Solar radiation forecasting based on convolutional neural network and ensemble learning

Solar radiation forecasting based on convolutional neural network and ensemble learning

Davide Cannizzaro [1], Alessandro Aliberti [1], Lorenzo Bottaccioli [1], Enrico Macii [2], Andrea Acquaviva [3] and Edoardo Patti [1]

[1] *Department of Control and Computer Engineering, Politecnico di Torino, 10129 Torino, Italy, email: {name.surname}@polito.it*

[2] *Interuniversity Department of Regional and Urban Studies and Planning, Politecnico di Torino, 10129 Torino, Italy, email: {name.surname}@polito.it*

[3] *Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi", Università di Bologna, 40126 Bologna, Italy, email: {name.surname}@unibo.it*

Abstract

Nowadays, we are moving forward to more sustainable energy production systems based on renewable sources. Among all Photovoltaic (PV) systems are spreading in our cities. In this view, new models are needed to forecast Global Horizontal Solar Irradiance (GHI), which strongly influences PV production. For example, this forecast is crucial to develop novel control strategies for smart grid management. In this paper, we present a novel methodology to forecast GHI in short- and long-term time-horizons, i.e. from next 15min up to next 24h. It implements machine learning techniques to achieve this purpose. We start from the analysis of a real-world dataset with different meteorological information including GHI, in the form of time-series. Then, we combined Variational Mode Decomposition (VMD) and two Convolutional Neural Networks (CNN) together with Random Forest (RF) or Long Short Term Memory (LSTM). Finally, we present the experimental results and discuss their accuracy.

Keywords: `Solar Radiation Forecast, Convolutional Neural Networks, Variational Mode Decomposition, Energy Forecast, Renewable Energy`

1. Introduction

To contrast the negative effects of climate changes, renewable energy is becoming a very attractive solution to produce clean and green energy. Renewable Energy Sources (RES) are increasingly cost-competitive compared to conventional fossil fuel-fired power plants. By the end of 2018, electricity generated from new wind power plants and Photovoltaic (PV) systems had become more economical than power from fossil fuel-fired plants in many places (REN21, 2019). However, the widespread propagation of RES, such as PV systems, is changing the electrical energy management in our cities. This contributes to the development of the so-called smart grids, in which strict rules deal with the electrical energy production, consumption and distribution. In this context, the availability of RES needs to be predicted at various time-horizons to keep the power grid in balance and an important role is played by electrical energy operators (Wan et al., 2015). For example, for an efficient conversion and utilisation of solar energy, the Global Horizontal Solar Irradiance (GHI) should be estimated and forecast through accurate methods and tools. In this context, new control strategies can be developed to maximise the consumption of energy when this is produced by RES. Demand-Response (DR) and Demand Side Management (DSM) are examples of application to achieve this purpose (Deng et al., 2015; Lin et al., 2017; Siano, 2014) that can take advantages from such prediction models.

In recent years, many models have been developed to increase the performance to forecast GHI. In (Monjoly et al., 2017), authors developed a model that uses signal decomposition and combines Autoregressive process and Neural Network to improve short-term GHI forecasts. In (Alanazi et al., 2017), a two-stage day-ahead solar forecasting method is developed. It implements Nonlinear Autoregressive Neural Network and Autoregressive Moving Average with Exogenous Input to break down the prediction into linear and nonlinear components and improve the accuracy of the final results. Currently, the research is focused on building models to predict only short-term or long-term forecasts

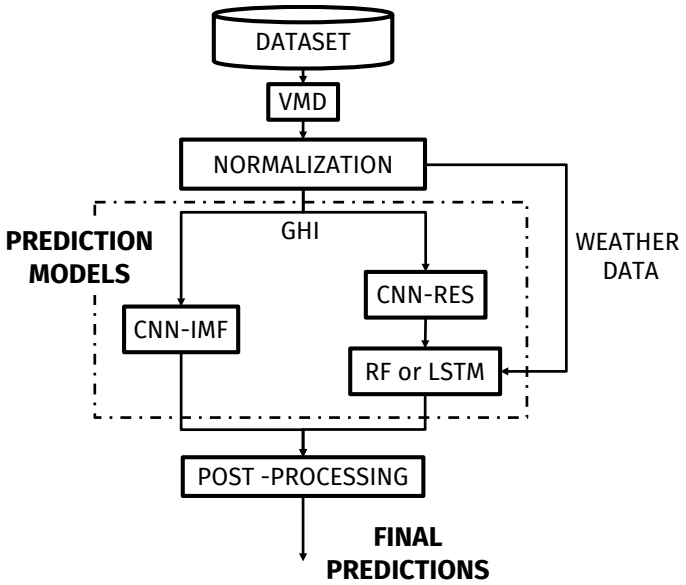


Figure 1: Schema of the proposed methodology.

separately.

Given the promising results using signal decomposition and hybrid models, we proposed an innovative methodology for both short- and long-term GHI predictions (from 15min to 24h) using a dataset with 15min or 1h time-resolution. Figure 1 presents the schema of our methodology for solar radiation forecasts using Variational Mode Decomposition (VMD) and a combination of Convolutional Neural Networks (CNN) and two regression models like Random Forest (RF) or Long Short-Term Memory (LSTM). Typically CNNs have applications in image and video recognition but, in this work, we use them for time-series forecasting in combination with the other models. For the decomposition we applied VMD, a recent developed algorithm that turns out to be more promising than Ensemble Empirical Mode Decomposition (EEMD) and Wavelet Decomposition (WD) (Lahmiri, 2017). Indeed, VMD does not require heavy computation and exhibits good robustness compared to WD and EEMD, as analysed by Dragomiretskiy and Zosso in (Dragomiretskiy & Zosso, 2014). The inputs given to our solution are processed through VMD, the decomposition returns a

set of sub-signals (or modes) that have different bandwidth and a residue that is the difference between input data and modes. As shown in Figure 1, the set of modes is given as input to a first CNN (i.e. CNN-IMF in the figure), while the residue to a second CNN (i.e. CNN-RES). The output of the CNN-RES is given as input, together with weather data, to the chosen regression model (i.e. RF or LSTM). The results from the regression model and CNN-IMF are post-processed to obtain the final prediction. We built a single set of models to be used to predict all the resolution periods requested, increasing the performance. We choose RF because it runs efficiently on large datasets, it is a classifier that evolves from decision trees and correlates them to form a better model. The output model of RF is the class that is the mean prediction of the individual trees. Moreover, as an alternative to RF, we tested also the LSTM due to its ability to predict time-series data (Hochreiter & Schmidhuber, 1997). LSTM belongs to a group of information processing techniques which can be used to find knowledge, patterns or models from a large amount of data in combination with other computational models or methodologies (Kamadinata et al., 2019). In literature, RF and LSTM showed promising results in forecasting solar radiation, in particular when they are involved in hybrid models that combine multiple learning models in a best-of-breed approach (Voyant et al., 2014; Srivastava & Lessmann, 2018). Srivastava and Lessmann propose in their paper an approach to create and forecast GHI with various models. It is noteworthy that satellite based GHI values suffer higher measurement error compared to pyranometer readings (Nottrott & Kleissl, 2010). In particular, forecast models based on satellite based GHI values cannot be validated to a level of accuracy higher than that of the measurements, this will lead to predictions with a certain error level as declared by the authors in their paper. Furthermore, due to these constraints in the use of satellite-derived data they can only perform a prediction with a time step of 3h and starting from 6h to 24h. Instead, our model has a much greater accuracy with a time step of 15min or 1h and a time horizon from next 15min to next 24h. In their paper, Srivastava and Lessmann clarify their aim in developing such a model. They propose a universally appli-

cable framework for forecast model comparisons that facilitates testing models at any geo-location where satellite data is available. They create a model to extract GHI data and then predict these data with various test models to verify their effectiveness. While our goal is completely different, we propose an hybrid model to accurately predict the next values of GHI from previous real GHI data and a set of meteorological parameters each of which is extremely important to the accuracy and final prediction of our model. We trained and tested the neural networks with a dataset consisting of six years of GHI and weather samples collected by a real weather station in our University campus. The CNN is a regularised version of multi-layer perceptrons exploiting a high number of regressors feed with GHI data pre-processed with VMD. LSTM and RF are trained using meteorological data provided by the weather station. Then, the results obtained are post-processed to provide the final GHI with the prediction range desired. The GHI predictions obtained by the proposed model can be given as input to tools for forecasting PV production, for example PVsim (Bottaccioli et al., 2017).

Two models are developed to test different time-resolutions (15min and 60min, a.k.a. M15 and M60, respectively) and prediction horizons (from 15min to 24h). The two models differ only on the granularity of both i) sampling time of measurements in the input dataset and ii) time-step of the output forecast. Thus, M15 has a 15min time-resolution which makes it suitable for sub-hourly predictions in the short-term; while, M60 has a 60min time-resolution suitable for hourly forecasts in the long-term.

The remaining part of the paper is organised as follows: Section 2 reviews literature solution on solar radiation forecast; Section 3 introduces the theoretical background behind our solution; Section 4 provides the performance indicators to evaluate the accuracy of GHI predictions. Section 5 describes the adopted structure and details all the steps performed to initialise, train and test the final model. Section 6 discusses the results obtained on solar radiation forecast and accuracy on GHI exploiting different structures. Finally, Section 7 discusses the concluding remarks.

2. Related works

For decades, solar radiation forecasting methods have been studied in literature using a variety of techniques. Hence, several solutions in literature analyse mathematical and physical models to forecast the solar radiation, such as stochastic models based on time-series (Kaplanis & Kaplani, 2010; Voyant et al., 2014; Badescu, 2008). However, these methodologies often are not sufficient in predicting GHI values due to time-series data that are non-stationary and non-linear (Madanchi et al., 2017; Nazaripouya et al., 2016). Several supervised machine learning techniques (Li et al., 2016a), Support Vector Regression methods (Jiang & Dong, 2016) and neural networks (Qing & Niu, 2018) have been introduced to forecast GHI using historical weather data as input information. In (Li et al., 2016b), authors exploit a deterministic model and a bootstrap multi-layer perceptron to forecast in a very short-term (i.e. 5, 10 and 15min) the Diffuse Horizontal Irradiance component (DHI). Generally, DHI is less investigated in modelling solar energy systems because DHI is part of GHI. Although (Li et al., 2016b) presents very interesting results, as the authors investigate how both cloud transmittance and cloud velocity impact on forecasting. However, this methodology is very limited because it allows only short-term prediction (from 5 to 15min). In addition, in order to have an high forecast fidelity, the authors themselves recommended to incorporate real-time sky and cloud data to the forecasting models and calibrate the cloud velocity derivations.

