UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Escola Superior d'Enginyeries Industrial,
Aeroespacial i Audiovisual de Terrassa

Study of electricity generation prediction and control systems in urban environments and energy communities

Document:

Report

Author:

Pablo Alexander Moreno Kübel

Director /Co-director:

Álvaro Luna Alloza / Gerard Laguna Benet

Degree:

Master in Automatic Systems and Industrial Electronics Engineering

Examination session:

Spring, 2022

MASTER FINAL THESIS

Study of electricity generation prediction and control systems in urban environments and energy communities

Pablo Alexander Moreno Kübel

June 2022

Acknowledgements

The author wants to show his gratitude to Jordi Cipriano and Gerard Laguna from the Cimne group for the introduction and the guidance in the exciting field of the application of statistical learning tools in electric energy generation, and also for their follow-up during the realization of the work and for all their suggestions that have contributed to improving the end result of this project.
The author wants also to show his gratitude to the director of this master's final thesis, Álvaro Luna, for his project proposal and for his dedication during the work realization.

Abstract

The scope of the project consists of the process of building, training and validating a model, developed in R language, capable of predicting the energy generation of a photovoltaic plant installation from the data available in the installation that is going to be used as inputs for the model. The model will be developed using statistical learning methods and can predict the forecasts for a specific number of hours introduced by the user, with a maximum of twenty-four hours previsions in advance. The model's predictions will be used in a model predictive control system to couple the forecasted energy generation with the consumption of a nearby building.

All the infrastructure for obtaining and processing data and for the execution of the code is installed and available and it doesn't need any change to be applied to the development of the project, so it isn't the scope of the project it's modification.

The process of building, training and validating the model will be performed with the available data collected during the period between 15/10/2019 and 31/01/2022.

The process of validating the model will be done with statistical indicators capable to evaluate the performance of the model objectively.

The process of using the model to produce real forecasts of energy generation will depend in the availability of the meteorological forecasts of the data of the installation that will be used as inputs of the model and it isn't the scope of this project to obtain this forecasts.

Resumen

El alcance del proyecto consiste en el proceso de construcción, entrenamiento y validación de un modelo, desarrollado en lenguaje R, capaz de predecir la generación de energía de una instalación de planta fotovoltaica a partir de los datos disponibles en la instalación que se va a utilizar como entradas para el modelo. El modelo se desarrollará mediante métodos estadísticos de aprendizaje y podrá predecir las previsiones para un determinado número de horas introducido por el usuario, con un máximo de veinticuatro horas de antelación. Las predicciones del modelo se utilizarán en un sistema de control predictivo del modelo para acoplar la generación de energía pronosticada con el consumo de un edificio cercano.

Toda la infraestructura para la obtención y procesamiento de datos y para la ejecución del código está instalada y disponible y no necesita ningún cambio para ser aplicado al desarrollo del proyecto, por lo que no es el alcance del proyecto su modificación.

El proceso de construcción, entrenamiento y validación del modelo se realizará con los datos disponibles recopilados durante el período comprendido entre el 15/10/2019 y el 31/01/2022.

El proceso de validación del modelo se realizará con indicadores estadísticos capaces de evaluar el desempeño del modelo objetivamente.

El proceso de uso del modelo para producir pronósticos reales de generación de energía dependerá de la disponibilidad de los pronósticos meteorológicos de los datos de la instalación que se usarán como entradas del modelo y no es el alcance de este proyecto obtenerlos.

Contents

List of Figures

List of Tables

Table of nomenclature

BHI	Beam Horizontal Irradiation
$CAMS$	Copernicus Atmosphere Monitoring Service
DHI	Diffuse Horizontal Irradiation
GHI	Global Horizontal Irradiation
POA	Plane of Array
PV	Photovoltaic
TOA	Top Of Atmosphere

Chapter 1

Introduction

1.1 Object

The aim of this Project is to use monitored data from a photovoltaic plant installation stored in a database to generate a model that relates different meteorological data available of the installation and the energy that it produces.
This model will estimate the next twenty-four hours energy generation.
The time step of the data available on the installation is one hour, for each time step that the model is capable to predict in advance is considered one horizon of prediction of the model, so in this case, considering that the time step is one hour and the aim is to predict the generation with twenty-four hours in advance, the objective is to build a model with twenty-four horizons of prediction, considering the zero horizon the one that predicts the immediate generation for the actual hour and the other twenty three horizons used to predict each of the next twenty three hours of the photovoltaic plant installation.

1.2 Justification

Currently, the model developed in the installation have the capability for making zero horizon's forecasts, i.e. is capable to predict the immediate energy generation for the next hour.
Once the model with capability of forecasting a maximum of twenty-four hours generation is developed, the model's predictions will be used in a model predictive control system to couple the forecasted energy generation with the consumption of a nearby building.

1.3 Specifications

The developed model will have as an input the number of horizons forecasted. Therefore, the model will self-adapt to the user horizon needs., in the project will be evaluated a maximum of twenty-four-hour horizons.
The model will be designed to reduce the RMSE given by the forecasts at the minimum value.

1.4 Background

Currently, in this photovoltaic installation, the data is collected and transferred to a database using the software NodeRed, and this stored data is used to preview the immediate photovoltaic plant generation for the actual hour, this corresponds with a model of a zero-horizon capability of prediction.
The data base used is an InfluxDB, which stores the data in temporal format.
All the infrastructure for obtaining and processing data and for the execution of the code to make the

immediate prediction of the energy production is already installed and available.
There is no need to modify the current infrastructure and install any additional device, only to change the code of the prevision to make it able to predict the energy generation with twenty-four horizons. To make it possible also have to be obtained and stored the forecasts of weather data with twenty-four horizons of prediction, i. e. the twenty-four hours weather data prediction updated every hour.

1.4.1 Installation data available

In order to build, train and evaluate the model, will be used the next data from different variables of the photovoltaic installation collected for each hour during the period between 15/10/2019 and 31/01/2022:

- Photovoltaic Plant Generation in kWh.
- Measured Plant Irradiance in W/m2 (source: measure of radiation in PV Plant)
- Measured Temperature in ºC. (source: measure of radiation of nearby weather station)
- Calculated Azimuth angle in º
- Calculated elevation angle in º
- Irradiation on horizontal plane (TOA) at the top of atmosphere (Wh/m2) (source: Copernicus Atmosphere Monitoring Service (CAMS)
- Clear sky global irradiation on horizontal plane (GHI) at ground level in Wh/m2 (source: CAMS)
- Clear sky beam irradiation on horizontal plane (BHI) at ground level in Wh/m2 (source: CAMS)
- Clear sky diffuse irradiation on horizontal plane (DHI) at ground level in Wh/m2 (source: CAMS)
- Clear sky beam irradiation on mobile plane (BNI) following the sun at normal incidence (Wh/m2) (source: CAMS)
- Global irradiation on horizontal plane at ground level (Wh/m2) (source: CAMS)
- Beam irradiation on horizontal plane at ground level (Wh/m2) (source: CAMS)
- Diffuse irradiation on horizontal plane at ground level (Wh/m2) (source: CAMS)
- Beam irradiation on mobile plane following the sun at normal incidence (Wh/m2) (source: CAMS)
- Global Plane Of Array (POA) radiation in Wh/m2 (source: computed with CAMS/PV data)
- Direct Plane Of Array (POA) radiation in Wh/m2 (source: computed with CAMS/PV data)
- Diffuse Plane Of Array (POA) radiation (Wh/m2) (source: computed with CAMS/PV data

1.5 Scope

The scope of the project consists of the process of building, training and validating a model, developed in R language, capable of predicting the energy generation of a photovoltaic plant installation from the data available in the installation that is going to be used as inputs for the model.
All the infrastructure for obtaining and processing data and for the execution of the code is installed and available and it doesn't need any change to be applied to the development of the project, so it isn't the scope of the project it's modification.
The process of using the model to produce real forecasts of energy generation will depend in the availability of the meteorological forecasts of the data of the installation that will be used as inputs of the model and it isn't the scope of this project to obtain this forecasts.
This project will cover the code changes needed to move from forecasting the immediate consumption in the next hour to forecasting the next twenty four hours.

1.6 State of the art

The first step is to develop a model that can be used to predict the energy production of the building, statistical learning (1) will be used to accomplish it.
The objective is to use the available data to calculate a function (f) which is the relation between the consumption which is considered the output (Y) of the model from certain entries (X) that will be explained later.

$$Y = f(X) + \epsilon \tag{1.1}$$

In (1.1) the term Y is the output, also called response, X are the inputs also called predictors, f is a function that gives the relation between the inputs and the output and, finally, ϵ is an error term.
In this case, the function obtained will be used as a predictor, i. e. when the future input data is introduced to the function it will return the predicted response of the future consumption, so the

$$\hat{Y} = \hat{f}(X) + \epsilon \tag{1.2}$$

Where $\hat{f}$ represents the estimate for f and $\hat{Y}$ represents the resulting prediction for Y.
The accuracy of $\hat{Y}$ as a prediction for the output Y depends on two quantities, the called reducible error and the irreducible error. The reducible error comes from the accuracy of f and can be reduced using the most appropriate statistical learning technique to estimate f.
The irreducible error comes from the term ϵ and it can't be reduced because Y can't be be predicted only using X it is also function of ϵ.
So, the aim of using Statistical Learning is to reproduce a model used to estimate f with accuracy to decrease as much as possible the reducible error having in mind that the response will always have an irreducible error that can't be eliminated.
The following steps will be followed during this project:

1. Build the model
2. Train the model
3. Validate the model

The process of building, validating and training the model will be used with an R package called Online forecasting which provides a generalized setup of data and models for online forecasting (2). This package has functionality for time-adaptative fitting of linear regression-based models.

Chapter 2

Model

2.1 Building the model

First, it must be decided the model that is going to be used, most statistical learning methods can be characterized either as parametric or non-parametric methods.

- **Parametric Methods**

Parametric methods involve a two-step model-based approach The first step is to assume some functional form of f, for example, a linear form. So, in this step have to be chosen form of the model that is going to be used to estimate f, which relates the output of the model and the different inputs or predictors chosen to estimate it.
The second step, after a model is selected, is to choose the procedure to fit or train the model with the available data.
The principal advantage of these methods is that they reduce the problem of estimating f to estimate a set of parameters.
One of the disadvantages is that the model that has been chosen possibly will not match the true unknown form of f, which can lead to poor estimations if the estimation is far from the reality.
But, if the model is too complex and the model chosen estimates excessively well the available data used to train it, the model can follow the errors or noise introduced in this available data and not be accurate in estimating the output when the new data is available, this problem is called overfitting.
In conclusion, it is necessary to obtain a flexible model which can fit different functional forms.

- **Non-Parametric Methods**

These methods don't make assumptions about the functional form of f, the objective of this method is to estimate f so it gets as closely as possible to the data points without being so rough or so wiggly.
One advantage of non-parametric methods is that don't make assumptions about the form of f, so don't have the danger to do a bad estimation of f due to the approach of its form with a resulting f so much different from the reality.
The principal disadvantage of these methods is that they need a very large number of observations to fit the model and obtain a good estimation of f due to this type of methods are more complex and doesn't reduce the problem to only estimating a set of parameters as do the parametric models.
In this case, will be used a parametric model for two reasons, the first one is to obtain a simpler model that have more interpretability and the other one is because the lack of available data for 24 hours previsions to fitting the model, so a parametric model, that needs less quantity of data to fit an train the model, will be used. Specifically, will be used a type of parametric model called linear regression.
The linear regression method is used when the relation between inputs and outputs is lineal and is used to predict a quantitative response.
A linear regression model can be described as:

$$Y_{t+k|t} = \beta_{0,k} + \beta_{1,k}u_{1,t+k|t} + ... + \beta_{m,k}u_{m,t+k|t} + \epsilon_{t+k|t} \tag{2.1}$$

The terms of equation (2.1) corresponds to:

- $Y_{t+k|t}$ corresponds to the response variable for a given time point (t) and for an specific horizon (k) used to make the prevision of the response evaluated in the time point t.
- $u_{m,t+k|t}$ corresponds to the input variable for a given time point (t) and for an specific horizon (k) used to make the prevision of the response evaluated in the time point t. Different number of inputs can be used to predict the response in this case is used an m number of inputs.
- $\epsilon_{t+k|t}$ represents the difference between the model prediction and the observed value for the k-step horizon for a given time point (t)
- $\beta_{0,k}$, $\beta_{1,k}$, $\beta_{m,k}$ are the coefficients estimated for each horizon (k) that gives the linear relation between the output and each input used to preview the response. There is one coefficient for each input and each horizon used to do the forecasts.

The onlineforecast package offers two options to estimate the coefficients used to relate the model inputs and the output:

- Least Squares (LS) method: in this method the coefficients are constant over the time and they don't change when new data is introduced into the model, so the model can be represented as seen before in equation (2.1):

  $$Y_{t+k|t} = \beta_{0,k} + \beta_{1,k}u_{1,t+k|t} + ... + \beta_{m,k}u_{m,t+k|t} + \epsilon_{t+k|t} \tag{2.2}$$

- Recursive Least Squares (RLS) method: in this method the coefficients change over the time, and they are updated when new data is introduced into the model, this model can be represented as shown in (2.9):

  $$Y_{t+k|t} = \beta_{0,k,t} + \beta_{1,k,t}u_{1,t+k|t} + ... + \beta_{m,k,t}u_{m,t+k|t} + \epsilon_{t+k|t} \tag{2.3}$$

  In this case $\beta_{0,k,t}$, $\beta_{1,k,t}$, $\beta_{m,k,t}$ are the coefficients estimated for each horizon (k) and updated each time point (t)

The method used in the present project is the RLS method with the aim that the model can be able to adapt and track the changes that the PV installation can suffer over time, such as panel degradation or meteorological changes over that time.
The RLS method, as described before, is a linear method, so the relation between inputs and output must be linear. The package presents a two-step modeling procedure to obtain an approach to model non-linear functional relations between the inputs and the outputs of the model:

- **First stage. Transformation:** the inputs are mapped by some function, potentially into a higher dimensional stage i. e. A linear approach of one non-linear input may result in two or more equivalent inputs in the system to perform this approximation. The functions that the package offers to do these transformations are:

  Low-pass filtering, ls(): low-pass filtering for modelling linear dynamics as a simple RC-model.

- **Second stage.Linear Regression Model:**one of the previous presented linear regression models is applied between the transformed inputs and the output.

When RLS model is used, the level of adaptability of the model can be controlled by setting the called forgetting factor (λ) in the RLS scheme to a value between 0 and 1. For $\lambda\bar{1}$ all past data is equally weighted and for $\lambda 1$ higher weight is put on recent data, the smaller value the faster the model adapts to data.

By optimizing the forgetting factor, the adaptability of the model can be tuned to optimally track the PV installation changes over time.
Finally, some transformation parameters or other parameters related to the regression scheme (e.g. the forgetting factor) should be optimized. These parameters are called "offline parameters" on the package description (**?**). The onlineforecasting package provides a setup to optimize these parameters using heuristic optimization. The default score, which is minimized, is the Root Mean Square Error (RMSE) of the previsions, for example, the forgetting factor can be obtained by solving:

$$min_{\lambda} = \frac{1}{n-k}\sum_{t=1}^{n-k}\left(Y_{t+k|t} - \hat{Y}_{t+k|t}\left(\lambda\right)\right)^{2} \tag{2.4}$$

The terms of equation (2.4) corresponds to:
$Y_{t+k|t}$ corresponds to the response variable for a given time point (t) and for an specific horizon (k).
$\hat{Y}_{t+k|t}\left(\lambda\right)$ corresponds to the prediction value for a given time point (t), for an specific horizon (k) and for a given forgetting factor ( ).
So, in this case, the value of the forgetting factor is searched to obtain the minimum value of the RMSE between the prediction and the response.
The data used to build the model is presented below.

2.1.1 Output of the model

The output correspond with the variable that want to be predicted, in this case the objective is to predict the PV plant generation given forecast data from different variables of the plant that will be used as inputs.

PV Plant Energy Generation

The PV plant Energy Generation will be used as model's output, the available data of this variable is shown below:

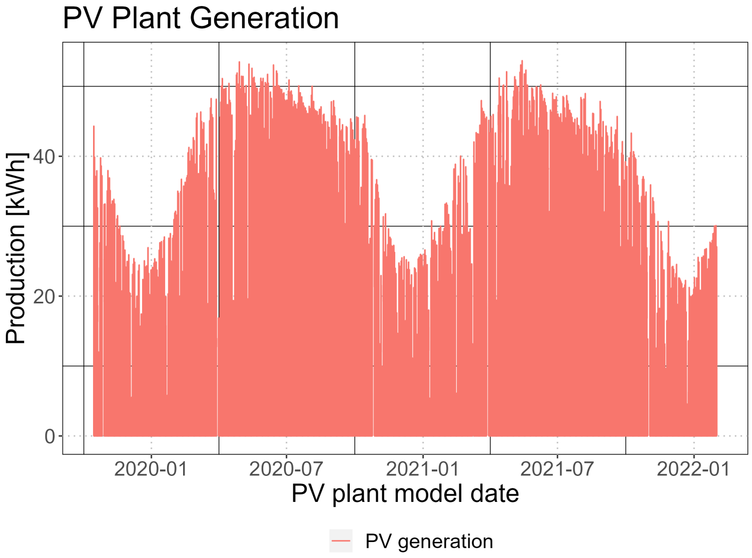


Figure 2.1: PV Plant Generation data available.

In Figure 2.1 is displayed the production of the plant in all the period of the available data, below ten days of production are displayed in order to observe the behaviour of the production for a given day:

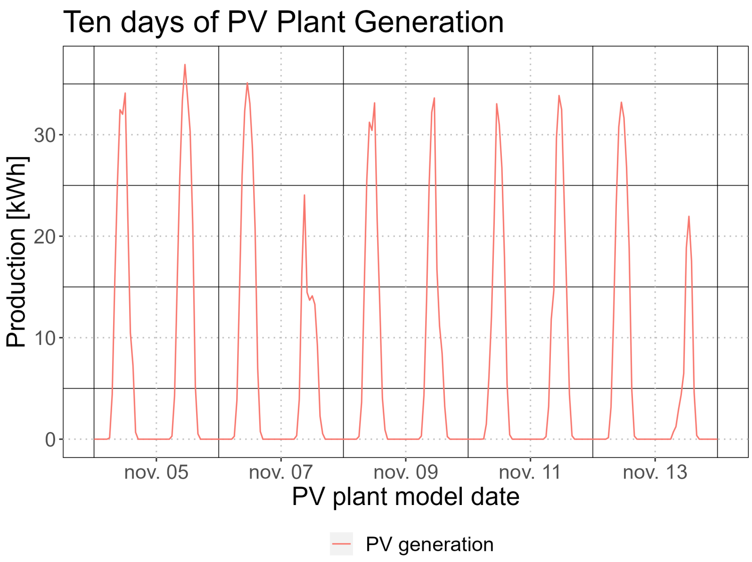


Figure 2.2: Ten days of the PV Plant Generation data available.

In Figure 2.2, ten days of the year 2019 PV plant's production can be observed.

