PANAMA

PRESCRIPTIVE SOLAR ANALYTICS & ADVANCED WORKFORCE MANAGEMENT

D3.2

PV generation forecasting models

Responsible Partner	University of Western Macedonia
Prepared by	Ioannis Panapakidis Despoina Kothona
Checked by WP Leader	Georgios Christoforidis
Verified by Reviewer #1	Rabia Şeyma Güneş
Verified by Reviewer #2	Onur Enginar
Approved by Project Coordinator	

Project PANAMA is supported under the umbrella of SOLAR-ERA.NET Cofound by the Austrian Research Promotion Agency (FFG), General Secretariat for Research and Technology (GSRT) and the Scientific and Technological Research Council of Turkey (TUBITAK).

Deliverable Record

Planned Submission Date	
Actual Submission Date	
Status and Version	Draft version 2

Version	Date	Author(s)	Notes
Draft version 1	28/02/2021	Ioannis Panapakidis Despoina Kothona	

Table of Contents

Table of Contents 3
List of Figures 4
Definition of Acronyms 5
EXECUTIVE SUMMARY 6
1 INTRODUCTION 7
2 PV GENERATION FORECASTING PROCESSES 7
3 MODEL DESCRIPTION 8
3.1 Long Short-Term Memory Units 8
3.2 Adam optimization algorithm 11
3.3 Evaluation metrics 11
4 DATA DESCRIPTION 12
5 FORECASTING MODELS 13
5.1 Model for 15min-ahead forecasting 16
5.2 Model for 1h-ahead forecasting 16
5.3 Model for 24h-ahead forecasting 17
6 Results 18
6.1 Results for 15min-ahead 18
6.2 Results for hour-ahead 20
6.3 Results for day-ahead 23
REFERENCES 27

List of Figures

Figure 1. 15-min ahead forecasting process……..7
Figure 2. Hour-ahead forecasting process……..8
Figure 3. Day-ahead forecasting process……..8
Figure 4. The structure of LSTM model……9
Figure 5. The architecture of an LSTM unit……9
Figure 6. Pearson's autocorrelation curve……13
Figure 7. General LSTM model……14
Figure 8. LSTM training process……14
Figure 9. Flow-chart of the operation of the forecasting model……15
Figure 10. Separation of days utilizing the clearness index (KT)……18
Figure 11. Forecasted and actual PV power generation……19
Figure 12. a) Boxplot of AE, b) Violin Plot of AE……19
Figure 13. Histograms of AE……20
Figure 14. Forecasted and actual PV power generation……21
Figure 15. Forecasted and actual PV power generation……21
Figure 16. a) Boxplot of AE, b) Violin Plot of AE……22
Figure 17. Histogram of AE……22
Figure 18. AE timeseries……23
Figure 19. Forecasted and actual PV power generation……24
Figure 20. a) Boxplot of AE, b) Violin Plot of AE……24
Figure 21. Histograms of AE……25
Figure 22. AE timeseries……26
Figure 23. AE timeseries Case#4……26

Definition of Acronyms

DL	Deep Learning
LSTM	Long Short-Term Memory
PV	PhotoVoltaics

Note: Mathematical symbols and terms are explained directly in the corresponding sections.

EXECUTIVE SUMMARY

The PV generation forecasting process refers to the prediction of future generation values using historical data such as previous generation and external variables. The latter refer to variables that influence the value of the target variable which is the generation. Deliverable 3.2 PV generation forecasting models is composed by three major parts. The first part refers to the description of the 15min, 1h and 24h ahead forecasting processes. Based on the time resolution, there are three different process which refer three different forecasting models. The mathematical formulation of the models is the subject of the second part. Finally, the third part presents results for the three forecasting horizons. To fully examine the potency of the models, various inputs combinations are examined that refer to the various cases.

1 INTRODUCTION

The development of PV generation forecasting model is based on DL. Specifically, the forecasting approach employs LSTM units and refers to three different forecasting horizons: a) 15min-ahead, b) hour-ahead and c) day-ahead. At each forecasting horizon several scenarios have been examined, in terms of the input data, in order to determine and detect the optimal inputs for each forecasting horizon.

2 PV GENERATION FORECASTING PROCESSES

Based on the forecasting horizon, PV generation forecasting is categorized into very short-term, short-term, medium-term and long-term. Very short-term forecasting involves time steps lesser than 1h. The 15min ahead forecasting belongs to this category. The forecasting model is properly trained to perform forecasts 15min ahead. Short-term forecasting refers to time steps from 1h to 168h ahead in the future. Medium-term includes time steps from 1 week (i.e., 168h) to 1 year (i.e., 8760h). Long term is concerned with time horizons from 1 year to several years ahead. The different horizons involve different tasks and issues present in power systems that utilize PV capacity. PV generation forecasting has the potential to address many of these issues and highlight the multiple roles of PV capacity in contemporary power systems and electricity markets.

Figure 1, Figure 2 and Figure 3 show the 15min-ahead, 1h-ahead and 24h-ahead processes, respectively. Let $h = 1,2,\dots,H$ be an indicator denoting the current time step, that is 15min. Also, let $d = 1,2,\dots,D$ be an indicator denoting the current day, where D is the number of days that the available data covers. The prediction of future generation values is held taking into account historical values of the target variable and external ones. An input selection scheme is implemented to select the proper historical values. The scope is select the number and types of inputs that will result in robust forecasting. For instance, in the 15min-ahead process, inputs closest to the target value are always selected. if the current time step is 12:00 am and the target value is 12:15 pm, the inputs will refer to the two closest values, i.e., generations of 11:45 am and 12:00 am. As it will be shown to the following Sections, these values exhibit the highest similarity to the target time step. The process continues for all 15min-ahead instances using the pervious two 15min instances as inputs. Hence, if the scope is to predict the generation of the next hour, four separate forecasts will take place, one per 15min time steps.

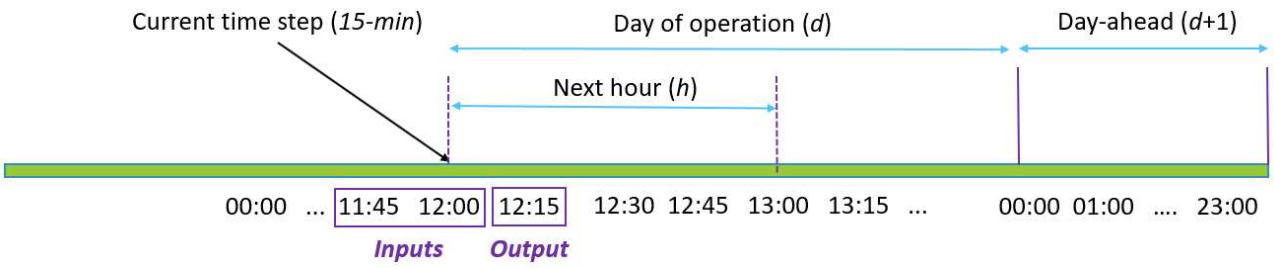

Figure 1. 15-min ahead forecasting process.

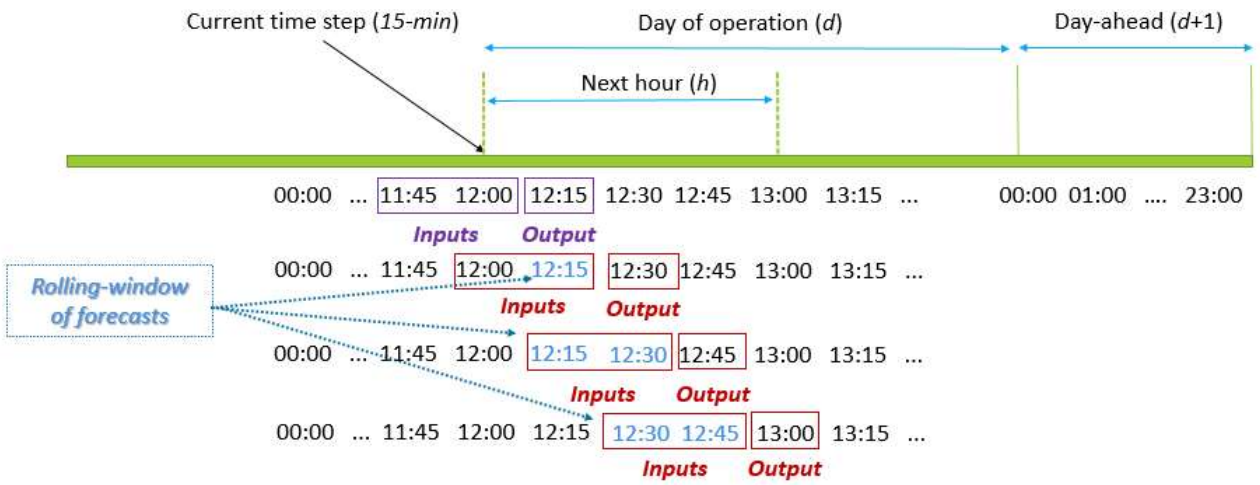

Figure 2. Hour-ahead forecasting process.

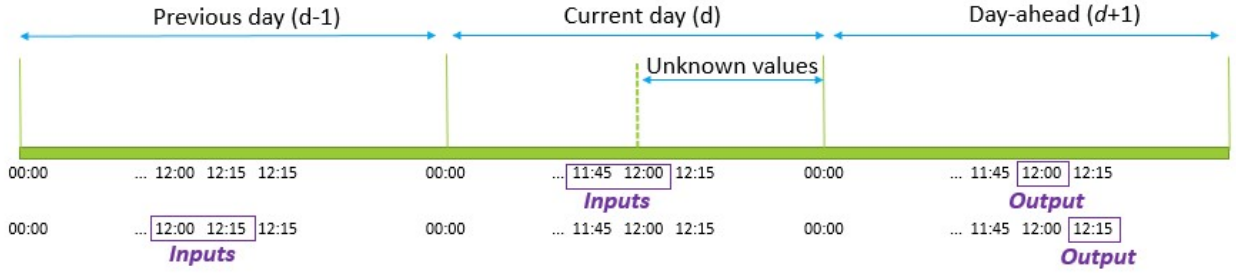

Figure 3. Day-ahead forecasting process.