In (Voyant et al., 2017), Voyant et al. reviewed several machine learning methods based on i) Artificial Neural Networks (ANN); ii) Autoregressive Integrated Moving Average networks (ARIMA); iii) different predictor ensemble methodologies; iv) Support Vector Machine; v) regression trees and vi) Random Forests (RF). In general, the authors state that in terms of quality of prediction in variable conditions, ANN and ARIMA are equivalent. However, ANNs are very efficient in solving non-linear problems (Zhang et al., 1998), while ARIMA assumes that the series are generated from linear processes, thus not suitable for

most real-world problems that are non-linear (Khashei et al., 2009). Moreover, the predictor ensemble methods are more promising than simple predictors. Finally, Support Vector Machine, regression trees and RF present very promising potentials. Currently, numerous studies have been developed using RF to forecast GHI (Benali et al., 2019).

Other studies try to reduce the complexity of forecasting problem and to analyse the correlation between the input data. In (García-Hinde et al., 2018), García-Hinde et al. try to improve the prediction accuracy of solar irradiance obtaining a robust sets of features by performing dimensionality reduction on the input variables. Real data from the weather station and forecasting model, that produce a very large number of variables, are used as input. In (Rodrigues Júnior & Serra, 2017), authors suggested dividing the complex problem of seasonal forecasting (like solar irradiance) into sub-problems easier to deal, forecasting separately each undetectable component. They propose a methodology based on a Neuro-Fuzzy Takagi-Sugeno (NF-TS). The NF-TS considers the undetectable components extracted from the time-series data to evolve because they present dynamic behaviours that are simpler to forecast. In (Galván et al., 2017), instead of providing a single measure for the prediction, lower and upper bound of solar radiation are used (prediction interval). The algorithm uses neural networks to create complex non-linear models whose outputs are the upper and lower limits of the prediction intervals. Numerous machine learning algorithms are tested to increase the performance in forecasting solar irradiation. In (Cao & Lin, 2008), the authors established a Diagonal Recurrent Wavelet Neural Network so as to carry out forecasting of the hourly global solar irradiance. The authors themselves recommended studying further the application of more weather information to forecast solar irradiation. Other algorithms are tested in literature like Support Vector Regression and regression trees that show advantages in GHI forecasting but they are not suitable to deal with a large volumes of incoming data. Unlike them, CNN allows to have fewer weights and shared parameters. CNN is particularly suitable to deal with two-dimensional arrays and uncertainties due to weather in time-series data like wind (Wang et al.,

2017b) and PV power (Wang et al., 2017a). To improve the effectiveness of the CNNs, a time-series decomposition can contribute to have an higher prediction accuracy when utilised to pre-process the data, such as VMD (Sun et al., 2018; Majumder et al., 2018). In (Zang et al., 2018), the authors developed a set of hybrid models that are a combination of different neural networks and they use them to forecast PV power production with different time-resolutions. In (Aliberti et al., 2018), authors developed a Nonlinear Autoregressive (NAR) neural network for short-term forecast of GHI and Photovoltaic energy prediction.

In literature, the models proposed focus only on predicting a short range of prediction horizons. Moreover, such models develop only short- or long-term predictions. Our methodology overcomes these problems. With respect to literature solutions, we propose a novel ensemble model based on the combination of multiple methodologies to forecast GHI by exploiting also the other meteorological measurements we considered as input parameters. The model takes advantages of a recently developed decomposition algorithm, VMD, that performs better than previous techniques like Ensemble Empirical Mode Decomposition and Wavelet Decomposition (Lahmiri, 2017) in both computational performances and robustness in decomposition (Dragomiretskiy & Zosso, 2014). The novelty of our approach is that there are not other examples of models in literature developed to predict GHI data for both short- and long-term (from 15 minutes to 24 hours) using a combination of multiple models and weather information. In its core, our methodology consists of a combination of two Convolutional Neural Networks and two regression models (i.e. Random Forest and Long Short-Term Memory). To improve the performance in the forecasts, our methodology distinguishes the input data between the four meteorological seasons. Both training and test have been performed by exploiting a dataset consisting of six years of real weather information sampled every 15min by a real weather station in our University campus. The meteorological measures we consider are: i) GHI, ii) air temperature, iii) relative humidity, iv) air pressure, v) wind speed and vi) cloud coverage. Whilst, solutions like (Zhang et al., 1998; Khashei et al., 2009; Benali et al., 2019; Aliberti et al., 2018) only exploit GHI

time-series. Finally, with this methodology, we developed two models to test two different time-resolutions, 15min and 60min respectively, with a prediction horizon ranging from next 15min up to 24h. Thus, our methodology allows both sub-hourly and hourly predictions in short- and long-term. It is worth noting that these two models differ only on the time-resolution of measurements in the input dataset that affects the time-step in the output forecast (i.e. predictions every 15min or 60min, respectively).

3. Theoretical Background

In this section, we analyse in detail the main algorithms used to create the model proposed in Section 5. We used the VMD, as decomposition algorithm, and a combination of i) CNNs and RF, ii) CNNs and LSTM to finally forecast the GHI.

3.1. Variational Mode Decomposition

As proposed by (Dragomiretskiy & Zosso, 2014), the role of the Variational Mode Decomposition (VMD) is to decompose an input signal into a discrete number of sub-signals, that have specific sparsity properties. The use of VMD allows to properly cope with noise, to determine the relevant bands adaptively and to estimates the corresponding modes concurrently while balancing errors. VMD leads to a decomposition of a signal f into its K principal modes, u_k, called Intrinsic Mode Function (IMF) based on modulation criteria. Each signal is expressed as:

$$u_k(t) = A_k(t)cos[\phi_k(t)]$$

where A_k is a non-negative envelope and ϕ_k is a non-decreasing function with $\omega_k(t) := \phi(t)'_k$ the centre pulsation of the mode. In a nutshell, the mode can be considered to be a pure harmonic signal with amplitude A_k and instantaneous frequency ϕ_k. To obtain the final modes, it is required to solve the following minimisation problem:

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \quad s.t. \sum_{k=1}^{K} u_k = f \qquad (1)$$

where $\{u_k\} := \{u_1, \cdots, u_K\}$ and $\{\omega_k\} := \{\omega_1, \cdots, \omega_K\}$ are shorthand notations for the set of all modes and their centre frequencies, δ is the Dirac distribution, π is the mathematical constant and j the imaginary unit. To address the reconstruction constraint, a quadratic penalty term (α) and Lagrangian multipliers (λ) are introduced in order to render the problem unconstrained. Therefore, the augmented Lagrangian $\mathcal{L}$ is set-up as follows (Nocedal, 2006):

$$\mathcal{L}(\{u_k\}, \{\omega_k\}, \lambda) := \alpha \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_{k=1}^{K} u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_{k=1}^{K} u_k(t) \right\rangle$$

The solution to the original minimisation problem (1) is now found as the saddle point of the augmented Lagrangian $\mathcal{L}$ in a sequence of iterative sub-optimisations called Alternate Direction Method of Multipliers (Bertsekas, 2014). In order to achieve the final solution, the iterative process is applied to u_k and ω_k, taking u_k as example, this problem can be solved in spectral domain:

$$\hat{u}_k^{n+1} = arg \min_{\hat{u}_k, u_k \in X} \left\{ \alpha \| j(\omega - \omega_k)[(1 + sgn(\omega))\hat{u}_k(\omega)] \|_2^2 + \left\| \hat{f}(\omega) - \sum_i \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2} \right\|_2^2 \right\} \quad (2)$$

where n is the number of iterations. Firstly, we select the number of K modes, typically this step is done observing the central frequency of the IMFs. We observe the decomposition of the modes with different values of K, when the value selected is too high, duplicated modes with a similar central frequency appear (Over-segmentation); instead, when the value chosen is too low some components are not correctly decomposed and they are contained in other modes (Under-segmentation). Then, the VMD algorithm is applied: the signal is extended by mirroring and analysed through a Fast Fourier transform to allow the decomposition in the frequency domain. The signal is reconstructed and the spectrum analysed to check the consistency of the decomposition. The de-

composition of the signal in K modes plus a residue is obtained as:

$$f(t) = \sum_{k=1}^{K} u_k(t) + res(t)$$

where $u_k(t)$ is the k^{th} IMF and $res(t)$ is the residue.

3.2. Convolutional Neural Network

In deep learning, a Convolutional Neural Network (CNN) is a feed-forward deep neural network. CNNs are regularised versions of multi-layer perceptrons inspired by biological processes. Firstly researched by Hubel and Wiesel (Hubel & Wiesel, 1968), CNN has applications in image and video recognition, image classification and natural language processing (Krizhevsky et al., 2017). A CNN consists of an input layer and an output layer with multiple hidden layers in the between. The hidden layers are typically a set of convolutional layers with a Rectified Linear Unit (ReLU) as activation function, followed by a flatten layer and fully connected layers (dense layers). All these layers are stacked sequentially to create the CNN, there are more layers than the previous described such as dropout but the use may not be suitable for all the situations and worse the model results.

During the development of the network, some parameters have to be selected manually such as number of filters, filter shape, batch size, iteration epoch and initial learning rate. These parameters, named hyper-parameters, can be determined through optimisation algorithm like grid search, random search or gradient-based optimisation. In our methodology, grid search is applied to obtain the optimal parameters for the task.

3.3. Random Forest regressor

Random Forest (RF) regression is a classifier that evolves from decision trees, it combines the predictions from multiple decision trees algorithms together to obtain more accurate results than any individual model. This technique uses multiple models to obtain better performance and is called model ensembling. In RF, all the base trees are constructed independently using a deterministic

algorithm by selecting a random set of input data from the training-set. Firstly researched by Ho (Ho, 1995), RF runs efficiently on large datasets and can handle thousands of input variables without variable deletion. In RF, typically three parameters need to be optimised: i) *number of estimators* - they represent the number of trees in the forest; ii) *max depth* - it restricts the depth of each individual tree to prevent over-fitting (if set to none there is no restriction); iii) *min samples leaf* - it sets the minimal size of the terminal nodes of the trees. These parameters can be optimised through grid search cross-validation method to improve the predictions of the model.

3.4. Long Short-Term Memory Neural Network

Long Short-Term Memory (LSTM) is an artificial Recurrent Neural Network architecture capable of learning long-term dependencies. LSTM is introduced by Hochreiter and Schmidhuber (Hochreiter & Schmidhuber, 1997), it is explicitly designed to avoid the long-term dependency problems that led to the vanishing gradient problem emerged as a major obstacle to Recurrent Neural Network performance (Hochreiter, 1998). All Recurrent Neural Networks have the form of a chain of repeating modules of neural network called cells. Typically, an LSTM is composed of a set of cells (hidden layer) and a fully connected layer (dense layer). In an LSTM cell, three gates are present: i) *Input gate* - it chooses the input value that should be used to modify the memory. ii) *Forget gate* - it chooses the details to be discarded from the cell. iii) *Output gate* - the input and the memory of the cell are used to decide the output.