2.1.2 Inputs of the model

Global plane of Array (POA) Irradiation

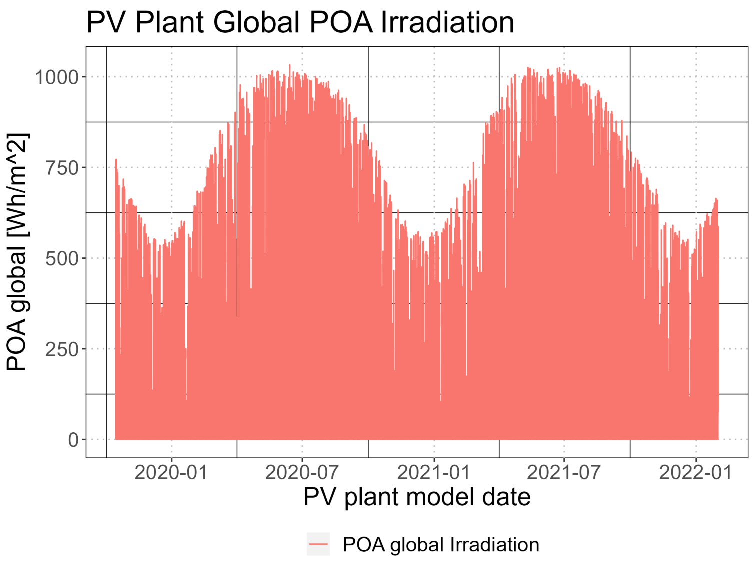


Figure 2.3: PV Plant Global POA Irradiation data available.

In Figure 2.3 is displayed the Global POA Irradiation in all the period of the available data, below it's shown the Global POA Irradiation of the same period shown in Figure 2.2:

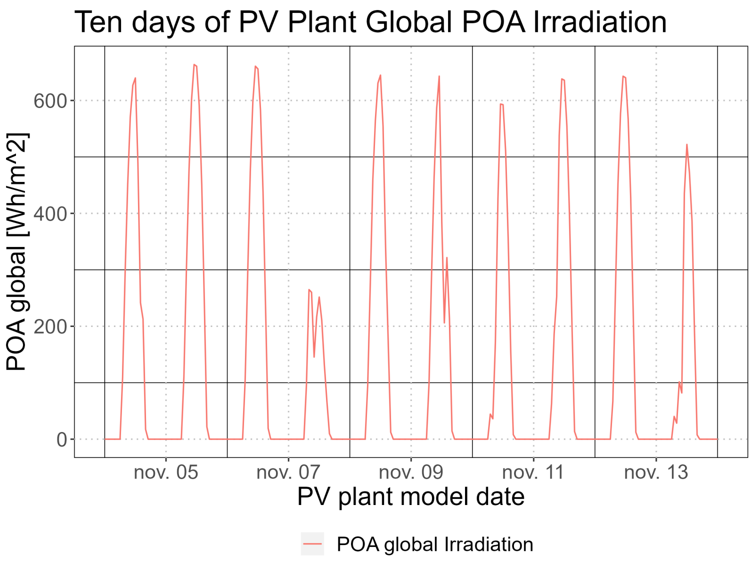


Figure 2.4: Ten days of the Global POA Irradiation PV Plant Generation data available.

The mathematical relation between Energy Generation and Global POA Irradiation will be obtained in the resulting model after training it as the relation between output and first input.

Direct plane of Array (POA) Irradiation

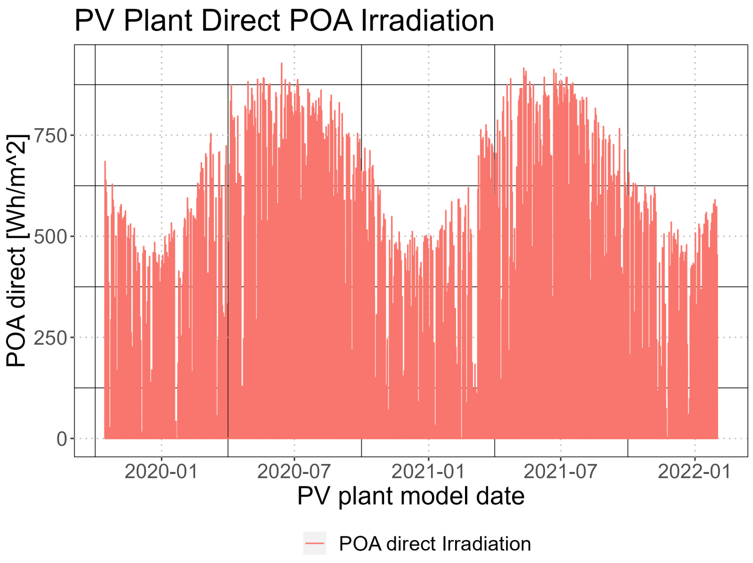


Figure 2.5: PV Plant Direct POA Irradiation data available.

In Figure 2.5 is displayed the Direct POA Irradiation in all the period of the available data, below is shown the Global POA Irradiation of the same period shown in Figure 2.2:

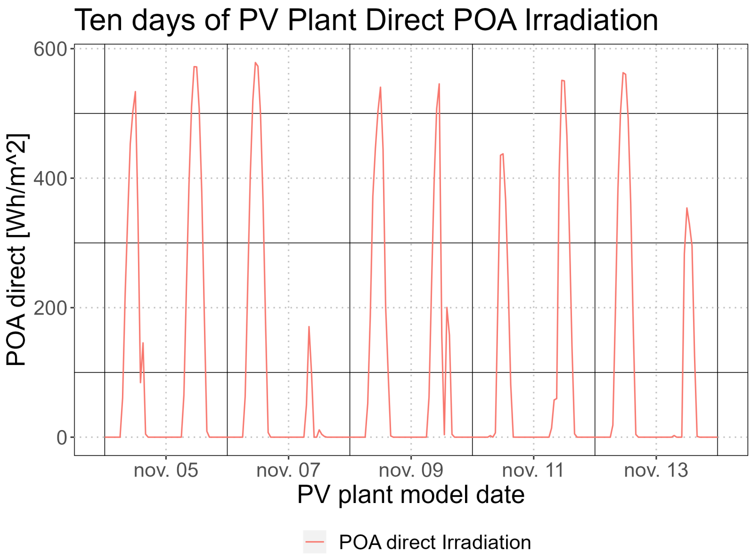


Figure 2.6: Ten days of the Direct POA Irradiation PV Plant Generation data available.

The mathematical relationship between Energy Generation and Direct POA Irradiation will be obtained in the resulting model after training it as the relation between output and second input.

Temperature combined with Global POA

It's wanted to find the relation between the ambient temperature and the PV plant generation because this temperature affects to the the temperature of the photovoltaic panels responsible for generating energy, and the panel's temperature affects the energy that these can produce.
In other hand, as the model used is an RLS, the relation obtained between the generation and the temperature will be linear, but it isn't always true because when there isn't Irradiation at the plant the panels stops producing energy but the temperature not necessary goes to zero degrees, it depends on the season.
So, in resume, it can still being temperature at the plant when there is no production and if the relation between this two variables obtained is linear, this can introduce an error giving previsions of energy generation where there isn't Irradiation. To avoid this error instead of using the temperature as an input, it will be used the product of the temperature and the Global POA of the plant, so when there isn't Irradiation this factor will return a value of zero and this will reduce the error that is introduced by using this variable as an input to predict generation.
The temperature of the PV plant is shown below:

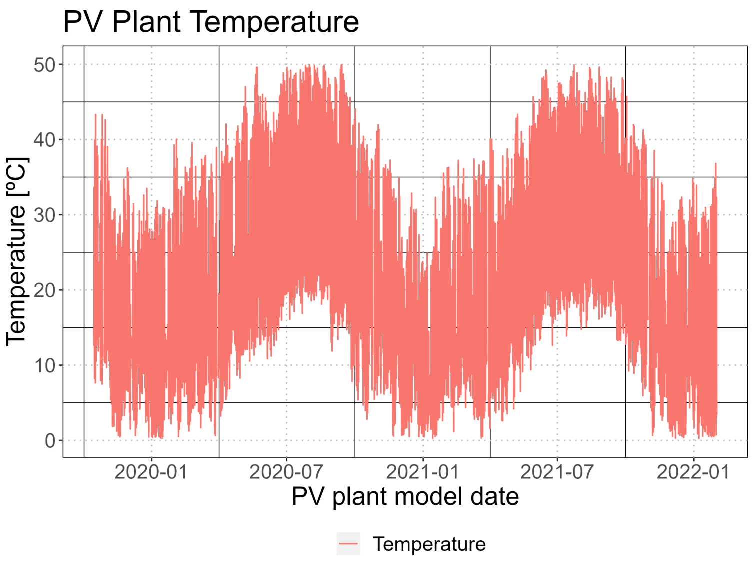


Figure 2.7: PV Plant Temperature data available.

In Figure 2.7 is displayed the temperature in all the period of the available data and can be observed that the temperature trend changes according to the season in which the data is obtained, below is shown the temperature of the same period shown in Figure 2.2:

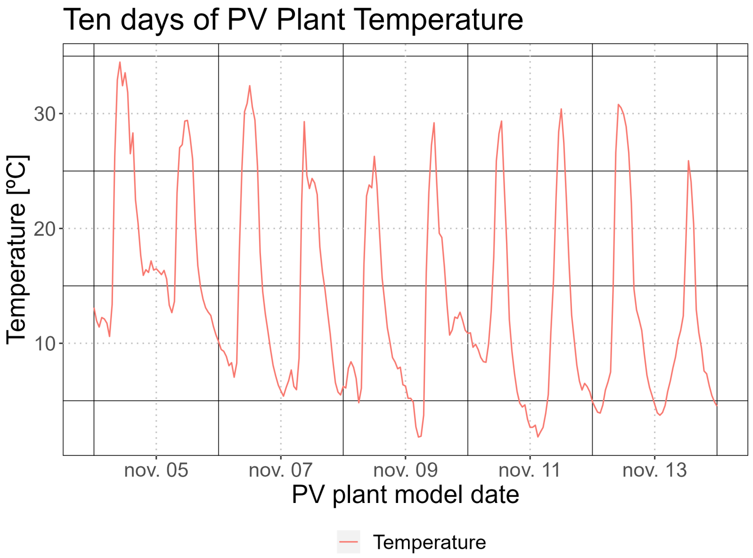


Figure 2.8: Ten days of the Temperature data available.

Next, the graph of the result of the multiplication between Temperature and Global POA is shown:

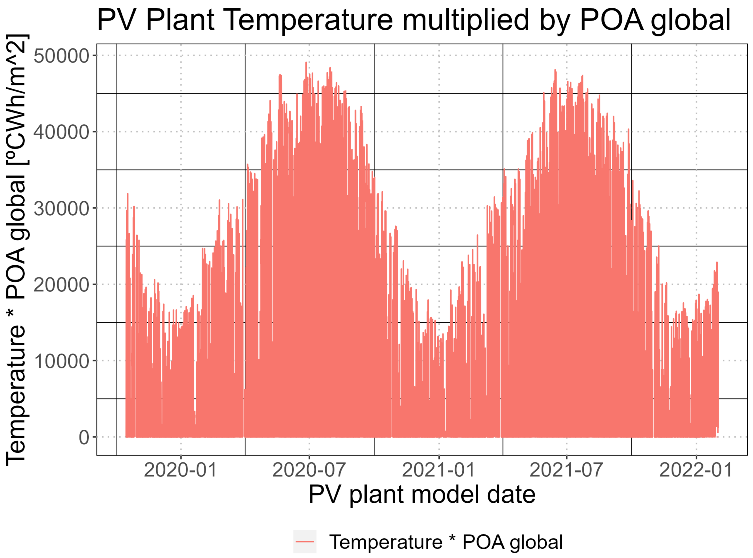


Figure 2.9: Product between the PV Plant Temperature and the Global POA data available.

To observe the effect that has this combination between these two variables, ten days of the result of this product are shown below:

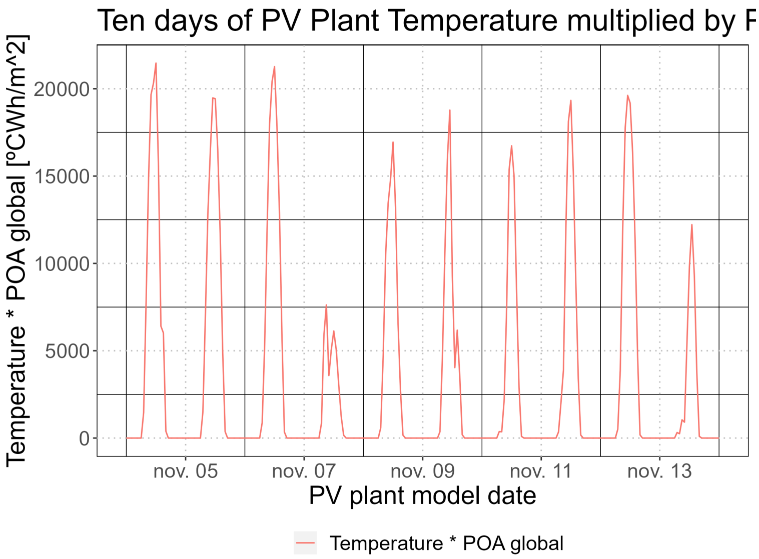


Figure 2.10: Ten days of the product between the PV Plant Temperature and the Global POA data available.

As can be seen in Figure 2.10 with this operation is possible to find the relation between temperature and generation by filtering the effect of the temperature in the prediction when there isn't Irradiation.
The mathematical relation between Energy Generation and the multiplication between Global POA Irradiation and Temperature will be obtained in the resulting model after training it as the relation between output and third input.
Finally, this effect can be observed by comparing the values obtained by the multiplication and the PV plant energy generation:

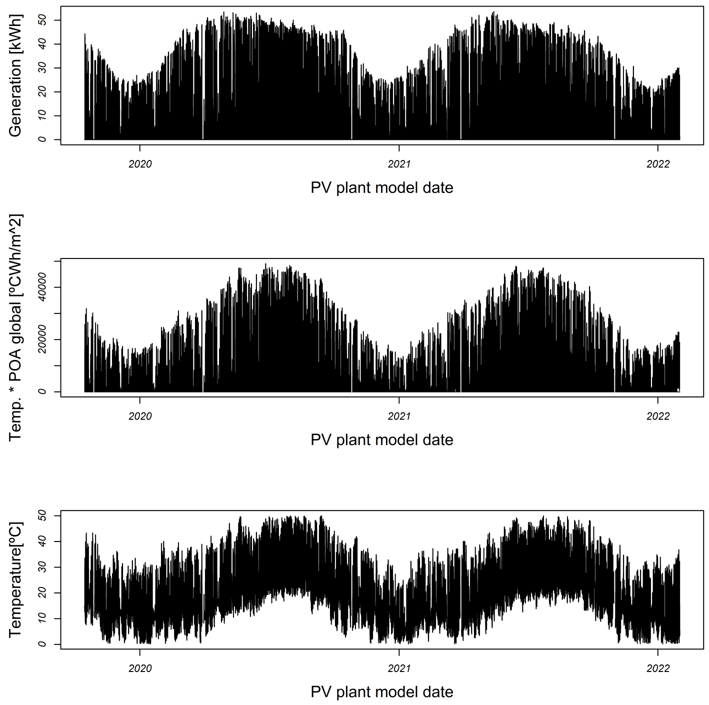


Figure 2.11: Comparison between Temperature without POA Global multiplication and Temperature with POA Global multiplication and Energy

And ten days of this comparison are shown in Figure 2.12 for a better understanding of this better linear related variable:

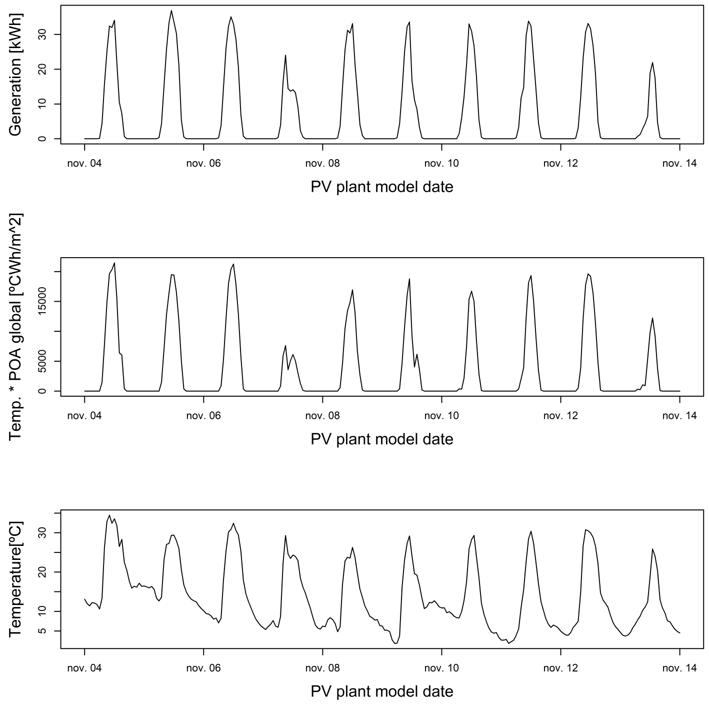


Figure 2.12: Ten days of comparison between Temperature without POA Global multiplication and Temperature with POA Global multiplication and Energy

Azimuth Angle

In the same way as the temperature, the relation between energy generation and the azimuth angle of the sun isn't linear for the whole day, if it isn't Irradiation at the plant the panels will not generate energy although the value of the azimuth angle isn't zero. So, in the same way that has been done with the temperature, the azimuth angle will be multiplied with Global POA values for each time instant and the product between these two variables will be used as an input to the model. First of all, it's shown the graphic of the azimuth angle available data:

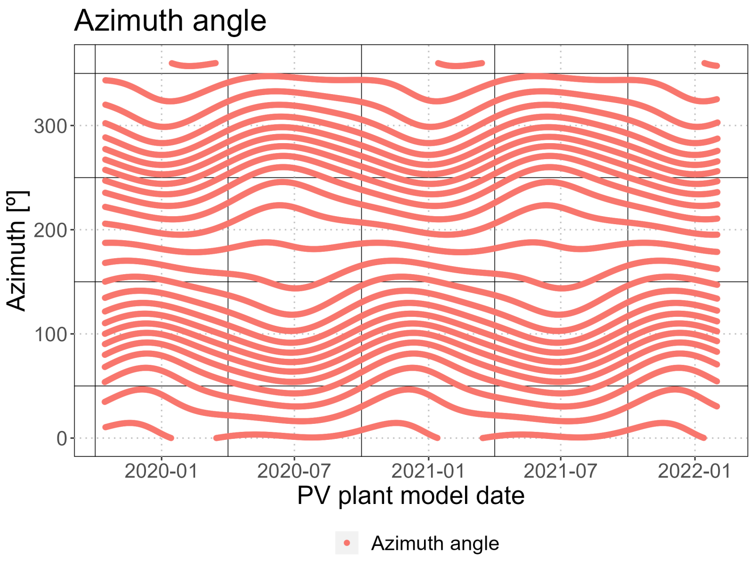


Figure 2.13: PV Plant Azimuth angle data available.

In Figure 2.13 is displayed the data points of the azimuth angle in all the period of the available, below is shown the azimuth angle of the same period shown in Figure 2.2:

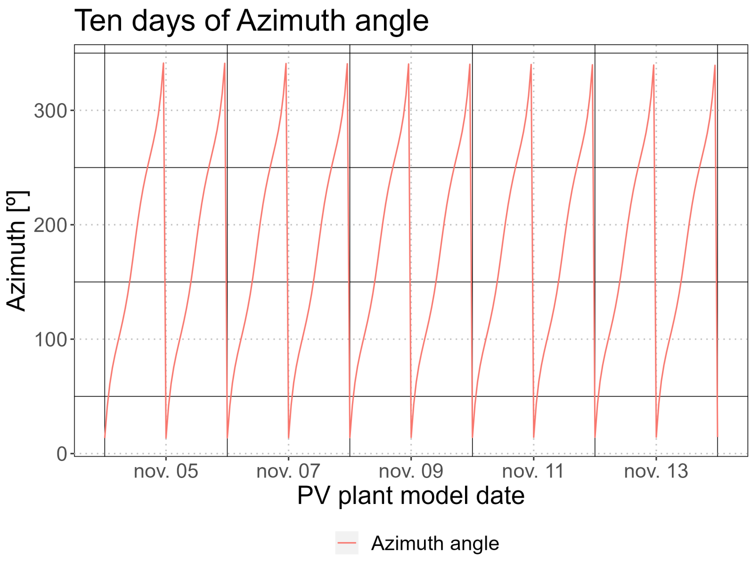


Figure 2.14: Ten days of the Azimuth angle data available.