In the 1h-ahead process, a different approach is followed. At the current time step, the forecasting outputs four values corresponding to the 15min time steps that correspond to the hour. In this case, the forecasting of the next step takes as inputs the forecasts, i.e., the outputs of the pervious steps. The day ahead forecasting model provides predictions for the PV generation of the next day with a time resolution of 15min. The forecasting process starts at 12:00 am of the current day d and the aim is to derive the generation of the 00:00-23:00 am of the next day.

3 MODEL DESCRIPTION

3.1 Long Short-Term Memory Units

The LSTM is a type of Recurrent Neural Network used in the field of DL [1]. The LSTM can deal with timeseries data and its main advantage is its ability to detect the dependencies between different features, such as PV power time series and solar irradiation timeseries. Moreover, the LSTM units are capable to keep the significant information of the previous time steps, by capturing time correlation between data.

The structure of the LSTM is depicted in the Figure 4. For the sake of simplicity, we assume that we utilize only the three previous time steps as inputs. As it is obvious, from the unfold version, the LSTM model consists of three LSTM units (U(0), U(1), U(2)), each unit for each time lag, respectively. Moreover, it is noticeable that the information derived from each unit is used as input at the next one.

This has as a result the final prediction (y(2), of the last unit (U(2)), to be based on the information of the previous time steps.

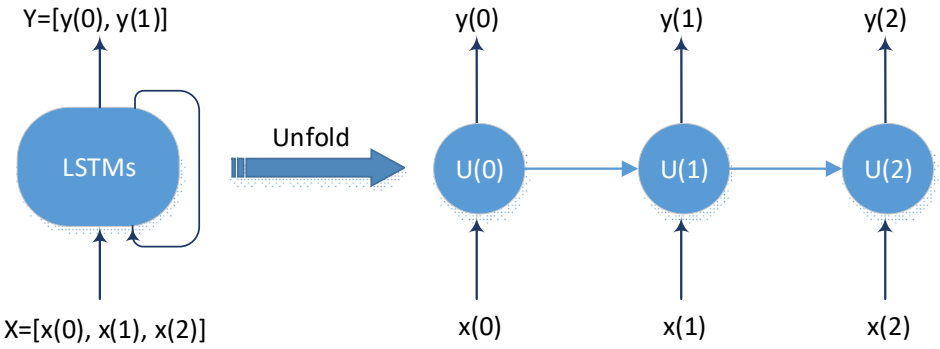

Figure 4. The structure of LSTM model.

In order to fully understand the kind of information that passes through the units, Figure 5 illustrates the architecture of the LSTM unit. Each unit consists of: a) the forget gate f_t, b) the input gate i_t, c) the output gate o_t and d) the cell state C_t.

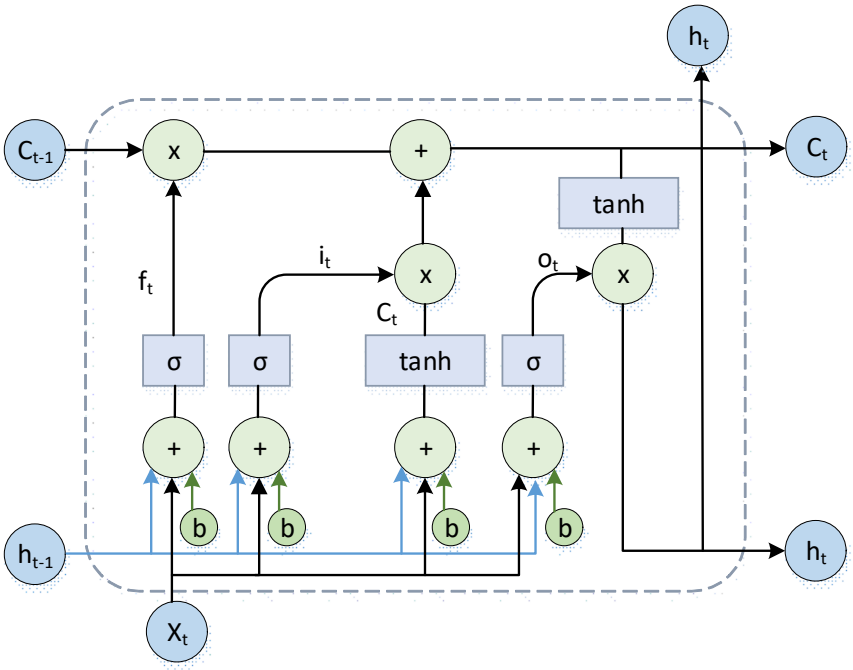

Figure 5. The architecture of an LSTM unit.

The cell state (C) allows the information to pass through the next cells. However, at each unit the kind of information that passes through are filtered and decided by the forget gate layer, which applies a sigmoid function at the hidden state (h_{t-1}) of the previous cell and the information of the current time step (x_t). The forget gate produces a number within a range 0 and 1 and is calculated as:

$$f_t = \sigma(W_{fx} x_t + W_{fh} h_{t-1} + b_f) \quad (1)$$

where σ is the sigmoid activation function of the forget gate, W_{fx} and W_{fh} are weight matrices of the forget gate layer of the current step input x_t and the hidden state h_{t-1}, , respectively and the b_f is the bias of the forget gate.

After the process of filtering, the next step is the decision of the new information that will be stored in the cell state. This is implemented in two stages. At the first one the input gate layer determines which values will be updated, by applying a sigmoid activation function. The second stage is based on a $\tanh$ layer which produces o vector of new candidate values that could be added to the cell state. The results of the two layers derive from the following equations:

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \tag{2}$$

$$\tilde{C}_t = \tanh(W_{cx}x_t + W_{ch}h_{t-1} + b_c) \tag{3}$$

where σ is the sigmoid activation function of the input gate, W_{ix} and W_{ih} are the weight matrices of the input gate layer of the current step input x_t and the hidden state h_{t-1}, respectively and the b_i is the bias of the input gate.

The results of the forget gate layer, the input gate layer and the $\tanh$ layer formulate the final cell state C_t vector, which will pass to the next unit, according to the following equation:

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t \tag{4}$$

Finally, the output of the unit h_t is derived from the output gate layer. Again, the estimation of the unit's output requires two steps. At the first one the tanh function is applied at the cell state vector, while at the second one, the output gate layer (o_t) is activated and a sigmoid function is applied at the information of the previous hidden state (h_{t-1}) and the current inputs x_t. The estimation of the output gate layer o_t and the final output h_t is based on the following equations:

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o) \tag{5}$$

$$h_t = o_t \cdot \tanh(C_t) \tag{6}$$

In order to train the LSTM model, it is essential to transform the time series into multiple samples. Depending on the way we create the samples the model can recognize univariate and multivariate time-series. Equations (7) and (8) express the sample's structure for univariate and multivariate time-series respectively:

$$\mathbf{S} = \begin{pmatrix} X^1_{t-n} & \cdots & X^1_t \end{pmatrix} \tag{7}$$

$$\mathrm{S} = \begin{pmatrix} X^1_{t-n} & \cdots & X^k_{t-n} \\ \vdots & \vdots & \vdots \\ X^1_t & \cdots & X^k_t \end{pmatrix} \tag{8}$$

where X stands for the variable, k denotes the number of variables and n is the number of previous time steps used in the forecasting.

3.2 Adam optimization algorithm

From the above analysis it is obvious that each LSTM unit contains four different type of weights, since it consists of four different layers (cell state layer, input gate layer, output gate layer and forget gate layer). If we consider that in a forecasting model we use more than one layers of LSTM units, in our case 64 layers, then the number of weights is increased exponentially. So, the selection of the weights' optimization algorithm in the training phase is important.

For our model Adam optimizer has been employed [2]. Adam, which is a variant of stochastic gradient decent optimization algorithm, has been widely used as optimization method in DL. The algorithm exploits the merits of momentum and RMSprop methods by using estimations of the first and second moments. The weights update rule is expressed as:

$$w_{t+1} = w_t - \frac{\alpha}{\sqrt{\hat{\upsilon}_t} + \varepsilon} \cdot \hat{m}_t \tag{9}$$

where α is the learning rate, ε is a constant value that prevents the zero division, $\hat{\upsilon}_t$ and $\hat{m}_t$ are the bias corrections and are denoted as:

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t} \tag{10}$$

$$\hat{\upsilon}_t = \frac{\upsilon_t}{1-\beta_2^t} \tag{11}$$

where m is the exponential moving average of gradients and υ is the exponential moving average of squared gradients denoted as:

$$m_t = \beta_1 \upsilon_{t-1} + (1-\beta_1)\left(\frac{\partial L}{\partial w_t}\right) \tag{12}$$

$$\upsilon_t = \beta_2 \upsilon_{t-1} + (1-\beta_2)\left(\frac{\partial L}{\partial w_t}\right)^2 \tag{13}$$

where $\partial L / \partial w_t$ is the gradient of L, the loss function to minimize.

3.3 Evaluation metrics

The assessment of the forecasting accuracy is based on a set of four evaluation indexes that measures the forecasting error [3]:

1. Absolute Error (AE), which is the absolute difference between the actual and the forecasted PV production.
2. Mean Absolute Error (MAE), which is the average of AE.

3. The Normalized Root Mean Square Error (NRMSE), that describes the standard deviation between the forecasting errors.
4. Mean Absolute Range Normalized Error (MARNE), which defines the absolute difference between the actual and the predicted PV power, normalized to the maximum power.

