Forecasting of PV plant output using hybrid wavelet-based LSTM-DNN structure model

Juan Ospina¹,², Alvi Newaz¹,², M. Omar Faruque¹,²
¹Department of Electrical and Computer Engineering, FAMU-FSU College of Engineering, 2525 Pottsdamer St, Tallahassee, FL 32310, USA
²Center for Advanced Power Systems (CAPS), 2000 Levy Ave, Tallahassee, FL 32310, USA
E-mail: jospina@fsu.edu

Abstract: This paper proposes a novel forecasting model designed to accurately forecast the PV power output for both large-scale and small-scale PV systems. The proposed model uses available temperature data, approximate and detailed coefficients obtained from the decomposed PV power time series using the stationary wavelet transform (SWT), and statistical features extracted from the historical PV data. The model is comprised of four long–short–term memory (LSTM) recurrent neural networks (RNN) designed to perform multi-step forecasting on the individual approximate and detailed coefficients decomposed by the SWT and a final deep neural network (DNN) designed to perform the next time step PV power forecast. The DNN makes use of the reconstructed values estimated by the four LSTM networks together with temperature data and statistical features to predict the final forecasted value of the next time step PV power. 30-min resolution data from a 12.6 MW PV system located in the state of Florida are used for testing and evaluating the proposed method against several models found in the literature. The results obtained suggest that the proposed model improved the forecasting accuracy significantly in the metrics used to compare with other models while reducing the number of features needed to perform the forecasting operation.

1 Introduction

As the cost of green energy sources decreases and other environmental issues arise due to the use of fossil fuels, new ways of producing renewable energy are being encouraged by governments and organisations in charge of promoting responsible use of our natural resources [1]. One of the most proliferating means of generating clean and renewable energy nowadays is the use of solar power. In the USA, the Sunshot Vision Study has reported that by 2030 almost 14% of the energy consumed in the USA will be generated from solar power only and by 2050 these numbers will nearly double to 27% [2]. Another example of this support is the Executive Order S21-09 in California which is calling for a 33% of its entire energy generation to come from renewable energy resources by the year 2020. Nonetheless, one major challenge for the integration of PV power in the electrical system and the overall energy market is its stochastic nature. The power generated from large-scale PV plants has the drawback of being very intermittent, affecting directly the ability to bid in the energy market and the consideration of PV power as a reliable energy-generation resource. For this reason, accurate forecasting of the available PV power in medium- to short-term scales could be a game changer in the integration of PV systems, by providing a more reliable estimation of the power being generated.

Typically, solar power enters into the picture as a non-dispatchable component, which cannot be accounted as a reliable resource for entering the energy market and any considerable forecast inaccuracy in its generation could cause substantial economic losses and power reliability issues in the whole electrical power system. For that reason, accurate models for solar power forecasting have the potential of being a paradigm shift in the overall energy market through medium- and short-term forecasting by allowing optimal control systems to use its information for unit commitment, economic dispatch applications, energy transactions, unit maintenance, and energy management decision-making [3].

This paper presents a novel hybrid wavelet-based LSTM-DNN structure model designed to accurately forecast the PV power generated by a PV plant in the next 30-min. The proposed model only makes use of available historical PV power data and estimated temperature to accurately predict the next time step PV power generation of the system. It is important to note that the model can be categorised as an adaptive model since its forecasting horizon can be modified and adapted to perform multi-step forecast (e.g. 30 min, 1 h, … n hours) and is only dependent on the lowest time resolution of the data available. In other words, the proposed model is not locked-in to a particular forecasting horizon, since it can be adapted to any time resolution as long as it is equal or greater than the time resolution of the available historical data.

This paper is organised as follows: Section 2 gives an overview of the current solar and PV power forecasting techniques. Section 3 reviews the stationary wavelet transform (SWT) and describes how it is used in the proposed forecasting model. Section 4 provides a description of the proposed model and describes the methodology used for designing the forecasting model. Section 5 shows the training and validation of the results together with a description of the metrics used for evaluating the performance of the model, and Section 6 presents the conclusions.

2 Overview of solar forecasting techniques

This section focuses on reviewing some of the novel solar irradiance and PV power forecasting models which have been proposed in the last few years. Some relevant models related to power consumption and load forecasting applications are also mentioned for completeness.