In the following graphic is shown the new variable obtained by multiplying the azimuth by the global POA data:

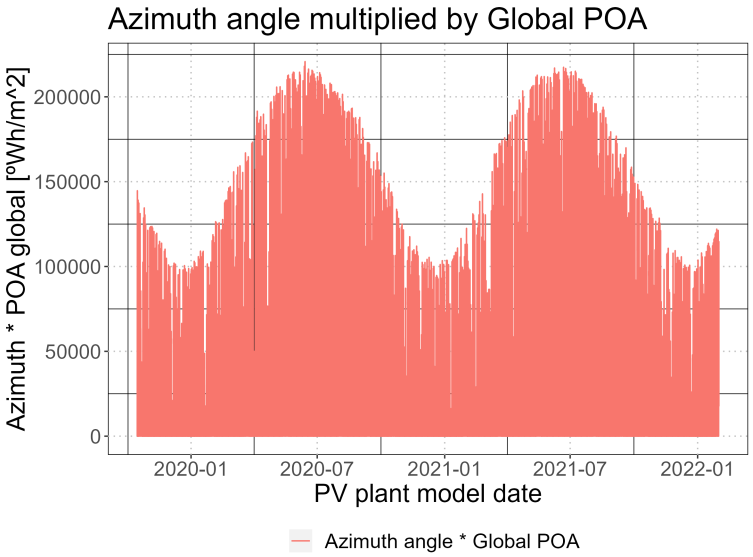


Figure 2.15: Azimuth angle data available multiplied by Global POA

And the same ten days that Figure 2.14 are displayed in Figure 2.16:

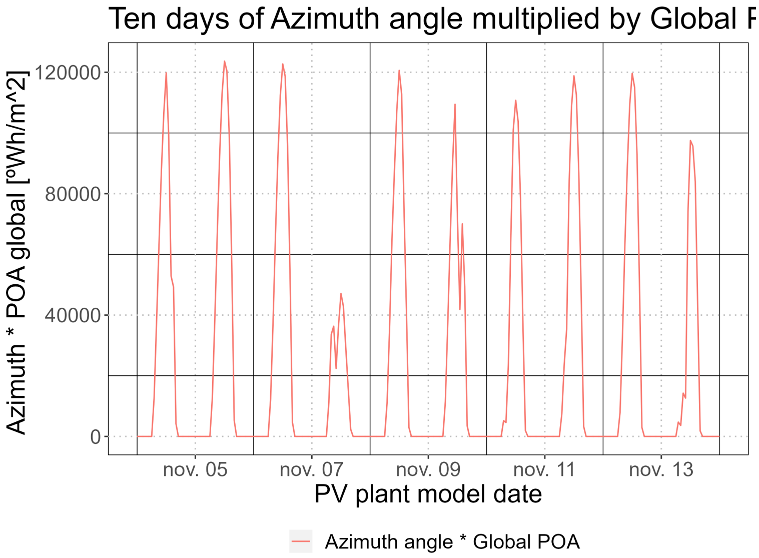


Figure 2.16: Ten days of Azimuth angle data available multiplied by Global POA

Finally, to see that the the the new variable obtained from the multiplication between the azimuth and the POA Global is more strictly linear, in Figure 2.17 are shown this new variable, the azimuth and the energy generation of the PV Plant:

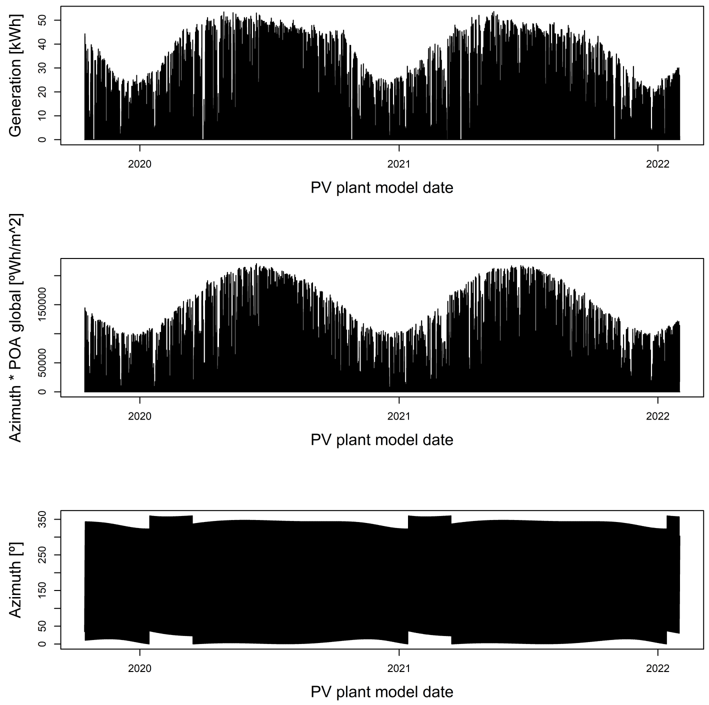


Figure 2.17: Comparison between Azimuth without POA Global multiplication and Azimuth with POA Global multiplication and Energy

And ten days of this comparison are shown in Figure 2.18 for a better understanding of this better linear related variable:

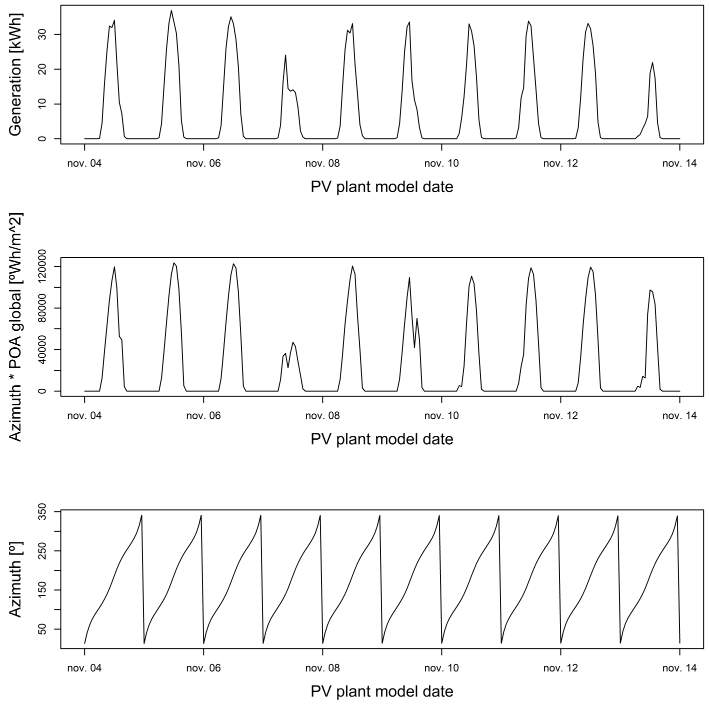


Figure 2.18: Ten days of comparison between Azimuth without POA Global multiplication and Azimuth with POA Global multiplication and Energy

On other hand, as can be seen in Figure 2.13 and Figure 2.14 this variable is periodic over the year. Therefore, in order to linearize it, the azimuth angle will be expressed using the Fourier Transformation with two harmonics, considering that each harmonic has two coefficients.

The Fourier transformation will be performed with the Fourier Series transformation function available in the onlineforecasting package previously commented. Doing this transformation the azimuth angle input will be transformed into four inputs corresponding with the first harmonic coefficients (sin1 and cos1) and the second harmonic coefficients (sin2 and cos2) , so as commented previously, the one-dimensional azimuth input will be transformed into a function with four-dimensional stage corresponding with the Fourier series coefficients.

This transformation means that the aim is to find the relationship between power generation (output) and the first two harmonics of the azimuth angle function.

The variables $sin1$ and $cos1$ corresponds with the first harmonic coefficients and $sin2$ and $cos2$ correspond with the second harmonic coefficients as shown in equation (2.5):

$$Azimuth(t) = a_0 + \sum_{i}^{\infty}(a_n cos\frac{n\pi t}{T} + b_n sin\frac{n\pi t}{T}) \tag{2.5}$$

The approach with two harmonics is given in expression (2.6)

$$Azimuth(t) = (a_1 cos\frac{2\pi t}{T} + b_1 sin\frac{2\pi t}{T}) + (a_2 cos\frac{4\pi t}{T} + b_2 sin\frac{4\pi t}{T}) \tag{2.6}$$

It will be created another model that also includes the sun elevation angle, this model is created with the objective to compare the results of predicting the energy generation with a model including only the sun azimuth angle and the model including both sun azimuth angle and sun elevation angle.

Elevation Angle

The Sun Elevation Angle's data points of all the period of available data is displayed in Figure 2.19:

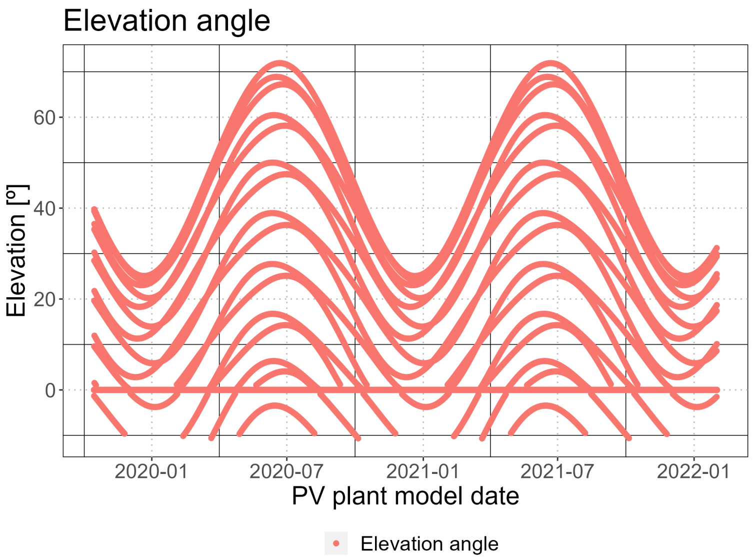


Figure 2.19: PV Plant Sun Elevation angle data available.

This variable will be treated in the same way as the azimuth angle, first of all, this variable will be multiplied by Global POA for each time instant and the product between these two variables will be used as an input to the model with to have an input that is linear related with the energy generation.

After this treatment, the comparison with the elevation angle, the elevation angle multiplied by global POA, and the energy generation is shown below:

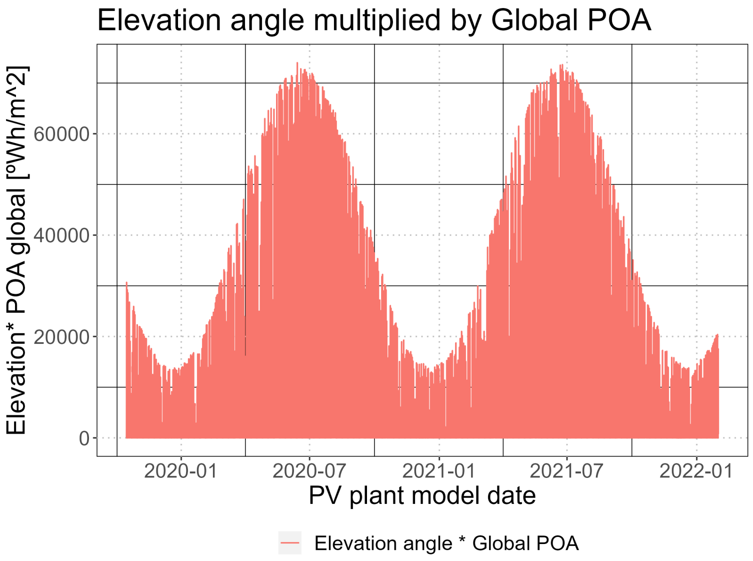


Figure 2.20: PV Plant Sun Elevation angle combined with POA Global Irradiation.

Finally, to see that the new variable obtained from the multiplication between the azimuth and the POA Global is more strictly linear, in Figure 2.21 are shown this new variable, the azimuth and the energy generation of the PV Plant:

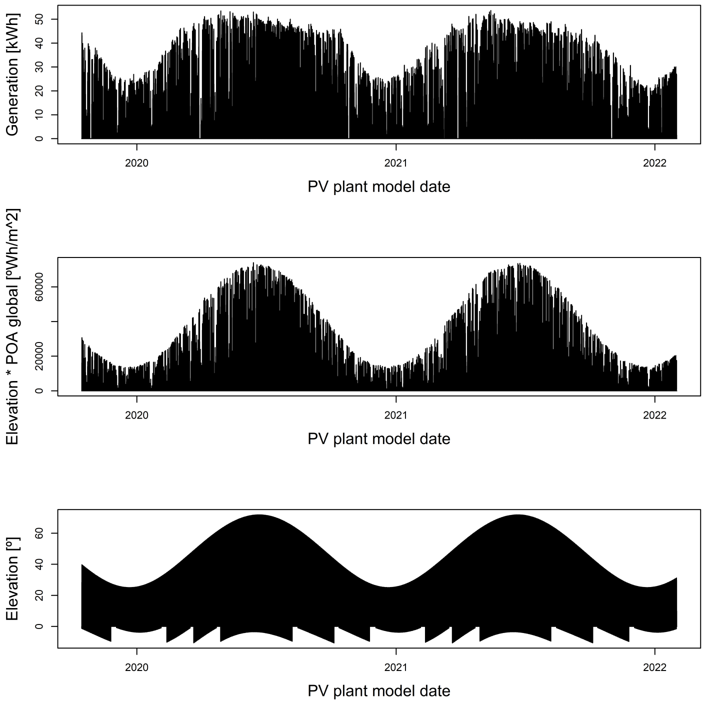


Figure 2.21: Comparison between Elevation without POA Global multiplication and Elevation with POA Global multiplication and Energy

And ten days of this comparison are shown in Figure 2.22 for a better understanding of this better linear related variable:

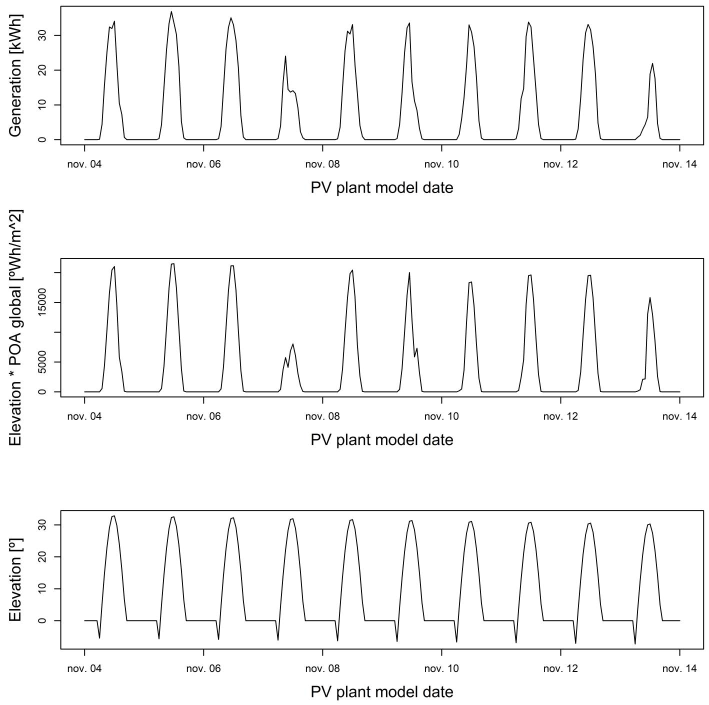


Figure 2.22: Ten days of comparison between Elevation without POA Global multiplication and Elevation with POA Global multiplication and Energy

In the Figure 2.22 can be seen that the relation between elevation angle and energy generation is more linear than the case of the azimuth angle, but also can be seen that when the plant isn't generating energy, the elevation angle takes negatives values that can introduce error to the model, so the variable obtained from the product between elevation angle and Global POA will be used as an input.

In Figure 2.22 is also observed that the elevation angle over the days is a periodic function so, in the same way done with azimuth angle, this variable will be expressed using the Fourier Transformation with two harmonics, considering that each harmonic has two coefficients.

This transformation means that the aim is to find the relationship between power generation (output) and the first two harmonics of the elevation angle function.

The variables $sin1$ and $cos1$ corresponds with the first harmonic coefficients and $sin2$ and $cos2$ corresponds with the second harmonic coefficients as shown in equation (2.7):

$$Elevation(t) = a_0 + \sum_{i}^{\infty}(a_n cos\frac{n\pi t}{T} + b_n sin\frac{n\pi t}{T}) \tag{2.7}$$

The approach with two harmonics is given in expression (2.8)

$$Elevation(t) = (a_1 cos\frac{2\pi t}{T} + b_1 sin\frac{2\pi t}{T}) + (a_2 cos\frac{4\pi t}{T} + b_2 sin\frac{4\pi t}{T}) \tag{2.8}$$

Relation between the output and the inputs of the model

In order to have a visual orientation of the relation between the energy generation and the inputs in the model without the sun elevation angle input, they are shown in the same image in figure 2.23:

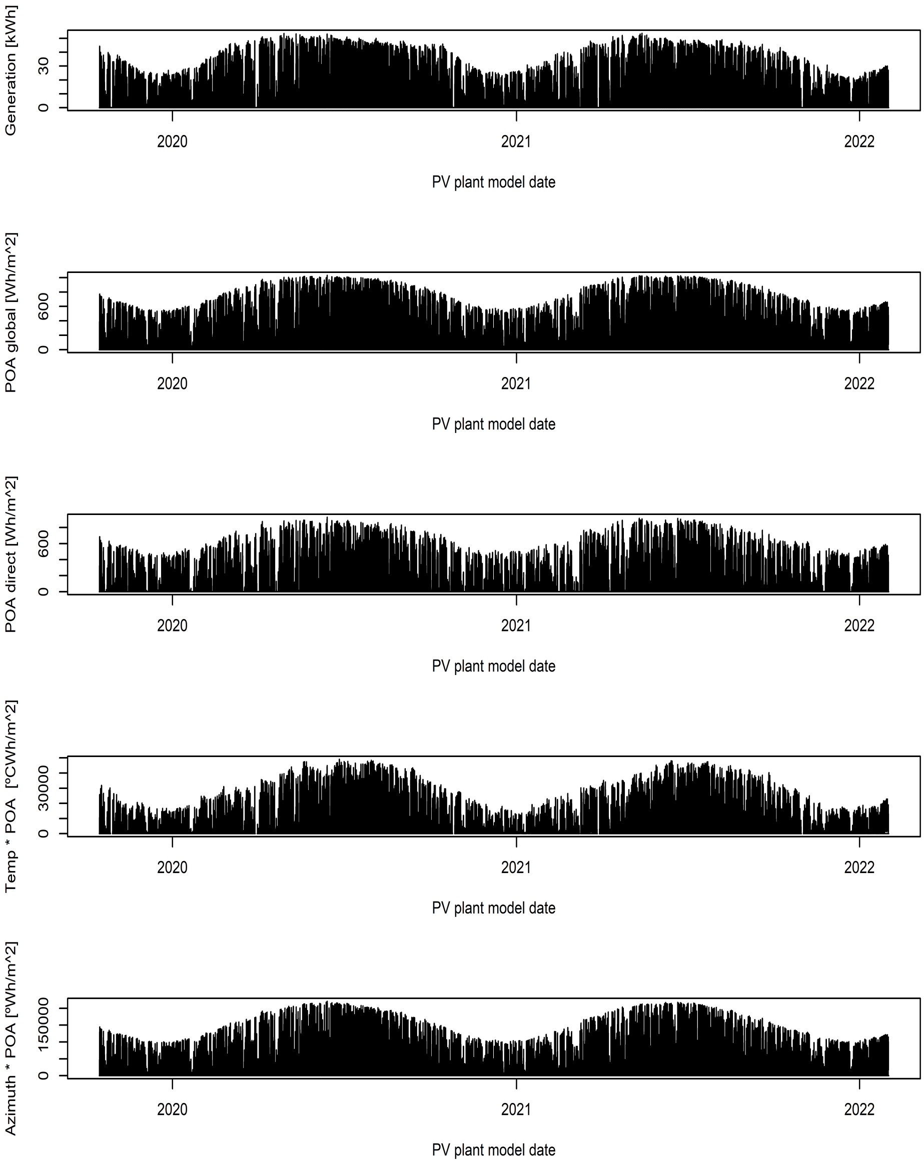


Figure 2.23: Otuput and Inputs of the model without elevation angle added as input.