4. Performance Indicators

In this study, we adopted the following performance indicators to evaluate the accuracy of our methodology: i) Root Mean Square Error (RMSE), ii) Normalised Root Mean Square Error percentage (NRMSE %), iii) Mean Absolute Error (MAE) and iv) coefficient of determination (R^2). RMSE is calculated as:

$$RMSE = \sqrt{\frac{\sum_{n=1}^{N}(\hat{y}_n - y_n)^2}{N}} \quad [W/m^2]$$

where N is the sample size, $\hat{y}_n$ is the n^{th} predicted value and y_n is the n^{th} real value of GHI.

NRMSE is given as follows:

$$NRMSE = \frac{100\%}{\hat{y}_{max} - \hat{y}_{min}} RMSE \quad [W/m^2]$$

where $\hat{y}_{max}$ and $\hat{y}_{min}$ are respectively the maximum and minimum value of the predicted samples.

MAE is defined as:

$$MAE = \frac{\sum_{n=1}^{N} |\hat{y}_n - y_n|}{N} \quad [W/m^2]$$

Finally, R^2 is given as:

$$R^2 = \frac{\sum_{n=1}^{N} (\hat{y}_n - \bar{y})^2}{\sum_{n=1}^{N} (y_n - \bar{y})^2}$$

where $\bar{y}$ is the mean value of real GHI. With low RMSE, NRMSE and MAE the model is more accurate. R^2 values range between 0 and 1; where 0 indicates no correlations and 1 a perfect correlation.

5. Proposed model

Figure 1 reports the schema with the main steps of our methodology to forecast GHI. In its core, it exploits VMD to decompose the GHI into a set of sub-signals (or modes) and a residue (i.e. difference between input data and modes). After a data normalization, the set of modes and residue are given as input to CNN-IMF and CNN-RES, respectively. The output of the CNN-RES, together with the other meteorological measures in the dataset, feeds one of the chosen regression models (i.e. RF or LSTM). Finally, the outputs of both CNN-IMF and regression model are combined together and post-processed providing the final GHI forecast. Two proposed models to forecast GHI, called M15 and M60, are built with the same structure and they use the same dataset differing only on the time-resolution. M15 is suitable for sub-hourly forecasts with a granularity of 15min; while, M60 has a granularity of 1h. These models are used for short- and long-term forecasts (up to next 24h). The rest of this section presents in detail the dataset and the proposed methodology.

5.1. Dataset and pre-processing

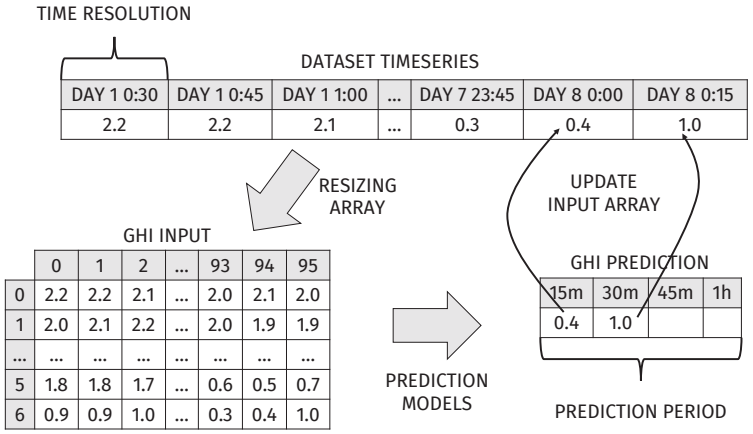
The dataset used in this study consists of six years of meteorological data collected by a real weather station in our University Campus in Turin, Italy, from December 2009 to November 2015 with a time-resolution of 15min. It is composed by six time-series one for each of the following meteorological measures: i) GHI, ii) air temperature, iii) relative humidity, iv) air pressure, v) wind speed and vi) cloud coverage. This dataset has been split into training- and test-set with 175301 and 35041 samples for each meteorological measure, respectively. In particular, training-set ranges from December 2009 to November 2014, test-set from December 2014 to November 2015. Both training- and test-set report data for the full four seasons. We fill missing data with the mean value of previous and next samples in the time series, thus allowing to keep the data aligned to the overall behavior. In (Layanun et al., 2017), the authors study the accuracy of filling the missing GHI samples with mean values. The results show that this method provides decreased errors and leads to a better solar forecasting performance compared to other imputation techniques. In (Demirhan & Renwick, 2018), the authors compare the accuracy of different imputation methods for missing solar irradiance data with different time resolutions. They identify the linear interpolation the most reliable and robust approach for the imputation of minutely solar irradiance values.

To feed both CNNs, after applying VMD, the dataset is normalised between 0 and 1 using the standard normalisation obtained with min-max scaler (LeCun et al., 2012) and then it is split according to meteorological seasons. Meteorological seasons are different from the astronomical seasons. Indeed for temperate areas in the northern hemisphere, spring begins on March 1^{st}, summer on June 1^{st}, autumn on September 1^{st}, and winter on December 1^{st}. This normalisation is needed to have the same range of values for each of the inputs given to the CNNs. Without normalisation, the ranges of our GHI signals decomposed would be different one from another, and thus the learning rate would cause corrections in each dimension that would differ from one another. With the meteorological seasons, the statistically colder, warmer and intermediate

months are those identified by these periods, with the months at extreme thermal averages (usually January and July) that fall in the middle of the respective meteorological season. With this division, extract seasonal statistics from the monthly ones becomes much easier and both of them are very useful for forecasting. As discussed previously, the dataset is composed of weather elements (i.e. air temperature, relative humidity, air pressure, wind speed and cloud coverage) and GHI. As input of both CNNs we use only the GHI values, while, based on the correlation analysis, the weather elements are used to feed the regression model (RF or LSTM) in combination with the output of CNN-RES. To feed the GHI time-series to the CNNs, it is necessary to create a 2D input vector and then to decompose it through VMD. VMD is required to decompose in frequency the behaviour of the GHI. Then the CNNs can fully analyze the signals and extract the relevant features. The input data are built to create a 2D array as shown in Figure 2, one input array is composed by 7 days with 96 GHI points corresponding to a time-resolution of 15min in 24h. Choosing 7 days is a good compromise that allows to maintain the size of the array small and to find repetitive patterns between data and daily correlation. To generalise, the input vector is composed by $n_{day} \times p_{GHI}$, where n_{day} is the number of day chosen and p_{GHI} is the number of GHI points that depends by the chosen time-resolution. To forecast GHI, we use an iterative approach. The proposed model has a single output, so it can only predict one-step ahead. For subsequent steps, the predicted value for time $t+1$ is used as one of the inputs for the prediction at time $t+2$, and so on. The input data are shifted of one resolution point and the process is repeated until the prediction period is reached (i.e. following a sliding window approach).

It is worth to say that to train the CNN with 1h time-resolution, the data are reduced taking the arithmetic mean of the values in 1h (i.e. from 0:00 to 0:45 to obtain the GHI value at 0:00) and obtaining 24 samples. Then, the same process used for 15min time-resolution is applied to build the input vector. To apply VMD we need to set reasonably the number of K modes of the decomposition. To avoid aliasing problems a K value not too high is needed, therefore $K = 9$

is chosen after observing the central frequency of the signal decomposition, as reported in (Zang et al., 2018).



	0	1	2	...	93	94	95
0	2.2	2.2	2.1	...	2.0	2.1	2.0
1	2.0	2.1	2.2	...	2.0	1.9	1.9
...	...	...	...	...	...	...	...
5	1.8	1.8	1.7	...	0.6	0.5	0.7
6	0.9	0.9	1.0	...	0.3	0.4	1.0

Figure 2: Schema to create the input data (15min time-resolution)

Because all the days have similar trends independently from the season, a period of GHI time series is taken as example, which represents the first day of December 2009 with a time-resolution of 15min, and its VMD results are shown in Figure 3. Each IMF has its own central frequency, the sum of all the modes and the residue gives the original GHI signal. The frequencies of the modes increase from the first IMF (low frequency) to the last IMF (high frequency). As previously explained, the original GHI time series is decomposed into 9 IMFs and 1 residue.

In order to make the best use of the meteorological data, we analysed the correlation between the IMFs, the residue and the weather elements. We used the standard Pearson correlation coefficient calculated using the following equation:

$$r_{xy} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum x_i^2 - n\bar{x}^2)}\ \sqrt{(\sum y_i^2 - n\bar{y}^2)}}$$

where n is the sample size, x_i and y_i are the individual sample points indexed with i and $\bar{x}$ and $\bar{y}$ are the sample mean. The correlation is strong when r_{xy} is

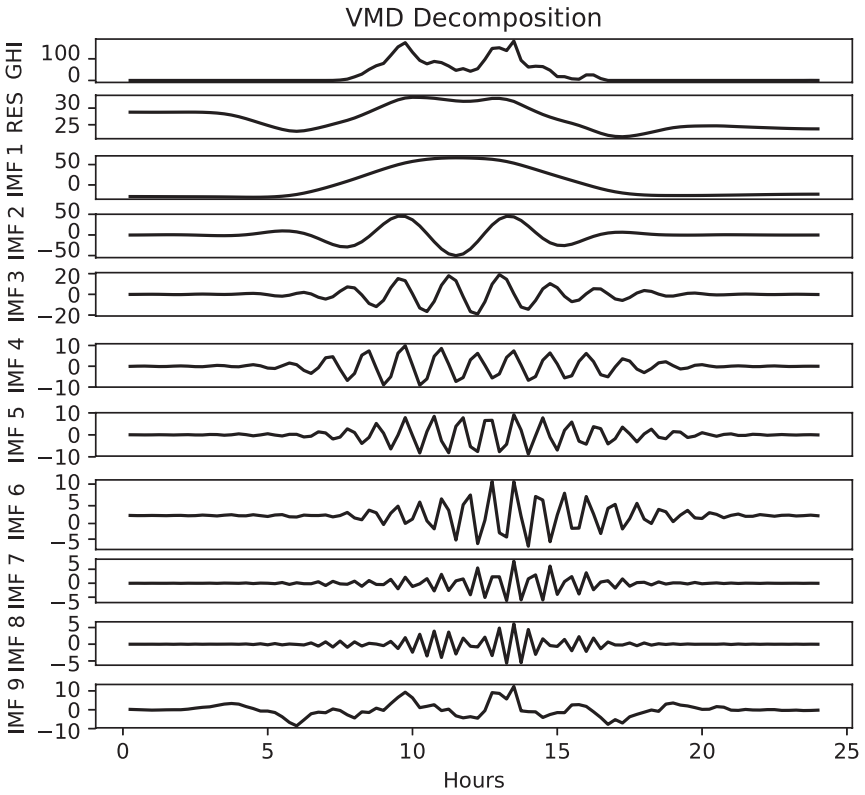


Figure 3: VMD results for the first day in December 2009 (15min time-resolution)

closed to ± 1; if it is near to 0 there is no correlation.

