Article

Trade-Off between Precision and Resolution of a Solar Power Forecasting Algorithm for Micro-Grid Optimal Control

Jean-Laurent Duchaud [1],*, Cyril Voyant [1,2], Alexis Fouilloy [1], Gilles Notton [1]
and Marie-Laure Nivet [1]

[1] Science for the Environment, University of Corsica, UMR CNRS 6134, 20000 Ajaccio, France;
    cyrilvoyant@gmail.com (C.V.); fouilloy_a@univ-corse.fr (A.F.); notton_g@univ-corse.fr (G.N.);
    nivet_m@univ-corse.fr (M.-L.N.)
[2] Radiotherapy Unit, Castelluccio Hospital, BP 85, 20177 Ajaccio, France
*   Correspondence: duchaud_jl@univ-corse.fr

check for
updates

Abstract: With the development of micro-grids including PV production and storage, the need for efficient energy management strategies arises. One of their key components is the forecast of the energy production from very short to long term. The forecast time-step is an important parameter affecting not only its accuracy but also the optimal control time discretization, hence its efficiency and computational burden. To quantify this trade-off, four machine learning forecast models are tested on two geographical locations for time-steps varying from 2 to 60 min and horizons from 10 min to 6 h, on global irradiance horizontal and tilted when data was available. The results are similar for all the models and indicate that the error metric can be reduced up to 0.8% per minute on the time-step for forecasts below one hour and up to 1.7% per ten minutes for forecasts between one and six hours. In addition, it is shown that for short term horizons, it may be advantageous to forecast with a high resolution then average the results at the time-step needed by the energy management system.

Keywords: short-term solar forecasting; machine learning techniques; statistical models; performance evaluation

1. Introduction

The photovoltaic market has been growing in recent years at an increasing rate. At the end of 2019, the total installed power in the world exceeded half a million megawatts (627 GWp) with the installation of 115 GWp during 2019 [1,2].

IEA [1] and EurObser [2] estimate the PV contribution to be between 2.6% and 3% of global electricity production today. This annual theoretical penetration level varies from 0.1% for Norway to 14.8% for Honduras [1]. Figure 1 shows the theoretical PV integration rate for the main PV countries and the cumulated PV installed capacity for the top 10 [1,3]. In some countries this average penetration rate is important and can reach even higher levels on a monthly or daily basis, inducing energy management problems discussed further below.

As the share of electricity produced by PV systems increases, this intermittent and stochastic energy production needs to be fully integrated into electricity grids. The variability of the solar source makes the PV production hard to manage for the grid operator. It requires additional and complex actions to balance the system as the electricity generation must equal the electricity demand at all times [3–5].
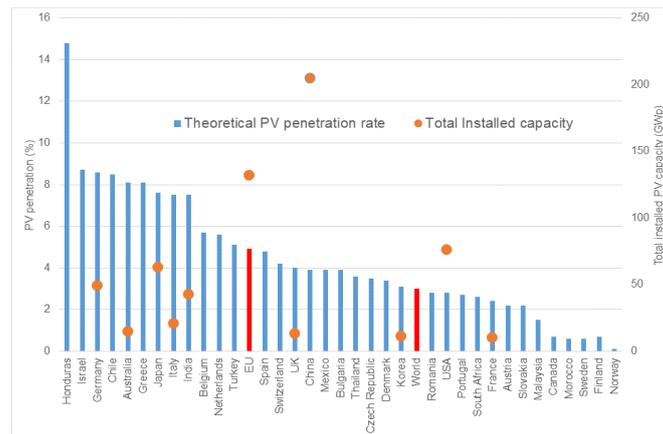
Figure 1. Theoretical penetration rate (annual basis) and cumulated PV capacity [1].

These difficulties in integrating solar energy production into electrical grid have a cost. This is all the more important as the electrical grid is small and not interconnected with large ones, as is the case for islands. This over cost due to the intermittence of the solar resource is discussed and estimated in the literature: from 6.16 to 8.47 €/MWh [6] to around 10 €/MWh [7]. A review on the complexity of the intermittent renewable energy integration, its consequences on the electricity management and on the over cost has been proposed [8]. It is today generally accepted that an increase of the intermittent renewable systems should be achieved with a common development of the energy storage means, the smart electrical grid and the forecasting methods for renewable production and users consumption.

This paper is focused on the latter: forecasting is necessary to estimate the reserves, schedule the power system, avoid congestion, optimally manage the storage and trade in the electricity market [9,10]. A good forecasting tool has a positive impact on the reduction of the integration costs of PV plants, on the decrease of the average annual operating costs, on the reserve shortfalls and on the increase of the percentage reduction in curtailments of PV systems. The higher the integration level of PV systems in the electrical network is, the more important are the effects of the forecasting [8].

The energy supplier must manage his system with various temporal horizons from one second to one year and consequently needs the corresponding forecast models [11]. To help obtain the best dispatch schedule, a set of algorithms, usually gathered in an Energy Management System (EMS), uses an optimization solver to plan the power exchanges between the grid components. The most common are (Mixed Integer) Linear Programming [12,13], Dynamic Programming [14] or evolutionary algorithms [15], used in the scope of a Model Predictive Control procedure [16,17]. As they work on the future production and consumption, they are very dependent on the forecast resolutions and their error: low resolution forecasts tend to have less error due to the averaging of the data but they increase the time between optimal points. On the contrary, high resolution forecasts enable a more frequent optimization but with a greater error at each point. Furthermore, more points drastically increase the computational burden of the optimizer.

A good illustration of the necessity to have efficient forecasting models with various temporal horizons and resolutions in an Energy Management System was demonstrated during the Tilos project. The H2020 Tilos project aimed to make a small Greek island (about 500 inhabitants) autonomous for its electricity supply by maximizing the use of renewable energies. The project successfully came to an end in 2019 and the energy of the island is actually produced by a 160 kWp PV plant coupled to a 800 kW wind turbine. The system is completed by a Na-NiCl$_2$ battery bank (800 kW–2880 kWh) and a backup diesel generator [18]. The grid EMS monitors the operation of the various components and controls the energy production and storage means according to the results provided by the forecasting platform [19].

The models developed are able to provide reliable estimations of the future load demand, wind and solar production. More specifically, some of the PV forecasting models use machine learning

methods similar to the ones presented in this paper. Each parameter is estimated at different horizons to cover the entire band of dispatch scheduling (i.e., from 10 min up to 48 h ahead) with different time resolutions (i.e., 1-min, 10-min, hourly and daily). The results are fed to the EMS in order to produce the dispatch schedules: day-ahead, intra-day and short-term dispatches. Among other things, this system enabled Tilos Island to produce up to 90% of its electricity needs from renewable sources in December 2019 while aiming at a renewable energy share around 70–75% on an annual basis.