The formulation of the evaluation metrics is defined in Equations (14) – (17):

$$\mathrm{AE} = \left| \mathrm{P}_a(i) - P_f(i) \right| \tag{14}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} \left| \mathrm{P}_a(i) - P_f(i) \right| \tag{15}$$

$$\mathrm{NRMSE} = \frac{1}{N}\sqrt{\sum_{i=1}^{N} \frac{\left(\mathrm{P}_a(i) - P_f(i)\right)^2}{\max\left(P_a\right)}} \tag{16}$$

$$\mathrm{MARNE} = \frac{1}{N}\sum_{i=1}^{N} \frac{\left| \mathrm{P}_a(i) - P_f(i) \right|}{\max\left(P_a\right)} \times 100 \tag{17}$$

where P_a is the actual produced power, P_f is the forecasted one and N is the total number of the forecasted time steps.

4 DATA DESCRIPTION

The data which are utilized for the training and the evaluation of the forecasting models are obtained from a PV plant located in Southern Greece. The nominal power of the system is 15 kW and the recording frequency of the produced power is 15min. Additionally, data about the modules' temperature and solar irradiation are also available with the same recording frequency. Moreover, the clearness indexes (K_T) of the dataset's days are known. The collected data covers a period from 02/01/2012 to 27/04/2014. However, the data from 30/12/2013 to 27/04/2014 are utilized for the validation of the models.

Since we deal with time series data and the forecasted value is based on the previous time steps, a Pearson's autocorrelation analysis of the PV power has been conducted [4]. Given measurements $P_1, P_2, \ldots, P_N$ at time $t_1, t_2, \ldots, t_N$. the lag k autocorrelation of P_t observation is given cy the following function:

$$\rho_{\mathrm{PP}}(k) = \frac{\mathrm{E}\left[(P_t - \mu_P)\left(\overline{P_{t+k} - \mu_P}\right) \right]}{s^2} \tag{18}$$

where s is the standard deviation of the power timeseries. The result of the above function gives a value between within a range [-1,1]. The closer to one the coefficient is, the more correlated the values are.

The recording frequency of the data is 15min, so we have 96 observation within a day. The analysis' results, which are presented in Figure 6 and refer to two sequential days, indicate that the most

correlated values of the observation Pt are the two previous time steps, P_{t-1} and P_{t-2}, for an intra-day analysis. Moreover, the most correlated values of the observation P_t with the previous day are the P_{t-96} and P_{t-97} values.

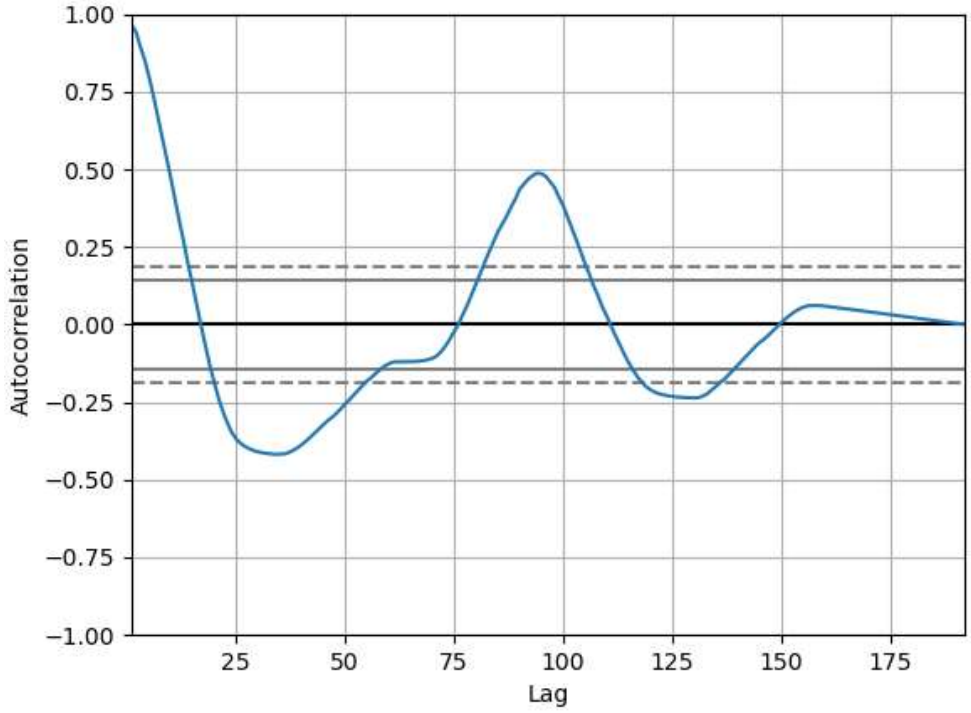

Figure 6. Pearson's autocorrelation curve.

5 FORECASTING MODELS

The implemented forecasting models refer to three different forecasting horizons: a) 15min-ahead, b) hour-ahead and c) day-ahead. Moreover, at each forecasting horizon several scenarios have been examined with different training periods and different input variables. In all cases, the models consist of 64 LSTM layers and each layer has one or two LSTM units, considering the number of previous time steps. The general structure of the LSTM model is presented in Figure 7.

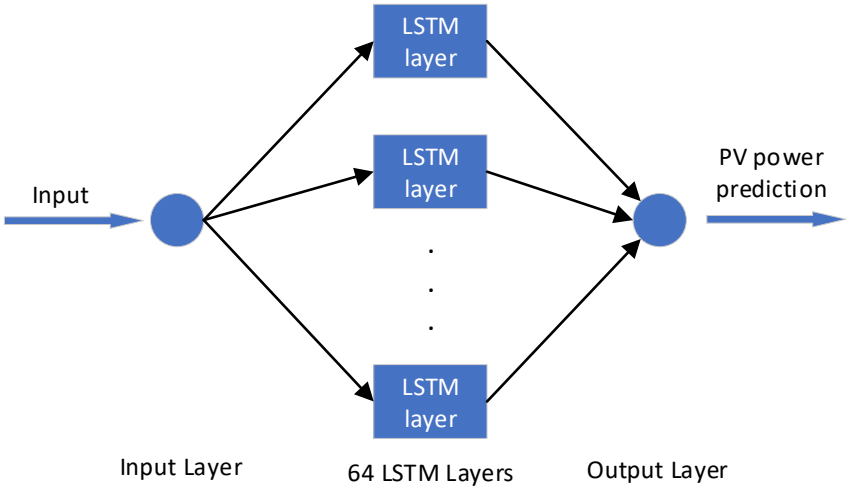

Figure 7. General LSTM model.

For the training of the model two basic parameters should be specified, the threshold of the MAE and the maximum number of epochs, i.e., the maximum number of the training iterations. These two parameters are used as convergence criteria for the training phase. Figure 8 represents the model's training process. Figure 9 presents the flowchart of the operation of the model.

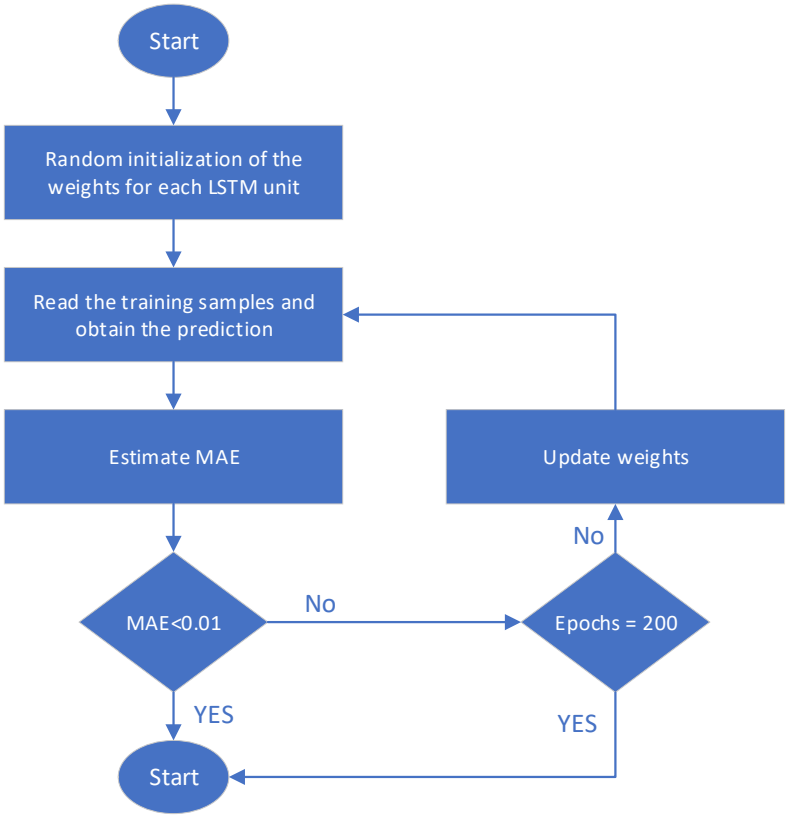

Figure 8. LSTM training process.