Table 1 reports the results of the correlation analysis for one week in December 2009 (reported as example). We can notice a strong correlation between the residue and the air temperature (≥ 0.9). Whilst, there is a moderate correlation (≥ 0.2) with the other meteorological measures. Moreover, the residue shows a particular correlation with air temperature (0.949) and cloud coverage (0.494). The IMF signals instead show almost no correlation with the meteorological data (< 0.2). Therefore, we choose to use together the feature extracted from the last layer of the CNN-RES and the meteorological data to feed the regression model (one among RF or LSTM) and compute the second part of the GHI.

Table 1: Correlations between decomposed components and weather data.

Component	Temp.	Hum.	Pres.	Wind sp.	Cloud cov.
Residue	**0.949**	**0.314**	**−0.266**	**−0.200**	**0.494**
IMF 1	0.853	0.043	0.082	0.022	0.215
IMF 2	0.008	0.038	0.003	−0.154	0.030
IMF 3	0.026	−0.027	0.005	−0.069	0.006
IMF 4	0.006	−0.010	0.004	−0.031	−0.003
IMF 5	0.007	0.005	0.003	−0.026	0.007
IMF 6	0.007	−0.004	−0.001	−0.008	0.007
IMF 7	0.006	0.011	−0.001	−0.009	0.003
IMF 8	0.001	0.009	0.001	−0.011	0.001
IMF 9	0.343	−0.057	0.030	−0.133	0.241

5.2. CNN-IMF and CNN-RES

The CNN possesses excellent abilities in analyzing linear and nonlinear features using a convolutional process. CNNs are a particular type of neural network for processing data with grid-like topology. CNNs are suitable for time-series data, that can be ordered as a 1D grid or 2D grid-like image data. As discussed in Section 3, the hyper-parameters of both CNNs have to be chosen. One of them is the pooling layer that reduces the dimensions of the data by combining the outputs of neurons at one layer into a single neuron in the next layer. The pooling layer is widely utilised in image processing but it is not suitable for regression problems like GHI forecasts. This is demonstrated by the following analysis. We tested two different types of pooling layers: max-pooling and average-pooling with pool size of 2×2 or 1×2, then the pooling layer is removed for comparison. Those pooling layers are added after the first convolutional layer. In Table 2, we reported the comparison between various pooling layers. The results reported are obtained using only the GHI from the dataset and to verify which pooling layer is most suitable to our case. To obtain the final result, we also need the contribution of the RF or LSTM models with the

Table 2: Comparisons of CNNs with different pooling layers

Model	Pooling layer	RMSE $[W/m^2]$	MAE $[W/m^2]$
CNN-IMF	max-pooling (2×2)	0.004	0.048
	average-pooling (2×2)	0.005	0.049
	max-pooling (1×2)	0.001	0.021
	average-pooling (1×2)	0.001	0.021
	no-pooling	**0.001**	**0.020**
CNN-RES	max-pooling (2×2)	0.006	0.054
	average-pooling (2×2)	0.006	0.056
	max-pooling (1×2)	0.002	0.019
	average-pooling (1×2)	0.002	0.019
	no-pooling	**0.001**	**0.019**

weather elements. For both CNNs, according to the comparison, the best results are obtained using no pooling layer (RMSE: 0.001 for both CNNs, MAE: 0.020 and 0.019 for CNN-IMF and CNN-RES, respectively), this is probably due to the characteristic of the GHI signal. With pool size (2×2), both the pooling methods resulted in information loss and reduced prediction accuracy (RMSE ≥ 0.005 and MAE ≥ 0.048), whereas average-pooling performed worse than max-pooling. Otherwise, with pool size (1×2) the results are comparable with the no pooling CNN (RMSE: 0.001 and 0.002, MAE: 0.021 and 0.019 for CNN-IMF and CNN-RES, respectively). Thus to achieve the high accuracy, no pooling layers are utilised in both CNNs. Table 3 and Table 4 report the final structure including all the other possible hyper-parameters for both CNN-IMF and CNN-RES for M15 and M60, respectively.

5.3. RF and LSTM

RF and LSTM are two alternative regression models combined with CNN-RES in our methodology. Thus the output of CNN-RES, together with the other meteorological data, is given as input to RF or LSTM in both M15 and

Table 3: Structure of CNN-IMF

Layer	CNN 15min (M15)		CNN 1h (M60)	
	Param	Train	Param	Train
Conv2D 1	Input: $7 \times 96 \times 9$ Filters: 16 Kernel size: 3×3 Padding: valid Activation: ReLU	1312	Input: $7 \times 24 \times 9$ Filters: 16 Kernel size: 3×3 Padding: valid Activation: ReLU	1312
Conv2D 2	Filters: 32 Kernel size: 2×2 Padding: valid Activation: ReLU	2080	Filters: 32 Kernel size: 2×2 Padding: valid Activation: ReLU	2080
Flatten	None	0	None	0
Dense 1	Units: 6 Activation: None	71430	units: 6 Activation: None	16134
Dense 2	Units: 1 Activation: None	7	units: 1 Activation: None	7
Total	None Epochs: 20 Batch size: 32	74829	None Epochs: 20 Batch size: 32	19533

M60 models.

The RF model gives more robust performances in different seasons than other models and produces accurate predictions that do not overfit the data. As shown in (Sun et al., 2016), RF models are indicated to forecast solar radiation estimation but the performances reduce when a huge number of weather parameters is involved. Instead, LSTM models exploit the temporal and spatial dependence of data, which represents a major asset in utilizing contextual information. LSTM

Table 4: Structure of CNN-RES

Layer	CNN 15min (M15)		CNN 1h (M60)	
	Param	Train	Param	Train
Conv2D 1	Input: $7 \times 96 \times 1$ Filters: 16 Kernel size: 3×3 Padding: valid Activation: ReLU	160	Input: $7 \times 24 \times 1$ Filters: 16 Kernel size: 3×3 Padding: valid Activation: ReLU	160
Conv2D 2	Filters: 32 Kernel size: 2×2 Padding: valid Activation: ReLU	2080	Filters: 32 Kernel size: 2×2 Padding: valid Activation: ReLU	2080
Flatten	None	0	None	0
Dense 1	units: 6 Activation: None	71430	units: 6 Activation: None	16134
Dense 2	units: 1 Activation: None	7	units: 1 Activation: None	7
Total	None Epochs: 20 Batch size: 32	73677	None Epochs: 20 Batch size: 32	18381

models have shown potential to deal with a large number of parameters, such as in the case of solar radiation forecast (Srivastava & Lessmann, 2018). A small set of parameters with the most significant weather data is selected to exploit both RF and LSTM models. As discussed in the next Section 6, the RF model is more suitable for dataset with a granularity of 15min while the LSTM model works better using a dataset with a granularity of 1h. The parameters to choose needed by RF are the number of estimators, max depth and min samples leaf.

Table 5: Structures for both RF and LSTM

Model	Parameters	Trainable parameters
Random Forest	Number of estimators: 100 Max depth: None Min samples leaf: 1 Criterion: MSE	None
LSTM	Units: 100	45200
Dense 1	Units: 1	101
Total	Epochs: 100 Batch size: 10	45301

Whilst, LSTM only need the number of units. All the chosen parameters and the final structures of RF and LSTM are reported in Table 5.

5.4. Post-processing

To obtain the final prediction, we need to post-process the output from CNN-IMF and the regression model (RF or LSTM). We sum the partial results coming from them to obtain the complete prediction and then we denormalise it. The data are denormalised using the reverse of min-max algorithm used for normalisation, as discussed in Section 5.1. The final result is the GHI forecast.

5.5. Creation and utilisation of the proposed model

As discussed in Section 3, the core of our solution consists on VMD and both CNNs. As shown in Figure 4, the dataset is composed of two parts: i) GHI and ii) Weather elements. The GHI data array ($n_{day} \times p_{GHI}$) is fed to the VMD algorithm, the signal is decomposed in the number of K modes chosen and a residue. Because CNNs are typically used for image processing, the IMFs signals are built together to create a 3D image using K as further dimension ($n_{day} \times p_{GHI} \times K$) and the resulting array is used as input to the channels of CNN-IMF. The output of the proposed model has only one dimension that corresponds to the sum of all the K IMFs. As advantage, using CNN allows

reducing the complexity feeding the model with a 3D array of all the IMFs, instead of creating a model for each IMF signal. As shown in Figure 4, channels work in parallel to train the CNN-IMF using all the IMFs, only a CNN is used to obtain the first part of the final GHI prediction. The residue, that is not involved in the previous part is used to train a second CNN (i.e. CNN-RES). The remaining part of the signal varies over days based on the weather conditions and climate changes. Therefore, CNN-RES is trained to extract model feature from the signal. The final layer of the CNN-RES contains the major characteristic of the signal. As shown in Figure 4, the regression model (RF or LSTM) is fed with the final layer output of CNN-RES and the second part of the dataset, the meteorological data (i.e. i) air temperature, ii) relative humidity, iii) air pressure, iv) wind speed and v) cloud coverage), to predict the second part of the GHI. Finally, the prediction outputs of the two parts are post-processed, as described in Section 5.4. As shown in Figure 2, the proposed model forecasts the GHI until the desired prediction period is reached (e.g. 15min, 1h or 24h), the time-resolution depends on the chosen model (M15 or M60). If the prediction period is longer than the chosen resolution, the algorithm uses the previous GHI predictions to build the next input arrays. Figure 5 shows the flowchart of the applied algorithm. The algorithm runs until the prediction time is reached and the output of all the steps compose the final prediction.