Authors in [4] present a solar energy forecasting model designed to forecast the solar energy generation in J/m² of solar farms. The model consists of forecasting energy values based on numerical weather predictions and convolutional neural networks (CNN). The output of the CNNs is then processed using Gaussian process regression in order to obtain the proper output based on the values given by solar farms. Similarly, in [5], authors present an adaptive framework based on a combination of data analytics approaches and machine-learning techniques to perform a solar energy forecast for a day-ahead prediction horizon. The model presented here consists on a nine-step process designed to extract relevant data, detect synoptic events, perform correlation analysis, process these features using an adaptive artificial neural network (ANN) and Monte–Carlo uncertainty analysis, and finally aggregate and fine-tune the data in order to present it to the user. In
researchers present an analogue ensemble approach designed to forecast regional PV power with an hourly resolution by utilising data such as weather forecasts, power measurements, and astronomical data. The primary strategy implemented here consists of a clustering and blending strategy applied to improve the accuracy of the solar PV forecast. Authors claim a significant reduction in the normalised mean square error (NRMSE), of around 13.8–61.21%, when compared to three baselines: North American Mesoscale Forecast System, the Global Forecast System, and the Short-Range Ensemble Forecast.

Researchers in [7] propose a novel approach of estimating the PV maximum generation by relying on the measurements of the DC voltage, current, and cell temperature of a specific PV array. The PV plant power is forecasted by indirectly estimating the irradiance of the system using these three estimators. According to the authors, this approach can accurately reconstruct the PV power generation of the PV array even during curtailment periods and can outperform other pyranometer-based methods that use simple sensing systems. In [8], authors present four different PV power forecasting models: a radial basis function neural network (RBFNN), k-nearest neighbours (kNN), weighted kNN (WkNN), and least square support vector machine (LS-SVM). PV power forecast is performed in different sites including Braedstrup (Denmark), and Trondheim (Norway), and a similar approach is taken in [10], where researchers present a type-1 and interval type-2 Takagi-Sugeno-Kang (TSK) fuzzy systems for modelling and predicting the PV power output of a PV plant.

In [9], authors propose a novel Takagi-Sugeno (T-S) fuzzy model-based PV power short-term forecasting. Here, researchers use a fuzzy C-means algorithm and the recursive least squares method to identify the antecedent and consequent parameters of the system. The model is tested using data from a 433 kW PV plant located at the St. Lucia campus of the Queensland University in Australia. A similar approach is taken in [10], where researchers propose a type-1 and interval type-2 Takagi-Sugeno-Kang (TSK) fuzzy systems for modelling and predicting the PV power output of a PV plant.

In [11], Sheng et al. present a short-term solar power forecasting model based on weathered Gaussian process regression. The main contribution of this paper is the introduction of an innovative method that employs the weighted Gaussian process regression approach to reduce the weight that high potential outliers can have in the data to be analysed. Yang et al. [12] present a weather-based hybrid forecasting algorithm designed to predict the day-ahead PV power output using three stages: a classification, a regression, and a forecasting stage. According to the authors, this procedure improves results obtained from classic SVM and ANN models. In [13], Asrari et al. present a hybrid gradient-descent and meta-heuristic approach designed to improve the accuracy and computational burden of an ANN model by selecting the optimal set of parameters for PV power prediction.

LSTM RNNs have also been proposed in the literature for forecasting the power generation of a PV plant. In [14], authors present an LSTM model designed to predict the PV power output of a plant located in Hae-Nam, Korea. Similar approaches are used in [15, 16], where authors in [15] make use of multi-layer perceptrons (MLP), deep belief networks, autoencoders, and LSTMs to forecast the PV power generated from 21 different facilities, while authors in [16] present a deep LSTM models designed to capture the temporal changes in the PV power data. The use of LSTMs was also explored by Qing et al. in [17], where a training performed a day-ahead solar irradiance forecast based on prediction structure that forecasts multiple outputs simultaneously with trained LSTM networks. The primary goal of the proposed model is to predict the next day irradiance values based only on features such as month, day of the month, hour of the day, temperature, dew point, humidity, visibility, wind speed, and weather type.