Numerous references present forecasting models using different algorithms, time-steps and horizons. For example, in [20], the authors propose a Neural Network forecasting algorithm for the hourly day-ahead consumption. The results are useful for energy trading in the day-ahead market and in a certain amount in the intra-day but the forecast lacks the resolution for optimal control of a micro-grid storage. The control algorithm will indeed be able to adapt quicker to weather variations if it has infra-hour forecasts.

Heinemann et al. [9] wrote an article where the best uses of the various forecasting methods are discussed and, more recently [21], obtained the same conclusions. Figure 2 summarizes the best usage domains for each of the main existing methods (Time series, Numerical Weather Prediction, satellite and image based methods) according to the forecasting time horizon and the temporal resolution. For example, the bottom left corner represents the high resolution and short term forecasts needed for management of the spinning reserve. The best suited algorithms use image recognition from cameras pointed at the sky to predict the cloud position and infer the future PV production.

Lauret et al. [22] compare several forecasting models in three locations with a one hour time-step and a maximum horizon of six hours. In [23], the models range from 5 to 30 min for the horizon and 5 to 10 min for the time-steps. The results obtained in these and in other publications are hardly comparable since the forecasts have been done with different input data and with unmatching time-steps and horizons. Furthermore, the error metrics used are often different and it does not help the analysis [24,25]. That is the reason why the present paper aims to compare, on the same location and with the same forecast model, the forecast error at different horizons and time-steps.
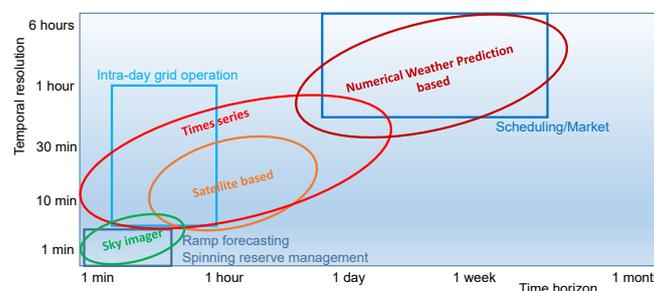
Figure 2. Forecasting time horizon versus temporal resolution according to Diagne et al. [21].

The study presented in this paper is focused on machine learning models for the future estimation of global horizontal and tilted irradiance. It intends to quantify the trade-off between the forecast resolution and the error at different horizons. The objective is to help the designer of an EMS to chose the best time-steps for the different forecast models needed.

The first section presented the need for the solar production prevision and the state of the art of the short term forecasting models with machine learning. The second section describes the input data used in the study and the methodology employed to train the forecast models, obtain a prevision and compute the error metrics. The third section presents the results obtained in five steps. First the models used and the test locations are compared. The model performances are then presented in error versus time-step graphs for the tested time parameters and are fitted by a first-order polynomial. The improvement offered by averaging the forecasted values is also discussed. Finally, with a view to use those results in a micro-grid EMS, the methodology is tested on tilted irradiance.

2. Materials and Methods

This section presents the global principles of the study. The first subsection details the test locations and the time series used. The second subsection presents the clear sky model needed to obtained stationary data. The forecast models used are then briefly described in the third subsection along with the reference models. The fourth subsection presents the different steps needed to train and use the models and the fifth explains how the error metric is computed. Finally, the methodology used to quantify the trade-off is proposed.

2.1. Input Weather Data

The datasets used in this study come from two sites with different meteorological conditions, both equipped with pyranometric sensors. The values measured are the global irradiance (horizontal for both stations and tilted for Ajaccio) for three years, at a time-step of one minute. In both cases one year has been used for training and validation while two years were used for testing. The first site is located in Corsica (France), a mountainous Mediterranean Island, near the town of Ajaccio where the George Peri Research Center is located. This station, near the sea, presents a great solar potential with a climate characterized by a warm and dry summer and a wet and mild winter. The second site is the great solar furnace of Odeillo (Pyrénées Orientales, France), located in the mountains, about 100 km from the sea. In this meteorological station, situated in altitude, the climate is rather unstable. It often presents a high nebulosity and rainfalls are still present even during the driest months.

The different conditions are characterized by the variability of the datasets [26]. The Table 1 summarizes the characteristics of the two sites and gives more information on the datasets.

Table 1. Test sites and global solar irradiance time series characteristics.

	Ajaccio	**Odeillo**
Latitude	41°54′46″ N	42°29′37″ N
Longitude	8°39′11″ E	2°1′45″ E
Altitude	5 m AMSL	1650 m AMSL
Variability [26]	0.196 (weak)	0.503 (strong)
Climate type [27]	Warm Mediterranean (Csa)	Temperate Oceanic (Cfb)
Annual irradiation	1.67 MW h m^{-2}	1.63 MW h m^{-2}
Training year	2015	2001
Testing years	2016–2017	2002–2003
Missing points	0.47%	0.19%

2.2. Clear Sky Modeling and Index

The majority of forecast models based on machine learning needs stationary inputs [28]. According to this requirement, the irradiance time series is made stationary during the data preprocessing phase. To do so, the measured irradiance is divided by the theoretical irradiance under clear sky conditions. Several clear sky models exist in the literature [29,30]. In this paper, the SOLIS model, developed by Mueller [31], is chosen due to its excellent results in the considered locations. It is built from radiative transfer models combined with Beer Lambert function integrated on whole solar spectra and with data provided by satellites (ER-2/ENVISAT). The Global Horizontal Irradiance under Clear Sky conditions $GHI^{cs}(t)$ is defined by Equation (1).

$$GHI^{cs}(t) = H_0 . e^{\frac{-\tau}{sin^b . h_s(t)}} sin(h_s(t)) \qquad (1)$$

with $h_s(t)$ the solar elevation, H_0 the irradiance at the top of the atmosphere, τ the atmospheric depth and b a fitting parameter. The last two parameters depend on the sites and their values are available on [32].

The clear sky index $K_t(t)$ used in the forecast models is assumed to be stationary enough and is defined as the ratio of the measured Global Horizontal Irradiance GHI to the GHI^{cs}:

$$K_t(t) = \frac{GHI(t)}{GHI^{cs}(t)} \tag{2}$$

2.3. Forecast and Reference Models

The study has been done considering four machine learning forecast models among the most used in the literature [33]. Throughout this paper, the output of the forecasting models will be marked with a $\frown$.