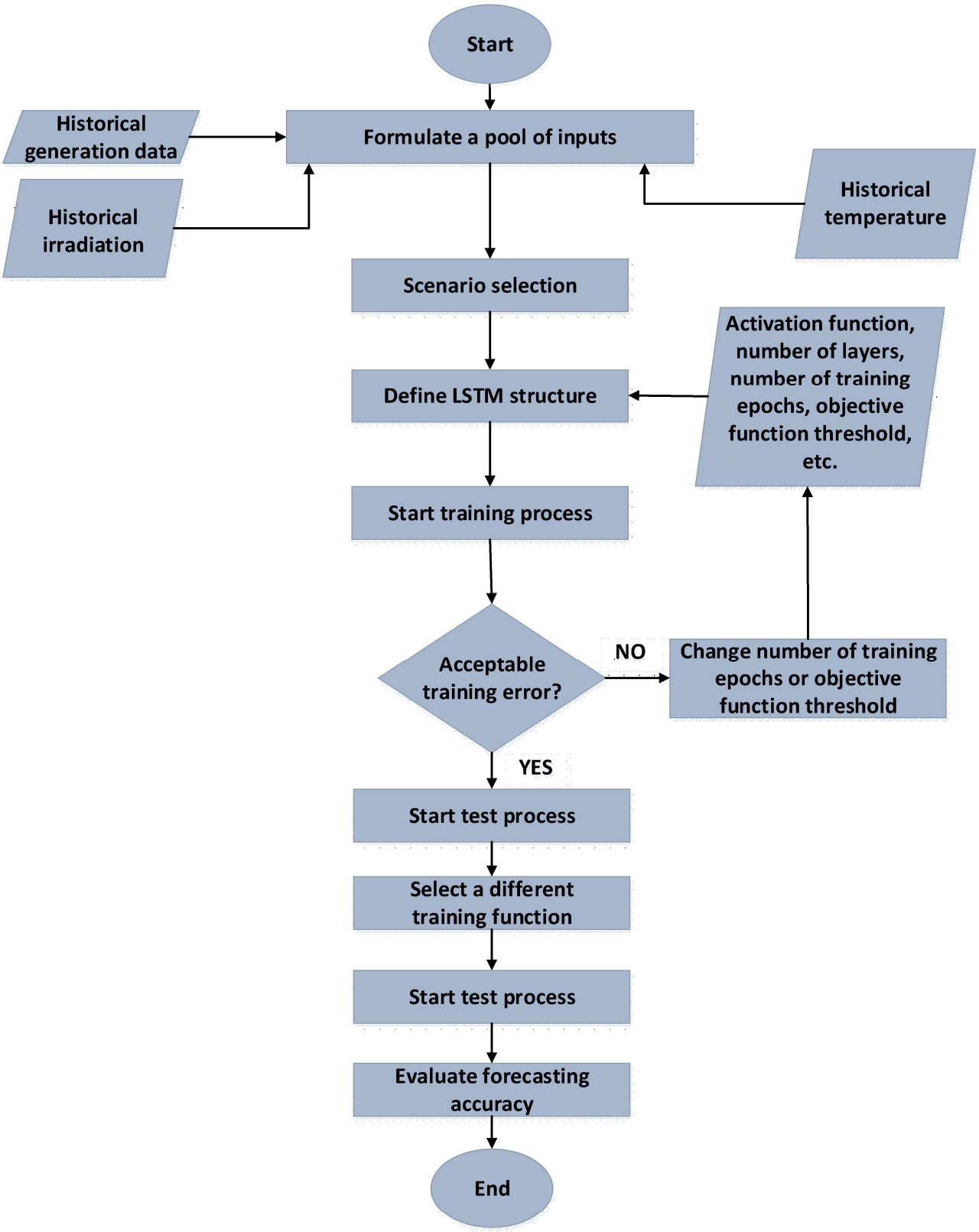

Figure 9. Flow-chart of the operation of the forecasting model.

5.1 Model for 15min-ahead forecasting

Since the available data for the training of the model covers a period of two years, in order to determine the optimal training period three basic cases have been examined considering only the PV power historical data:

- **Case#1:** The model takes as input the two previous time steps of the power and give as output the next time step. For the training of the model the data which are utilized covers a period from 02/01/2012 to 29/12/2013.
- **Case#2:** The model takes as input the 2 previous time steps and give as output the next time step. For the training of the model the data which are utilized covers the periods 02/01/2012 - 29/04/2012 and 31/12/2012 – 28/4/2013.
- **Case#3:** The model takes as input the 2 previous time steps and give as output the next time step. For the training of the model the data which are utilized covers a period from 02/01/2012 to 29/12/2013.

After determining the most suitable training period, 02/01/2012 to 29/12/2013, different scenarios have been implemented in terms of the extraterrestrial variables, i.e. solar irradiation and module's temperature.

- **Case#4:** Alongside with the PV power, solar irradiation and panel's temperature are also included. Moreover, information about the previous day is also utilized. For instance, in order to predict the PV power output at 12:00am the PV production, the solar irradiation and the module's temperature at 12:00am of the previous day are used additionally as input. The model takes as input the 2 previous time steps and give as output the next time step.
- **Case#5:** It is similar to the Case#4 but the PV power of the previous day is excluded. The test dataset is transformed as input samples and predicted values. The input samples consist of the two previous time steps of PV power, solar irradiation and panel's temperature. However, the main difference is that the ward method is applied to the train samples and divided them into clusters. Accordingly, each cluster is used to train different forecasting model. Finally, we find the centroid of the cluster with the minimum Euclidian distance from each test sample and the prediction is obtained from the model which was trained with the samples of the specific cluster.

5.2 Model for 1h-ahead forecasting

From the 15min ahead forecasting model derived that the training period of two years (31/12/2012 to 29/12/2013) provide more accurate results compared to the other two training periods but the difference is minor. Based on this, since the increment of the training samples significantly increases the computational burden, the training data for the hour ahead forecasting model covers the period from 31/12/2012 to 29/12/2013.

At the present forecasting horizon eight different cases have been examined:

- **Case#1**: The model gets as input the three previous lags of the PV power, the solar irradiation and the module's temperature and provides the forecasted PV power of the next four time steps.
- **Case#2**: It is similar to Case#1 but instead of three previous observations only the two previous are used as input. Figure 1 represents the Case#2 procedure.

- **Case#3**: The model's sample consists of the two previous time steps of the PV power generation, the solar irradiation and the module's temperature and provides the prediction of the first time step (15min-ahead). Afterwards, the PV power prediction of the first time step, is used as input in order to predict the second time step. The procedure is executed iteratively until the prediction of the next hour is achieved. In this case we assume that weather predictions about the solar irradiation are available. Moreover, for the forecast of the next time step, the PV power of the previous day is also utilized. For instance, in order to predict the PV power output at 11:00 am the PV production, the solar irradiation and the module's temperature at 11:00 am of the previous day are used additionally as input. Figure 2 illustrates the forecasting process of the case.
- **Case#4**: It is similar to the Case#3 but the PV power of the previous day is excluded. The test dataset is transformed as input samples and predicted values. The input samples consist of the two previous time steps of PV power, solar irradiation and panel's temperature. However, the main difference is that the ward method is applied to the train samples and divided them into three clusters. Accordingly, each cluster is used to train different forecasting model. Finally, we find the centroid of the cluster with the minimum Euclidian distance from each test sample and the prediction is obtained from the model which was trained with the samples of the specific cluster.

5.3 Model for 24h-ahead forecasting

For the day-ahead forecasting five different cases have been examined:

- **Case#1**: In order to predict the PV generation of the next day the model takes as inputs the two previous time-steps PV power, the solar irradiation and the panels' temperature of the current day, if prediction refers to the PV power until 12 am, or of the previous day, if the prediction refers to the PV power after 12 am.
- **Case#2:** It is similar to Case#1 but the model also takes into account and K_T index prediction of the next day. According to the information of the K_T index, a former analysis has been implemented. The power data of the training set are separated into three categories according to the K_T index: a) mostly cloudy, b) partly cloudy and c) clear. Afterwards, three PV power profiles derived from each category, by calculating the mean value of the PV power generation. Since, the recording frequency the data is 15min the PV power profiles are vectors with size 1x96. Finally, the samples of the model, despite the PV power, the solar irradiation and the module's temperature, include a fourth feature that refers to the PV power profile of the category the predicted day belongs. The process of the days' separation is presented in Figure 9.
- **Case#3:** The samples of the model are constructed according to the K_T of the next day. The days again are separated into three categories, but it takes as inputs the two previous time steps of the PV power, the solar irradiation and modules temperature of the previous day which lies in the same category.
- **Case#4**: It is similar to Case#3 but according to K_T the days of the training set are separated into three categories and three different LSTM models are trained.
- **Case#5**: it is similar to Case#2 with the main difference that every time we have information about the actual PV power of the day we predicted, the day is added to the category that belongs, according to the K_T and the PV power profile is recalculated.

The general structure of the forecasting process is illustrated in Figure 9.

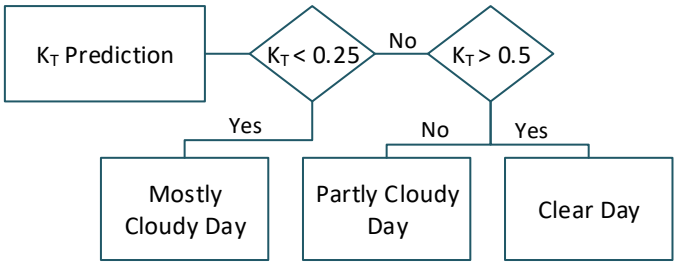

Figure 10. Separation of days utilizing the clearness index (K_T).

6 Results

6.1 Results for 15min-ahead

At Table 1 the results of the 15min-ahead forecasting are presented. At the first three cases, as it has been mentioned before, only the PV power is utilized in the forecasting process and their basic difference is at the training period. In Case#4 and Case#5 alongside with the PV power data, information about solar irradiation and panels' temperature are utilized and the forecasting accuracy has been significantly improved. Specifically, the utilization of the clustering method, in Case#4, has proved that the division of the data and the training of different forecasting models can further improve the predictions' accuracy. Moreover, in Figure 10 the actual and the forecasted PV power production of eleven sequential days are presented for each case.

Table 1. Evaluation indexes of 15min-ahead forecasting.