It is important to mention that LSTMs have also been widely used in other forecasting applications, such as power consumption and short-term load forecasting, and present a similar structure as the ones used in solar power applications. For example in [18, 19], Kong et al. present two different load forecasting models based on LSTM networks that are designed to forecast aggregated residential and meter-level power consumption based on appliance learning. In [19], authors explore the difficulty of conducting single-meter load forecasting and demonstrate how the aggregation of individual load forecasting data reduces computational costs. Specifically, this paper presents the following research contributions:
1. A novel PV power forecasting model, designed to perform short-term forecasts (30 min), that demonstrates several advantages when compared to other comparable models including: (a) fewer measured features required to forecast the PV power output (only the historical PV power signal and predicted temperature are used as inputs to the model), (b) significant increase in forecasting performance, outperforming all five other models tested here, and (c) minor complexity and computational cost increase, making it suitable for real-time applications, such as optimal control algorithms that require PV forecasts in a short-term resolution in order to execute control actions in real-time.

2. A novel forecasting model that uses a combination of the SWT, four LSTM (long–short–term memory) Networks, and a Deep Neural Network (DNN) to accurately forecast the PV power for the next time step. The model is defined in a way that it can be easily adapted to perform multi-step forecasts, i.e. \((t + 1, t + 2, \ldots, t + 12)\).

3. A real-world test of the model is conducted by performing forecasts using real data obtained from a 12.6 MW PV plant located in the state of Florida and publicly available data from other databases facilitated by the National Renewable Energy Laboratory (NREL).

### 3 Wavelet transform

The highly fluctuating nature and non-stationary behaviour of the original PV power signal are good indications of how the use of the WT can be useful for recognising the frequency and temporal patterns in order to improve the forecast error differences. The basic definition of the WT gives us the capability of representing patterns in order to improve the forecast error differences. The WT can be useful for recognising the frequency and temporal analyses of the signal due to the fact that it is a shift-variant transform. These shift-variant results arise from the use of the scaling filter and the wavelet function, which in turn can be used for obtaining information about the time and frequency domains of the signal. The WT is most commonly defined in two categories: the continuous WT (CWT) and the discrete WT (DWT) [25]. The CWT is defined by (1), where the \(*\) denotes the complex conjugate of the set of wavelets chosen, and \(s\), \(t\), and \(\tau\) are the scale and translation dimensions of the signal. The set of wavelets is defined by (2), where the set is generated from a mother wavelet, \(\psi\), and specific values for the translation and scale of the signal, including an energy normalisation factor, \(1/\sqrt{\delta_s}\). In other words, just as the Fourier Transform (FT) decomposes the respective signal into sine and cosine signals, the CWT decomposes the non-stationary signal into a series of wavelets with a different combination of scales and translations [29].

\[
\gamma(s, t) = \int_{-\infty}^{\infty} f(t) \psi^*_s(t) \, dt \quad (1)
\]

\[
\psi_s(t) = \sqrt{\delta_s} \psi\left(\frac{t - \tau}{s}\right) \quad (2)
\]

However, the high computational complexity of calculation and high redundancy of the CWT make it impractical for the desired application. According to (1), in the CWT, the wavelet transform is constantly being calculated by shifting a continuously scalable function over a signal and obtaining the correlation between these two signals. In consequence, the obtained wavelet coefficients will be highly redundant and, for most functions, the CWT will have no analytical solution, and its solution must be calculated numerically. In order to address these issues, the DWT is introduced. The DWT analyses the signal at different frequency bands with various resolutions by decomposing the original signal into detailed and approximation coefficient values. To achieve this, the DWT makes use of two set of functions called the scaling function, which is associated with a low-pass filter, and the wavelet function, associated with a high-pass filter. The DWT can be defined by (3), where \(j\) and \(k\) are integers and \(\tau_k\) and \(\delta_k\) are the discretised scales and translations. Subsequently, the scaling function \(\phi(2^j t - k)\) and wavelet function \(\psi(2^j t)\) of the DWT can be described by (4) and (5).