The persistence of the weather conditions is one of the simplest forecasting models [22]. It uses the clear sky model (Section 2.2) and can be written:

$$\widehat{GHI}(t+\ell) = GHI(t).\frac{GHI^{cs}(t+\ell)}{GHI^{cs}(t)} = K_t(t).GHI^{cs}(t+\ell) \tag{3}$$

where $\widehat{GHI}(t+\ell)$ is the forecast at the horizon ℓ, $GHI(t)$ is the present data, $GHI^{cs}(t+\ell)$ and $GHI^{cs}(t)$ the clear sky value at the horizon ℓ and at the instant t respectively. In this paper, this model is referred as Scaled Persistence but it can be found in the literature as Persistence or Smart Persistence. This kind of model gives good performances at very short term horizons but its accuracy decreases quickly as the horizon increases. However, due to its simplicity, this model is often used as a naive reference to compare the performance of more complex models.

Another simple and naive model, sometimes used to characterize a site, is the Climatology [34]. In this case, the clear sky irradiance at the desired horizon is multiplied by the averaged clear sky index $\overline{K_t}$, computed over the historical data, as presented in Equation (4) below.

$$\widehat{GHI}(t+\ell) = \overline{K_t}.GHI^{cs}(t+\ell) \tag{4}$$

To ease the reading, the machine learning models used during this study will not be described in detail here. For further information one can check the dedicated part in literature [33].

- The first model is the Auto Regressive Moving Average (ARMA) [35,36]. It is a relatively old model developed in the 1970s using a combination of auto regressive model with a moving average model. Despite its simplicity, it offers very good performances in stationary time series analysis and is often used in numerous domains [37]. Although it is not a machine learning method, it can be assimilated as so.
- The second model is based on artificial neural network model with the Multi Layer Perceptron (MLP) [38,39]. The chosen architecture is a MLP with one hidden layer and one output layer. The training algorithm is the Levenberg–Marquardt, which is well suited to the solar time series forecasting.
- The third model from the regression trees family is the Random Forest (RF) model, developed by Breiman [40]. It is an ensemble model based on the aggregation of simple regression tree results in order to obtain the overall forecast [41]. The input dataset is divided in samples and each tree grows independently. This improvement by addition of randomness enables one to build more robust models and reduces the risks of over training.
- The last model is a Support Vector Regression model (SVR). It is a kernel based method derived from the Support Vector Machine (SVM) to treat regression tasks [22,42]. The adaptation of this type of model to forecast time series gives some good results and is widely used in this research area.

2.4. Data Processing

Before being used for the training or the forecasting, the raw measurements must go through several steps to ensure the good operation of the models. The preprocessing phases are detailed below:

1. Quality check: the measured data is verified to make sure it is usable. Sometimes errors can occur in the time series due to issues in the acquisition chain. The quality check procedure is based on the GEOSS project [43] and checks, for example, if the measurement values range between 0 and the extraterrestrial irradiance.
2. Data filtering: selects the useful part of the data set. During this step, the night hours are deleted, based on the solar angle h_s. At dawn and at dusk neighboring items can create masks on the pyranometers that can disturb the training process. Furthermore, uncertainties associated with pyranometers are typically higher for low solar angle [44]. For those reasons and according to the solar masks of the two sites, any measurement where ($h_s \leq 8°$) is excluded from the data set.
3. Stationarization: the irradiance time series is, by nature, dependent on the seasons and periodic. However, many time series forecasting models require stationary data to be used [45]. A good method consists of using a model giving the solar irradiance in clear sky conditions and calculating the clear sky index for each point of the data set as explained in Section 2.2.
4. Averaging: The data is averaged to the desired model time-step, denoted Δt. The global horizontal irradiance GHI_i represents the averaged value of the measured irradiance in $[t_i, t_i + \Delta t]$. This step is the main focus of this paper. It defines the resolution of the prediction and, as it will be demonstrated, it has an impact on the forecast error.

Once the data is cleaned and averaged, it can be used to train the model or to estimate the future irradiance. Prior to the training phase, the number of historical data that the Time Series Forecasting model takes as input must be defined. To determine this number (denoted p), a method usually used in the case of linear models, consists in using the Partial AutoCorrelation Factor (PACF) [46]. The order p is then the lag value after which the PACF crosses the upper confidence interval for the first time. These p lags will act as the model input while forecasting the time series. Although the PACF has originally been used with ARMA models and is not the best feature selection method for the machine learning models, it is not the focus of this paper and the method is assumed to be applicable to all the models presented here.

After the parameters have been selected, the model is trained on historical data. Each model type has a dedicated training method, but generally speaking it consists of splitting the datasets in two parts—training (80%) and validation (20%)—and using an optimization algorithm to find the coefficients minimizing the forecast error on the validation subset. This step is realized with the corresponding algorithms available in Matlab. Preliminary tests have shown that using one year for training was enough to avoid weather seasonality and to consider enough input points.

For the forecasting phase, the averaged data is used as input of the model. p points are needed to compute a prevision. The mathematical formulation of the ℓ-horizon forecast (corresponding to the q-step ahead forecasting with $\ell = q.\Delta t$ and $p, q \in \mathbb{N}_+^*$) is given by:

$$K_t(t + \ell) = f\left(K_t(t), K_t(t - \Delta t), \dots, K_t\left(t - (p - 1).\Delta t\right)\right) + \epsilon(t + \ell) \tag{5}$$

$$K_t(t + \ell) = \widehat{K}_t(t + \ell) + \epsilon(t + \ell) \tag{6}$$

where K_t is the measured Clear Sky Index, $\widehat{K}_t$ is the forecasted CSI, ℓ the horizon, p the number of historical data, Δt the model time-step, $\epsilon(t + \ell)$ a random white noise and f a representation of the model.

The training and forecasting phases can be summed up in a flow chart presented in Figure 3.
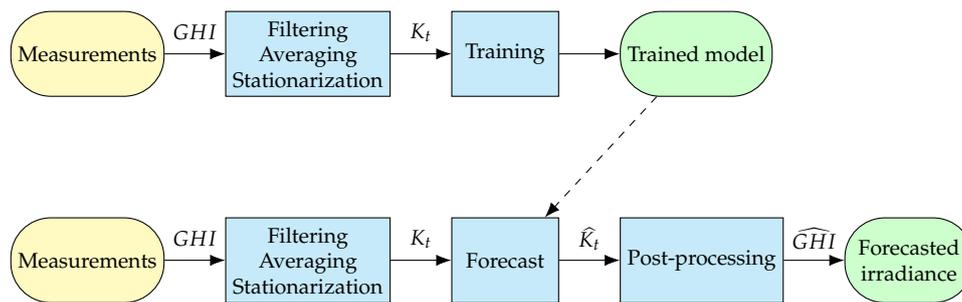
Figure 3. Training and forecast block diagrams.