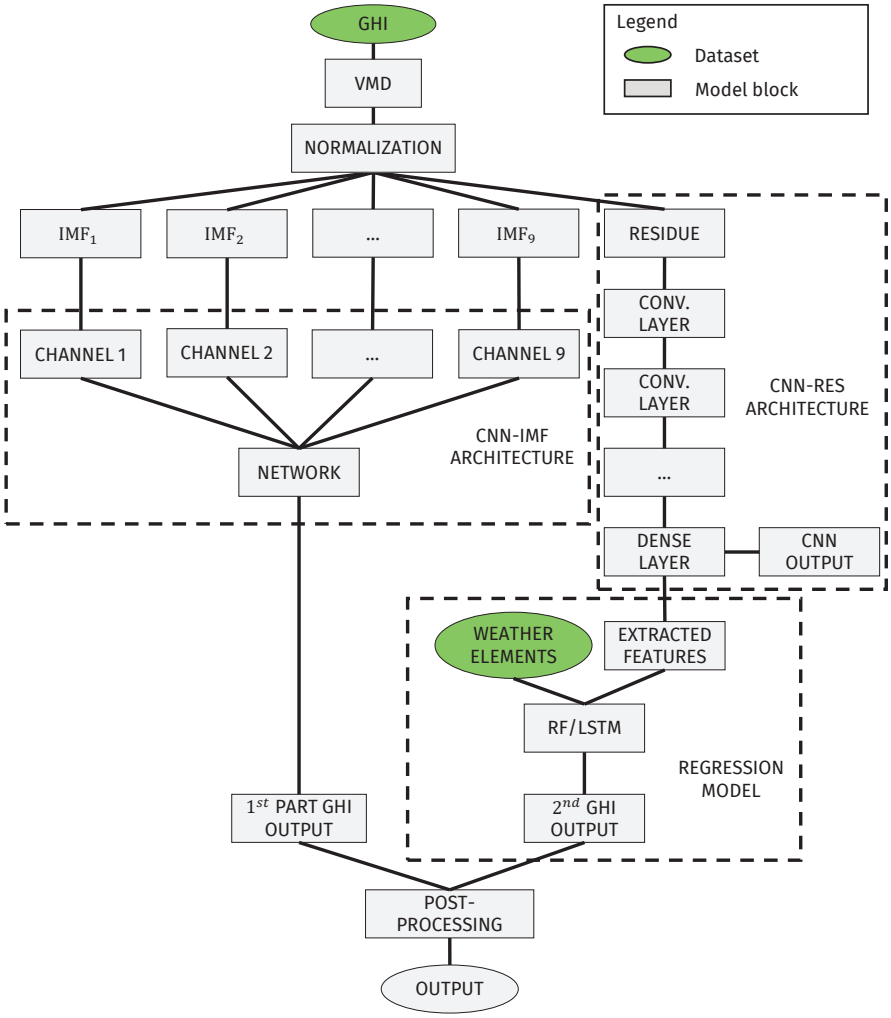


Figure 4: Detailed schema of the proposed methodology

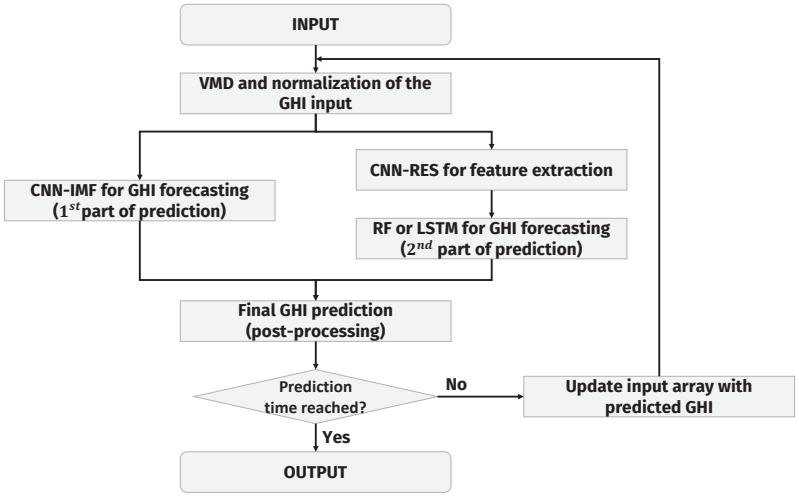


Figure 5: Flow chart of our model to forecast GHI.

6. Experimental results and analysis

This section presents the experimental results of both M15 and M60 models (see Section 5) in forecasting GHI in short- and long-term (i.e. from next 15min up to next 24h). We also evaluate and compare the performance in exploiting either RF or LSTM, as they can be exploited alternatively. To assess the accuracy of our models, we adopted the performance indicators presented in Section 4. First, we analysed each individual season and then the whole year. Finally, we compare our M15 and M60 models with two models in literature (Aliberti et al., 2018) based on Nonlinear Autoregressive neural network (NAR).

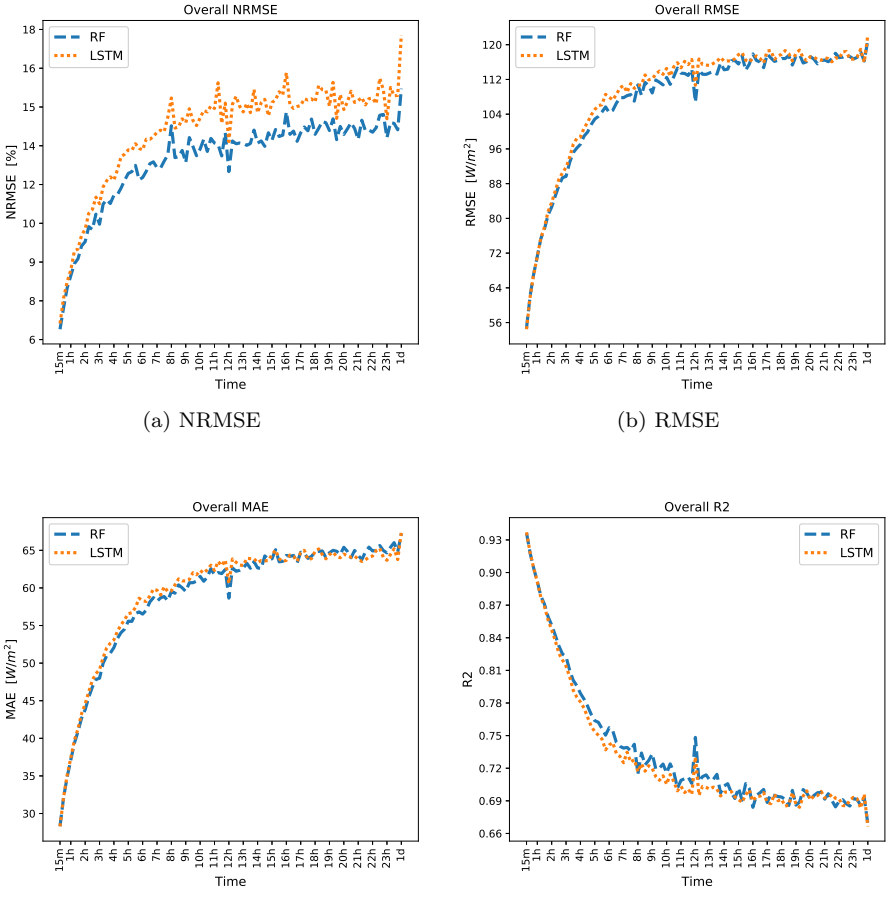
6.1. Performance assessments of the M15 model

Table 6 reports the performance indicators obtained by forecasting GHI with the M15 model for different time-horizons. As described in Section 5, M15 performs its analysis with 15min time-resolution. In general, we can notice that M15 better performs during the winter because there are small variations in GHI trends, making easier the learning phase of our model and consequently increasing the accuracy during tests. Indeed, varying the time-horizon from next 15min up to next 24h, R^2 decreases from 0.96 to about 0.63. On the other hand, summer provides worse results because of high variations in GHI trends among sunny, cloudy and rainy days. Thus, R^2 varies from 0.90 for the next 15min to about 0.66 for next 24h. Autumn and spring have similar GHI trends, thus their performances are quite comparable. Finally, considering the whole year, R^2 decreases from about 0.94 to 0.67; RMSE varies from 54.464 W/m^2 to 121.066 W/m^2; MAE increases from 28.286 W/m^2 to 67.208 W/m^2; NRMSE rises from 6.401% to 15.711%.

As reported in Table 6 and shown in Figure 6, there are no remarkable differences in exploiting RF or LSTM in terms of RMSE, MAE and R^2. Indeed, M15 provides similar results. However, if we consider NRMSE, we can notice that RF provides slightly better performance than LSTM, especially for long-term forecast, e.g. 15.711% and 17.779%, respectively, at next 24h. In general,

Figure 6 shows that the gap between RF and LSTM becomes almost constant for predictions with time-horizons between 6h and 24h (e.g., ~ 0.1 for R^2 and $\sim 2\%$ for NRMSE).

Finally, Figure 7 and Figure 8 show the trends for one week, taken as example, for each meteorological season at next 15min and next 24h, respectively. Real GHI samples are represented by the blue dashed-line, while forecasts with RF and LSTM are depicted by orange and green dashed-lines, respectively.



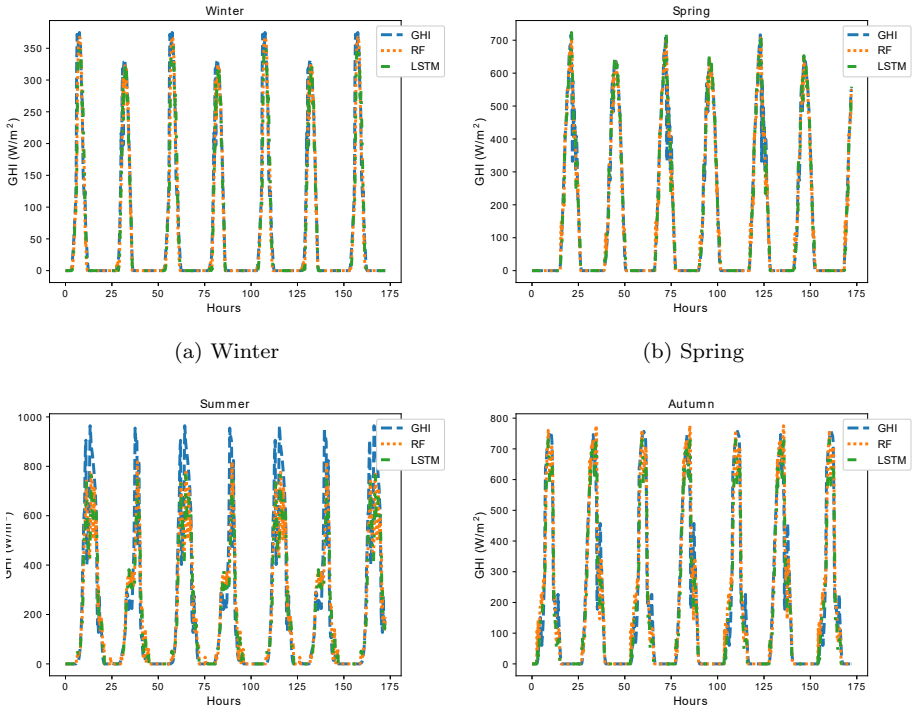
(a) NRMSE (b) RMSE

(c) MAE (d) R^2

Figure 6: Overall performance trends for M15

Table 6: Performance indicator of M15 (i.e. 15min time-resolution) at different time-horizons.