\[
\psi_j,(t) = \frac{1}{\sqrt{\delta_k}} \psi\left(\frac{t - k\tau_j}{\delta_k}\right) \quad (3)
\]

\[
\phi(2^j t) = \sum_{k} h_{j-1}(k) \phi(2^{j-1} t - k) \quad (4)
\]

\[
\psi(2^j t) = \sum_{k} g_{j-1}(k) \psi(2^{j-1} t - k) \quad (5)
\]

Using these equations, the signal \(f(t)\) can be expressed as (6), where \(\lambda_{j-1}(k)\) and \(\gamma_{j-1}(k)\) are the coefficients obtained by taking the inner products of the signal, \(f(t)\), with the scaling and wavelet function, respectively, as seen in (7) and (8).

\[
f(t) = \sum_{k} \lambda_{j-1}(k) \phi(2^{j-1} t - k) + \sum_{k} \gamma_{j-1}(k) \psi(2^{j-1} t - k) \quad (6)
\]

\[
\lambda_{j-1}(k) = \langle f(t), \phi_j,(t) \rangle \quad (7)
\]

\[
\gamma_{j-1}(k) = \langle f(t), \psi_j,(t) \rangle \quad (8)
\]

The coefficients \(h(k)\) in (4) are known as the scaling filter and the coefficients \(g(k)\) in (5) are called the wavelet filter. These two sets of coefficients are associated with the decomposition or filtering process in which the signal is decomposed into different frequency bands. As seen in Fig. 1, the decomposition process of the signal \(f(t)\) begins with the filtering of the signal via a high-pass and a low-pass filter, followed by the subsampling of the resultant signal by 2. This decomposition procedure, commonly known as sub-band coding, halves the time resolution, by only taking half of the samples, while doubling the frequency resolution, by spanning half of the previous signal frequency band. Fig. 1 depicts the signal decomposition/reconstruction procedure. As observed, the low-pass filter removes the higher frequency components of the signal and outputs the approximate coefficients while the high-pass filter outputs the detailed coefficients of the signal.

#### 3.1 Stationary wavelet transform (SWT)

Nonetheless, the DWT presents a significant potential problem for the analysis of the signal due to the fact that it is a shift-variant transform [30]. These shift-variant results arise from the use of the sub-sampling operation on the DWT algorithm and result in wavelet coefficients that are highly dependent on their location on the sub-sampling lattice. In other words, as the analysis of the signal becomes more certain in the frequency components, the analysis of the time domain becomes less certain. Small shifts in
the original signal have the potential of causing large changes in the wavelet coefficients and large changes in the reconstructed waveform [30]. To avoid this issue, the sub-sampling operation of the signal is removed from the filtering operation in the DWT and only the up-sampled operation is kept at each level of the decomposition. This modification in the DWT is commonly known as the SWT. The use of the stationary data resulted in more skilful forecasts when using the LSTM networks to predict the individual decomposed coefficients of the signal. Fig. 2 depicts the scheme used for the SWT.

The mother wavelet or wavelet function chose to perform the SWT decomposition process was the nearly symmetrical wavelet, symlet 2. This wavelet was the one that gave the best forecasting performance when compared with the Daubechies family, Coiflet family, Morlet, and Haar wavelets [31].

The symlet wavelets have the characteristic of being orthogonal and near symmetric, ensuring minimal phase distortion. These symmetrical wavelets can be identified based on the order $N$, which indicates the number of vanishing moments of the wavelet for a given support width of $2N − 1$. Fig. 3 shows the corresponding scaling function and wavelet function of the symlet with $N = 2$, most commonly known as symlet 2. A more thorough mathematical description of the symmetrical wavelet equations and coefficients can be found in [32].

4 Methodology: hybrid LSTM-DNN model

The essential concept behind the operation of the proposed hybrid wavelet-based LSTM-DNN model consists in the use of a DNN trained with features obtained from the statistical analysis of the power signal, the predicted temperature value, and a fabricated forecast value obtained by the reconstruction of the forecasted output coefficients given by individual LSTMs. The LSTM networks are trained to perform multi-step forecast of the individual approximate and detailed coefficients of the SWT decomposed power signal for the next $24$ h. These values are then reconstructed into a forecasted power signal using the inverse SWT (ISWT), and the respective value is used as a feature to the DNN. The main idea behind using the SWT decomposition in the model is based on the ability to capture and forecast only the most relevant information, in the time and frequency domain, from the historical signal. By decomposing the historical PV power signal into four different SWT coefficients (three detailed and one approximated), LSTMs are able to improve the overall signal forecast based on their ability to predict these individual components instead of forecasting the aggregated PV power signal.

