	Time in hours when the Tx occurs
HourminTn	Time in hours when the Tn occurs
HourminSunset	Time in hours when the sunset occurs
HourminSunrise	Time in hours when the sunrise occurs
EnergyT	The integral of the half hourly temperature values of a day
VPD	Vapor Pressure Deficit
DeltaT	Tx - Tn
ea	Actual vapor pressure
es	Saturation vapor pressure
HourminSunset - HourminTx	Difference between HourminSunset and HourminTx
HourminSunrise - HourminTn	Difference between HourminSunrise and HourminTn

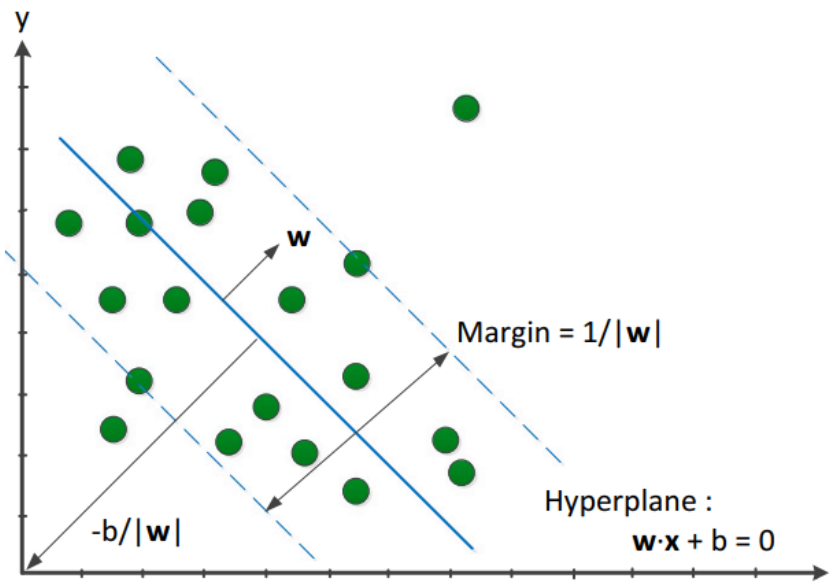
2. Materials and methods

Multilayer Perceptron

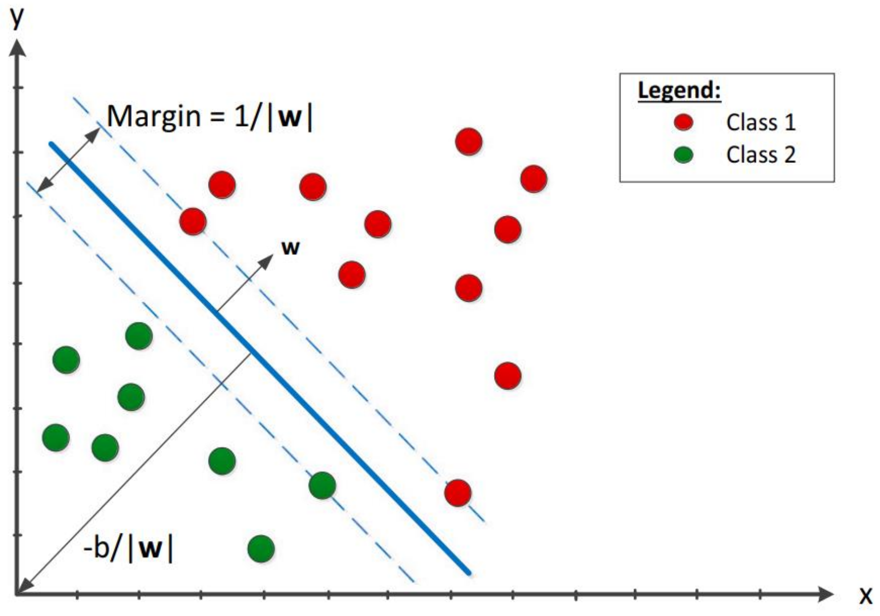


2. Materials and methods

Support Vector Machine



Support Vector Machine for regression tasks

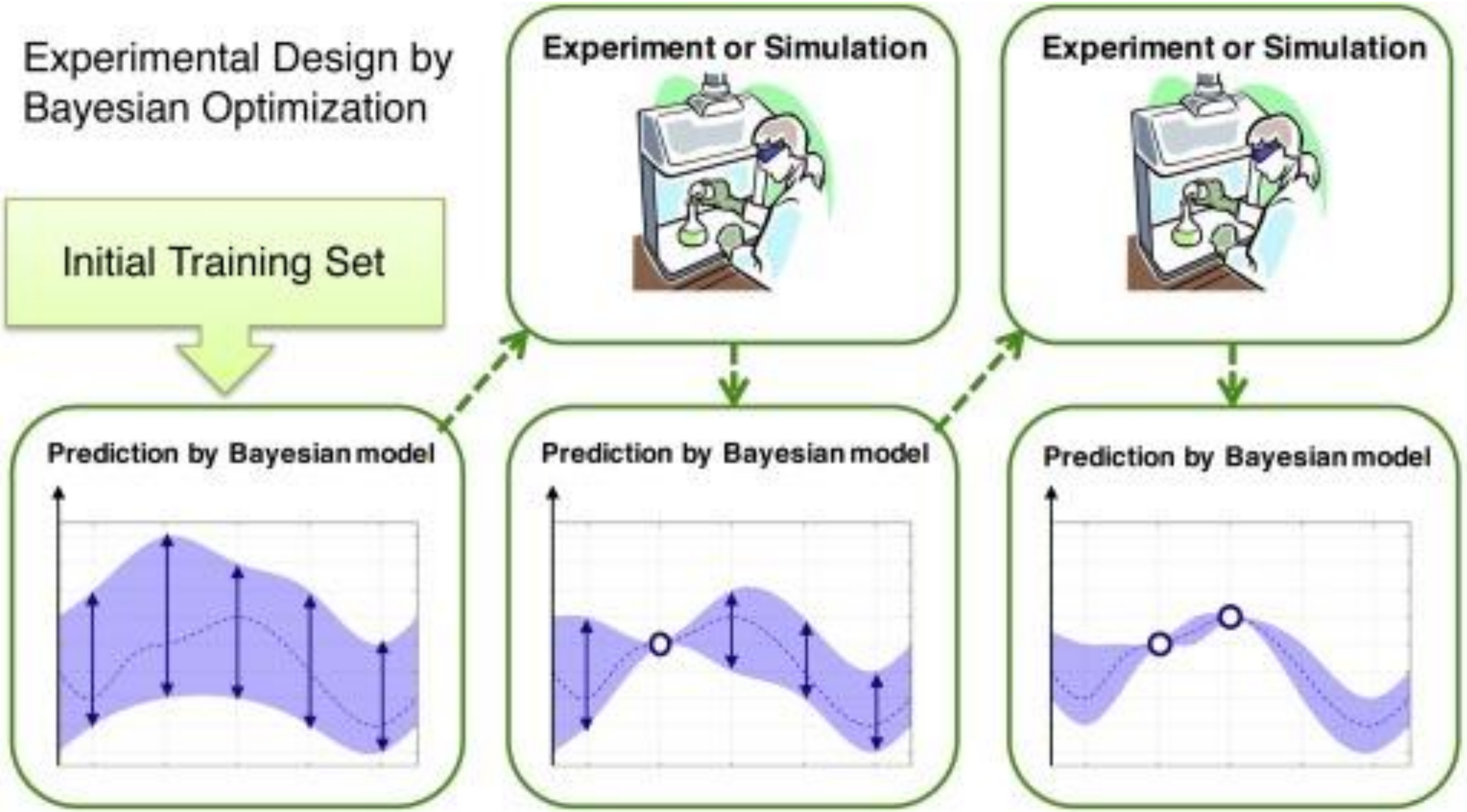


Support Vector Machine for classification tasks

$$min\frac{1}{2}||w||^2 s.t. \quad y_i(\boldsymbol{x_i} \;\; \boldsymbol{w} + b) - 1 \quad \geq 0 \quad \forall i$$