Cases	MAE	NRMSE	MARNE
Case#1	0.440	6.338	2.635
Case#2	0.453	6.348	2.712
Case#3	0.450	6.348	2.695
Case#4	0.416	5.992	2.491
Case#5	0.407	6.272	2.440

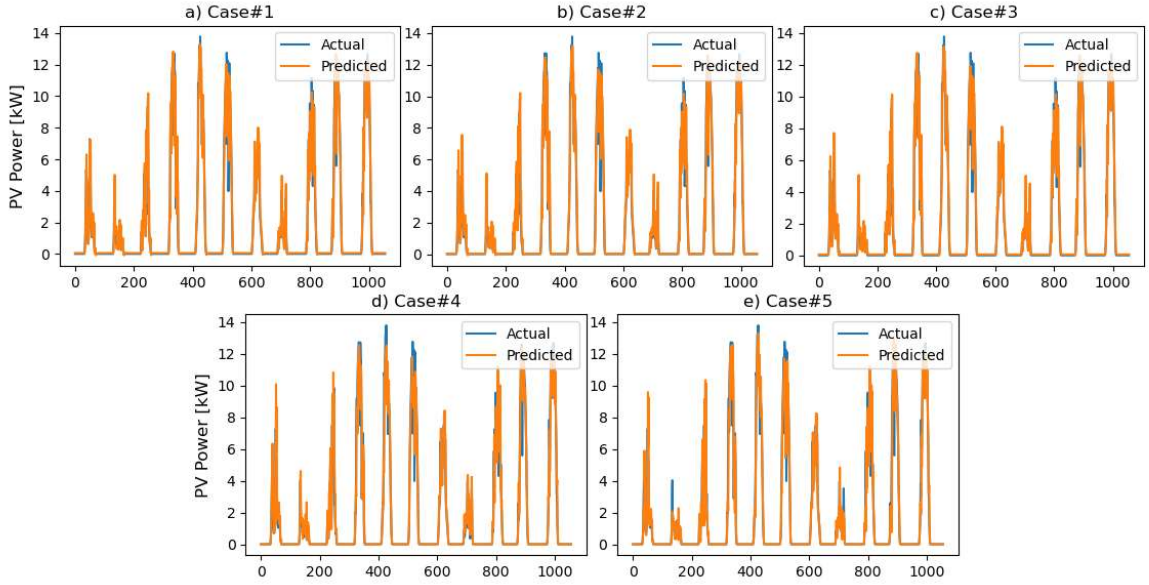

Figure 11. Forecasted and actual PV power generation.

Additionally, in Figure 11 the AE of the predictions is presented for each case with a boxplot and a violin plot respectively. The boxplot is very useful in terms of outliers detection. In our case as outlier is defined a large deviation between the actual and the predicted PV power. Violin plot on the other hand, contains density information and is used to effectively describe the distribution of the data and its probability density. Moreover, the frequency distribution of the AEs can be observed more clearly in Figure 12, where the histogram of each case is presented.

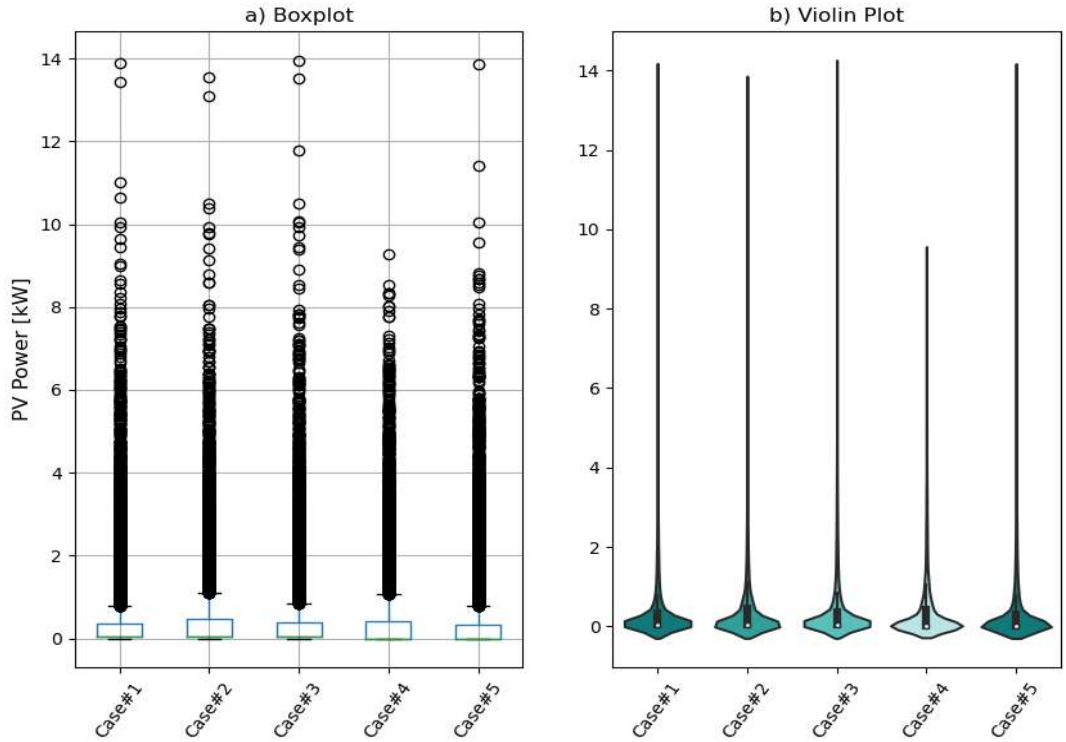

Figure 12. a) Boxplot of AE, b) Violin Plot of AE.

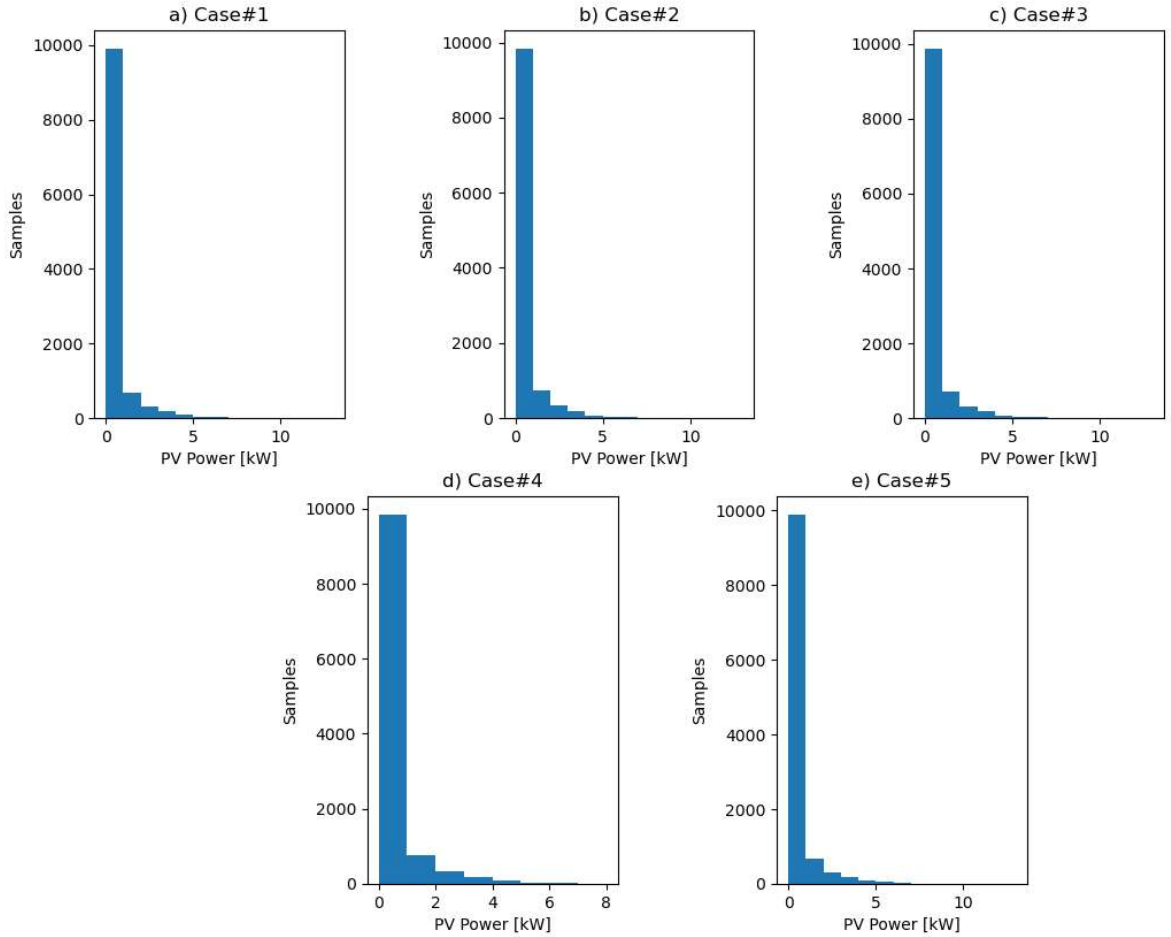

Figure 13. Histograms of AE.

6.2 Results for hour-ahead

The results of the hour-ahead forecasting are presented in the Table 2. As it is obvious, Case#4 provides more accurate predictions compared to the other cases. So, when the clustering method is employed, we can significantly improve the forecasting accuracy of the model.

Table 2. Evaluation indexes of hour-ahead forecasting.

Cases	MAE	NRMSE	MARNE
Case#1	0.643	8.563	3.849
Case#2	0.674	9.160	4.035
Case#3	0.527	7.483	3.157
Case#4	0.493	6.922	2.952

In Figure 13, the actual and the predicted PV power of two sequential days for each case are presented. It is noticeable that in Case#4, even for cloudy days, we can obtain predictions with small deviations between real and predicted power.