Below is shown the output and the inputs of the same period shown in Figure 2.2:

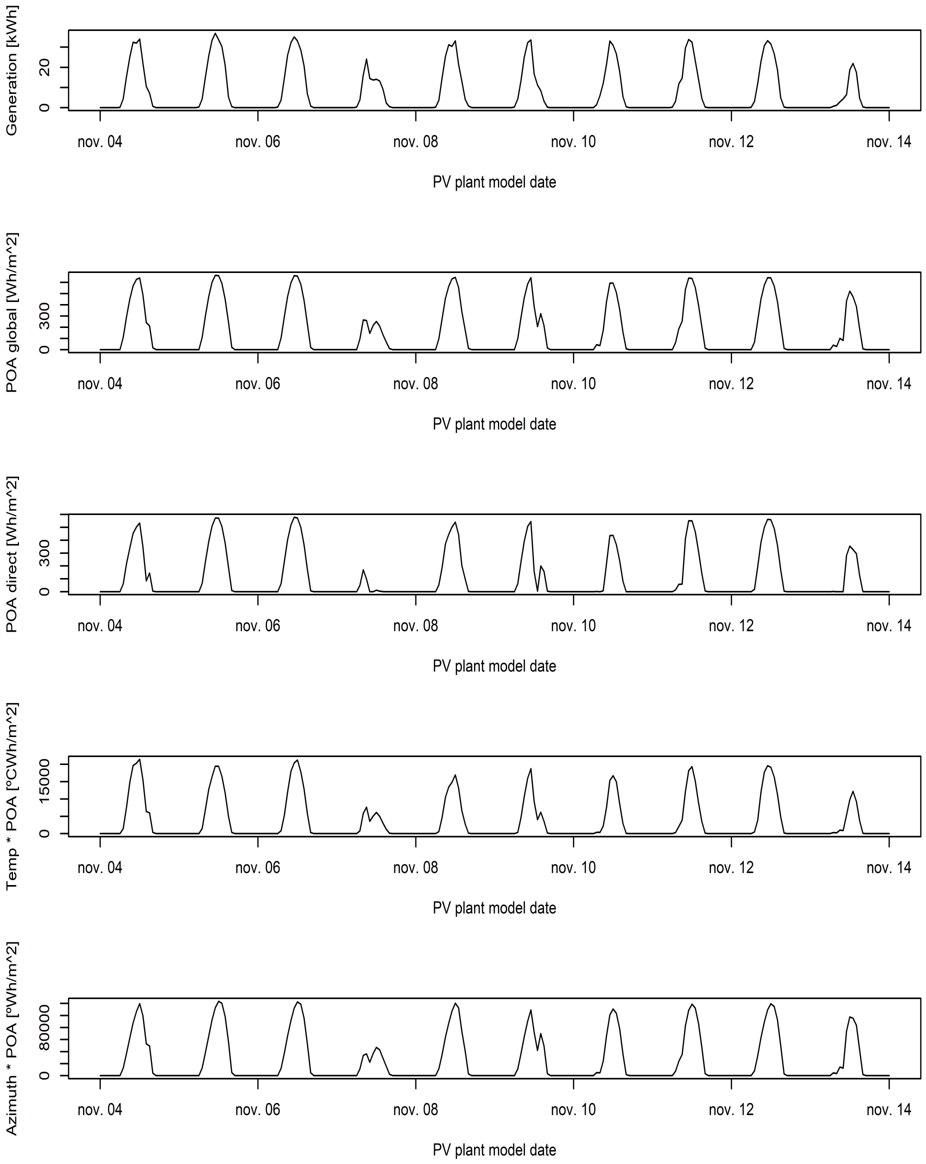


Figure 2.24: Ten days of Otuput and Inputs of the model without elevation angle added as input.

Finally, the mathematical relation between this variables will be obtained in the resulting model after training it as the relation between output and fourth, fifth, sixth and seventh inputs as for each instant of time:

$$\begin{aligned} G_{t+k|t} = \beta_{0,k,t} + \beta_{1,k,t} POA, g_{t+k|t} + \beta_{2,k,t} POA, d_{t+k|t} + \beta_{3,k,t} PT_{t+k|t} + \\ + \beta_{4,k,t} PA, sin1_{t+k|t} + \beta_{5,k,t} PA, cos1_{t+k|t} + \beta_{6,k,t} PA, sin2_{t+k|t} + \\ + \beta_{7,k,t} PA, cos2_{t+k|t} + \epsilon_{t+k|t} \end{aligned} \tag{2.9}$$

Where G corresponds to the PV plant energy generation, POA, g corresponds to the global POA irradiation, POA, d corresponds to the direct POA Irradiation, PT corresponds to the global POA Irradiation multiplied by the temperature, $PA, sin1$, $PA, cos1$, $PA, sin2$ and $PA, cos2$ corresponds to the two firsts coefficients of the Fourier transformation of the sun azimuth angle multiplied by the global POA irradiation.

And in order to have a visual orientation of the relation between the energy generation and the inputs in the model with the sun elevation angle input, they are shown in the same image in figure 2.25:

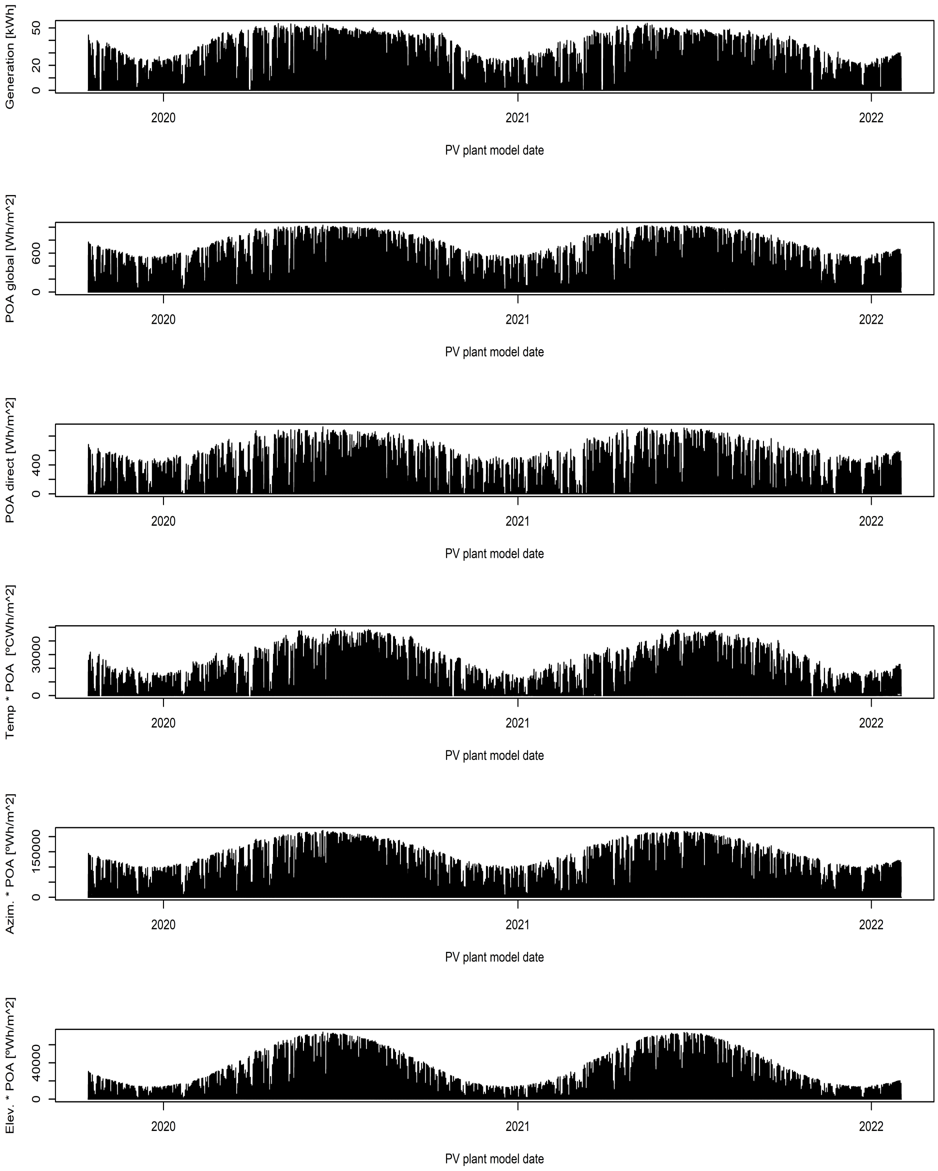


Figure 2.25: Otuput and Inputs of the model with elevation angle added as input.

Below is shown the output and the inputs of the same period shown in Figure 2.26:

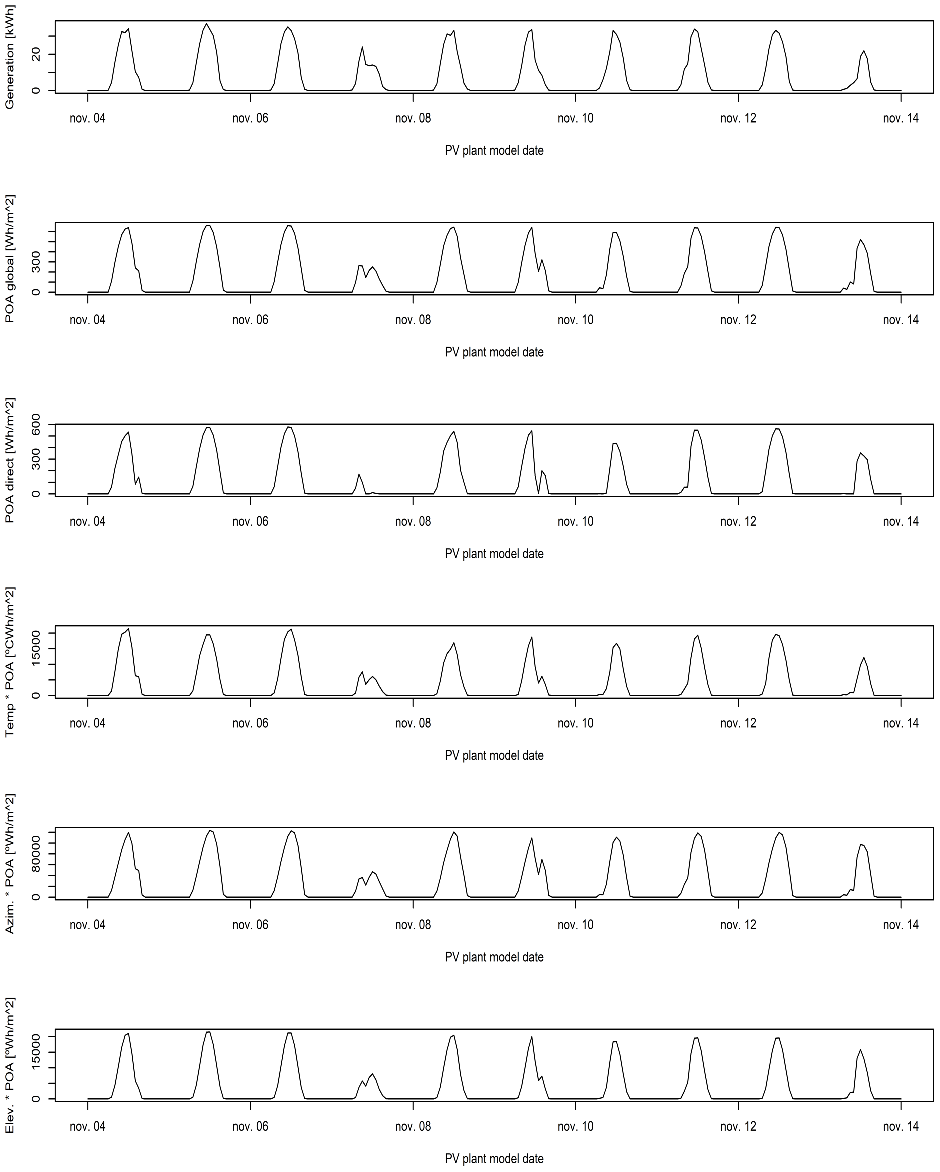


Figure 2.26: Ten days of Otuput and Inputs of the model with elevation angle added as input.

Finally, the mathematical relation between these variables will be obtained in the resulting model after training it as the relation between output and fourth, fifth, sixth, seventh, eighth, ninth, tenth and eleventh inputs for each instant of time:

$$\begin{aligned} G_{t+k|t} = \beta_{1,k,t}POA,g_{t+k|t} + \beta_{2,k,t}POA,d_{t+k|t} + \beta_{3,k,t}PT_{t+k|t}+ \\ +\beta_{4,k,t}PA,sin1_{t+k|t} + \beta_{5,k,t}PA,cos1_{t+k|t} + \beta_{6,k,t}PA,sin2_{t+k|t}+ \\ +\beta_{7,k,t}PA,cos2_{t+k|t} + \beta_{8,k,t}PE,sin1_{t+k|t} + \beta_{9,k,t}PE,cos1_{t+k|t}+ \\ +\beta_{10,k,t}PE,sin2_{t+k|t} + \beta_{11,k,t}PE,cos2_{t+k|t} + \epsilon_{t+k|t} \end{aligned} \tag{2.10}$$

Where G corresponds to the PV plant energy generation, POA,g corresponds to the global POA irradiation, POA,d corresponds to the direct POA Irradiation, PT corresponds to the global POA Irradiation multiplied by the temperature, $PA,sin1$, $PA,cos1$, $PA,sin2$ and $PA,cos2$ corresponds to the two firsts coefficients of the Fourier transformation of the sun azimuth angle multiplied by the global POA irradiation and $PE,sin1$, $PE,cos1$, $PE,sin2$ and $PE,cos2$ corresponds to the two firsts coefficients of the Fourier transformation of the sun elevation angle multiplied by the global

POA irradiation.

2.1.3 Setting up the model

The first step is to create a new object, this can be done with the following instruction:

- model <- forecastmodel$new

After this step, is needed to add the output and the inputs of the model.
Adding the output:

- model$output <- "OutputName"
  In this case, the aim is to forecast the power Generation of the photovoltaic installation, so it set as the model output.

Adding the inputs:

- model$add_inputs (
  Input_1 = "Input_1_Name"
  Input_2 = "Input_2_Name"
  ...
  Input_n = "Input_n_Name")
  As mentioned before, the inputs for the model without elevation angle combined with global POA are:
  - Input 1: Plane of Array (POA) global
  - Input 2: Plane of Array (POA) direct
  - Input 3: Temperature multiplied wit Global POA
  - Input 4: Azimuth angle sin1 multiplied by Global POA
  - Input 5: Azimuth angle cos1 multiplied by Global POA
  - Input 6: Azimuth angle sin2 multiplied by Global POA
  - Input 7: Azimuth angle cos2 multiplied by Global POA

  An the inputs for the model in wich is added the elevation angle combined with global POA are:
  - Input 1: Plane of Array (POA) global
  - Input 2: Plane of Array (POA) direct
  - Input 3: Temperature multiplied by Global POA
  - Input 4: Azimuth angle sin1 multiplied by Global POA
  - Input 5: Azimuth angle cos1 multiplied by Global POA
  - Input 6: Azimuth angle sin2 multiplied by Global POA
  - Input 7: Azimuth angle cos2 multiplied by Global POA
  - Input 8: Elevation angle sin1 multiplied by Global POA
  - Input 9: Elevation angle cos1 multiplied by Global POA
  - Input 10: Elevation angle sin2 multiplied by Global POA

  - Input 11: Elevation angle cos2 multiplied by Global POA

The next step is to add the regression step parameters, in this case only will be added the forgetting factor previously described:

- model$add_regprm("rls_prm(lamada=lamda_value")

As mentioned before, the package offers the possibility to optimise the offline parameters as the forgetting factor, so the optimization of this parameter will be done with the next instruction:

- model$add_prmbounds(lambda = c(lower, init, upper))
  The parameter bounds for the optimisation are described by:
  - Lower: lower value that can take the optimised parameter.
  - Init: initial value that takes the optimised parameter.
  - Upper: upper value that can take the optimised parameter.

The result of the optimisation is stored in:

- model$regprm
  So, adding this instruction in the model will return the value optimised for the forgetting factor stored in regprm.

Finally, the number of horizons used in the model must be added, this is done with the next instruction:

- model$kseq <- first_horizon:last_horizon
  In this case, the first horizon is equal to zero for the immediate prediction and the last horizon is equal to twenty-four, corresponding to the twenty-four hour energy generation forecast.

2.2 Training the model

The next step is to fit the rls model in order to train it and make it capable of predict the energy generation of the PV plant from the forecast data of the inputs.
The instruction to fit the model is the following:

- fit <- rls_fit(prm, model, Dtrain, srcorefun, returnanalysys)
  In this case the arguments of the instruction correspond to:
  - prm: optimized parameters, in this case correspond the forgetting factor (λ) optimized.
  - model:the model wich is wanted to train, in this case the model previously described.
  - Dtrain: data used to train the model, this correspond to the actual value of each input and their next twenty-four hours values for each time point, in equation (2.12 it's expressed the matrix of the training data for an specific time point. The Dtrain matrix used in the model without the elevation angle combined with global POA used as input is:

$$Dtrain = \begin{pmatrix} output_{1,0|t} & output_{1,1|t} & output_{1,2|t} & \dots & output_{1,23|t} \\ input_{1,0|t} & input_{1,1|t} & input_{1,2|t} & \dots & input_{1,23|t} \\ input_{2,0|t} & input_{2,1|t} & input_{2,2|t} & \dots & input_{2,23|t} \\ input_{3,0|t} & input_{3,1|t} & input_{3,2|t} & \dots & input_{3,23|t} \\ input_{4,0|t} & input_{4,1|t} & input_{4,2|t} & \dots & input_{4,23|t} \\ input_{5,0|t} & input_{5,1|t} & input_{5,2|t} & \dots & input_{5,23|t} \\ input_{6,0|t} & input_{6,1|t} & input_{6,2|t} & \dots & input_{6,23|t} \\ input_{7,0|t} & input_{7,1|t} & input_{7,2|t} & \dots & input_{7,23|t} \end{pmatrix} \tag{2.11}$$

And the Dtrain matrix used in the model with the elevation angle combined with global POA used as input is:

$$Dtrain = \begin{pmatrix} output_{1,0|t} & output_{1,1|t} & output_{1,2|t} & \cdots & output_{1,23|t} \\ input_{1,0|t} & input_{1,1|t} & input_{1,2|t} & \cdots & input_{1,23|t} \\ input_{2,0|t} & input_{2,1|t} & input_{2,2|t} & \cdots & input_{2,23|t} \\ input_{3,0|t} & input_{3,1|t} & input_{3,2|t} & \cdots & input_{3,23|t} \\ input_{4,0|t} & input_{4,1|t} & input_{4,2|t} & \cdots & input_{4,23|t} \\ input_{5,0|t} & input_{5,1|t} & input_{5,2|t} & \cdots & input_{5,23|t} \\ input_{6,0|t} & input_{6,1|t} & input_{6,2|t} & \cdots & input_{6,23|t} \\ input_{7,0|t} & input_{7,1|t} & input_{7,2|t} & \cdots & input_{7,23|t} \\ input_{8,0|t} & input_{8,1|t} & input_{8,2|t} & \cdots & input_{8,23|t} \\ input_{9,0|t} & input_{9,1|t} & input_{9,2|t} & \cdots & input_{9,23|t} \\ input_{10,0|t} & input_{10,1|t} & input_{10,2|t} & \cdots & input_{10,23|t} \\ input_{11,0|t} & input_{11,1|t} & input_{11,2|t} & \cdots & input_{11,23|t} \end{pmatrix} \tag{2.12}$$

The term $output_{1,0|t}$ corresponds to the current (horizon 0) value of the output 1, corresponding to the PV Plant Energy Generation, at the time instant evaluated, the term $output_{1,1|t}$ corresponds to value of the output 1 for the next hour (horizon 1) at time instant evaluated, the term $output_{1,2|t}$ corresponds to value of the output 1 for the second hour (horizon 2) from the time instant evaluated and the term $output_{1,23|t}$ corresponds to value of the output 1 for the twenty-four-hour (horizon 23) from the time instant evaluated.