2. Materials and methods

Bayesian Optimization

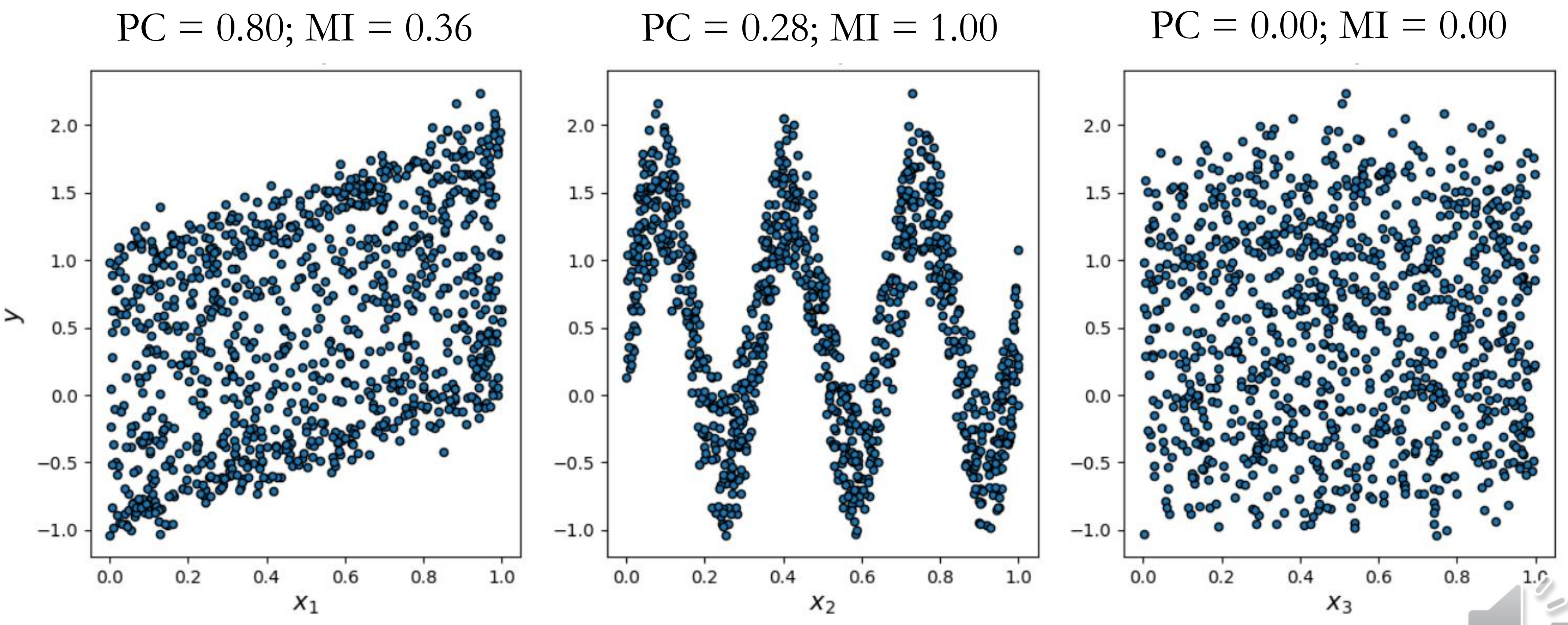


2. Materials and methods

Pearson Correlation and Mutual Information Coefficient

$$Pearson\ Correlation\ (PC) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

$$MI(X;Y) = H(X) - H(X|Y)$$



2. Materials and methods

Methodology

Data collection

Data preprocessing

Quality assurance tests (Estévez et al., 2011 ; 2016).

For Mutual Information and Pearson correlation, choose the different input features based on the different proposed values (0.2, 0.3, 0.4, 0.5, 0.6, 0.7 and 0.8).

Split into train (70% full dataset), validation (20% train dataset) and test (30% full dataset).

For each input feature configuration (one for every correlation function and value; a total of 14 different configurations).

For each model (MLP and SVM).

Hyperparameter selection using Bayesian Optimization

Build the final model.

Asses the model performance using the testing dataset

Save results

2. Materials and methods

Statistical analysis

$$MBE = \frac{1}{n}\sum_{i=1}^{n} x_i - y_i$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