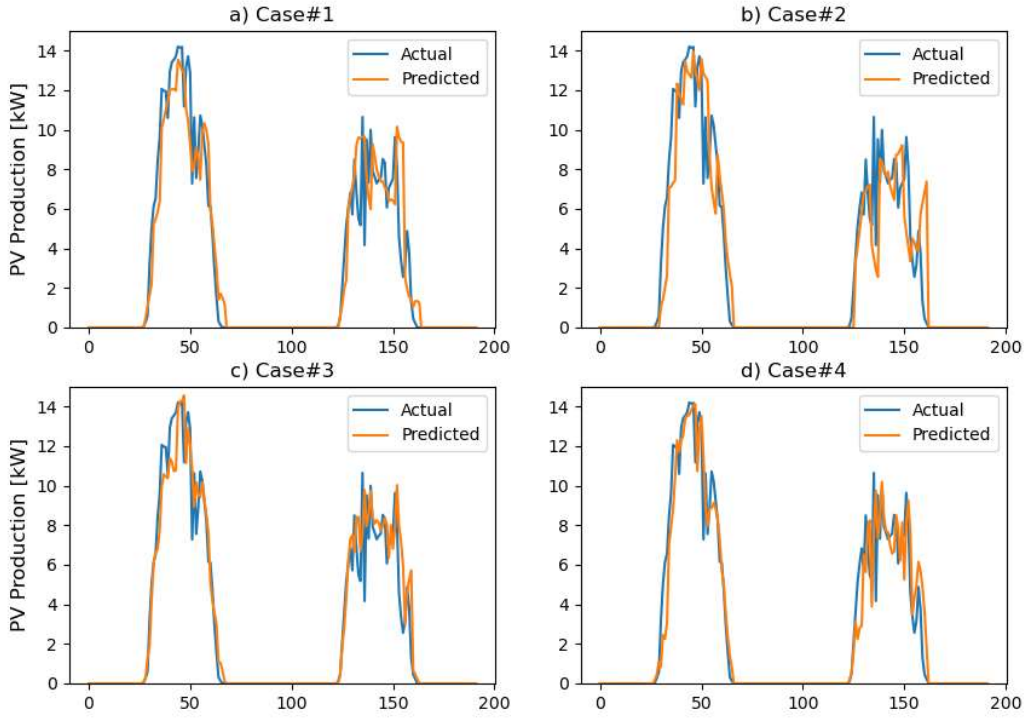

Figure 14. Forecasted and actual PV power generation.

Moreover, in order to fully examine the forecasting potential of the clustering model in Case#4, different number of clusters have been tested. Table 3 illustrates the evaluation metrics when the number of clusters varies from two to four. The results indicate that for the specific training set the optimal number of clusters is three. In Figure 14 the actual and the forecasted power of the two sequential days, which were examined previously, are presented again for Case#4 in terms of the different number of clusters.

Table 3. Evaluation indexes of hour-ahead forecasting.

Cases	MAE	NRMSE	MARNE
Case#4_2	0.606	8.485	3.628
Case#4_3	0.493	6.922	2.952
Case#4_4	0.505	7.327	3.293

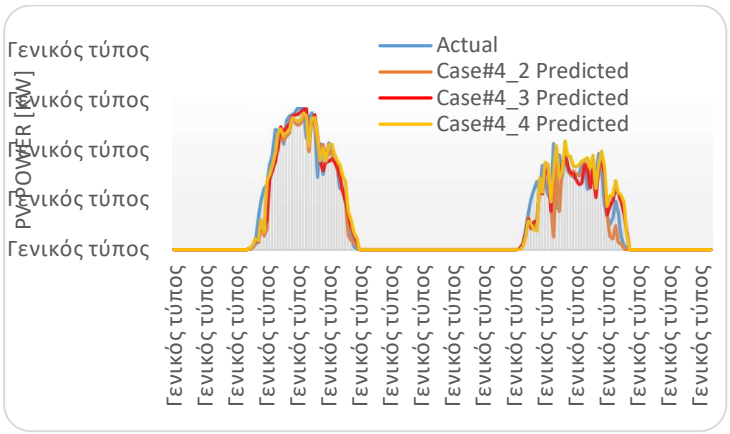

Figure 15. Forecasted and actual PV power generation.

In the Figures 15 and 16 the boxplot, the violin plot and the histogram of each case are presented. In Case#4 the maximum AE is almost 11kW, while in other cases exceeds the 13kW.

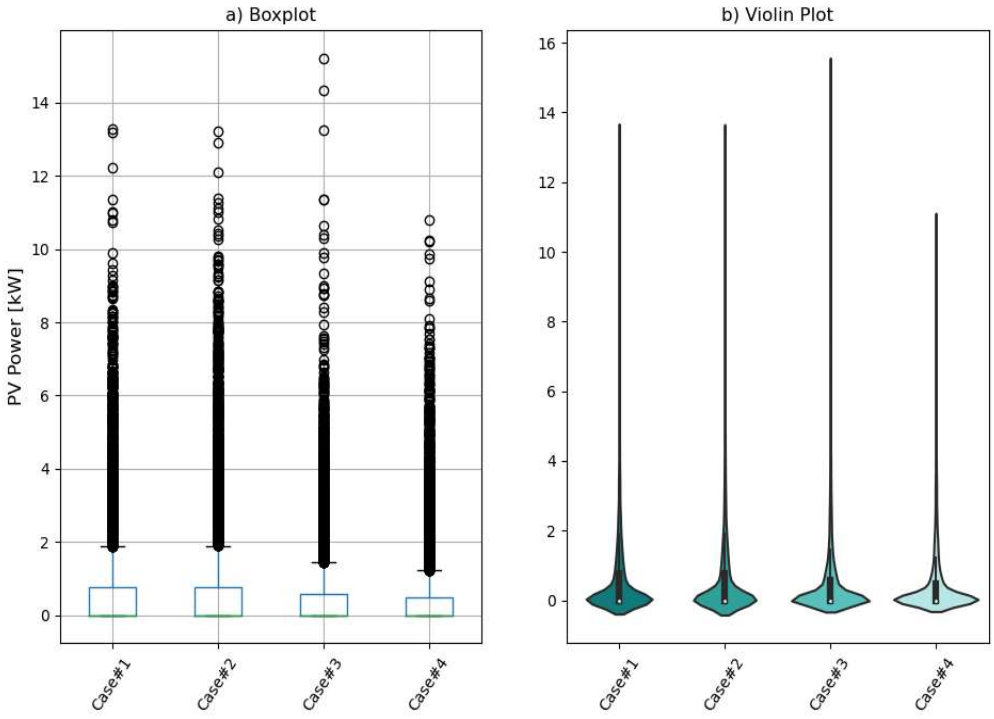

Figure 16. a) Boxplot of AE, b) Violin Plot of AE.

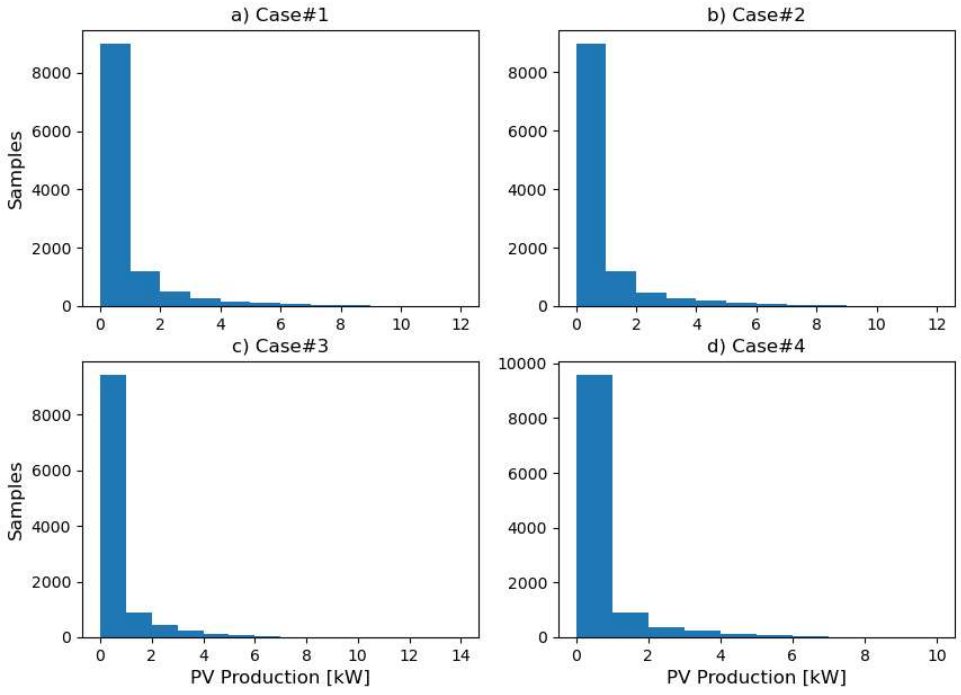

Figure 17. Histogram of AE.

Moreover, Figure 17 depicts the AE of the total training test in order to fully examine the performance of the models. It is obvious that Case#4 provide more accurate results compared to the other three cases.