2.5. Model Error Evaluation

When the training is complete, the model is tested over two years worth of data. The measurements go through steps 1 to 4 of the preprocessing (Section 2.4) then are used as input for the model. It yields a matrix of q forecasts of the Clear Sky Index $\widehat{K}_t$ done every Δt over the testing period. Those $\widehat{K}_t$ are multiplied by the corresponding Clear Sky values to give the forecasted irradiance $\widehat{GHI}$, represented as the "Post-processing" phase in Figure 3.

Hence for a given time t_i, the models use as input the p past measurements $\left[GHI_{i-p+1}, \ldots, GHI_i\right]$ and returns q forecasts $\left[\widehat{GHI}_{i+1}, \ldots, \widehat{GHI}_{i+q}\right]$. To evaluate the models' performances, the forecasts $\widehat{GHI}_k$ are compared to the measured value GHI_k and the Normalized Root Mean Squared Error is computed from Equation 7, where $\langle \quad \rangle$ represents the mean function.

$$nRMSE = \frac{\sqrt{\left\langle \left(GHI - \widehat{GHI}\right)^2 \right\rangle}}{\langle GHI \rangle} \tag{7}$$

2.6. Experimental Methodology

Each of the selected models is configured with a maximum horizon of 6 h and for time-steps of 10, 15, ..., 60 min. Those models are trained over a year and then tested on two years. This gives 11 series of nRMSE values, one for each time-step length containing the error for each horizon. The series do not have the same number of points: the one corresponding to a 10-min time-step has 36 values and the one corresponding to a 60-min time-step has only 6. To ease the graphic representation, all the values are interpolated to match the 10-min time-step series.

To improve the optimal control of a given micro-grid, infra-hour forecasts can be used to obtain more precise set-points of the system. Hence, a second set of models has been trained and tested for the very short term forecasting. In this case, the time-step varies from 2 min to 10 min and the horizon from 10 min to one hour.

3. Results

The results are presented in five subsections. In the first one, the four forecasting models are compared and the best one is chosen to be detailed in the remainder of the paper. The second subsection compares the results for one configuration on the two locations. The main results are presented in the third subsection, where the metrics are computed for several time-steps and horizons for the global horizontal irradiance. The fourth subsection discusses briefly the advantages of doing several forecasts with a small time-step rather than one unique forecast. Finally, the methodology is applied on tilted irradiance in Ajaccio.

3.1. Forecast Model Comparison

All the models mentioned in Section 2.3 have been tested with the same methodology. Figure 4 presents a selection of results in Odeillo. Due to the large number of cases tested and to avoid overloading the paper, only the results for 20 to 60 min time-steps for one site are plotted.
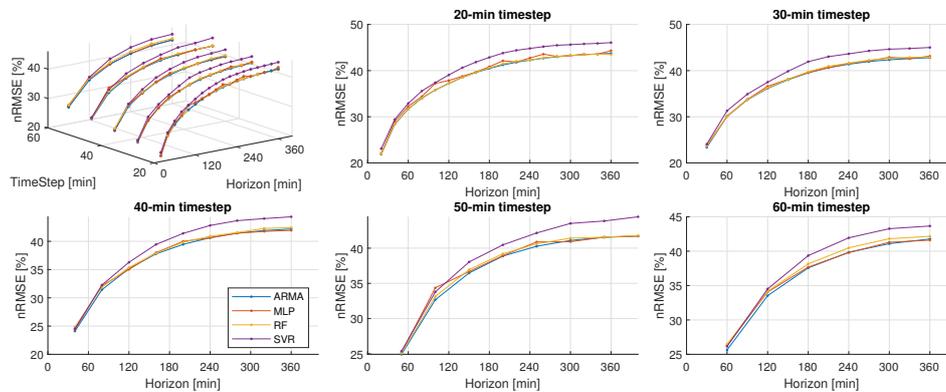
Figure 4. Comparison of the forecasting models for different time-steps and horizons at Odeillo.

Auto Regressive Mobile Average, Multi Layer Perceptron and Random Forest models have similar performances over the studied range. Support Vector Regression models are however slightly less performing. As those results have already been observed in [26], where the performances of eleven forecast models are compared, they will not be further detailed here. The models' performances could be improved by fine-tuning their meta-parameters but this would go beyond the scope of this paper.

Although the models behave in the same way, most of the times ARMA offers the best performances. Furthermore, it is the easiest model to implement and the faster to run. In [22], the authors found similar conclusions. The short term evolution of the irradiance is indeed quasi-linear and using more complex methods does not increase the performances. If more data is used for the training, nonlinear models can have better performances than ARMA but the amount of data needed is rarely available. Hence, for the rest of the study and for the sake of concision, only the results obtained with ARMA models will be presented.

3.2. Test Location Comparison

The study has been done in two sites to verify its validity. Although it is not enough to generalize the findings, it increases the confidence on the results. Figure 5 compares the results for a 30 min time-step in the two locations. As expected, the forecast error increases with the horizon for ARMA and Scaled Persistence models and it is lower in Ajaccio than in Odeillo due to the high variability of the irradiance in mountain areas. This is consistent with the results in [26], where the models' performances on four datasets with different weather variability have been compared.

The Scaled Persistence forecast offers relatively good results for the very short term but after less than three hours, it has worse results than the Climatology model [34]. ARMA is always the best model, even at 6-hour horizon where its error tends to a horizontal asymptote.

As mentioned in Section 2.6, the error metrics are computed over a 2-year period. To illustrate that this time-span is enough to obtain a converged value, the evolution of the nRMSE for the ARMA model with a 30-min time-step and an horizon of one hour is plotted in Figure 6.

To plot this figure, the errors are taken in random order, then the cumulative nRMSE is computed. This means that 2000 points on the horizontal axis correspond to 2000 points taken randomly in the two-year test period. If they had not been randomized, the first 2000 points would have corresponded to the start of the test period and then depended on seasonality variation. The metrics convergence is a complex problem and will not be further detailed in this paper and will be subject to future works.
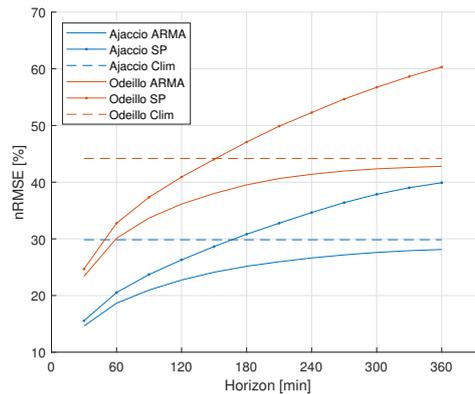
Figure 5. nRMSE for the two sites and for Auto Regressive Moving Average (ARMA), Scaled Persistence and Climatology models for a 30 min time-step.
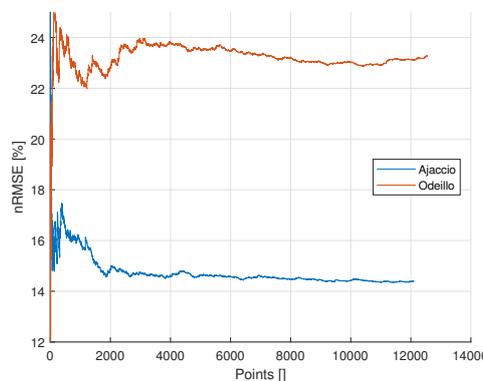
Figure 6. Convergence of the nRMSE for ARMA model versus the number of points taken randomly over the two-year test period.