The term $input_{1,0|t}$ corresponds to the current (horizon 0) value of the input 1, corresponding to the Global POA, at the time instant evaluated, the term $input_{1,1|t}$ corresponds to value of the input 1 for the next hour (horizon 1) at time instant evaluated, the term $input_{1,2|t}$ corresponds to value of the input 1 for the second hour (horizon 2) from the time instant evaluated and the term $input_{1,23|t}$ corresponds to value of the input 1 for the twenty-four hour (horizon 23) from the time instant evaluated.

It is considered the same for the other inputs ($Input_2$ to $Input_7$) in the model without the sun elevation angle as input and ($Input_2$ to $Input_1 1$) in the model with the sun elevation angle as input.

Each instant point corresponds to one hour of the data available to build the model previously described.

- returnanalysis: is used to obtain a list with the values of the horizons to perform the analysis of the results.

2.3 Validating the model

Once the model is built and trained, the next step is to perform an analysis of the results given by the model in order to validate it.

First of all, the horizons obtained by the model which doesn't include the sun elevation angle are shown in Figure 2.27:

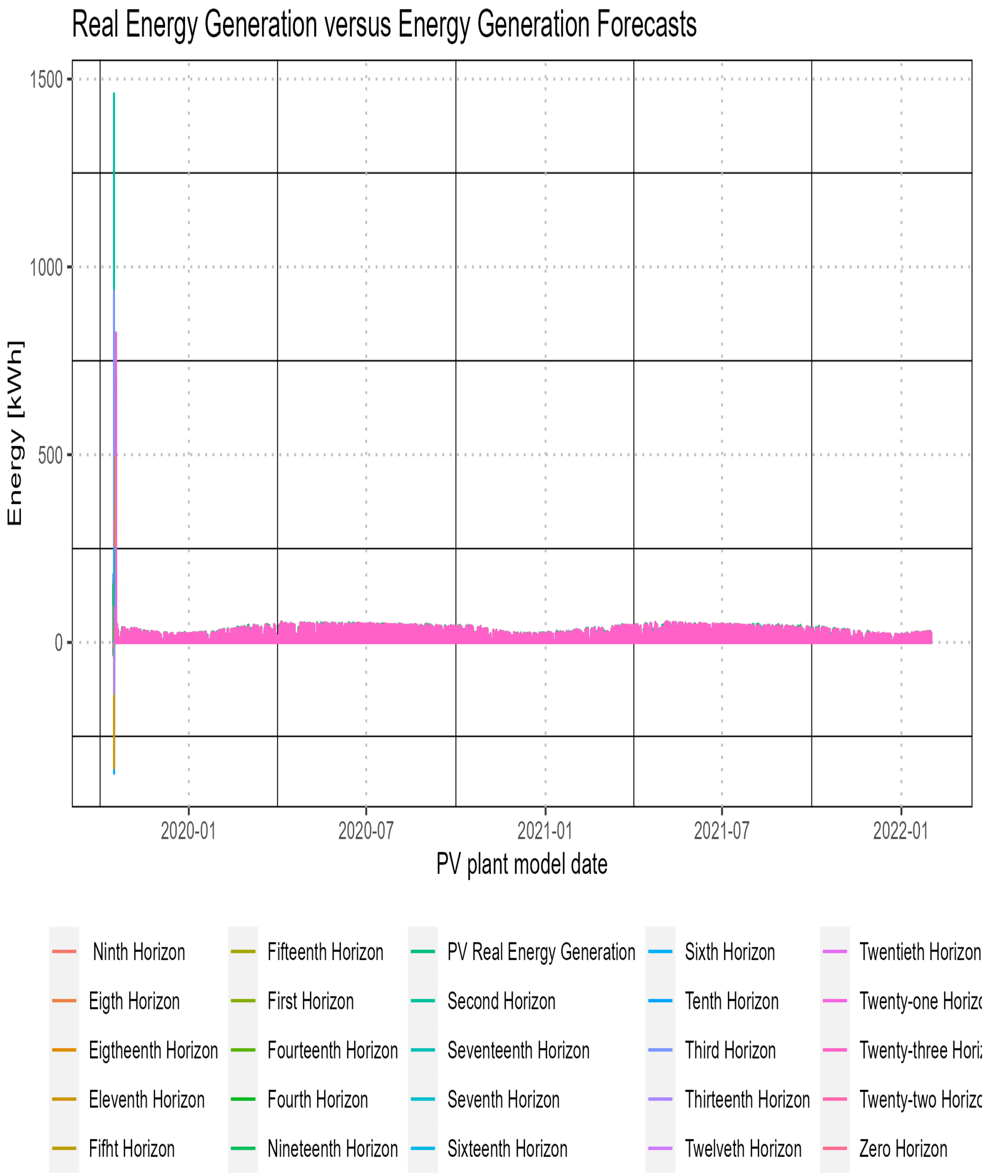


Figure 2.27: All forecasts predicted by the model without sun elevation angle.

In this case, the zero horizon is the immediate prediction of the energy generation done the last hour, the first horizon is the prediction of energy generation that was done one hour before the last, the second horizon is the prediction of energy generation that was done two before and the same for each horizon until the twenty-third horizon which corresponds with the prediction of energy generation that was done twenty-four hours before. In Figure 2.27 is observed that the predicted values aren't correct for the early values of the forecasts, ten days of the forecasts of the year 2019 are shown below:

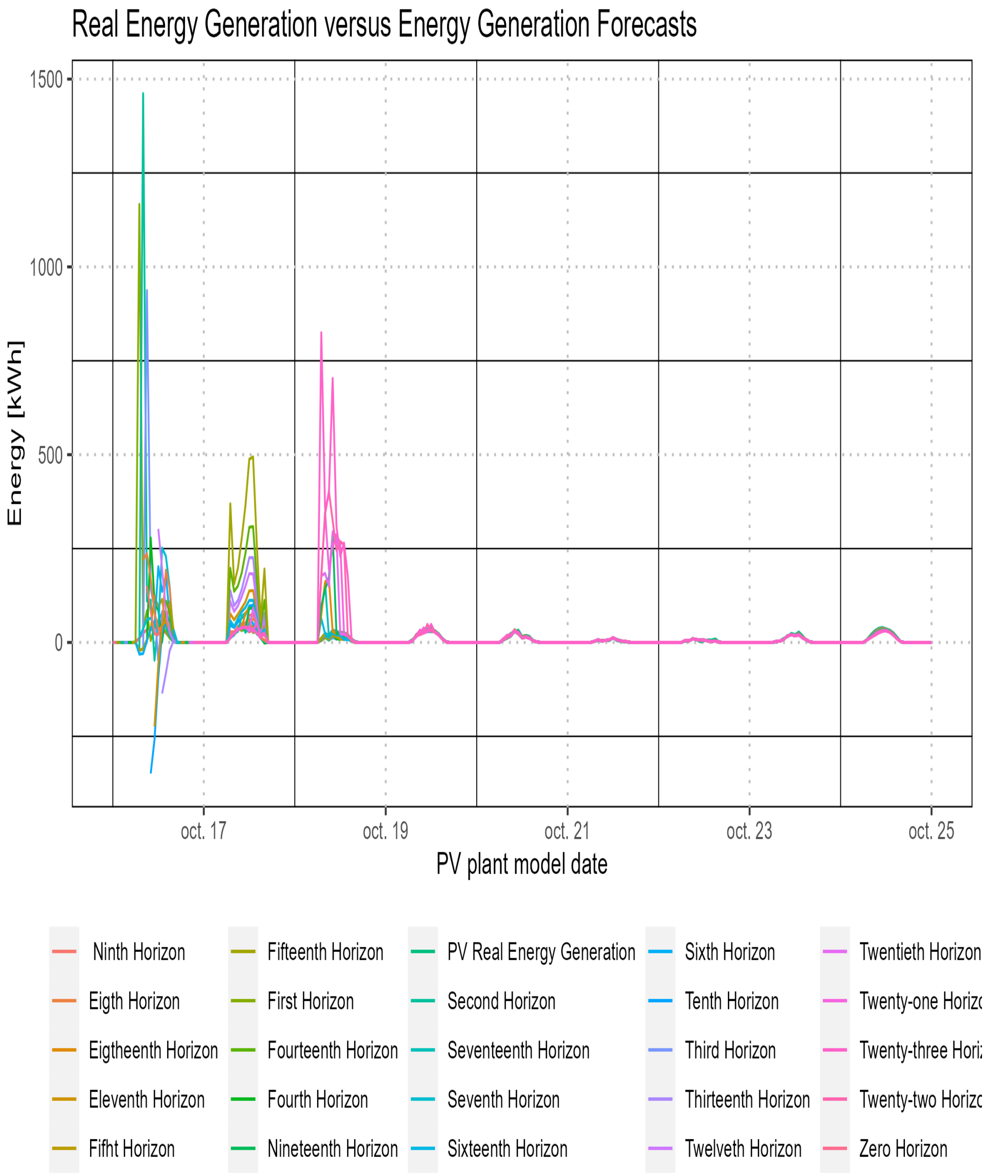


Figure 2.28: Ten days of 2019 forecasts predicted by the model without sun elevation angle.

These models need long amounts of data to be trained so, in the early days when the model hasn't received enough data to be capable of making correct predictions of the energy generation, the forecasts given by the model have values that are incorrect as can be seen in Figure 2.28, giving peak values close to 1500 kWh for the predicted generation when the real value is smaller than 60 kWh. If these early values given by the bad prediction of the model are filtered, the trend of forecasts is much more similar to the real energy generation as shown below in Figure 2.27:

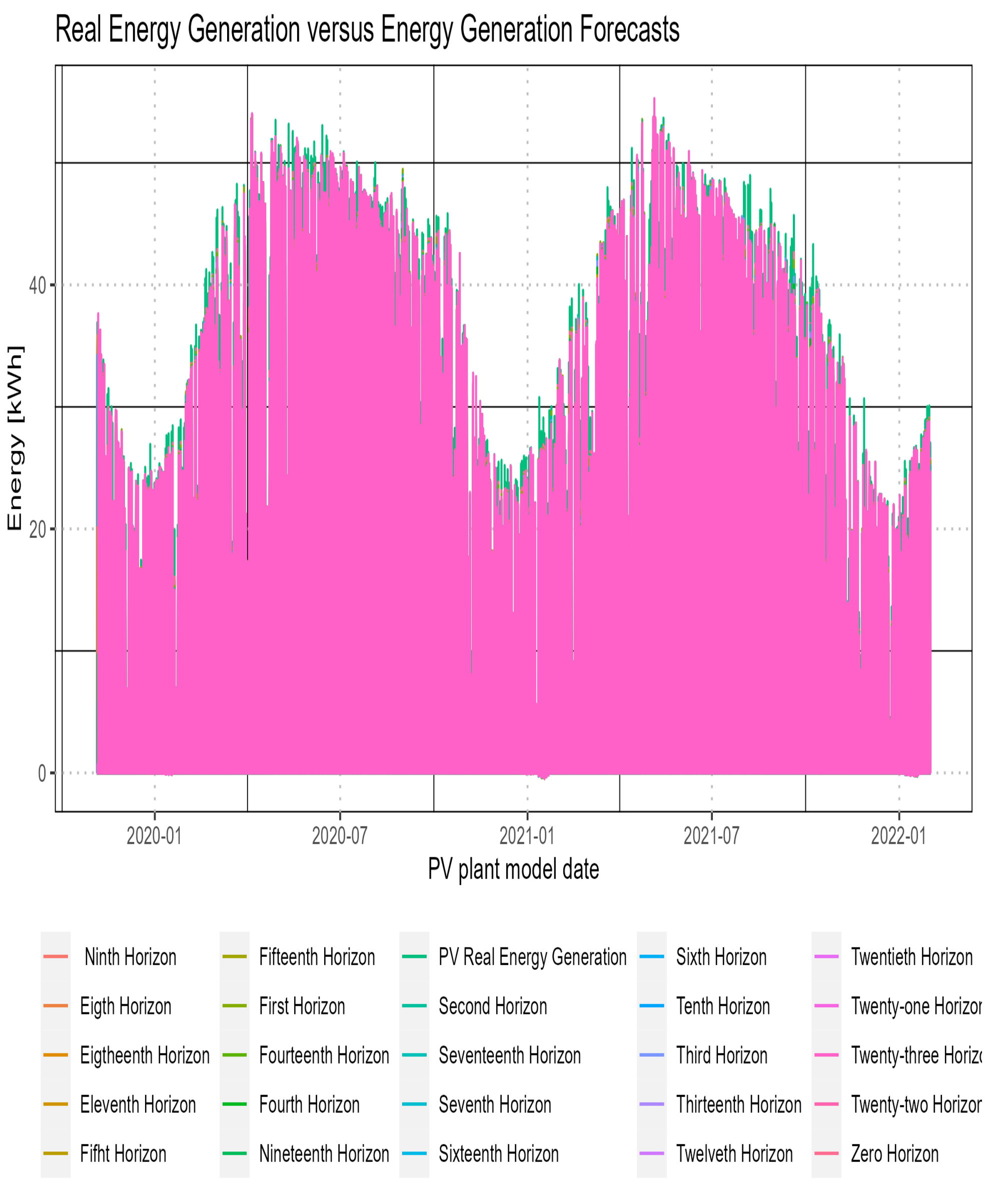


Figure 2.29: All forecasts predicted by the model without sun elevation angle filtering early incorrect predictions.

To compare the energy generation forecasts obtained with the real value of energy generation, only the most indicative forecasts are selected to plot the results:

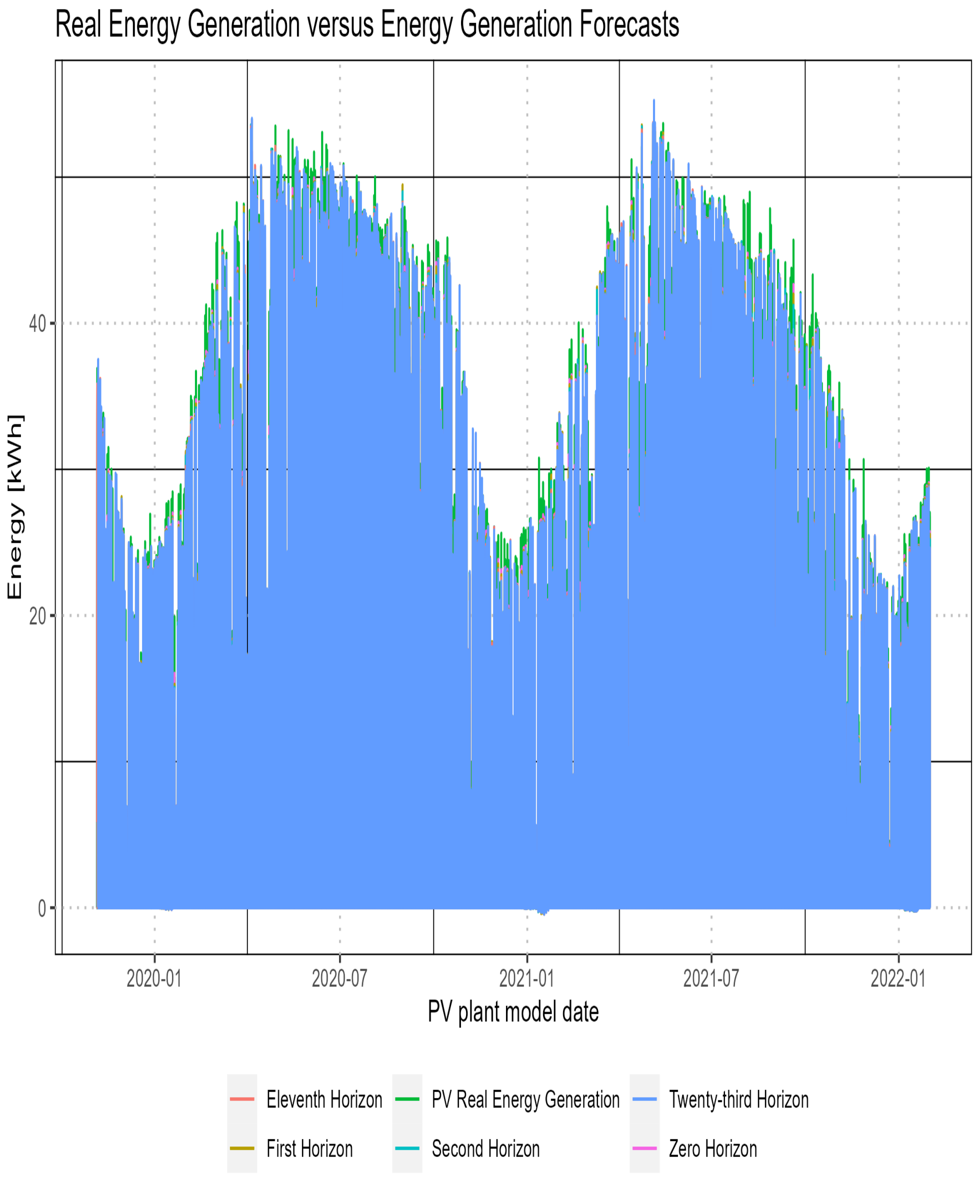


Figure 2.30: Indicative forecasts predicted by the model without sun elevation angle filtering early incorrect predictions.