Time resolution	Time-horizon	Season	RMSE $[W/m^2]$		MAE $[W/m^2]$		NRMSE [%]		R^2	
			RF	LSTM	RF	LSTM	RF	LSTM	RF	LSTM
15m	15m	winter	24.314	24.040	10.527	10.477	3.978	4.076	0.960	0.960
		spring	64.190	62.142	31.402	30.240	6.680	6.709	0.942	0.944
		summer	89.493	90.591	52.878	53.922	9.647	10.319	0.900	0.898
		autumn	41.076	41.084	18.433	18.505	5.297	5.385	0.944	0.944
		year	54.768	54.464	28.310	28.286	6.401	6.622	0.936	0.936
	30m	winter	28.313	27.976	12.213	12.142	4.632	4.771	0.945	0.946
		spring	74.273	72.019	36.545	35.341	7.729	7.878	0.922	0.927
		summer	98.627	99.986	58.405	59.649	10.632	11.540	0.879	0.876
		autumn	47.341	47.717	21.432	21.643	5.953	6.255	0.926	0.925
		year	62.138	61.924	32.149	32.194	7.237	7.611	0.918	0.918
	45m	winter	31.400	30.973	13.637	13.589	5.137	5.282	0.932	0.934
		spring	81.859	78.784	40.250	39.085	8.476	8.457	0.905	0.912
		summer	104.691	107.164	62.078	63.915	11.823	12.233	0.864	0.857
		autumn	50.975	51.592	23.583	23.868	6.574	6.987	0.914	0.912
		year	67.231	67.128	34.887	35.114	8.002	8.240	0.904	0.904
	1h	winter	33.140	32.897	14.369	14.353	5.422	5.610	0.925	0.926
		spring	87.294	84.075	43.275	42.125	9.049	8.888	0.892	0.900
		summer	109.150	111.714	65.105	66.984	12.449	12.894	0.852	0.845
		autumn	55.479	56.354	25.702	26.054	6.977	7.387	0.898	0.895
		year	71.266	71.260	37.113	37.379	8.474	8.695	0.892	0.891
	3h	winter	48.721	49.419	21.013	21.262	7.971	8.427	0.837	0.832
		spring	106.852	103.522	55.520	54.383	10.782	10.436	0.839	0.848
		summer	131.294	139.784	80.215	84.867	14.978	16.134	0.785	0.756
		autumn	71.790	74.183	35.214	36.013	8.122	10.024	0.829	0.818
		year	89.664	91.727	47.991	49.131	10.463	11.255	0.823	0.814
	12h	winter	60.163	60.834	25.939	26.496	9.843	10.374	0.752	0.746
		spring	132.763	129.236	70.254	68.955	12.909	12.789	0.751	0.764
		summer	153.428	167.490	97.040	103.618	17.503	19.332	0.707	0.651
		autumn	80.855	85.127	41.317	43.393	9.745	12.149	0.784	0.760
		year	106.802	110.672	58.638	60.616	12.500	13.661	0.748	0.730
	24h	winter	73.906	72.896	32.699	32.908	13.226	14.633	0.625	0.635
		spring	155.058	144.480	83.022	78.514	14.719	14.297	0.660	0.705
		summer	164.377	183.571	104.970	113.210	23.940	29.265	0.664	0.581
		autumn	90.922	87.841	48.140	44.937	10.959	12.920	0.726	0.745
		year	121.066	122.197	67.208	67.392	15.711	17.779	0.669	0.666



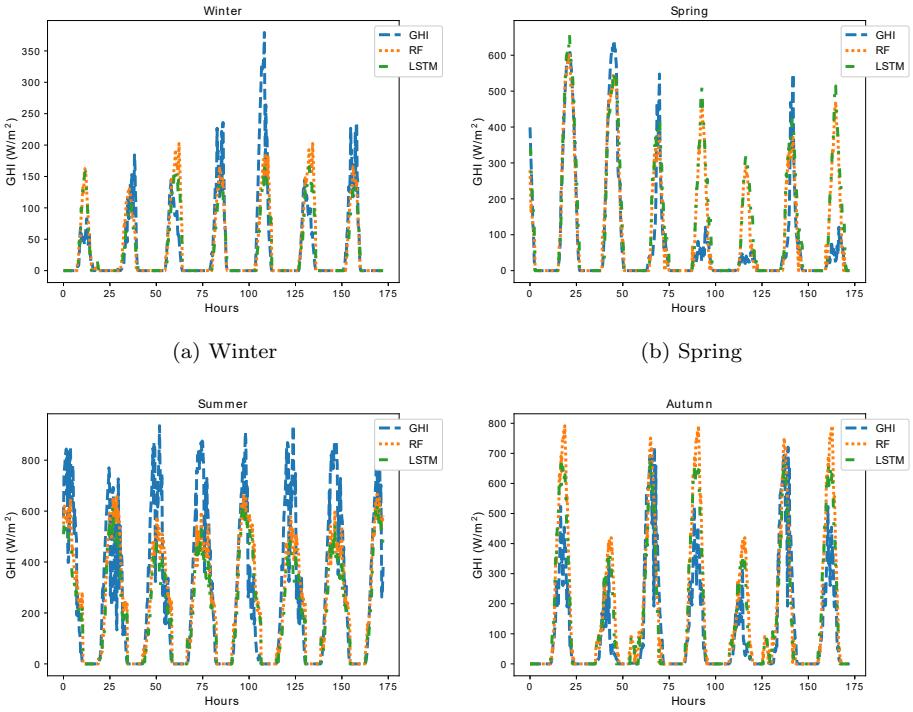
(a) Winter

(b) Spring

(c) Summer

(d) Autumn

Figure 7: GHI Forecast trends for M15 at next 15min



(a) Winter

(b) Spring

(c) Summer

(d) Autumn

Figure 8: GHI Forecast trends for M15 at next 24h

6.2. Performance assessments of the M60 model

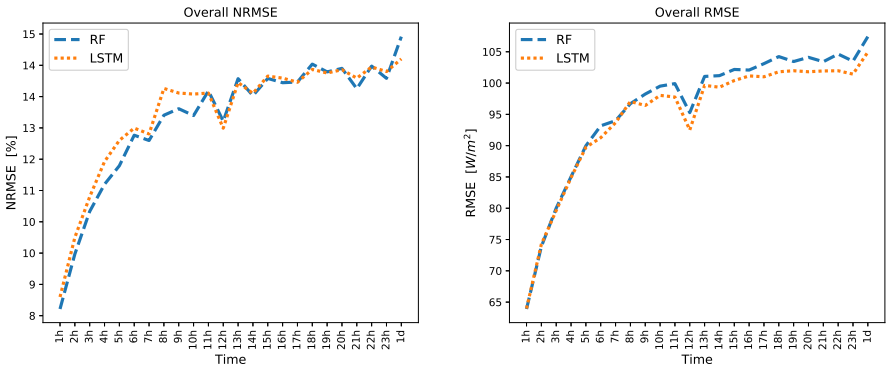
Table 7 reports the performance indicators given by the M60 for GHI predictions at different time-horizons, from next 1h up to next 24h. Differently than M15, M60 performs its forecasts exploiting a 60min time-resolution (see Section 5). As for M15, M60 provides best results in forecasting GHI in winter, while worse performances are given during summer for the very same reasons discussed in Section 6.1. Indeed, R^2 decreases from 0.93 for next 1h to about 0.73 for next 24h in winter, while it drops from approx. 0.90 to about 0.79 in summer. Also in this case, forecasts obtained in autumn and spring are quite comparable because of their similarities in GHI trends. Finally, considering the whole year, R^2 falls from about 0.91 to 0.74; RMSE rises from 63.899 W/m^2 to 104.962 W/m^2; MAE varies from 35.126 W/m^2 to 58.355 W/m^2; NRMSE increases from 8.171% to 14.565%.

Differently than M15, M60 better performs with LSTM than RF for long-term forecasts. This is highlighted by the R^2 in Figure 9 for time-horizons over the 10h. Whilst, from 1h to 10h, RF slightly outperforms LSTM in the NRMSE; indeed, the difference is $\sim 0.3\%$, which is negligible. More in general, results point out that, over the 10h, the error rate is almost constant (e.g., ~ 0.5 for R^2 and $\sim 2\%$ for NRMSE).

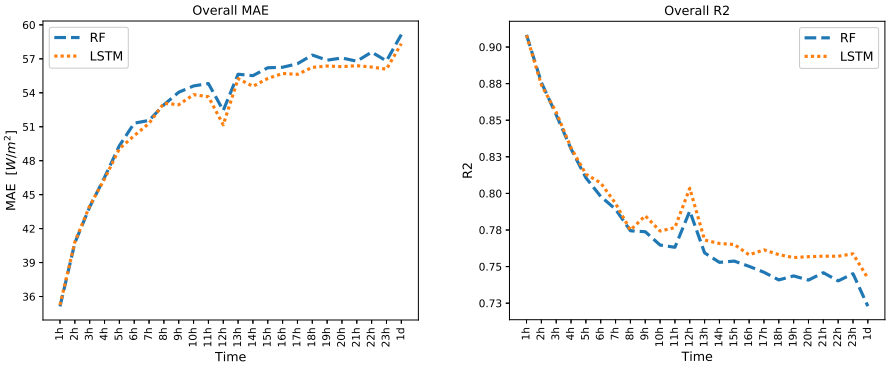
Finally, Figure 10 and Figure 11 report the trends for one week for each meteorological season at 1h and 24h time-horizon, respectively. The blue dashed-line represents the real GHI samples, orange and green dashed-lines depict forecasts with RF and LSTM, respectively.