3.3. Errors for Horizontal Irradiance

Figure 7 presents the results obtained with ARMA for Ajaccio and Odeillo with a time-step varying from 10 min to one hour and a maximum horizon of 6 h or slightly more. It is indeed not possible to reach exactly 360 min with some time-steps (25 min for example) so the maximum horizon is given by the first step reached after 6 h. The results are interpolated to match the 10-min time-step so they can be plotted as a surface and better compared.

Figure 8 presents the results for Ajaccio. The increase of the error with the horizon, already shown in Figure 5, can still be seen. The novelty here is the slight reduction of the nRMSE when the time-step increases. The short term unpredictable variations of the irradiance are indeed smoothed by the averaging, leading to a smaller variability and an "easier" forecast. This is illustrated in Figure 9, where the look of an irradiance measurement is plotted with averaging window lengths of 10, 30 and 60 min. The 60-min window plot is "smoother" and closer to the data trend than the algorithm tends to forecast.

The 3D figures are difficult to exploit and only presented for the sake of illustration. To quantify the nRMSE decrease with the resolution, the surface is sliced at six horizons, from one hour to six hours. The points obtained are plotted in Figure 10a,b. Note that some points are derived by interpolation; as we already said, it is indeed not possible to obtain a 60-min horizon with a 25-min time-step. The results may be fitted with a first order polynomial with relatively good correlation ($R^2 > 0.92$). The fit coefficients are presented in Table 2. With a second order polynomial, the fitting offers better results ($R^2 > 0.98$) but in the scope of choosing the time-step for a micro-grid optimal control algorithm, a quadratic fit seems to be less relevant.
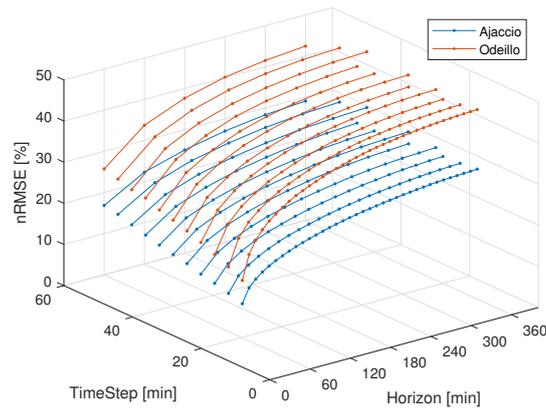
Figure 7. Error of ARMA models for the different time-steps and horizons in Ajaccio and Odeillo.
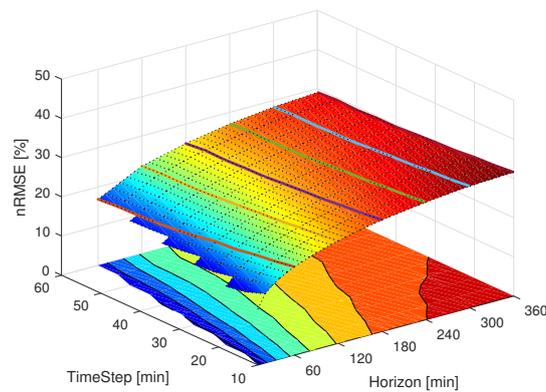
Figure 8. Interpolated results for Ajaccio. The colored lines are extracted to Figure 10a.
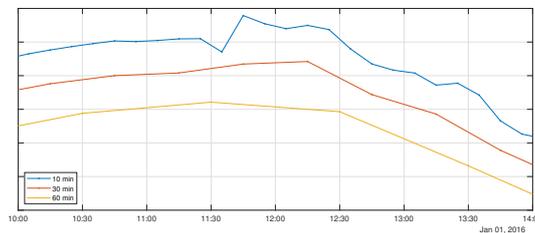
Figure 9. Same irradiance measurement averaged with a 10/30/60-min window. The plots have been offset for better readability.
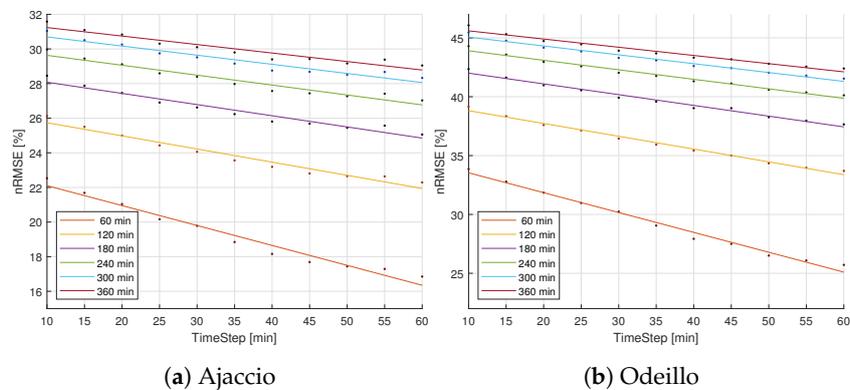
(**a**) Ajaccio　　　　　　　　　　　(**b**) Odeillo

Figure 10. Error (nRMSE) vs. time-steps (Δt) for time-step 10 to 60 min and horizons 1 to 6 h.

Table 2. Fitting results for Ajaccio and Odeillo, horizons 1 to 6 h.

Horizon	Ajaccio			Odeillo		
	a	b	R^2	a	b	R^2
60	−0.115	23.3	0.968	−0.169	35.2	0.988
120	−0.076	26.5	0.959	−0.109	39.9	0.990
180	−0.065	28.7	0.949	−0.091	42.9	0.983
240	−0.057	30.2	0.940	−0.081	44.7	0.980
300	−0.053	31.2	0.929	−0.075	45.8	0.976
360	−0.049	31.7	0.918	−0.070	46.3	0.964

The results differ for each horizon. For short term horizons, increasing the resolution has a significant effect on the error: for example, in Odeillo at 60 min, a 10-min reduction in the time-step results in a 1.7% nRMSE increase. This is less pronounced for long term horizons; e.g., 0.8% per 10 min at 6 h in Odeillo. As mentioned before, the nRMSE is lower in Ajaccio. The trade-off between resolution and error is similar; however, the coefficients are slightly lower.