In our case, only four different SWT coefficients are chosen due to the fact that these were the ones that contain the most useful information for the forecasting process of the analysed signal. According to the tests conducted, it was observed that adding more coefficients, such as the approximate 1 and 2, or decomposing the signal further increased the complexity and training time significantly while adding almost zero or negative effects on the forecast results. The number of decomposition levels depends primarily in the particular signal being analysed, that is why, if the proposed model is used for another application, such as load forecasting, tests should be made in order to find the optimal decomposition levels for both the approximate and detailed decomposition levels. Fig. 4 shows the overall design of the proposed hybrid wavelet-based LSTM-DNN model. As observed in the figure, the training and validation of the proposed model can be divided into four major steps:

- **Step 1**: Data extraction, sliding window processing, and SWT decomposition.
- **Step 2**: Training and validation of LSTM RNNs models and ISWT reconstruction.
- **Step 3**: Statistical features extraction.
- **Step 4**: DNN training and validation via feature selection.

4.1 Data extraction, sliding window processing, and SWT decomposition

Before applying the selected sliding window technique and performing the SWT decomposition, the historical PV power data is pre-processed with the objective of removing any outlier or incorrect measurements present in the data. After this pre-processing step is completed, a sliding window technique that divides the data into blocks of $24$ h (or $48$ values in 30-min resolution) is used to process data for later decomposing each block using the SWT. A sliding window technique is a data processing method designed to divide a datastream $x$ into several chunks of data blocks, $k$. In our particular case, $k = 48$. The current block window of size $48$ is then decomposed into approximate and detailed coefficients by passing the block through the SWT decomposition. As mentioned before, the mother wavelet used in the decomposition procedure is the symmetrical wavelet $2$ (symlet2). After several tests using different wavelets, this was the wavelet that gave the best performance in terms of less error. Similarly, it was discovered that the optimal level for this particular PV signal was the decomposition level $3$. This means that the coefficients used for this particular forecasting application are the...
detailed 1, detailed 2, detailed 3, and approximate 3. Other applications could benefit from the use of a different mother wavelet and a different decomposition level.

4.2 Training and validation of LSTM RNNs models and ISWT reconstruction

An LSTM network can be defined as a type of RNN structured to solve the vanishing gradient problem present in regular RNNs. LSTMs are explicitly designed to avoid the long-term dependency problem by helping to preserve a more constant error allowing the RNNs to continue to learn over many time steps. An LSTM contains special blocks called memory blocks, which in part contain memory cells with recurrent connections, that allow storing the temporal state of the network, and other units called gates that control the flow of information in the network. Fig. 5 depicts the general architecture of a LSTM block. As seen in the figure, the LSTM block has the ability to add or remove information to the cell state, $C_t$, by updating its value according to the resultant operations obtained from the ‘forget gate layer’ and the ‘input gate layer’. The output of the block, $y_t$, is computed using a filtered version (passed through tanh) of the current state of the cell ($C_t$) and the output computed from the inputs ($o_t$). Succeedly, the LSTM network will compute an output sequence $y_1, \ldots, y_T$ from an input sequence $x_1, \ldots, x_T$ using the following equations:

\[
\begin{align*}
    f_t &= \sigma(W_f \times [C_{t-1}, y_{t-1}, x_t] + b_f) \\
    i_t &= \sigma(W_i \times [C_{t-1}, y_{t-1}, x_t] + b_i) \\
    o_t &= \sigma(W_o \times [C_{t-1}, y_{t-1}, x_t] + b_o) \\
    C_t &= f_t \times C_{t-1} + i_t \times C_t^\phi \\
    y_t &= o_t \times h(C_t)
\end{align*}
\]

where $W$ is used to denote the weight of the matrices (e.g. $W_f$ is weight matrix from the forget gate layer), $b$ refers to a bias vector (e.g. $b_f$ is the bias vector from the forget gate layer), $f$ is the forget gate layer value, $C$ is the cell state, $i$ is the input gate layer value, and $o$ is the output gate value. Function $\sigma$ represents the sigmoid gate activation function, $g$ is the input activation function (tanh), and $h$ is the output activation function (tanh).