Table 7: Performance indicator of M60 (i.e. 60min time-resolution) at different time-horizons

Time resolution	Time-horizon	Season	RMSE $[W/m^2]$		MAE $[W/m^2]$		NRMSE [%]		R^2	
			RF	LSTM	RF	LSTM	RF	LSTM	RF	LSTM
1h	1h	winter	31.739	31.420	14.862	14.642	5.502	5.726	0.929	0.931
		spring	82.591	80.132	47.378	46.199	9.282	9.009	0.901	0.906
		summer	88.222	90.350	52.735	54.066	10.501	11.140	0.899	0.894
		autumn	53.046	54.011	25.528	26.108	7.399	8.037	0.904	0.901
		year	63.899	63.978	35.126	35.254	8.171	8.478	0.908	0.908
	3h	winter	43.389	42.278	19.800	19.056	7.521	7.363	0.868	0.874
		spring	106.027	100.538	60.989	59.263	13.112	12.109	0.836	0.853
		summer	105.532	109.510	63.380	65.662	13.079	14.622	0.855	0.844
		autumn	65.108	66.204	31.476	32.065	8.947	10.059	0.855	0.850
		year	80.014	79.633	43.911	44.012	10.665	11.038	0.854	0.855
	12h	winter	57.074	51.697	25.573	23.054	9.893	9.449	0.771	0.812
		spring	132.854	121.156	77.804	71.629	17.170	14.562	0.743	0.787
		summer	118.344	122.719	71.020	73.697	14.667	16.300	0.818	0.805
		autumn	72.811	74.457	35.352	36.288	10.156	10.856	0.819	0.811
		year	95.271	92.507	52.437	51.167	12.972	12.792	0.788	0.804
	24h	winter	68.838	62.266	31.694	28.379	12.354	11.432	0.667	0.728
		spring	153.807	139.886	89.516	83.129	20.750	16.813	0.656	0.716
		summer	125.567	132.757	74.702	79.477	15.640	17.633	0.795	0.771
		autumn	81.556	84.939	40.741	42.436	11.800	12.384	0.773	0.754
		year	107.442	104.962	59.163	58.355	15.136	14.565	0.723	0.742

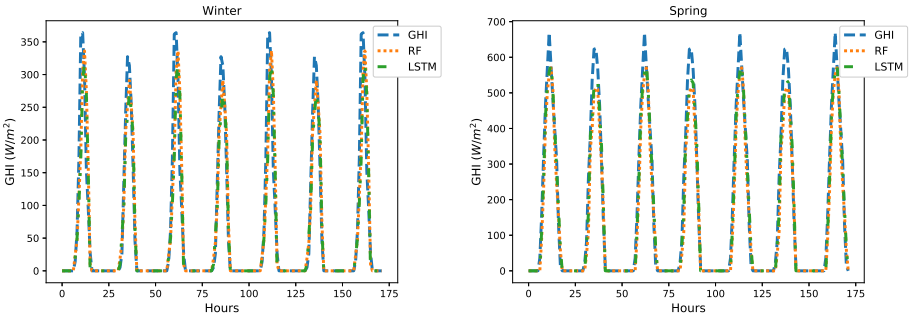


(a) NRMSE (b) RMSE



(c) MAE (d) R^2

Figure 9: Overall performance trends for M60



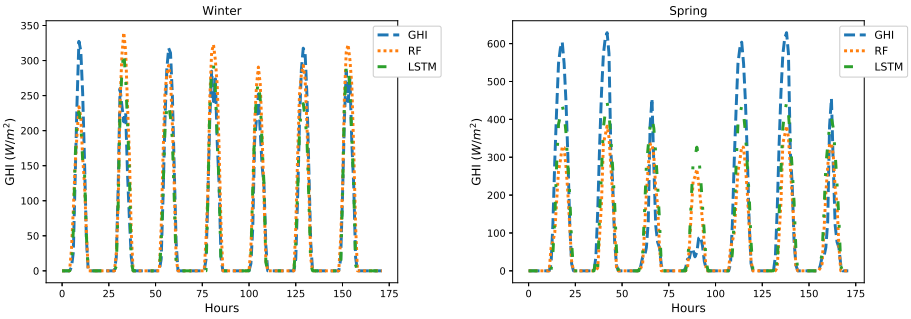
(a) Winter

(b) Spring

(c) Summer

(d) Autumn

Figure 10: GHI Forecast trends for M60 at next 1h



(a) Winter

(b) Spring

(c) Summer

(d) Autumn

Figure 11: GHI Forecast trends for M60 at next 24h

6.3. Comparison with NAR-based model in literature

In this section, we compare the performance of our models with the literature solution proposed in (Aliberti et al., 2018), where authors presented a Nonlinear Autoregressive neural network (NAR) for short-term GHI forecasts. To perform a fair comparison, we implemented and trained two NAR models, NAR-15 and NAR-60, following the methodology provided in (Aliberti et al., 2018) that have been trained and tested with our dataset with 15min and 60min time-resolution, respectively. Hence, we compare our M15 with NAR-15 and our M60 with NAR-60. As reported in (Aliberti et al., 2018), both NAR-15 and NAR-60 only need GHI trends in input, neglecting the other meteorological elements.

Table 8 summarises this comparison, reporting the results of the performance indicators in forecasting GHI for different time-horizons. Considering 15min time-resolution, we can notice that our M15, in both RF- and LSTM-based implementations, improves the forecast in terms of error rate and time-horizon. Indeed, NAR-15 provides good results in the very short-term, i.e. up to next 45min. In forecasts after the next 1h, our M15 outperforms NAR-15 in all the indicators. As can be noted, the NRMSE trend for NAR-15 reports some fluctuations due to wrong forecasts that can be considered as outliers. These errors are present in particular when the prediction is closed to the max value of the real GHI. In detail, R^2 for NAR-15 decreases from about 0.96 to 0.72 (M15-RF: from 0.94 to 0.85, M15-LSTM: from 0.94 to 0.85); RMSE varies from 47.70 W/m^2 to 123.64 W/m^2 (M15-RF: from 54.77 W/m^2 to 82.62 W/m^2, M15-LSTM: from 54.46 W/m^2 to 83.66 W/m^2); MAE increases from 17.19 W/m^2 to 77.45 W/m^2 (M15-RF: from 28.31 W/m^2 to 43.81 W/m^2, M15-LSTM: from 28.29 W/m^2 to 44.57 W/m^2); NRMSE rises from 5.24% to 6.43% (M15-RF: from 6.40% to 9.79%, M15-LSTM: from 6.62% to 10.27%).

Regarding 60min time-resolution (see Table 8), similar considerations can be done. We can notice that our M60 outperforms NAR-60 from the next 2h predictions (i.e. after two prediction steps). Indeed, NAR-60 performances decrease considerably in the long-term, making it suitable only for short-term forecasts. Also in this case, the NRMSE trend for NAR-60 is affected by fluctu-

Table 8: Comparison of our methodology with (Aliberti et al., 2018) at different time-horizons for both 15min and 60min time-resolutions

Time resolution	Time-horizon	RMSE [W/m^2]			MAE [W/m^2]			NRMSE [%]			R^2		
		RF-based	LSTM-based	NAR	RF-based	LSTM-based	NAR	RF-based	LSTM-based	NAR	RF-based	LSTM-based	NAR
15min	15m	54.768	54.464	47.700	28.310	28.286	17.185	6.401	6.622	5.239	0.936	0.937	0.958
	30m	62.138	61.924	64.830	32.149	32.194	26.493	7.237	7.611	7.442	0.918	0.918	0.922
	45m	67.231	67.128	75.284	34.887	35.114	33.661	8.002	8.240	5.145	0.904	0.904	0.895
	1h	71.266	71.260	84.777	37.113	37.379	40.800	8.474	8.695	4.386	0.892	0.891	0.866
	2h	82.620	83.661	123.636	43.808	44.572	77.450	9.789	10.272	6.433	0.852	0.847	0.716
60min	1h	63.899	63.978	61.305	35.126	35.254	28.094	8.171	8.478	6.261	0.908	0.908	0.928
	2h	73.880	74.249	85.131	40.739	40.823	49.093	9.576	10.000	10.283	0.876	0.874	0.861
	3h	80.014	79.633	114.732	43.911	44.012	81.799	10.665	11.038	14.427	0.854	0.855	0.748
	6h	93.170	91.267	230.197	51.314	50.209	171.747	12.613	12.795	19.419	0.798	0.807	−0.016
	12h	95.271	92.507	405.587	52.437	51.167	255.536	12.972	12.792	15.372	0.788	0.804	−2.153
	24h	107.415	104.942	432.790	59.141	58.340	298.978	15.132	14.563	12.909	0.723	0.742	−2.590

ations due to some outliers in predictions close to the max value of the real GHI. After 6h, NAR-60 performances are completely degraded. This is highlighted by R^2, which values are lower than 0. In detail, R^2 for NAR-60 decreases from about 0.93 to −2.59 (M60-RF: from 0.91 to 0.72, M60-LSTM: from 0.91 to 0.74); RMSE varies from 61.31 W/m^2 to 432.790 W/m^2 (M60-RF: from 63.90 W/m^2 to 107.42 W/m^2, M60-LSTM: from 63.98 W/m^2 to 104.94 W/m^2); MAE increases from 28.09 W/m^2 to 298.98 W/m^2 (M60-RF: from 35.13 W/m^2 to 59.14 W/m^2, M60-LSTM: from 35.25 W/m^2 to 58.34 W/m^2); NRMSE rises from 6.26% to 12.91% (M60-RF: from 8.17% to 15.13%, M60-LSTM: from 8.48% to 14.56%).

7. Conclusion

In this paper, we presented a methodology to forecast the Global Horizontal Solar Irradiance in short- and long-term, combining VMD, CNNs and two regression models (used alternatively). Following our methodology, we developed two models, M15 and M60, to test different time-resolutions (15min and 60min, respectively). The analysis of the experimental results highlighted that the performances of both models are comparable. In particular, M15 is suitable for sub-hourly forecasts in the short-term, while M60 for hourly previsions in the long-term.

GHI is the renewable source needed by PV systems to produce energy. To move forward more sustainable cities, next-generation energy distribution systems will include different RES distributed across the city. Consequently, current energy distribution systems must be enhanced with novel and smart control systems by taking advantages of rising ICT technologies, thus becoming smart grids. In this view, DR and DSM are examples of possible applications to match the energy produced by RES with the total energy requested by the city (Deng et al., 2015; Lin et al., 2017; Siano, 2014). Usually, DR strategies need sub-hourly information to work in the short-term (i.e. from next 15min), while DSM needs hourly information up to next 24h to perform load balancing. In this context, M15 and M60 alternatively combined with tools to estimate PV production, like PVsim (Bottaccioli et al., 2017), can unlock design and development of more accurate policies for DR and DSM, respectively.