Next, these horizons are displayed separated:

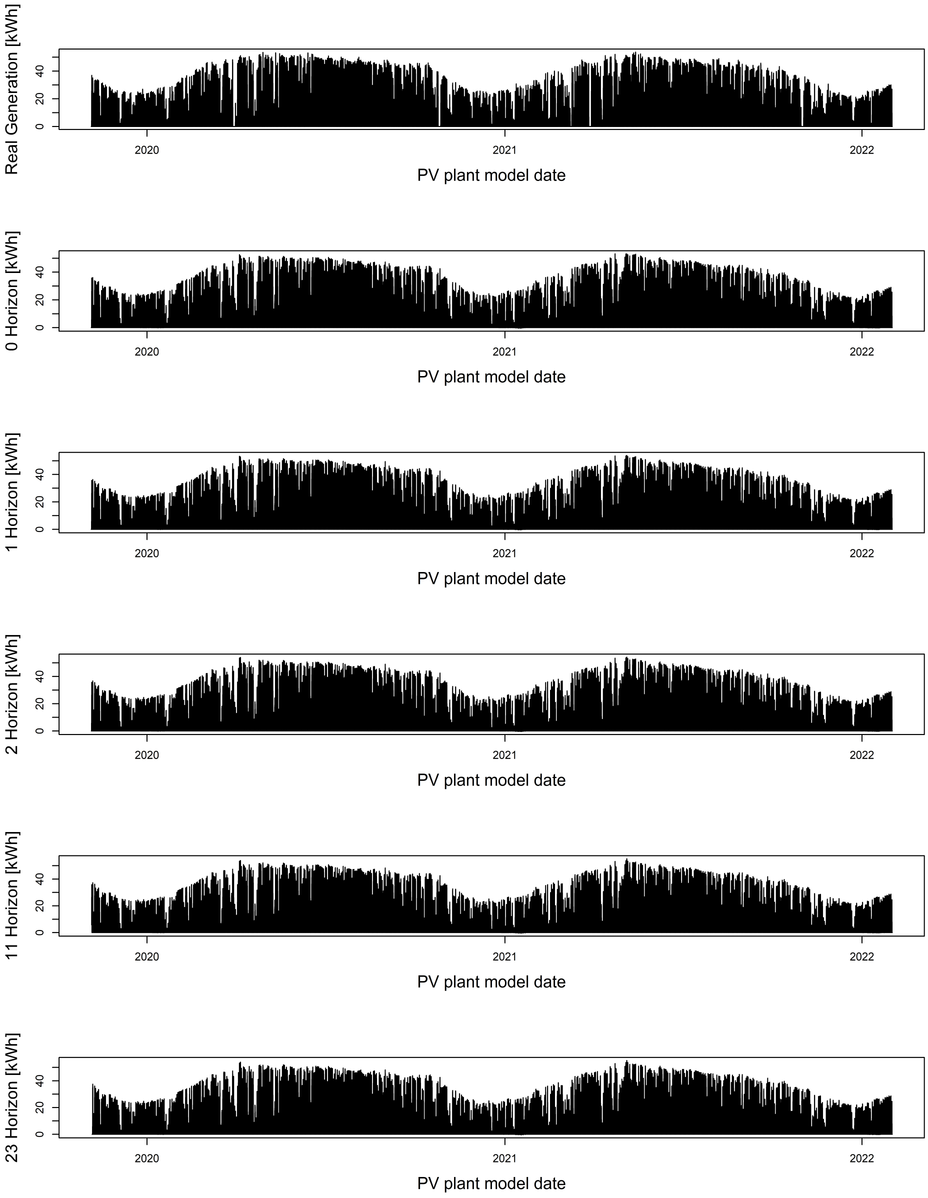


Figure 2.31: Indicative forecasts predicted by the model without sun elevation angle filtering early incorrect predictions shown separately.

The last days of the available predicted forecasts are shown in Figure 2.32.

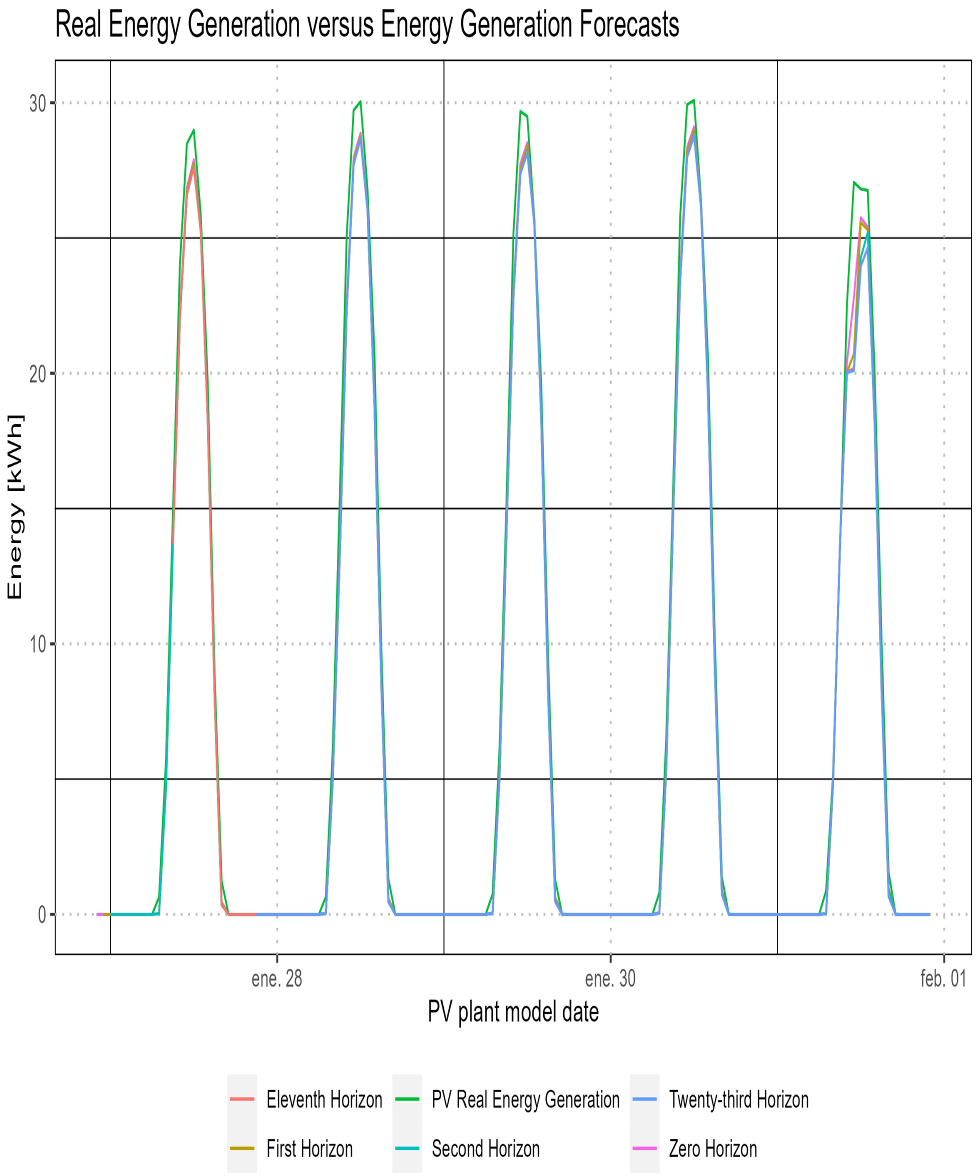


Figure 2.32: Last days of indicative forecasts predicted by the model without sun elevation angle filtering early incorrect predictions.

Next, these horizons are displayed separated:

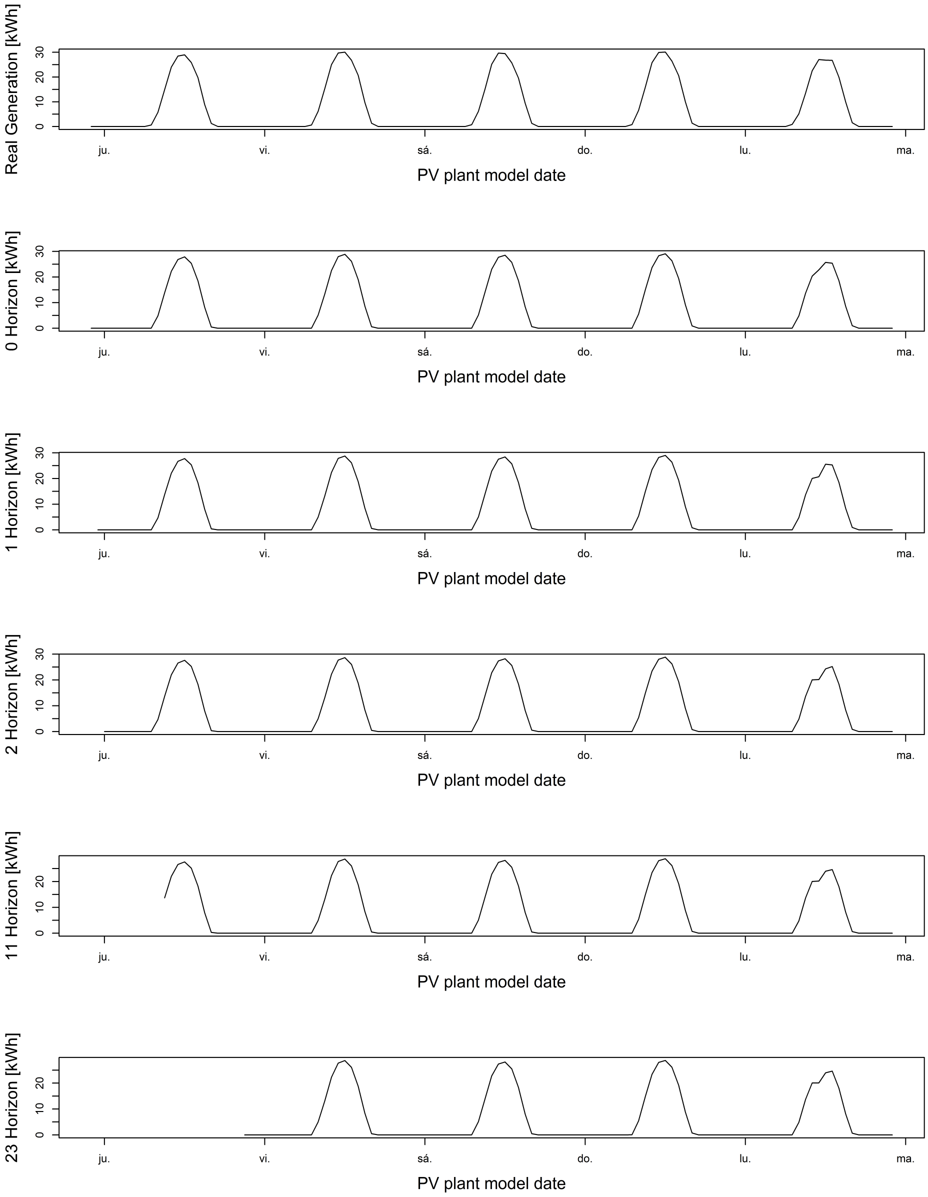


Figure 2.33: Last days of indicative forecasts predicted by the model without sun elevation angle filtering early incorrect predictions shown separately.

Next, the horizons obtained by the model which includes the sun elevation angle are shown in Figure 2.27, in this case, have been filtered the incorrect predictions of the early days in the same way done with the forecasts obtained in the model in which sun elevation angle isn't included:

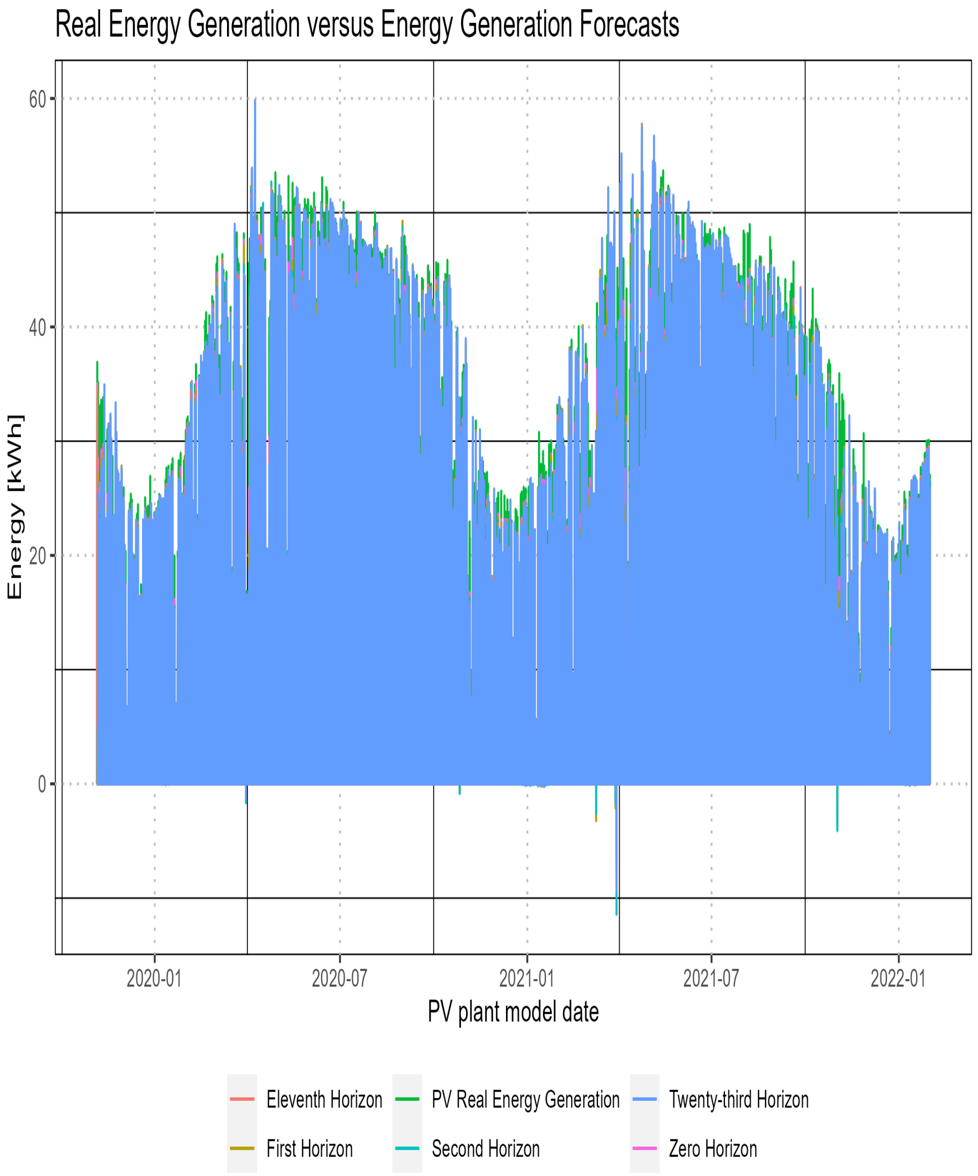


Figure 2.34: Indicative forecasts predicted by the model with sun elevation angle filtering early incorrect predictions.

Next, these horizons are displayed separated:

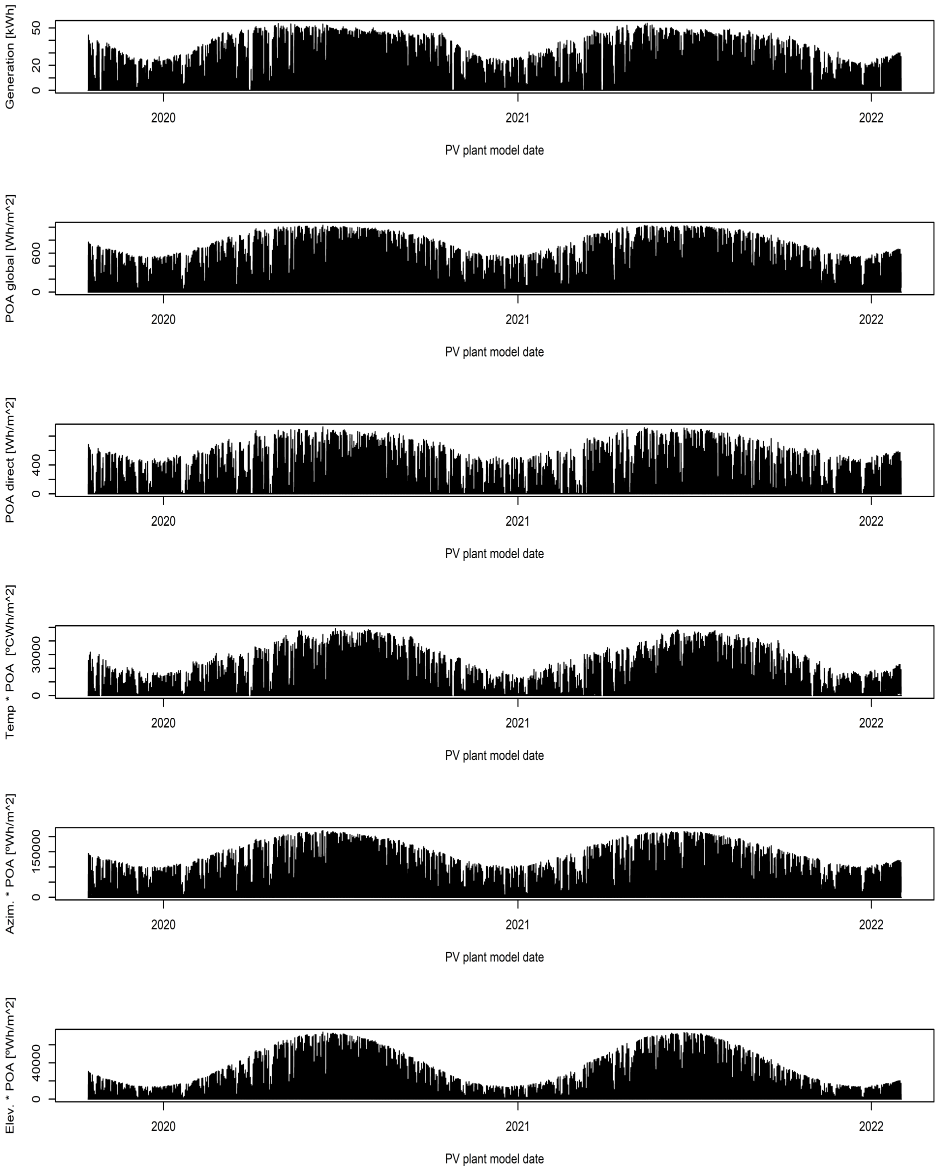


Figure 2.35: Indicative forecasts predicted by the model with sun elevation angle filtering early incorrect predictions shown separately.

The last days of the available predicted forecasts are shown in Figure 2.36.

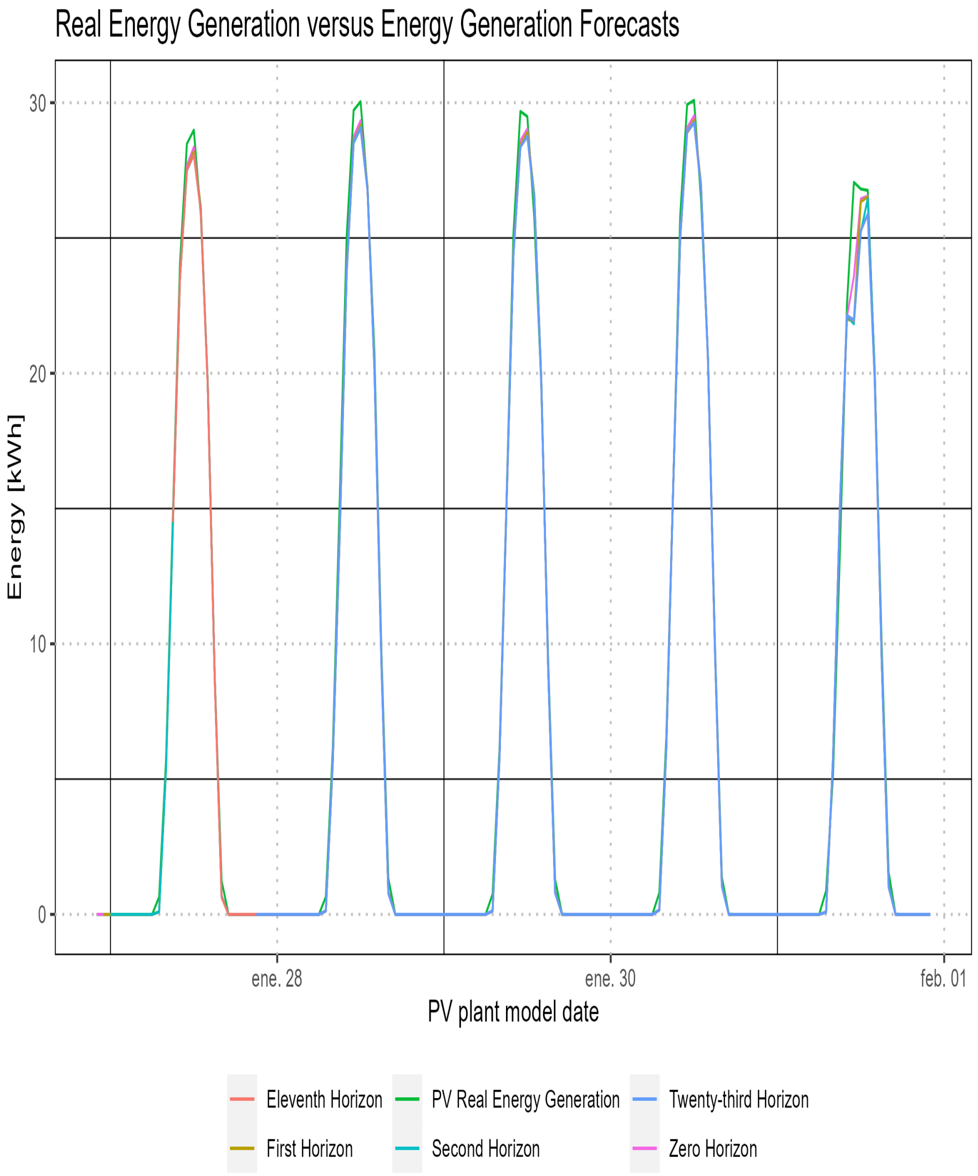


Figure 2.36: Last days of indicative forecasts predicted by the model with sun elevation angle filtering early incorrect predictions.