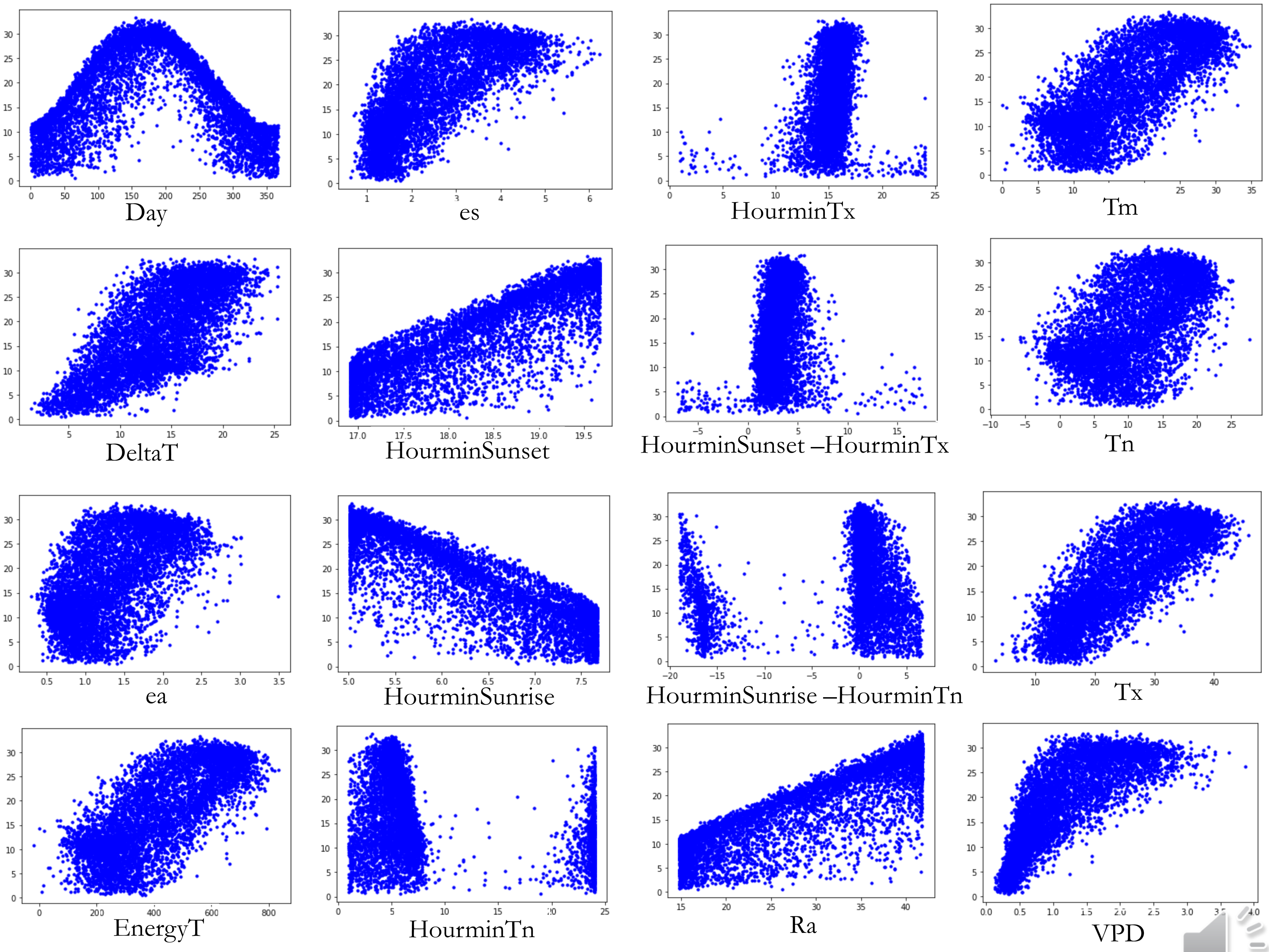
3. Results

Mutual Information coefficient and Pearson correlation

Features	Mutual Information	Pearson Correlation
DOY	0.9003	-0.1013
Tx	0.5823	0.8194
Tm	0.4838	0.7541
Tn	0.2772	0.5783
Ra	0.8534	0.8583
HourminTx	0.1567	0.2602
HourminTn	0.2606	-0.3107
HourminSunset	0.7525	0.8510
HourminSunrise	0.7648	-0.8489
HourminSunset - HourminTx	0.1707	0.2128
HourminSunrise - HourminTn	0.2303	0.2012
EnergyT	0.4847	0.7543
VPD	0.6776	0.8030
DeltaT	0.4262	0.7373
ea	0.2847	0.5844
es	0.4885	0.7549