In order to assess the extent to which those results are specific to horizon, we have reproduced this experiment for the very short term. Figure 11 shows the results for time-steps from 2 min to 10 min and horizons 10 min to 60 min. The fitting coefficients are presented in Table 3. At the very short term, the nRMSE is lower but the slope is greater than between one to six hour horizons. This means that for the now-casting, reducing the time-step can drastically increase the nRMSE, up to 0.8% per minute in Odeillo at 10 min horizon.
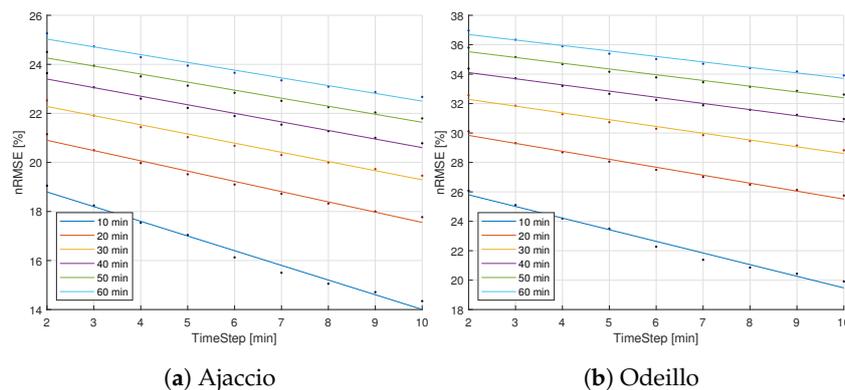
(**a**) Ajaccio (**b**) Odeillo

Figure 11. Error (nRMSE) vs. time-step (Δt) for time-steps 2 to 10 min and horizons 10 to 60 min.

Table 3. Fitting results for Ajaccio and Odeillo, horizons 10 to 60 min.

Horizon	Ajaccio			Odeillo		
	a	b	R^2	a	b	R^2
10	−0.599	20.0	0.982	−0.791	27.4	0.982
20	−0.419	21.7	0.985	−0.541	30.9	0.987
30	−0.374	23.0	0.983	−0.460	33.2	0.983
40	−0.349	24.1	0.981	−0.420	34.9	0.981
50	−0.328	24.9	0.978	−0.391	36.3	0.977
60	−0.316	25.7	0.977	−0.373	37.4	0.976

The one-hour horizon has been tested with time-steps between 2 and 10 min for very short term and time-steps from 10 to 60 min for short term forecasts. Figure 12 shows those results concatenated to illustrate the nRMSE reduction on a wide range. It shows the continuity between the results with a greater slope for the short time-step forecasts.
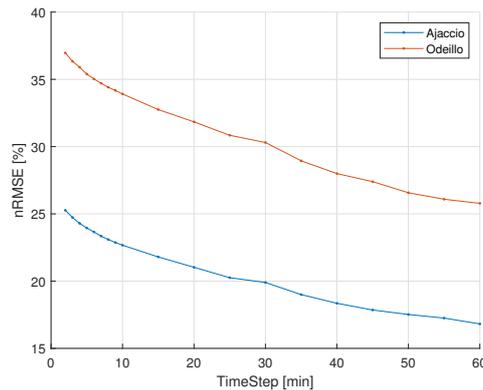
Figure 12. Very short term and short term results at one hour horizon.

3.4. Forecast Averaging

It is important to note that although the error for small time-steps is higher, it is lower on average. In other terms, to forecast the average irradiance between t and $t + 60$ min, it is better to forecast at several smaller steps rather than to do one forecast with a 60-min time-step. For example, six forecasts can be done with a 10-min time-step and averaged to obtain the hourly value.

To illustrate this statement, the forecast at 60 min of several models, with time-steps varying from 2 to 60 min, are compared in Ajaccio, still with the ARMA model. For the 2-min time-step, 30 forecasts are averaged, 20 forecasts for the 3-min model, etc. Their nRMSE is plotted in Figure 13. As stated before, small steps models can better estimate the average irradiance for the next hour. In the case of Ajaccio, reducing the time-step from 60 min to 5 min offers a 1.6% reduction of the nRMSE. However, going below 5 min does not improve the forecast accuracy. Yet for longer horizons this improvement is less significant.
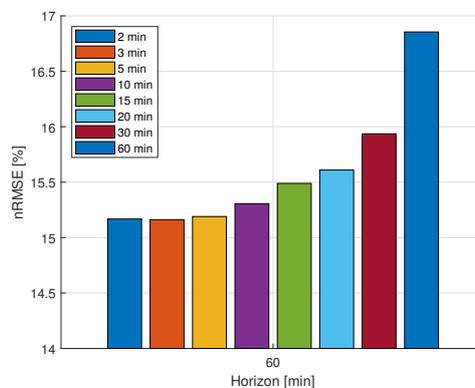
Figure 13. nRMSE at one hour for different time-steps in Ajaccio.

The improvement may be very useful at the sunrise hours. As no data is available at night, univariate models are forced to fetch irradiance values from the day before. Obviously, a lot of weather modifications can append during the night and there is only a small chance that the clear sky index will remain constant. A model with a 60-min time-step needs to wait 60-min after the sunrise so that the first historical data comes from the same day. On the contrary, a 10-min time-step model only needs 10 min after dawn. Since the first lag is the most significant data, it makes the forecast more accurate.

However, the minimum forecast time-step is given by the measurement interval of the input data. Unfortunately, the measurements are far too often available on an hourly basis, especially for older weather stations. The resolution they provide may not be sufficient to run an optimal control algorithm for the micro-grid.

3.5. Errors for Tilted Irradiance

The study has been done mainly with global horizontal irradiance to comply with most of the literature and because that is the most widely available. However, in reality, the PV panels are for the vast majority of cases, inclined to increase their production. That is why, for an operational use in a micro-grid, the most significant measurement is the global irradiance incident on the PV panels.

To assess the validity of the study on tilted data, the same methodology is applied to global irradiance incident on a 45° plane oriented to the south in Ajaccio. The ARMA model is this time configured to use the Global Tilted Irradiance (GTI) instead of the GHI.