Next these horizons are displayed separated:

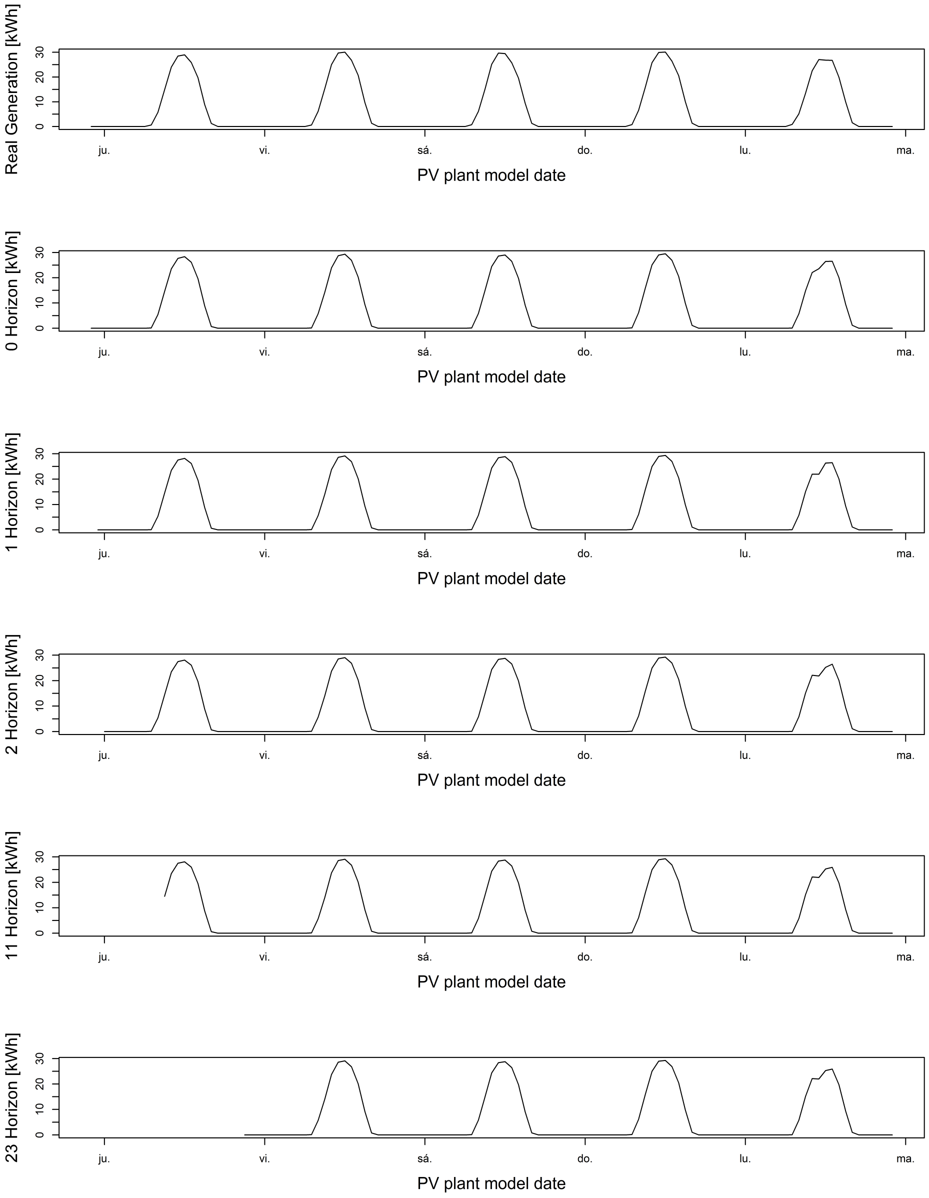


Figure 2.37: Last days of indicative forecasts predicted by the model with sun elevation angle filtering early incorrect predictions shown separately.

Plotting the energy generation forecasts and the real energy generation can be observed that the responses are very similar when the model is enough fitted and adjusted, this makes it difficult to see the differences between the forecasts generated with both models and the real energy generation.
It's also impossible to plot all the data and analyze each data point separately to evaluate the performance of the models forecasting the energy generation due to the big amount of data available.
This is why some statistical indicators are needed to evaluate the performance of the models.

2.3.1 Root Mean Square Error (RMSE)

The Root Mean Square Error (RMSE) indicates the quantity of error between two sets of data, in this case between each forecast and the real energy generation data. A higher RMSE indicates a larger difference between the model forecasts and the real generation data, so it's wanted the RMSE to be as small as possible.
First, to see the importance of optimizing the forgetting factor, the evolution of the RMSE values of

the model without the sun elevation angle as an input is shown in Figure 2.38:

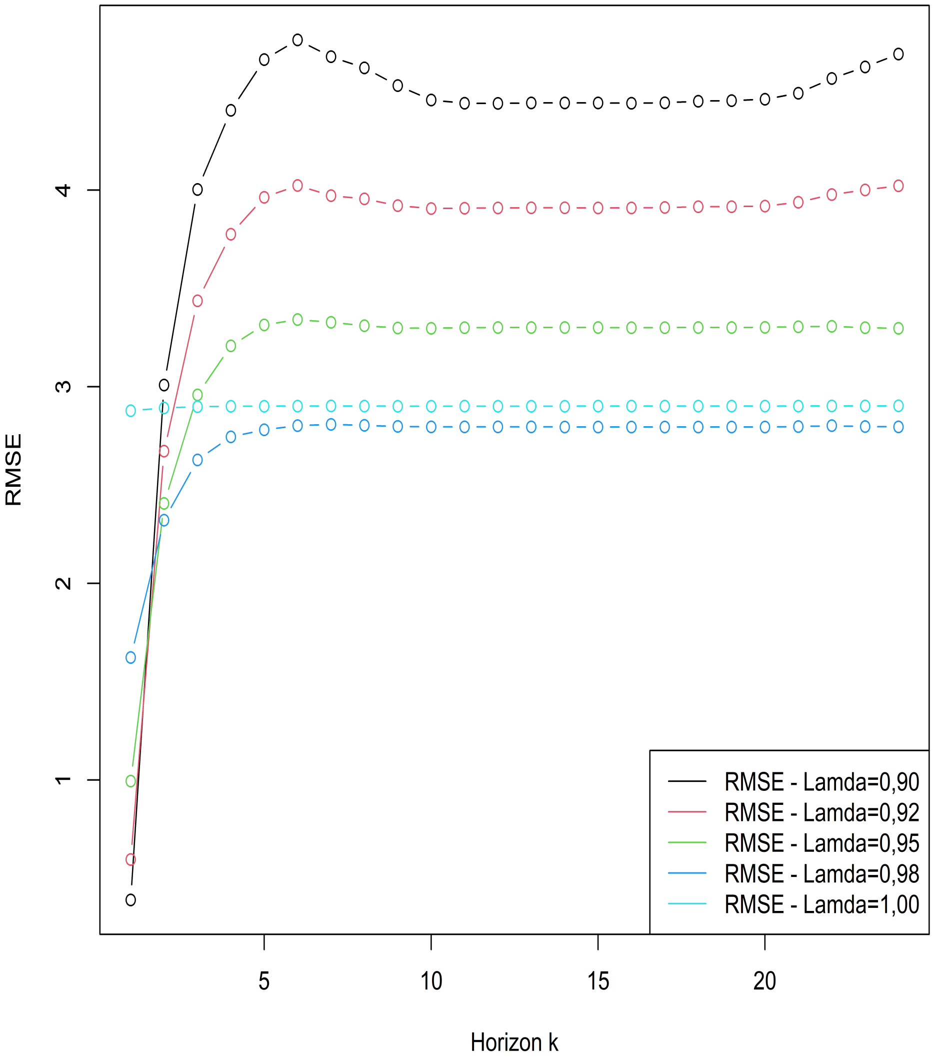


Figure 2.38: RMSE of the model without Sun Elevation angle with different lambda values.

The conclusion of the figure is that is important to search the value of the forgetting factor to obtain a model with the right accuracy of predictions, small values of the forgetting factor make the RMSE increase but also larger values makes it increase, so this value has to be optimized to a value that make it minimum.

From now on, is used the optimizer to obtain the value of the forgetting factor that minimizes the RMSE of the forecasts obtained by the model. The next table shows the values of the RMSE for both models with the forgetting facto optimized:

RMSE values		
Horizon of prediction	Model without Sun Elevation Angle as input	Model without Sun Elevation Angle as input
Horizon 0	2.061962	2.163397
Horizon 1	2.382208	2.455260
Horizon 2	2.537554	2.592977
Horizon 3	2.604079	2.645184
Horizon 4	2.636018	2.678356
Horizon 5	2.657345	2.696032
Horizon 6	2.668338	2.703084
Horizon 7	2.670885	2.706552
Horizon 8	2.670178	2.707944
Horizon 9	2.669657	2.708223
Horizon 10	2.669716	2.708504
Horizon 11	2.669675	2.708393
Horizon 12	2.669585	2.708323
Horizon 13	2.669541	2.708273
Horizon 14	2.669453	2.708199
Horizon 15	2.669336	2.708093
Horizon 16	2.669286	2.708094
Horizon 17	2.669220	2.708108
Horizon 18	2.669202	2.708728
Horizon 19	2.669578	2.709990
Horizon 20	2.672429	2.711421
Horizon 21	2.677636	2.716131
Horizon 22	2.681774	2.722243
Horizon 23	2.692152	2.730550

Table 2.1: RMSE values

In Figure 2.39 can be observed the RMSE for each forecast of both models, the model without sun elevation angle as an input and the model with the sun elevation angle as an input.

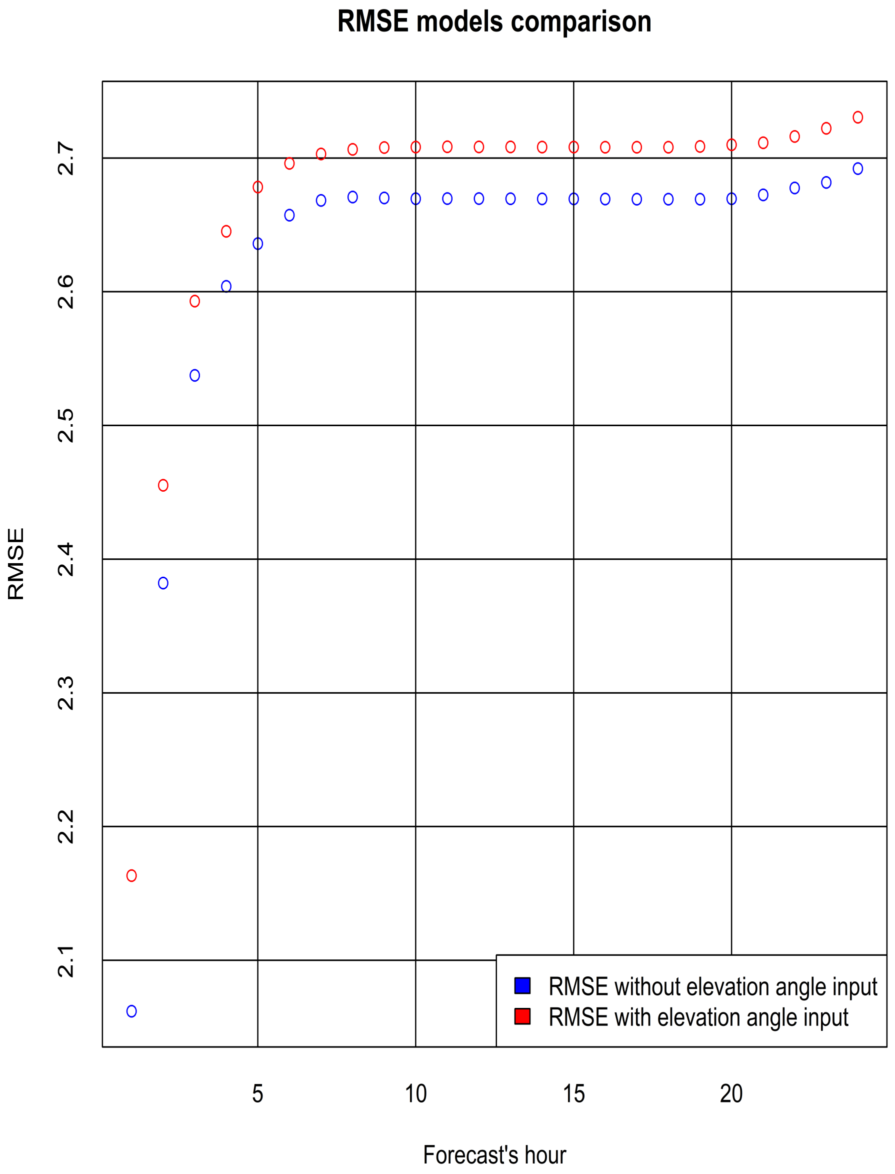


Figure 2.39: Models RMSE

As commented, as the data used isn't the real meteorological forecasts because they aren't available, there is used the historical data available to validate the model, this produces that the RMSE value is the one obtained with perfect predictions and it can change when the real data forecasts are available to use it to predict the model output.

The first conclusion that can be extracted is that in both models the RMSE grow with larger horizons and, as commented, if the uncertainty of the real meteorological forecasts were added to the model to make the predictions this would probably make them grow more the larger the horizon because the error of the meteorological predictions would be added to the model error.

Another conclusion the model without the sun elevation angle as input has a smaller RMSE on each horizon so, in conclusion, in terms of RMSE, to predict the energy generation twenty-four hours in advance the model without this input have a better performance in all horizon excluding the immediate energy generation prediction.

2.3.2 Coefficient of the Variation of the Root Mean Square Error (CVRMSE)

Coefficient of the Variation of the Root Mean Square Error (CVRMSE) indicates the quantity of error between two sets of data, is the same as RMSE but normalised to input values. A higher CVRMSE indicates a larger difference between the model forecasts and the real generation data, so it's wanted the CVRMSE to be as small as possible.
The next table shows the values of the CVRMSE for both models:

CVRMSE values		
Horizon of prediction	Model without Sun Elevation Angle as input	Model without Sun Elevation Angle as input
Horizon 0	0.2244438	0.2356255
Horizon 1	0.2593396	0.2673499
Horizon 2	0.2761203	0.2822461
Horizon 3	0.2832367	0.2878637
Horizon 4	0.2866661	0.2914228
Horizon 5	0.2889477	0.2933114
Horizon 6	0.2901098	0.2940556
Horizon 7	0.2903679	0.2944105
Horizon 8	0.2902796	0.2945437
Horizon 9	0.2902104	0.2945561
Horizon 10	0.2902022	0.2945691
Horizon 11	0.2901838	0.2945414
Horizon 12	0.2901625	0.29452504
Horizon 13	0.2901632	0.2945280
Horizon 14	0.2901790	0.2945461
Horizon 15	0.2902062	0.2945758
Horizon 16	0.2902386	0.2946177
Horizon 17	0.2902653	0.2946579
Horizon 18	0.2902681	0.2947328
Horizon 19	0.2903094	0.2948732
Horizon 20	0.2906109	0.2950315
Horizon 21	0.2911636	0.2955329
Horizon 22	0.2915878	0.2961799
Horizon 23	0.2927117	0.2970975

Table 2.2: CVRMSE values

In Figure 2.40 can be observed the CVRMSE for each forecast of both models, the model without sun elevation as an input and the model with the sun elevation as an input.

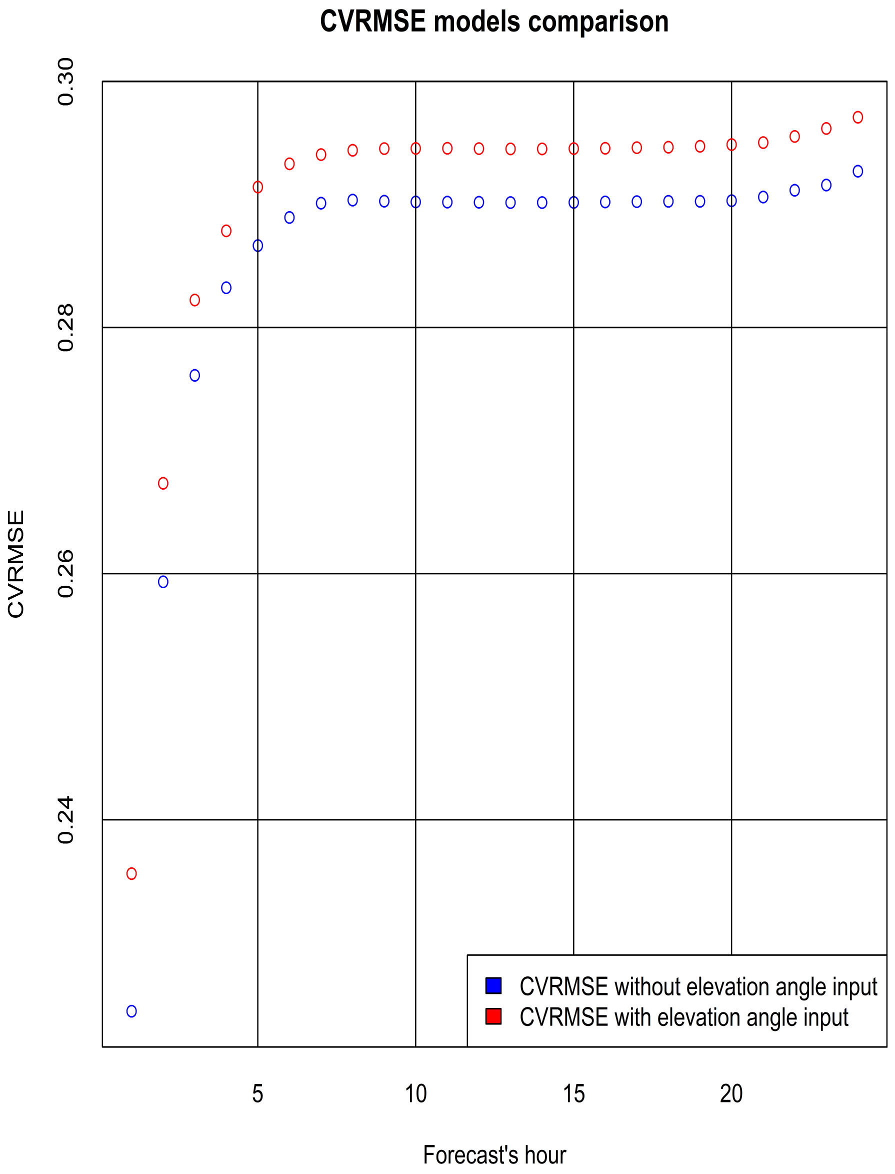


Figure 2.40: Models CVRMSE

The conclusions of this indicator are the same as the RMSE indicator, in both models, the CVRMSE grow with larger horizons, and as commented, if the uncertainty of the real meteorological forecasts were added to the model to make the predictions this would probably make them grow more the larger the horizon because the error of the meteorological predictions would be added to the model error.