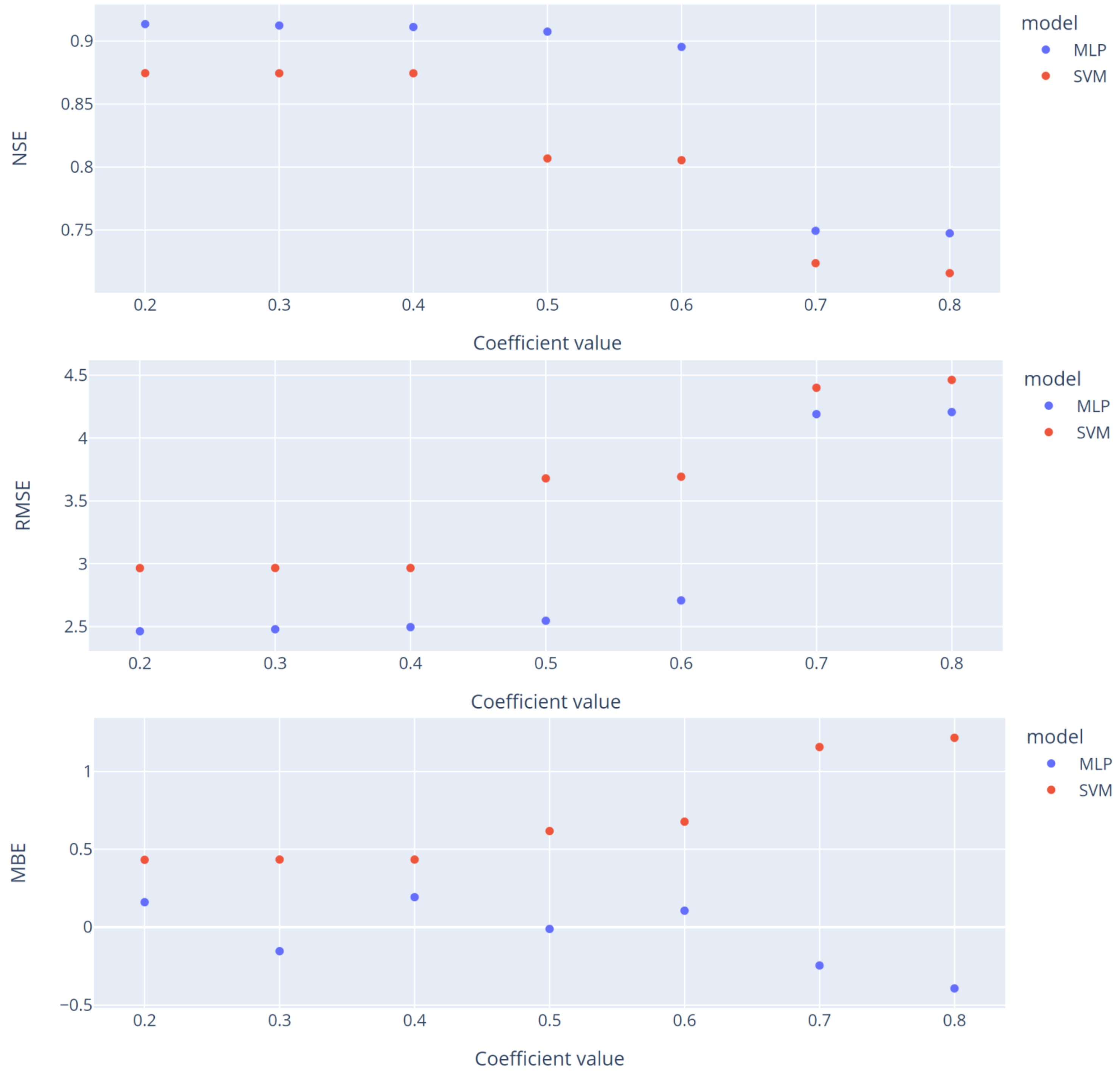
3. Results

Relation to different features



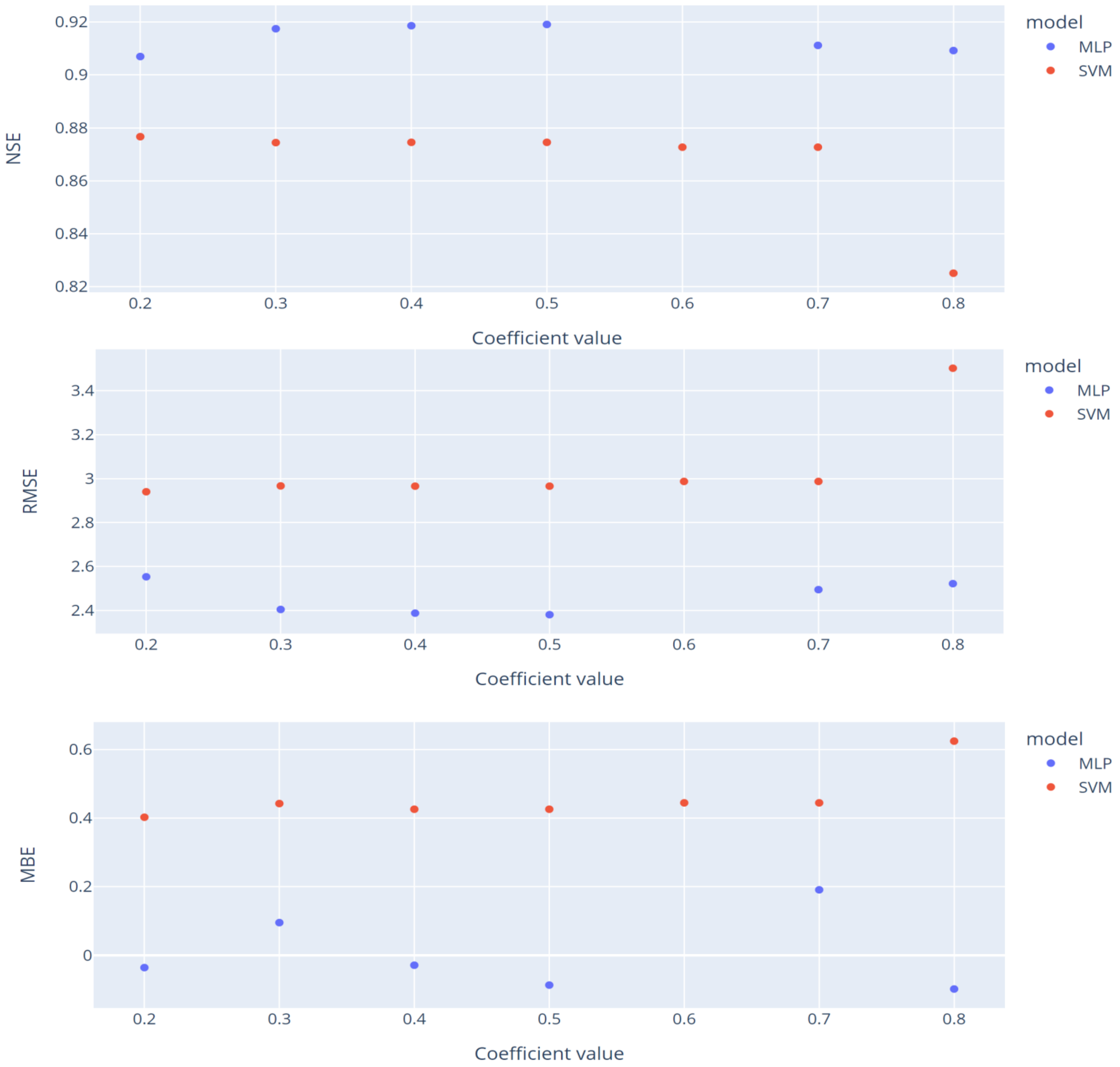
3. Results

Mutual Information coefficient results



3. Results

Mutual Information coefficient results



4. Conclusions

Mutual Information coefficient results

MLP

Multilayer Perceptron highly outperformed Support Vector Machine in all cases

Correlation Function

Both MI and PC obtained very accurate estimations not only using the features with higher correlation values

Ranking

The five best configurations were:

1. PC (0.5)
2. PC (0.4)
3. PC (0.3)
4. MI (0.2)
5. MI (0.3)

Statistics

The best RMSE =2.381 mm/d, NSE=0.919, and MBE = -0.086 mm/d

Strength

The developed methodology automate the feature selection

Future works

Extending this work to more meteorological stations and different climates

Thank you

Web page
juanantoniobellido.github.io

Twitter
@_jabellido_

Email
p22bejij@uco.es