Abbreviations

The following abbreviations are used in this manuscript:

Artificial Neural Network	ANN
Autoregressive Integrated Moving Average	ARIMA
Coefficient of determination	R^2

Convolutional Neural Network	CNN
Demand Side Management	DSM
Diffuse Horizontal Irradiance	DHI
Ensemble Empirical Mode Decomposition	EEMD
Global Horizontal Solar Irradiance	GHI
Intrinsic Mode Function	IMF
Long Short-Term Memory	LSTM
Mean Absolute Error	MAE
Neuro-Fuzzy Takagi-Sugeno	NF-TS
Normalised Root Mean Square Error	NRMSE
Photovoltaic	PV
Random Forest	RF
Renewable Energy Sources	RES
Rectified Linear Unit	ReLU
Root Mean Square Error	RMSE
Support Vector Machine	SVM
Variational Mode Decomposition	VMD
Wavelet Decomposition	WD

References

Alanazi, M., Mahoor, M., & Khodaei, A. (2017). Two-stage hybrid day-ahead solar forecasting. *2017 North American Power Symposium, NAPS 2017*, (pp. 1–6).

Aliberti, A., Bottaccioli, L., Cirrincione, G., Macii, E., Acquaviva, A., & Patti, E. (2018). Forecasting Short-term Solar Radiation for Photovoltaic Energy Predictions. In *Proceedings of the 7th International Conference on Smart Cities and Green ICT Systems - Volume 1: SMARTGREENS,* (pp. 44–53). INSTICC SciTePress. doi:10.5220/0006683600440053.

Badescu, V. (2008). *Modeling solar radiation at the earth's surface: Recent advances*. Springer. doi:10.1007/978-3-540-77455-6.

Benali, L., Notton, G., Fouilloy, A., Voyant, C., & Dizene, R. (2019). Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components. *Renewable Energy*, *132*, 871–884. doi:10.1016/j.renene.2018.08.044.

Bertsekas, D. P. (2014). *Constrained optimization and Lagrange multiplier methods*. Academic press.

Bottaccioli, L., Estebsari, A., Patti, E., Pons, E., & Acquaviva, A. (2017). A Novel Integrated Real-time Simulation Platform for Assessing Photovoltaic Penetration Impacts in Smart Grids. *Energy Procedia*, *111*, 780–789. doi:10.1016/j.egypro.2017.03.240.

Cao, J., & Lin, X. (2008). Application of the diagonal recurrent wavelet neural network to solar irradiation forecast assisted with fuzzy technique. *Engineering Applications of Artificial Intelligence*, *21*, 1255–1263.

Demirhan, H., & Renwick, Z. (2018). Missing value imputation for short to mid-term horizontal solar irradiance data. *Applied Energy*, *225*, 998 – 1012.

Deng, R., Yang, Z., Chow, M.-Y., & Chen, J. (2015). A survey on demand response in smart grids: Mathematical models and approaches. *IEEE Transactions on Industrial Informatics*, *11*, 570–582.

Dragomiretskiy, K., & Zosso, D. (2014). Variational mode decomposition. *IEEE Transactions on Signal Processing*, *62*, 531–544. doi:10.1109/TSP.2013.2288675.

Galván, I. M., Valls, J. M., Cervantes, A., & Aler, R. (2017). Multi-objective evolutionary optimization of prediction intervals for solar energy forecasting with neural networks. *Information Sciences*, *418–419*, 363–382.

García-Hinde, O., Terrén-Serrano, G., Hombrados-Herrera, M., Gómez-Verdejo, V., Jiménez-Fernández, S., Casanova-Mateo, C., Sanz-Justo, J., Martínez-Ramón, M., & Salcedo-Sanz, S. (2018). Evaluation of dimensionality reduction methods applied to numerical weather models for solar radiation forecasting. *Engineering Applications of Artificial Intelligence*, *69*, 157–167.

Ho, T. K. (1995). Random decision forests. *Proceedings of the International Conference on Document Analysis and Recognition*, *1*, 278–282.

Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *6*, 107–116.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*, 1735–1780.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, *195*, 215–43. doi:papers://47831562-1F78-4B52-B52E-78BF7F97A700/Paper/p352.

Jiang, H., & Dong, Y. (2016). A nonlinear support vector machine model with hard penalty function based on glowworm swarm optimization for forecasting daily global solar radiation. *Energy Conversion and Management*, *126*, 991–1002.

Kamadinata, J. O., Ken, T. L., & Suwa, T. (2019). Sky image-based solar irradiance prediction methodologies using artificial neural networks. *Renewable Energy*, *134*, 837–845. doi:10.1016/j.renene.2018.11.056.

Kaplanis, S., & Kaplani, E. (2010). Stochastic prediction of hourly global solar radiation for Patra, Greece. *Applied Energy*, *87*, 3748–3758. doi:10.1016/j.apenergy.2010.06.006.

Khashei, M., Bijari, M., & Raissi Ardali, G. A. (2009). Improvement of autoregressive integrated moving average models using fuzzy logic and artificial neural networks (anns). *Neurocomputing*, *72*, 956–967.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*, 84–90. URL: http://dl.acm.org/citation.cfm?doid=3098997.3065386. doi:10.1145/3065386.

Lahmiri, S. (2017). Comparing variational and empirical mode decomposition in forecasting day-ahead energy prices. *IEEE Systems Journal*, *11*, 1907–1910.

Layanun, V., Suksamosorn, S., & Songsiri, J. (2017). Missing-data imputation for solar irradiance forecasting in thailand. *2017 56th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, (pp. 1234–1239).

LeCun, Y. A., Bottou, L., Orr, G. B., & Müller, K.-R. (2012). Efficient backprop. In *Neural networks: Tricks of the trade* (pp. 9–48). Springer.

Li, J., Ward, J., Tong, J., Collins, L., & Platt, G. (2016a). Machine learning for solar irradiance forecasting of photovoltaic system. *Renewable Energy*, *90*, 542–553.

Li, M., Chu, Y., Pedro, H. T., & Coimbra, C. F. (2016b). Quantitative evaluation of the impact of cloud transmittance and cloud velocity on the accuracy of short-term dni forecasts. *Renewable energy*, *86*, 1362–1371.

Lin, C.-C., Deng, D.-J., Liu, W.-Y., & Chen, L. (2017). Peak load shifting in the internet of energy with energy trading among end-users. *IEEE Access*, *5*, 1967–1976.

Madanchi, A., Absalan, M., Lohmann, G., Anvari, M., & Reza Rahimi Tabar, M. (2017). Strong short-term non-linearity of solar irradiance fluctuations. *Solar Energy*, *144*, 1–9. doi:10.1016/j.solener.2017.01.008.

Majumder, I., Dash, P. K., & Bisoi, R. (2018). Variational mode decomposition based low rank robust kernel extreme learning machine for solar irradiation forecasting. *Energy Conversion and Management*, *171*, 787–806. doi:10.1016/j.enconman.2018.06.021.

Monjoly, S., André, M., Calif, R., & Soubdhan, T. (2017). Hourly forecasting of global solar radiation based on multiscale decomposition methods: A hybrid approach. *Energy*, *119*, 288–298.

Nazaripouya, H., Wang, B., Wang, Y., Chu, P., Pota, H. R., & Gadh, R. (2016). Univariate time series prediction of solar power using a hybrid wavelet-ARMA-NARX prediction method. In *Proceedings of the IEEE Power Engineering Society Transmission and Distribution Conference* (pp. 1–5). IEEE volume 2016-July. doi:10.1109/TDC.2016.7519959.

Nocedal, S. J., J; Wright (2006). *Numerical Optimization*. (2nd ed.). Berlin, Germany: Springer.

Nottrott, A., & Kleissl, J. (2010). Validation of the nsrdb–suny global horizontal irradiance in california. *Solar Energy*, *84*, 1816 – 1827. doi:https://doi.org/10.1016/j.solener.2010.07.006.

Qing, X., & Niu, Y. (2018). Hourly day-ahead solar irradiance prediction using weather forecasts by lstm. *Energy*, *148*, 461–468.

REN21 (2019). *Renewables 2019 Global Status Report*. Technical Report Paris, REN21 Secretariat.

Rodrigues Júnior, S. E., & Serra, G. L. d. O. (2017). A novel intelligent approach for state space evolving forecasting of seasonal time series. *Engineering Applications of Artificial Intelligence*, *64*, 272–285.

Siano, P. (2014). Demand response and smart grids - a survey. *Renewable and Sustainable Energy Reviews*, *30*, 461–478.

Srivastava, S., & Lessmann, S. (2018). A comparative study of lstm neural networks in forecasting day-ahead global horizontal irradiance with satellite data. *Solar Energy*, *162*, 232–247.

Sun, H., Gui, D., Yan, B., Liu, Y., Liao, W., Zhu, Y., Lu, C., & Zhao, N. (2016). Assessing the potential of random forest method for estimating solar

radiation using air pollution index. *Energy Conversion and Management*, *119*, 121–129.

Sun, S., Wang, S., Zhang, G., & Zheng, J. (2018). A decomposition-clustering-ensemble learning approach for solar radiation forecasting. *Solar Energy*, *163*, 189–199. doi:10.1016/j.solener.2018.02.006.

Voyant, C., Darras, C., Muselli, M., Paoli, C., Nivet, M. L., & Poggi, P. (2014). Bayesian rules and stochastic models for high accuracy prediction of solar radiation. *Applied Energy*, *114*, 218–226. doi:10.1016/j.apenergy.2013.09.051.

Voyant, C., Notton, G., Kalogirou, S., Nivet, M.-L., Paoli, C., Motte, F., & Fouilloy, A. (2017). Machine learning methods for solar radiation forecasting: A review. *Renewable Energy*, *105*, 569–582.

Wan, C., Zhao, J., Song, Y., Xu, Z., Lin, J., & Hu, Z. (2015). Photovoltaic and solar power forecasting for smart grid energy management. *CSEE Journal of Power and Energy Systems*, *1*, 38–46.

Wang, H., Yi, H., Peng, J., Wang, G., Liu, Y., Jiang, H., & Liu, W. (2017a). Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network. *Energy Conversion and Management*, *153*, 409–422. doi:10.1016/j.enconman.2017.10.008.

Wang, H. z., Li, G. q., Wang, G. b., Peng, J. c., Jiang, H., & Liu, Y. t. (2017b). Deep learning based ensemble approach for probabilistic wind power forecasting. *Applied Energy*, *188*, 56–70. doi:10.1016/j.apenergy.2016.11.111.

Zang, H., Cheng, L., Ding, T., Cheung, K. W., Liang, Z., Wei, Z., & Sun, G. (2018). Hybrid method for short-term photovoltaic power forecasting based on deep convolutional neural network. *IET Generation, Transmission & Distribution*, *12*, 4557–4567. doi:10.1049/iet-gtd.2018.5847.

Zhang, G., Eddy Patuwo, B., & Y. Hu, M. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, *14*, 35–62.