In the same way as the RMSE, the sun elevation angle as input have the smaller CVRMSE in each horizon son, in conclusion, in terms of CVRMSE, to predict the energy generation twenty-four hours in advance the model without this input has a better performance in all horizons.

2.3.3 Coefficient of determination ($\mathbf{R}^2$)

Is the proportion of the variation in the dependent variable that is predictable from the independent variables. It provides a measure of how well-observed outcomes are replicated by the model.

The range of this coefficient ranges from 0 to 1, the closest the value is to 1 the better the adjustment

of the model at the variable that it's trying to explain.
The values of the R^2 of the model predictions are obtained after the filtering of the early predictions of 2019 in which the model isn't correctly adjusted yet because there is not sufficient data to train it as commented before.
The next table shows the values of the R^2 for each horizon for both models:

R^2 values		
Horizon of prediction	Model without Sun Elevation Angle as input	Model without Sun Elevation Angle as input
Horizon 0	0.9783887	0.9761817
Horizon 1	0.9711426	0.9693324
Horizon 2	0.9672832	0.9658155
Horizon 3	0.9655708	0.9644368
Horizon 4	0.9647277	0.9635474
Horizon 5	0.9641596	0.9630688
Horizon 6	0.9638662	0.96287667
Horizon 7	0.9637974	0.9627824
Horizon 8	0.9638150	0.9627441
Horizon 9	0.9638278	0.9627363
Horizon 10	0.9638253	0.9627284
Horizon 11	0.9638256	0.9627310
Horizon 12	0.9638288	0.9627329
Horizon 13	0.9638307	0.9627344
Horizon 14	0.9638309	0.9627340
Horizon 15	0.9638283	0.96273082
Horizon 16	0.9638241	0.9627242
Horizon 17	0.9638216	0.9627183
Horizon 18	0.9638252	0.9627038
Horizon 19	0.9638184	0.9626719
Horizon 20	0.9637415	0.9626300
Horizon 21	0.9635992	0.9624986
Horizon 22	0.9634886	0.9623295
Horizon 23	0.9632020	0.9620911

Table 2.3: R^2 values

In Figure 2.41 can be observed the R^2 for each forecast of both models, the model without sun elevation as an input and the model with the sun elevation as an input.

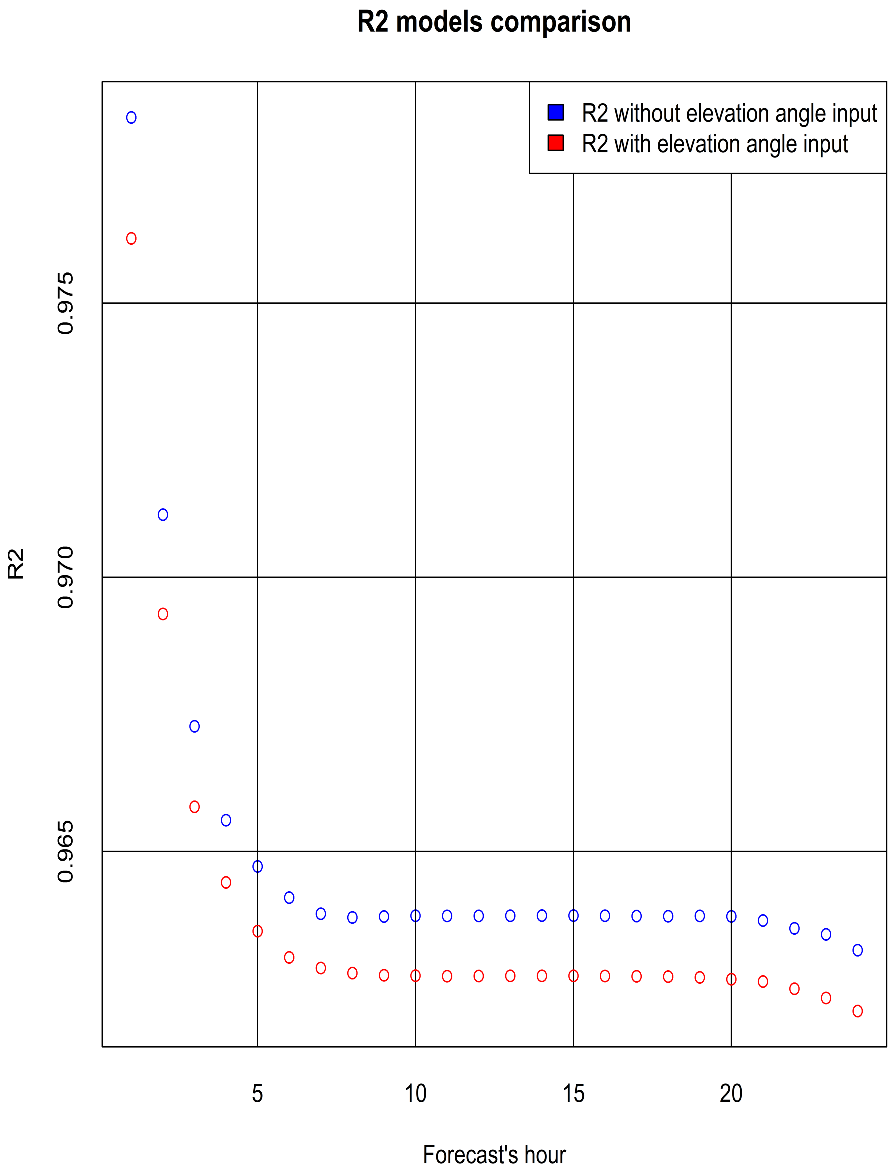


Figure 2.41: Models R^2

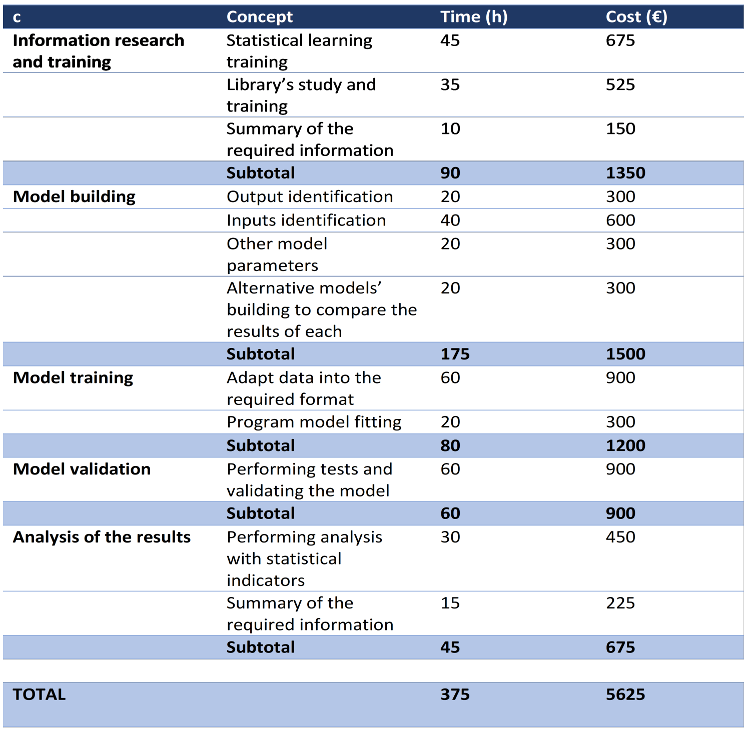
In the same way as with RMSE and CVRMSE, as the data used isn't the real meteorological forecasts because they aren't available, there is used the historical data available to validate the model, this produces that the R^2 value is the one obtained with perfect predictions and it can change when the real data forecasts are available to use it to predict the model output.
In the case of this indicator, the first horizon has a value closer to 1 which is the ideal value for this indicator and decreases with larger horizons in both models so, as happens the same as with the RMSE, for the immediate prediction the value of R^2 indicates that the predictions replicates more accurately the outcomes and with larger horizons the performance of the predictions decreases.
Also, the R^2 of the model without the sun elevation angle as input has a closer value to 1, so this model replicates more accurately the outcomes. In conclusion, also in terms of R^2, with the objective to predict the energy generation twenty-four hours in advance the model without sun elevation angle as input has a better performance in all horizons.

Chapter 3

Economic and enviroment aspects

3.1 Economic Balance

In Figure 3.1 is presented the budget for the project development. As it haven't been necessary to modify the infrastructure, the budget only refers to the hours worked.



c	Concept	Time (h)	Cost (€)
Information research and training	Statistical learning training	45	675
	Library's study and training	35	525
	Summary of the required information	10	150
	Subtotal	**90**	**1350**
Model building	Output identification	20	300
	Inputs identification	40	600
	Other model parameters	20	300
	Alternative models' building to compare the results of each	20	300
	Subtotal	**175**	**1500**
Model training	Adapt data into the required format	60	900
	Program model fitting	20	300
	Subtotal	**80**	**1200**
Model validation	Performing tests and validating the model	60	900
	Subtotal	**60**	**900**
Analysis of the results	Performing analysis with statistical indicators	30	450
	Summary of the required information	15	225
	Subtotal	**45**	**675**
TOTAL		**375**	**5625**

Figure 3.1: Economic balance

3.2 Enviroment aspects

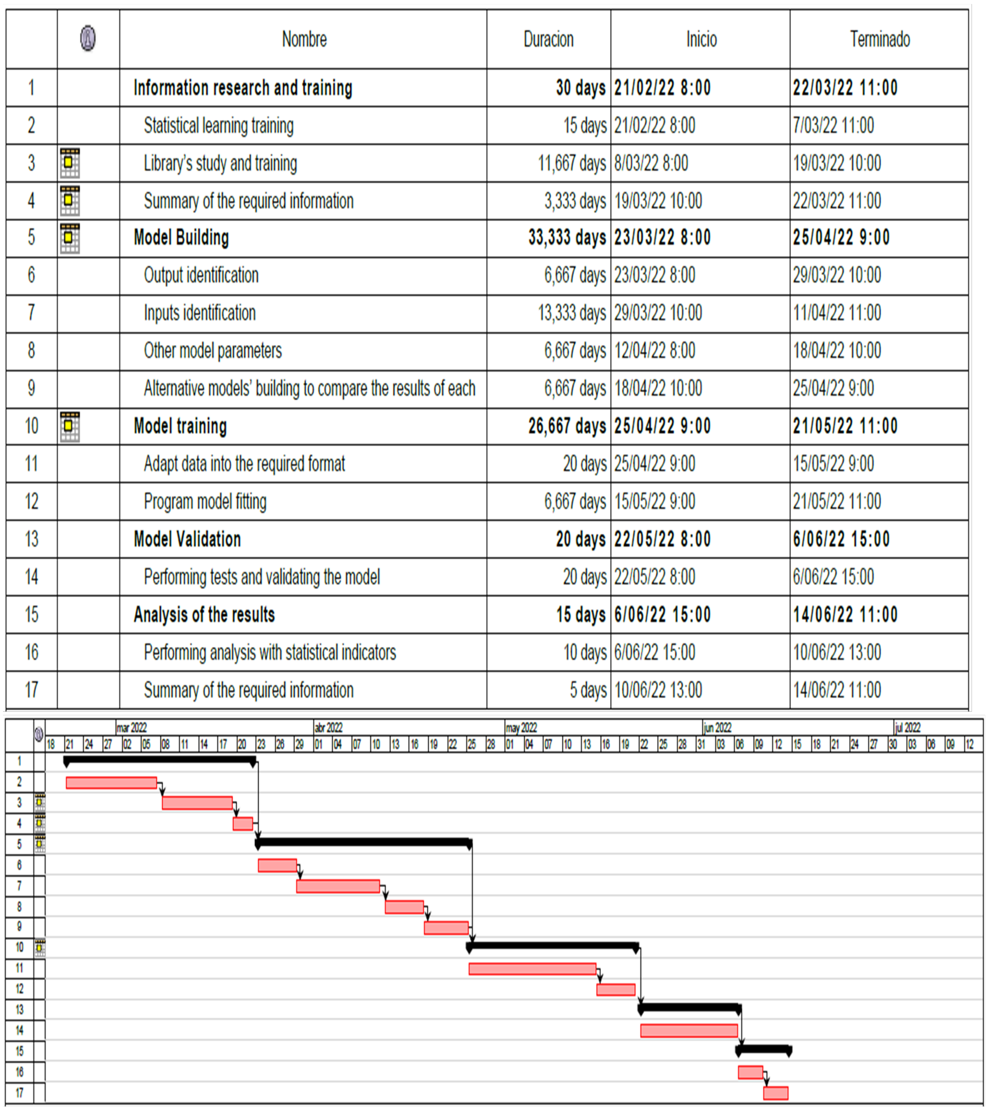
Considering that the project development doesn't imply a modification of the actual infrastructure of the installation, its development doesn't have any negative environmental impact.

Also, this project promotes the use of renewables and green energies, precisely photovoltaic solar energy generation, which pollutes less than traditional energy sources.

Chapter 4

Temporal developement

The tasks with their developement times and the Gantt diagram are shown in Figure 4.1:



	ⓘ	Nombre	Duracion	Inicio	Terminado
1		**Information research and training**	**30 days**	**21/02/22 8:00**	**22/03/22 11:00**
2		Statistical learning training	15 days	21/02/22 8:00	7/03/22 11:00
3	📅	Library's study and training	11,667 days	8/03/22 8:00	19/03/22 10:00
4	📅	Summary of the required information	3,333 days	19/03/22 10:00	22/03/22 11:00
5	📅	**Model Building**	**33,333 days**	**23/03/22 8:00**	**25/04/22 9:00**
6		Output identification	6,667 days	23/03/22 8:00	29/03/22 10:00
7		Inputs identification	13,333 days	29/03/22 10:00	11/04/22 11:00
8		Other model parameters	6,667 days	12/04/22 8:00	18/04/22 10:00
9		Alternative models' building to compare the results of each	6,667 days	18/04/22 10:00	25/04/22 9:00
10	📅	**Model training**	**26,667 days**	**25/04/22 9:00**	**21/05/22 11:00**
11		Adapt data into the required format	20 days	25/04/22 9:00	15/05/22 9:00
12		Program model fitting	6,667 days	15/05/22 9:00	21/05/22 11:00
13		**Model Validation**	**20 days**	**22/05/22 8:00**	**6/06/22 15:00**
14		Performing tests and validating the model	20 days	22/05/22 8:00	6/06/22 15:00
15		**Analysis of the results**	**15 days**	**6/06/22 15:00**	**14/06/22 11:00**
16		Performing analysis with statistical indicators	10 days	6/06/22 15:00	10/06/22 13:00
17		Summary of the required information	5 days	10/06/22 13:00	14/06/22 11:00

Figure 4.1: Tasks with their developement time

Chapter 5

Conclusions

- The resulting model is able to correctly generate the energy generation forecasts.
- The code is adaptable for the number of horizons that are wanted to predict, the number of horizons can be set based on the availability and the accuracy of the forecasts of the model's inputs.
- There have been built two models, one that includes the sun elevation angle as input and one that doesn't include it. The metrics used to analyze the forecasts indicate that the resulting model is capable to predict the energy generation accurately and that of the two built models, the one that can predict more accurately the generation of electrical energy is the one that does not include the sun elevation angle, i.e the model with fewer inputs.
- Adding more inputs to the model does not necessarily imply an improvement in the model performance, some inputs may introduce redundant data given by other inputs and also introduce more error to the model predictions by its linearization. Also, if the inputs are forecasts predicted by other models, they may have the error introduced by the model used to generate it, and adding more input forecasts can introduce this error to the model and be added to the error of the own model.
- The application of Statistical Learning methods is a useful tool that can be applied to systems in which the relation between inputs and outputs isn't known but it has to be elected the model correctly to obtain a correct response of it.
- To obtain acceptable results for the model, big amounts of data are needed to perform the model training otherwise the model can have a bad performance making predictions.
- The performance of the model does not necessarily increase by adding more inputs to the system, some of these inputs may be redundant and not provide new information or introduce more error to the model decreasing its performance.
- The data set obtained must be analyzed as a whole using some statistical indicators that provide information about its behavior.
- Using forecasts as inputs of the model makes it dependent on their performance to predict the data correctly, if these forecasts have an error making the prediction, this error will be added to the model decreasing its performance to produce its own forecasts.
- Using a linear model to produce forecasts implies that the relation between inputs and outputs has to be linear in another case an error produced by this non-linearity can affect the performance of the model predicting the desired variable.
- The transformation of non-linear variables into linear variables can introduce an error to the model if the approximation isn't correctly done.

Chapter 6

Future work

Once the twenty-four-hour meteorological forecasts of the installation are available, this data has to be used as inputs of the model in order to have a more realistic prediction.

When the energy production forecasts are predicted using the available meteorological forecasts, the analysis of the model results has to be done in order to evaluate the error introduced by the uncertainty of the data introduced as model inputs.

With the analysis results, the model may be adjusted to decrease the error of its predictions or to be more or less adaptable to data changes, this can be performed by the forgetting factor.

Also, a comparison between the two proposed models can be performed with the meteorological forecasts to analyze if their behavior is the same in both cases or have some differences in order to confirm the model election or change it.

Bibliography

[1] T. Hastie, R. Tibshirani, G. James, and D. Witten, "An introduction to statistical learning (2nd ed.)," *Springer texts*, vol. 102, p. 618, 2021.

[2] P. Bacher, H. G. Bergsteinsson, L. Frölke, M. L. Sørensen, J. Lemos-Vinasco, J. Liisberg, J. K. Møller, H. A. Nielsen, and H. Madsen, "onlineforecast: An R package for adaptive and recursive forecasting," pp. 1–36, 2021. [Online]. Available: http://arxiv.org/abs/2109.12915

[3] A. Smets, K. Jäger, O. Isabella, R. Van Swaaij, and M. Zeman, *Solar Energy - The physics and engineering of photovoltaic conversion, technologies and systems*, feb 2016.

[4] P. Bacher, "Building heat load forecasting," pp. 02–15, 2022. [Online]. Available: https://onlineforecasting.org/examples/building-heat-load-forecasting.html

[5] P. Bacher, H. Madsen, H. A. Nielsen, and B. Perers, "Short-term heat load forecasting for single family houses," *Energy and Buildings*, vol. 65, pp. 101–112, 2013.

[6] Bacher, Peder, "Setup of data for an onlineforecast model." [Online]. Available: https://onlineforecasting.org/vignettes/setup-data.html

[7] P. Bacher, "Solar power forecasting," aug 2021. [Online]. Available: https://onlineforecasting.org/examples/solar-power-forecasting.html

[8] Bacher, Peder, "Setup and use onlineforecast models," may 2022. [Online]. Available: https://onlineforecasting.org/vignettes/setup-and-use-model.html

[9] Bacher, Peder, "Setup of data for an onlineforecast model," may 2022. [Online]. Available: https://onlineforecasting.org/vignettes/setup-data.html

[10] P. Bacher, "Nice tricks with onlineforecast," may 2022. [Online]. Available: https://onlineforecasting.org/vignettes/nice-tricks.html

[11] Bacher, Peder, "Forecast evaluation," may 2022. [Online]. Available: https://onlineforecasting.org/vignettes/forecast-evaluation.html

[12] "Coeficiente de determinación - Wikipedia, la enciclopedia libre." [Online]. Available: https://es.wikipedia.org/wiki/Coeficiente_de_determinacion

[13] "Root-mean-square deviation - Wikipedia." [Online]. Available: https://en.wikipedia.org/wiki/Root-mean-square_deviation