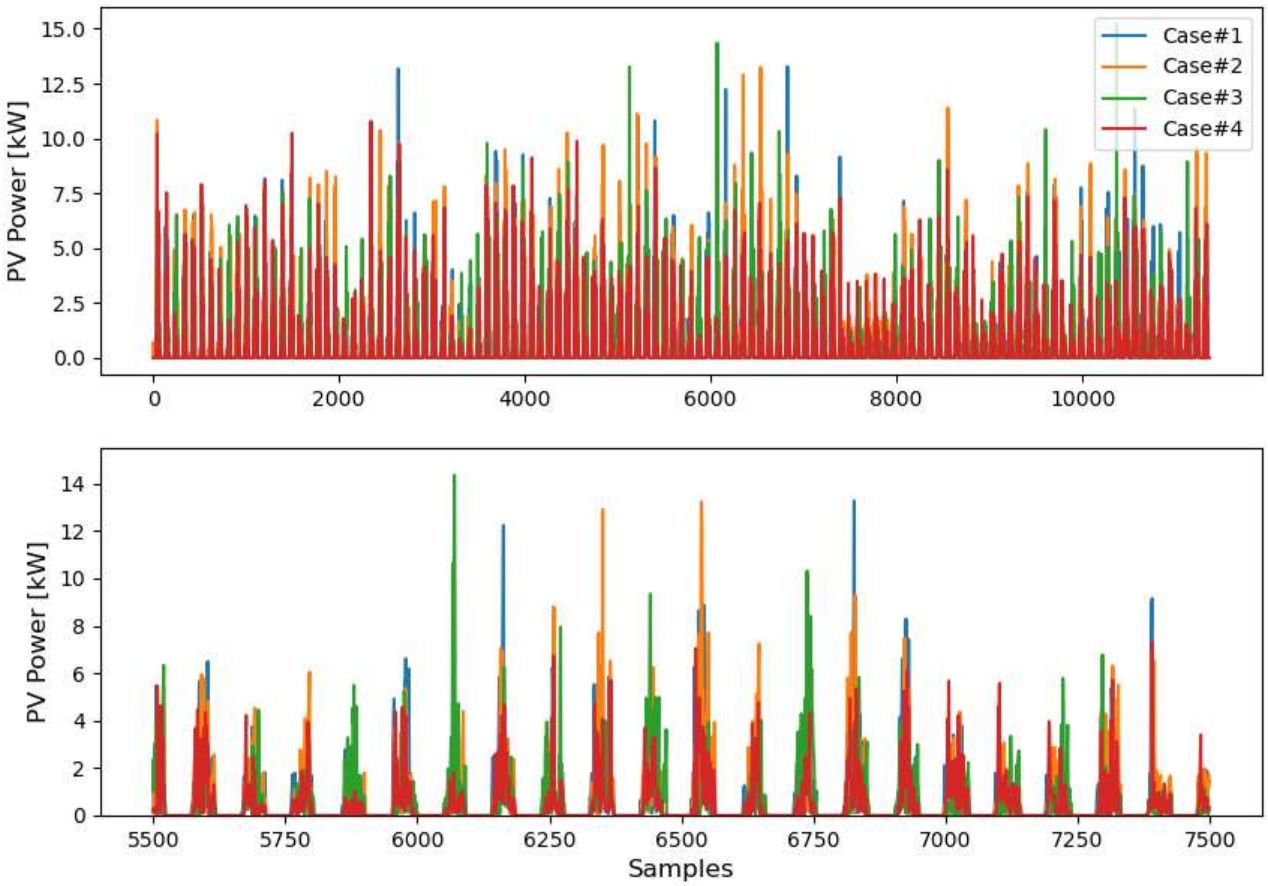

Figure 18. AE timeseries.

6.3 Results for day-ahead

For the day ahead forecasting five different cases have been examined. The evaluation metrics of the cases are presented in Table 4. Case#2 and Case#3, where the PV power profile of the cloudy, partly cloudy and sunny days is used as input, outperform. Specifically, when the PV profiles are updated, considering the actual power of the test set, we can achieve more accurate results.

Table 4. Evaluation indexes of day-ahead forecasting.

Cases	MAE	NRMSE	MARNE
Case#1	1.135	15.597	8.118
Case#2	0.848	10.623	5.077
Case#3	1.032	12.805	6.177
Case#4	0.914	11.721	5.474
Case#5	0.847	10.750	5.070

Figure 18 presents the actual and the forecasted PV power for each case. Moreover, the distributions of AEs of Case#3 and Case#5 have small deviations as it is obvious from Figure 29, which presents the boxplot and the violin plot of each case, and Figure 20, which depicts the histogram of the AE of each case.

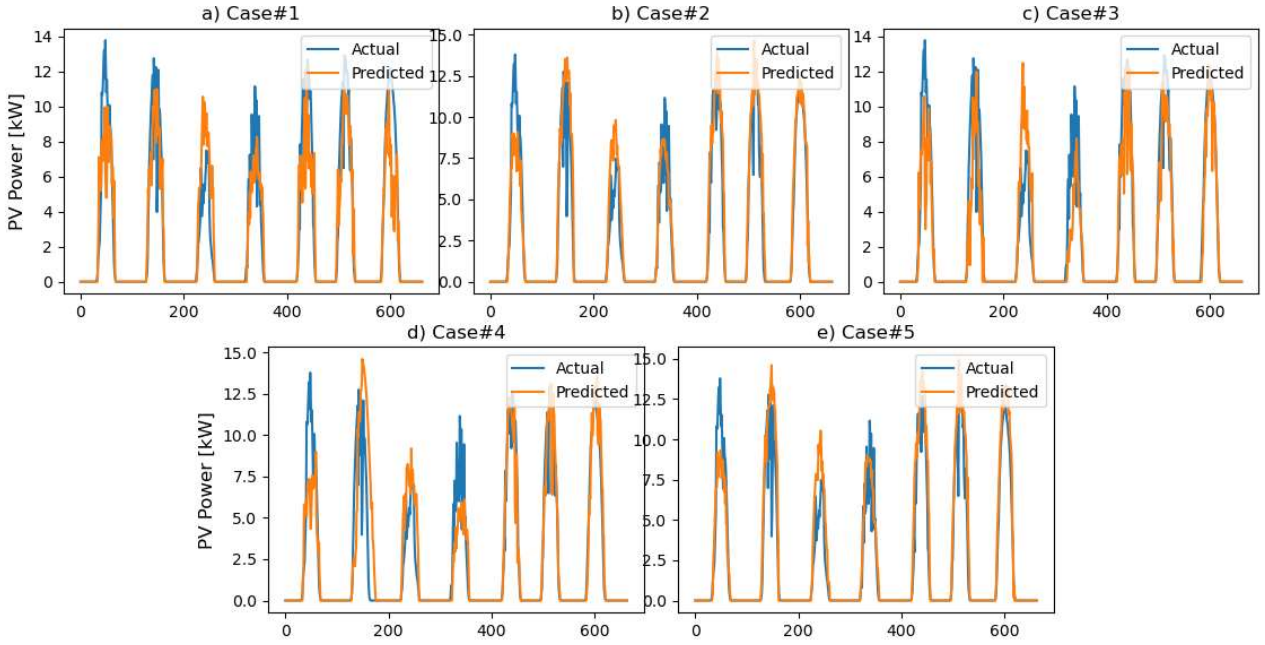

Figure 19. Forecasted and actual PV power generation.

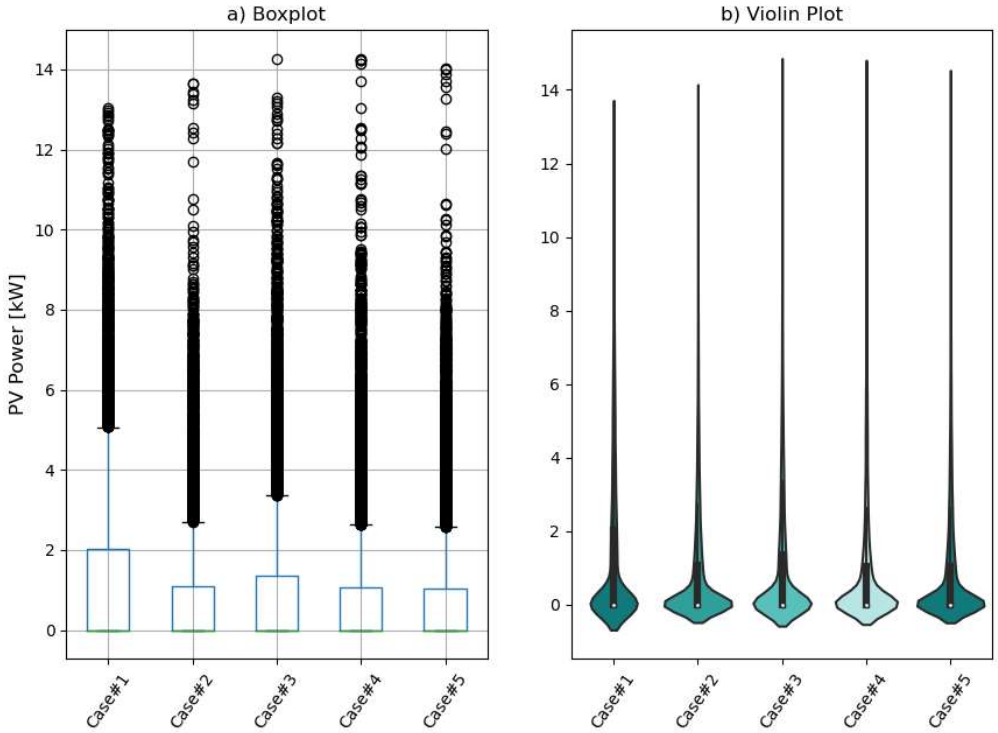

Figure 20. a) Boxplot of AE, b) Violin Plot of AE.