The trade-off between nRMSE and time-step is once again presented in Figure 14 and in the fitting coefficients in Table 4. The nRMSE is higher for the tilted irradiance than for the horizontal one. This phenomenon has already been observed in [47] and is mainly due to the difficulties of estimating the diffuse irradiance. The slope coefficients are, however, lower than for the horizontal irradiance, meaning that reducing the resolution would be less useful.
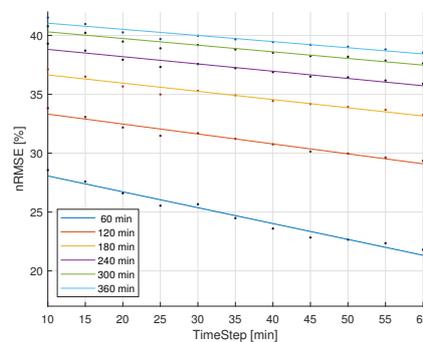
Figure 14. Error (nRMSE) vs. time-step (Δt) for tilted irradiance (45°) in Ajaccio.

Table 4. Fitting results for Tilted Irradiance (45°) in Ajaccio.

	Ajaccio		
Horizon	*a*	*b*	R^2
60	−0.135	29.4	0.971
120	−0.084	34.2	0.959
180	−0.070	37.3	0.945
240	−0.062	39.4	0.933
300	−0.056	40.9	0.926
360	−0.052	41.6	0.916

In addition, the remark on averaging the forecast results, mentioned in the previous section, still holds and should be taken into account when choosing the time-step.

4. Conclusions

This paper presented the evolution of the forecast error with respect to the forecast time-step. Four models — namely ARMA, MLP, RF and SVR— have been tested over two years on two sites: Ajaccio and Odeillo, in France. They present the same behavior and the model offering the best performances, ARMA, is presented in the rest of the study. The error metrics are computed for time-steps varying from two minutes to one hour and for horizons from ten minutes to six hours.

The results show that, for the Global Horizontal Irradiance, increasing the time-step can lead to a reduction in the nRMSE, up to 0.8% per minute at short term in Odeillo. The Global Tilted Irradiance forecasts are less accurate but their error can be reduced by 1.3% per 10-min time-step increase for the one hour horizon.

The methodology shown in this paper may help the designers of future EMS for integrated micro-grids to choose the best time-steps for theirs algorithms. They will have to balance the computational cost of high resolution optimizations and the forecast errors at small time-steps.

Future works will assess the convergence of the error metrics mentioned earlier as well as a methodology to better compare the error at different time-steps. The models will also be improved to increase their performances for tilted irradiance and at sunrises. The idea would be then to test adaptive time-steps to best fulfill EMS resolution requirements all day long.

References

1. Detollenaere, A.; Van Wetter, J.; Masson, G. *Snapshot of Global PV Markets 2020*; Technical Report; IEA—Photovoltaic Power Systems Programme: Paris, France, 2020.

2. *Photovoltaic Barometer 2019|EurObserv'ER*; European Union: Brussels, Belgium, 2020.

3. Heptonstall, P.; Gross, R.; Steiner, F. *The Costs and Impacts of Intermittency—2016 Update: A Systematic Review of the Evidence on the Costs and Impacts of Intermittent Electricity Generation Technologies*; UK Energy Research Centre Report; UK ERC: London, UK, 2017; ISBN 1-90314 404-3. [CrossRef]

4. Lara-Fanego, V.; Ruiz-Arias, J.A.; Pozo-Vázquez, D.; Santos-Alamillos, F.J.; Tovar-Pescador, J. Evaluation of the WRF model solar irradiance forecasts in Andalusia (southern Spain). *Sol. Energy* **2012**, *86*, 2200–2217, [CrossRef]

5. Notton, G. Importance of islands in renewable energy production and storage: The situation of the French islands. *Renew. Sustain. Energy Rev.* **2015**, *47*, 260–269. [CrossRef]

6. Lueken, C.A. Integrating Variable Renewables into the Electric Grid: An Evaluation of Challenges and Potential Solutions. Ph.D. Thesis, Carnegie Mellon University, Pittsburgh, PA, USA, 2012. [CrossRef]

7. Horin, C; Cohen, G; Apt, J. *The Costs of Solar and Wind Power Variability for Reducing CO_2 Emissions*; Working Paper; Carnegie Mellon Electricity Industry Center: Pittsburgh, PA, USA, 2011.

8. Notton, G.; Nivet, M.L.; Voyant, C.; Paoli, C.; Darras, C.; Motte, F.; Fouilloy, A. Intermittent and stochastic character of renewable energy sources: Consequences, cost of intermittence and benefit of forecasting. *Renew. Sustain. Energy Rev.* **2018**, *87*, 96–105. [CrossRef]

9. Heinemann, D.; Lorenz, E.; Girodo, M. Forecasting of solar radiation. In *Solar Energy Resource Management for Electricity Generation from Local Level to Global Scale*; Nova Publishing: Oldenburg, Germany, 2006.

10. Espinar, B.; Aznarte, J.L.; Girard, R.; Moussa, A.; Kariniotakis, G. Photovoltaic Forecasting: A state of the art. In Proceedings of the 5th European PV-Hybrid and Mini-Grid Conference, Tarragona, Spain, 29–30 April 2010.

11. Kostylev, V.; Pavlovski, A. Solar power forecasting performance—Towards industry standards. In Proceedings of the 1st International Workshop on the Integration of Solar Power into Power Systems, Aarhus, Denmark, 24 October 2011; pp. 1–8.

12. Bordin, C.; Anuta, H.O.; Crossland, A.; Gutierrez, I.L.; Dent, C.J.; Vigo, D. A linear programming approach for battery degradation analysis and optimization in offgrid power systems with solar energy integration. *Renew. Energy* **2017**, *101*, 417–430. [CrossRef]

13. Theo, W.L.; Lim, J.S.; Wan Alwi, S.R.; Mohammad Rozali, N.E.; Ho, W.S.; Abdul-Manan, Z. An MILP model for cost-optimal planning of an on-grid hybrid power system for an eco-industrial park. *Energy* **2016**, *116*, 1423–1441. [CrossRef]

14. Rigo-Mariani, R.; Sareni, B.; Roboam, X.; Turpin, C. Optimal power dispatching strategies in smart-microgrids with storage. *Renew. Sustain. Energy Rev.* **2014**, *40*, 649–658. [CrossRef]

15. Odeim, F.; Roes, J.; Heinzel, A. Power Management Optimization of an Experimental Fuel Cell/Battery/Supercapacitor Hybrid System. *Energies* **2015**, *8*, 6302–6327. [CrossRef]

16. Torreglosa, J.P.; García, P.; Fernández, L.M.; Jurado, F. Energy dispatching based on predictive controller of an off-grid wind turbine/photovoltaic/hydrogen/battery hybrid system. *Renew. Energy* **2015**, *74*, 326–336. [CrossRef]