Four LSTMs networks are trained to recognise the non-linear patterns of each coefficient (D1, D2, D3, A3) obtained from the SWT decomposition. Before training and validating each LSTM, each coefficient signal is normalised using min-max normalisation with the objective producing values between the range of −1 and 1.

The outputs of the LSTM are de-normalised and used to reconstruct the forecasted PV power signal using the inverse SWT (ISWT) procedure. The value obtained from this process is then used as a feature or predictor input to the DNN model.

4.3 Statistical features extraction

Besides the PV power predictor described above, five additional statistical features are used to represent the characteristics of the
PV power time series. These features are the Mean ($\mu$), Standard Deviation ($\sigma$), Variance (Var), Skewness (Skew), and Kurtosis (Kurt). These values are calculated from window signals created from the recorded PV power values at a specific time. For instance, if we need to forecast the PV power value at 1:00 pm, a statistical analysis is performed on the signal made up of historical values of the PV power at 1:00 pm. The $\mu$, $\sigma$, Var, Skew, and Kurt are then calculated from the signal.

4.4 DNN training and validation via feature selection

In this final stage, a DNN is trained using an exhaustive feature selection strategy designed to find the best forecast, i.e. the smallest root-mean-squared error (RMSE), for the validation set. The input layer of DNN contains $m$ features determined by the feature selection procedure, in which $m = P + T + n$. Here, $n$ is the number of relevant statistical features, $P = 1$ the forecasted PV power value obtained from the LSTMs, and $T = 1$ the predicted temperature value for the defined time. The number of hidden neurons and hidden layers of the DNN was determined empirically through a trial-and-error process. Based on the observations derived from the trial-and-error process, it was found that three hidden layers with 45 hidden neurons gave the best performance and neither higher number of neurons or hidden layers gave any improvement on the performance of the DNN. The output layer of the DNN consists of one neuron that represents the forecasted PV power value.

5 Training and validation of results

The databases used in this analysis were obtained from the Sunshine State Solar Grid Initiative (SUNGRIN) project data repository [33] (a DOE sponsored project), and the NREL (National Renewable Energy Laboratory) solar radiation database. The PV power data were obtained from a fixed ground-mounted solar plant located in northern Florida, with an AC power rating of 12.6 MW and a time resolution of 30 min. The database was pre-processed in order to separate the relevant data and select the predictive features used in the models proposed for the forecasting analysis. The models used to evaluate and compare the performance of the proposed hybrid wavelet-based LSTM-DNN model are the persistence (naive) model, the wavelet transform BPNN (WT+BPNN) [25], the wavelet transform (WT + RBFNN) [25], a DNN model, a support vector regressor (SVR) model, and an LSTM only model. The DNN and the SVR are modelled with the same features as the WT + RBFNN and the WT + BPNN, with the difference that they do not include any wavelet features. The exact features used by these models are: predicted temperature, temperature half-hour earlier, temperature 12 h earlier, temperature 20 h earlier, irradiance half-hour earlier, irradiance 12 h earlier, irradiance 20 h earlier, PV power half-hour earlier, PV power 12 h earlier, and PV power 20 h earlier. The LSTM only model is based on the structure that gave the best performance of all models proposed in [16]. Fig. 6 describes the methodologies used for each of the test models used to compare the proposed model.

The forecasting models were trained, validated, and tested using 35,089 samples (around 2 years' worth of data). In order to avoid overfitting, the dataset was split into a training set (70%), a validation set (15%), and a testing set (15%) using a technique called k-fold cross-validation [34]. The k-fold cross-validation sets were divided into $k = 3$ folds of three sets of multiple days:
Table 1 shows the results obtained from each of the forecasting models tested in each of the test datasets. As observed, the proposed hybrid wavelet-based LSTM-DNN model outperforms all the other models compared in almost every metric used. It is able to significantly reduce the MSE, MAPE, and nRMSE values when compared to other models such as the persistence, the SVR, the WT + RBFNN, the WT + BPNN, and the LSTM. It should be noted that the WT + RBFNN and the WT + BPNN did not perform as expected with the evaluated PV power signal. These differences could come from the fact that the PV power signal used in this study had a smaller time resolution (30-min instead of 1-h) than the one used in [25]. Moreover, even though the DNN model also showed promising forecasting results, metrics such as the R-squared and the FSS show an increase in forecasting performance when the hybrid wavelet-based LSTM-DNN model was used. It can also be observed that using an LSTM to directly forecast the PV power signal can be simpler to implement but gives worse results when compared to the proposed model. This comes from the fact that performing the SWT decomposition filters out unnecessary information, such as noise from the signal, thus improving the performance of the individual LSTMs and the overall performance of the model making it more resilient to noise. Fig. 7 show the R-squared plots for each forecasting model and each of the test cases evaluated in Table 1. Table 2 shows the average training times for each one of the forecasting models evaluated.