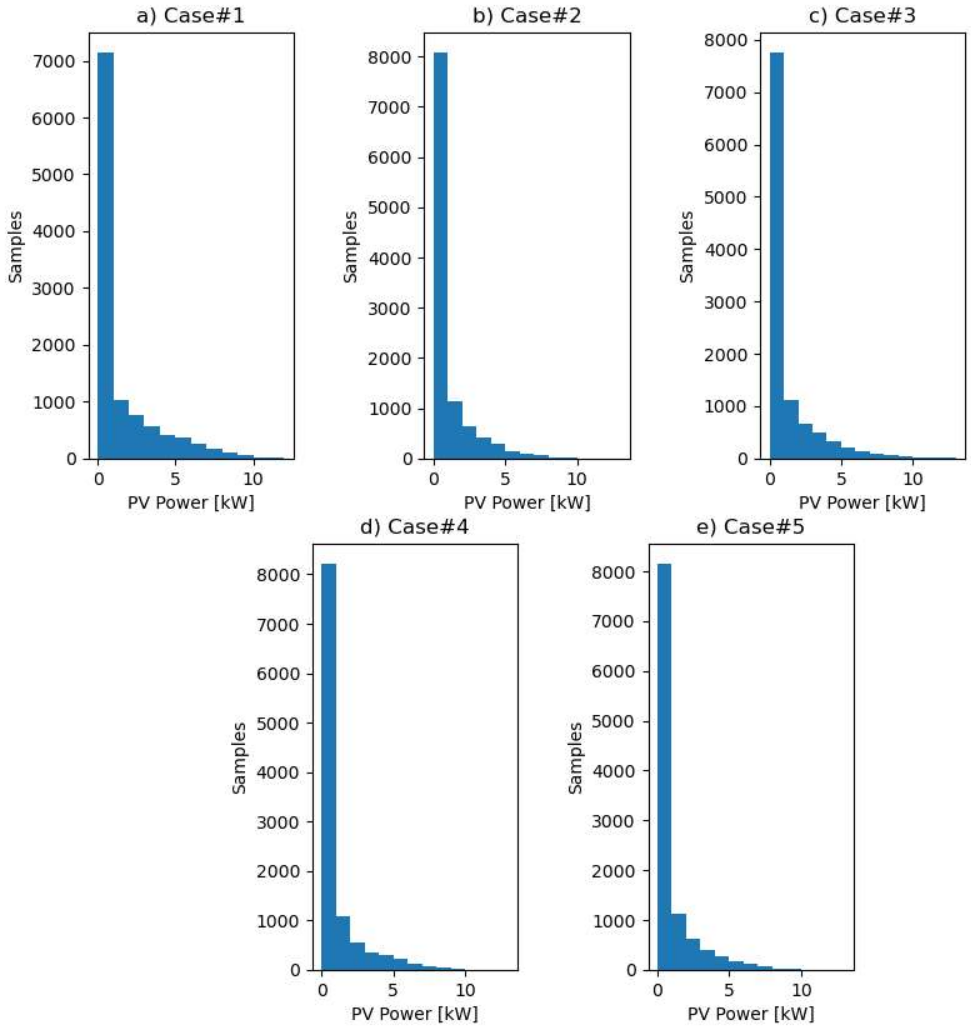

Figure 21. Histograms of AE.

At Figure 21 the timeseries of AE are presented. It is clear that the utilization of PV profiles in the forecasting model can reduce the AE. Since in Case#4 we separate the training and the test sets in three categories according to K_T and we trained three different LSTMs, the timeseries of AE are presented in Figure 22 for each LSTM model.

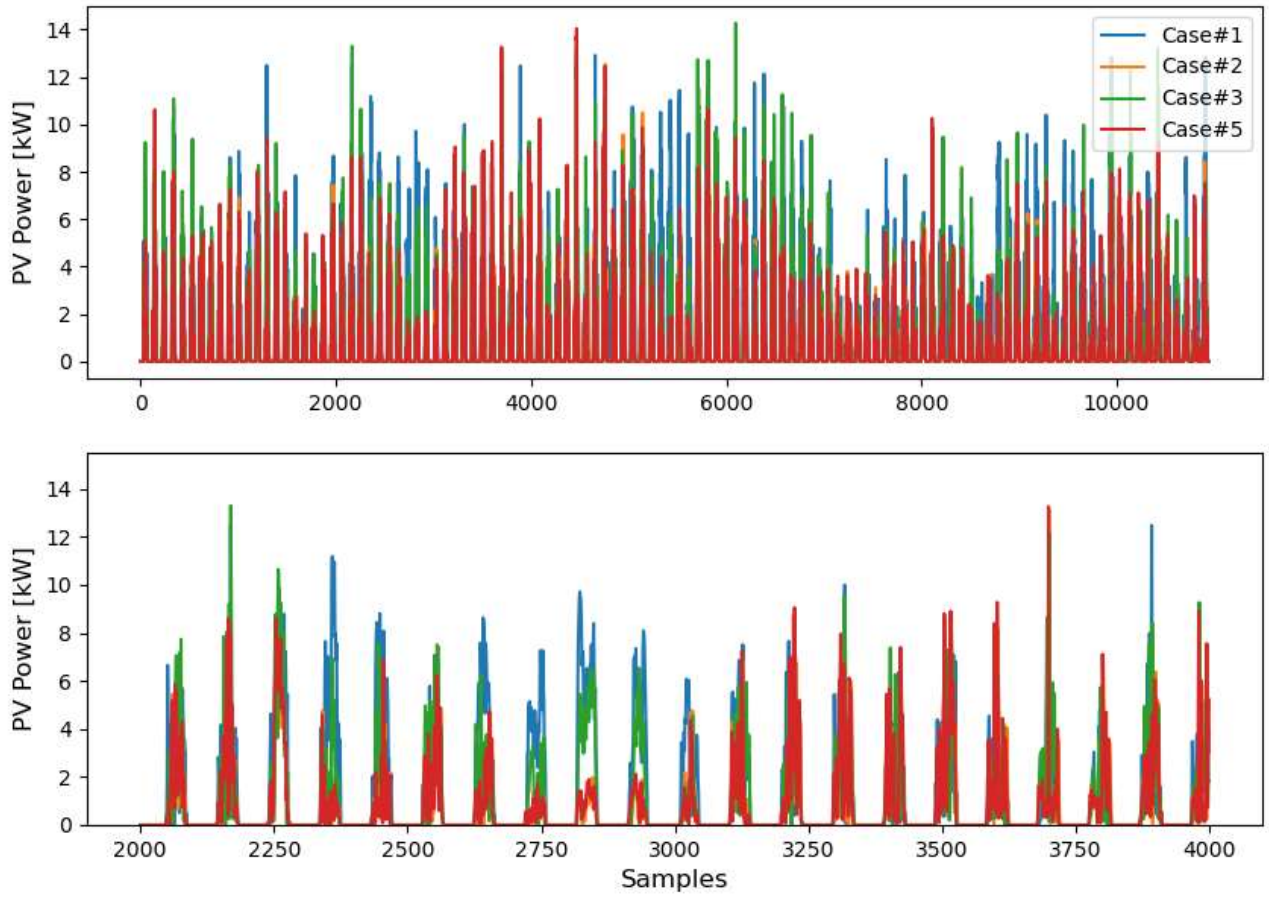

Figure 22. AE timeseries.

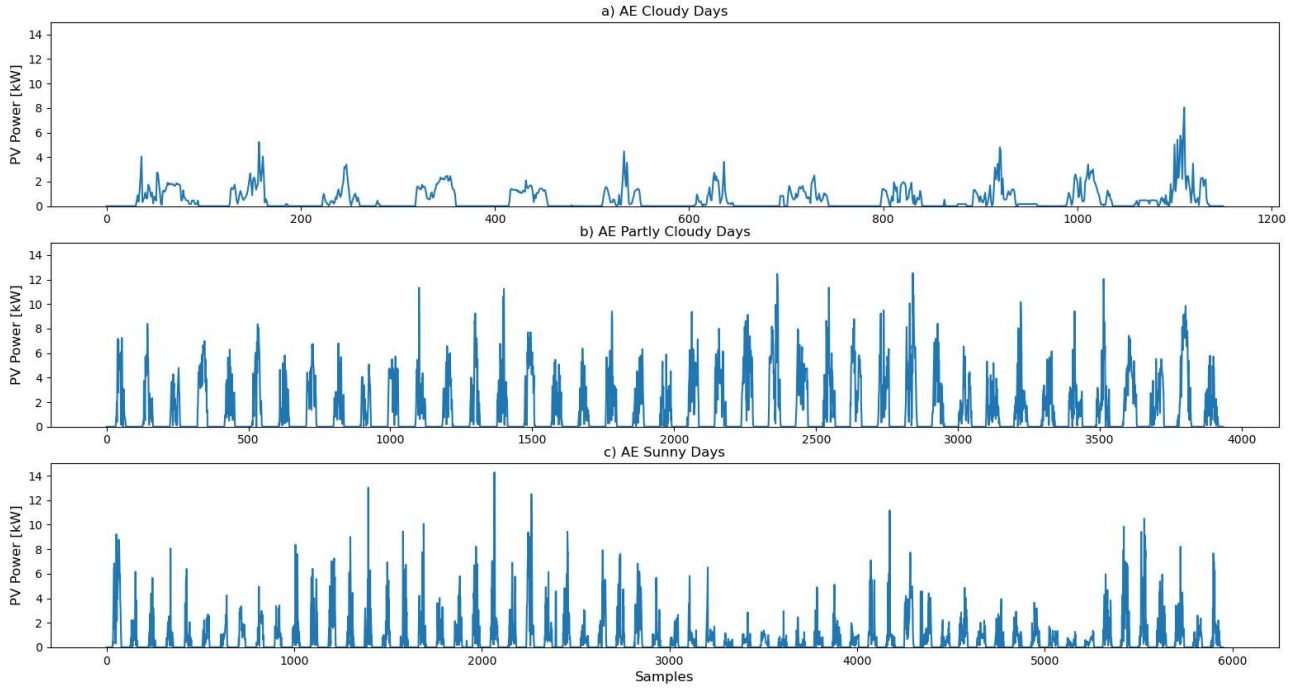

Figure 23. AE timeseries Case#4.

REFERENCES

[1] J. Brownlee, *Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs and LSTMs in Python*. Machine Learning Mastery, 2018.

[2] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," presented at the 3rd International Conference for Learning Representations, San Diego, 2015.

[3] R. Ahmed, V. Sreeram, Y. Mishra, and M. D. Arif, "A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization," *Renewable and Suistainable Energy Reviews,* vol. 124, May 2020.

[4] J. Brownlee, *Introduction to Time Series Forecasting With Python: How to Prepare Data and Develop Models to Predict the Future*. Machine Learning Mastery, 2017.