17. Velarde, P.; Valverde, L.; Maestre, J.; Ocampo-Martinez, C.; Bordons Alba, C. On the comparison of stochastic model predictive control strategies applied to a hydrogen-based microgrid. *J. Power Sources* **2017**, *343*, 161–173. [CrossRef]

18. Kaldellis, J.K.; Zafirakis, D. Prospects and challenges for clean energy in European Islands. The TILOS paradigm. *Renew. Energy* **2020**, *145*, 2489–2502. [CrossRef]

19. Alamo, D.H.; Medina, R.N.; Ruano, S.D.; García, S.S.; Moustris, K.P.; Kavadias, K.K.; Zafirakis, D.; Tzanes, G.; Zafeiraki, E.; Spyropoulos, G.; et al. An Advanced Forecasting System for the Optimum Energy Management of Island Microgrids. *Energy Procedia* **2019**, *159*, 111–116. [CrossRef]

20. Ahmad, A.; Javaid, N.; Mateen, A.; Awais, M.; Khan, Z.A. Short-Term load forecasting in smart grids: An intelligent modular approach. *Energies* **2019**, *12*, 164. [CrossRef]

21. Diagne, M.; David, M.; Lauret, P.; Boland, J.; Schmutz, N. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. *Renew. Sustain. Energy Rev.* **2013**, *27*, 65–76. [CrossRef]

22. Lauret, P.; Voyant, C.; Soubdhan, T.; David, M.; Poggi, P. A benchmarking of machine learning techniques for solar radiation forecasting in an insular context. *Sol. Energy* **2015**, *112*, 446–457. [CrossRef]

23. Pedro, H.T.; Coimbra, C.F. Nearest-neighbor methodology for prediction of intra-hour global horizontal and direct normal irradiances. *Renew. Energy* **2015**, *80*, 770–782. [CrossRef]

24. Yang, D.; Alessandrini, S.; Antonanzas, J.; Antonanzas-Torres, F.; Badescu, V.; Beyer, H.G.; Blaga, R.; Boland, J.; Bright, J.M.; Coimbra, C.F.; et al. Verification of deterministic solar forecasts. *Sol. Energy* **2020**. [CrossRef]

25. Ahlburg, D.A. Error measures and the choice of a forecast method. *Int. J. Forecast.* **1992**, *8*, 99–100. [CrossRef]

26. Fouilloy, A.; Voyant, C.; Notton, G.; Motte, F.; Paoli, C.; Nivet, M.L.; Guillot, E.; Duchaud, J.L. Solar irradiation prediction with machine learning: Forecasting models selection method depending on weather variability. *Energy* **2018**, *165*, 620–629. [CrossRef]

27. Wladimir, K. The thermal zones of the Earth according to the duration of hot, moderate and cold periods and to the impact of heat on the organic world. *Meteorol. Z.* **2011**, 351–360. [CrossRef]

28. Pazikadin, A.R.; Rifai, D.; Ali, K.; Malik, M.Z.; Abdalla, A.N.; Faraj, M.A. Solar irradiance measurement instrumentation and power solar generation forecasting based on Artificial Neural Networks (ANN): A review of five years research trend. *Sci. Total Environ.* **2020**, *715*, 136848. [CrossRef]

29. Cros, S.; Liandrat, O.; Sébastien, N.; Schmutz, N.; Voyant, C. Clear sky models assessment for an operational PV production forecasting solution. In Proceedings of the 28th European Photovoltaic Solar Energy Conference and Exhibition, Villepinte, France, 30 September–4 October 2013.

30. Yang, D. Choice of clear-sky model in solar forecasting. *J. Renew. Sustain. Energy* **2020**, *12*, 026101. [CrossRef]

31. Ineichen, P. A broadband simplified version of the Solis clear sky model. *Sol. Energy* **2008**, *82*, 758–762. [CrossRef]

32. NASA. GSFC Nasa Aeronet. Available online: https://www.earthobservations.org/geoss.php (accessed on 5 March 2020)

33. Voyant, C.; Notton, G.; Kalogirou, S.; Nivet, M.L.; Paoli, C.; Motte, F.; Fouilloy, A. Machine learning methods for solar radiation forecasting: A review. *Renew. Energy* **2017**, *105*, 569–582. [CrossRef]

34. Lauret, P.; Lorenz, E.; David, M. Solar Forecasting in a Challenging Insular Context. *Atmosphere* **2016**, *7*, 18. [CrossRef]

35. Box, G.; Jenkins, G. *Time Series Analysis: Forecasting and Control*; John Wiley & Sons: Hoboken, NJ, USA, 1976.

36. De Gooijer, J.G.; Hyndman, R.J. 25 years of time series forecasting. *Int. J. Forecast.* **2006**, *22*, 443–473. [CrossRef]

37. Rezaie-Balf, M.; Kisi, O.; Chua, L.H.C. Application of ensemble empirical mode decomposition based on machine learning methodologies in forecasting monthly pan evaporation. *Hydrol. Res.* **2018**. [CrossRef]

38. Kalogirou, S.A. Applications of artificial neural-networks for energy systems. *Appl. Energy* **2000**, *67*, 17–35. [CrossRef]

39. Mellit, A. Artificial intelligence techniques for modelling and forecasting of solar radiation data: A review. *Int. J. Artif. Intell. Soft Comput.* **2008**, 52–76. [CrossRef]

40. Breiman, L. *Random Forests*; Springer: Berlin, Germany, 2001; Volume 45, pp. 5–32.

41. Ibrahim, I.A.; Khatib, T. A novel hybrid model for hourly global solar radiation prediction using random forests technique and firefly algorithm. *Energy Convers. Manag.* **2017**, *138*, 413–425. [CrossRef]

42. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer Science & Business Media: Berlin, Germany, 1986; p. 324.

43. Global Earth Observation System of Systems. Available online: https://www.earthobservations.org/geoss.php (accessed on 25 January 2020).

44. Geiger, M.; Diabaté, L.; Ménard, L.; Wald, L. A web service for controlling the quality of measurements of global solar irradiation. *Sol. Energy* **2002**, *73*, 475–480. [CrossRef]

45. Voyant, C.; De Gooijer, J.G.; Notton, G. Periodic autoregressive forecasting of global solar irradiation without knowledge-based model implementation. *Sol. Energy* **2018**, *174*, 121–129. [CrossRef]

46. Pankratz, A. *Forecasting with Univariate Box—Jenkins Models: Concepts and Cases*; John Wiley & Sons: Hoboken, NJ, USA, 2009.

47. David, M.; Lauret, P.; Boland, J. Evaluating tilted plane models for solar radiation using comprehensive testing procedures, at a southern hemisphere location. *Renew. Energy* **2013**, *51*, 124–131. [CrossRef]