From the experiments and tests conducted, it can be observed that the proposed model gives advantages mainly by increasing the prediction accuracy while reducing the number of measurement features or historical data (temperature, irradiance, humidity etc not used in the hybrid model) needed for the forecasting model. It was observed that the proposed model did not increase the complexity and training time of the model significantly, making it a viable option to be used for real-world applications where the number of measurement features, such as irradiance and humidity, is difficult to obtain due to the necessity of owning the required specialised equipment and in large-scale PV systems applications where 2 to 3% differences in forecasted values can be significant for control operations.

On the other hand, it is important to discuss some of the limitations that the proposed model has when compared with other well-known forecasting models. These limitations are primarily related to the quality of the data used for training the model and its availability. The performance of the model could naturally decrease if the model does not have sufficient training data or if the

### Table 1: Comparison of forecasting performance of the proposed Hybrid WT + LSTM-DNN model with all six other models

<table>
<thead>
<tr>
<th>Model Type</th>
<th>MSE, kW</th>
<th>nRMSE</th>
<th>nMBE</th>
<th>R</th>
<th>MAPE, %</th>
<th>FSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>persistence (naive)</td>
<td>Spring</td>
<td>1857</td>
<td>0.110</td>
<td>0.000</td>
<td>0.942</td>
<td>5.673</td>
</tr>
<tr>
<td></td>
<td>Summer</td>
<td>2332</td>
<td>0.124</td>
<td>0.000</td>
<td>0.924</td>
<td>6.454</td>
</tr>
<tr>
<td></td>
<td>Fall</td>
<td>1080</td>
<td>0.084</td>
<td>0.000</td>
<td>0.948</td>
<td>3.942</td>
</tr>
<tr>
<td>DNN</td>
<td>Spring</td>
<td>1154</td>
<td>0.087</td>
<td>0.010</td>
<td>0.965</td>
<td>4.254</td>
</tr>
<tr>
<td></td>
<td>Summer</td>
<td>1545</td>
<td>0.101</td>
<td>0.006</td>
<td>0.950</td>
<td>5.097</td>
</tr>
<tr>
<td></td>
<td>Fall</td>
<td>847</td>
<td>0.075</td>
<td>0.022</td>
<td>0.967</td>
<td>3.635</td>
</tr>
<tr>
<td>SVR</td>
<td>Spring</td>
<td>1260</td>
<td>0.091</td>
<td>0.006</td>
<td>0.960</td>
<td>4.892</td>
</tr>
<tr>
<td></td>
<td>Summer</td>
<td>1536</td>
<td>0.100</td>
<td>0.004</td>
<td>0.949</td>
<td>5.271</td>
</tr>
<tr>
<td></td>
<td>Fall</td>
<td>948</td>
<td>0.079</td>
<td>0.029</td>
<td>0.964</td>
<td>4.149</td>
</tr>
<tr>
<td>LSTM</td>
<td>Spring</td>
<td>1747</td>
<td>0.104</td>
<td>−0.029</td>
<td>0.949</td>
<td>5.399</td>
</tr>
<tr>
<td></td>
<td>Summer</td>
<td>2204</td>
<td>0.1178</td>
<td>−0.022</td>
<td>0.928</td>
<td>6.348</td>
</tr>
<tr>
<td></td>
<td>Fall</td>
<td>958</td>
<td>0.077</td>
<td>0.0071</td>
<td>0.953</td>
<td>3.941</td>
</tr>
<tr>
<td>WT + BPNN</td>
<td>Spring</td>
<td>1891</td>
<td>0.111</td>
<td>0.006</td>
<td>0.939</td>
<td>6.404</td>
</tr>
<tr>
<td></td>
<td>Summer</td>
<td>2145</td>
<td>0.119</td>
<td>0.002</td>
<td>0.928</td>
<td>6.933</td>
</tr>
<tr>
<td></td>
<td>Fall</td>
<td>1011</td>
<td>0.081</td>
<td>−0.021</td>
<td>0.966</td>
<td>4.160</td>
</tr>
<tr>
<td>WT + RBFNN</td>
<td>Spring</td>
<td>2782</td>
<td>0.135</td>
<td>0.001</td>
<td>0.912</td>
<td>7.088</td>
</tr>
<tr>
<td></td>
<td>Summer</td>
<td>3610</td>
<td>0.154</td>
<td>0.005</td>
<td>0.887</td>
<td>8.119</td>
</tr>
<tr>
<td></td>
<td>Fall</td>
<td>1921</td>
<td>0.112</td>
<td>−0.010</td>
<td>0.903</td>
<td>5.312</td>
</tr>
<tr>
<td>hybrid WT + LSTM-DNN</td>
<td>Spring</td>
<td>973</td>
<td>0.080</td>
<td>−0.016</td>
<td>0.971</td>
<td>3.749</td>
</tr>
<tr>
<td></td>
<td>Summer</td>
<td>762</td>
<td>0.071</td>
<td>−0.005</td>
<td>0.975</td>
<td>3.761</td>
</tr>
<tr>
<td></td>
<td>Fall</td>
<td>660</td>
<td>0.066</td>
<td>0.015</td>
<td>0.971</td>
<td>3.555</td>
</tr>
</tbody>
</table>
available training data have a substantial amount of missing data points or falsified data points. Proper pre-processing procedures, such as filtering, cleaning, and false data checks, need to be performed to avoid high variance in the LSTM and DNN models. Other forecasting techniques that do not rely so much on historical data could potentially outperform the proposed model due to these issues. Further improvements of the proposed model could focus on the development of pre-processing techniques tailored to perform integrity checks on the training and testing data.

6 Conclusion

This paper proposed a novel hybrid wavelet-based LSTM-DNN forecasting model designed to forecast the PV power available in a PV system for a medium- to short-term forecasting period. The highly variable behaviour observed on the solar power series made it a good candidate for the application of wavelet analysis and the use of LSTMs and DNN networks for forecasting purposes. The results obtained (MSE, nRMSE, nMBE, R, MAPE, FSS) demonstrate that the model proposed was highly accurate when predicting the non-linear response of the solar power generation in a PV system and outperforms other evaluated models in terms of better prediction accuracy. The adoption of this type of models could become the turning point in the introduction of solar energy giving operators more control and information regarding the best way to use these renewable energy resources. Future work will focus on the exploration of this method for other forecasting problems (e.g. load forecasting), the exploration of a more sophisticated feature selection strategy, and the development of optimal predictive controls designed to optimise the management and consumption of renewable energy.

7 References


Table 2 Average training times for all forecasting models

<table>
<thead>
<tr>
<th>Models</th>
<th>Average training time</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNN</td>
<td>00:04:51</td>
</tr>
<tr>
<td>SVR</td>
<td>00:02:46</td>
</tr>
<tr>
<td>LSTM</td>
<td>00:03:26</td>
</tr>
<tr>
<td>WT + BPNN</td>
<td>00:04:33</td>
</tr>
<tr>
<td>WT + RBFNN</td>
<td>00:03:56</td>
</tr>
<tr>
<td>hybrid WT + LSTM-DNN</td>
<td>00:06:18</td>
</tr>
</tbody>
</table>

Fig. 7 Scatter R-squared plots for 30-minute ahead PV power forecasting. Each row represents the different test cases: Spring, Summer, Fall. Each column represent a different forecasting model. Staring from the left: column 1 = Persistence, column 2 = DNN, column 3 = SVR, column 4 = LSTM, column 5 = WT + BPNN, column 6 = WT + RBFNN, column 7 = Hybrid WT + LSTM-DNN


