

Solar Energy Production Forecast Using Standard Recurrent Neural Networks, Long Short-Term Memory, and Gated Recurrent Unit

Adrian-Nicolae Buturache¹, Stelian Stancu²

¹*Economic Cybernetics and Statistics Doctoral School, Bucharest University of Economic Studies
Piata Romana Square no. 6, 010374, Bucharest, Romania
E-mail. ad.buturache@yahoo.ro*

²*Informatics and Economic Cybernetics Department, Bucharest University of Economic Studies
Piata Romana Square no. 6, 010374, Bucharest, Romania
E-mail. stelian.stancu@csie.ase.ro*

crossref <http://dx.doi.org/10.5755/j01.ee.32.4.28459>

Solar radiation is among the renewable resources on which modern society relies to partially replace the existing fossil fuel-based energy resources. Awareness of how the energy is produced must complement awareness of how it is consumed. In the economic context, the gains derive from predictability across the entire supply chain. This paper represents a compressive study on how standard recurrent neural networks, long short-term memory, and gated recurrent units can be used to forecast power production of photovoltaic (PV) systems. This approach can be used for other use cases in solar or even wind power prediction since it provides solid fundamentals for working with weather data and recurrent artificial neural networks, being the core of any smart grid management system. Few studies have explored how these models should be implemented, and even fewer have compared the outcomes of different model types. The data used consist of weather and power production data with a one-hour resolution. The data were further pre-processed to unveil the maximum information. The most effective model parameters were selected to make the forecast. Solar energy plays a key role among other renewable energy sources in the European Union's climate action and the European Green Deal. Under these initiatives, important regulations are implemented and financial resources made available for those who possess the capabilities required to solve the open points. The much-needed predictability that gives the flexibility and robustness needed for deploying and adopting more renewable technologies can be ensured by utilizing a neural-based predictive approach.

Keywords: *Recurrent Neural Networks; Long Short-Term Memory; Gated Recurrent Unit; Renewable Energy; Photovoltaic Cells; Forecast; Smart Grids; Green Energy; Data-Driven Decision.*

Introduction

Renewable energy has attracted considerable interest owing to its sustainability and potential abundance. In recent decades, photovoltaic (PV) cells have been installed globally, demonstrating that this solution could help in reducing the dependence on conventional polluting and limited energy resources. This solution began as niche technology before attaining a global reach. The stochastic nature of solar irradiance and meteorological conditions leads to discontinuity in solar power production, impeding smooth and continuous supply to customers. Intermittency and wide fluctuations in energy supply have a negative impact on the entire grid infrastructure. Finding the balance between supply and demand is above all a matter of forecasting energy production. Therefore, several problems that hinder the adoption of solar energy can be resolved through the development of predictive tools. Forecasting PV power output is key to successfully integrating PV-based renewable energy solutions into existing power grids. The scientific community is developing methods at the same pace at which the industry evolves.

All technical aspects related to renewable energy forecasting must be further coupled with the economic implications. In addition to a continuous supply, the end user is ultimately seeking cheap energy, while the market

players seek greater profits. First, an accurate forecast will help solar farm operations teams to manage the assets in sync with the grid (Wan *et al.*, 2015) and develop a lean-oriented production (Halldorsson *et al.*, 2018). On the Transmission System Operator, a balance must be ensured in accordance with the produced and consumed energy (Gyalai-Korpos *et al.*, 2020). In the case of solar energy or any other renewable energy resources for which volatility is present, the gap between supply and demand is filled using other energy resources or stored energy. Additional costs are generated each time this mechanism is activated (Gyalai-Korpos *et al.*, 2020). An accurate forecast of the energy to be produced leads to improved economic performance (Kraas *et al.*, 2013) and the economic performance is inversely proportional with the error level (Granger & Pesaran, 1997). Moreover, the effects on the economic side are not caused only by the model performance but also by the way in which the production has been forecasted. Overestimation and underestimation have an asymmetric impact, and the model hyperparameter tuning must take the market conditions into consideration, as even this decision might create a bias (Botterud *et al.*, 2012). For energy dispatching, cost-oriented strategies that have energy production and consumption forecast at their core are also applied (Hanna *et al.*, 2014; Lee *et al.*, 2019; 164).

A performant model must be able to forecast a wide range of power outputs during short periods of time due to the day-night cycle, weather conditions, and solar elevation angle. It is still necessary to prototype and prove to customers that sufficiently flexible solutions are available to accommodate performant predictive analytics regardless of the technology, scale, geographical position, or other measurable factors. Optimal management of renewable power systems will make these technologies more profitable and efficient. In the last 170 years, energy demand has been on an ascending path, increasing by approximately 2.5 % per year (Jarvis *et al.*, 2012). By the end of 2050, it is projected that the increase in energy demand compared with 2018 will be close to 50 % (Kahan, 2019). Existing non-renewable technologies have been provided since the Industrial Revolution, given the need for considerable amounts of energy.

Two main issues emerge in relation to the existing system: negative environmental impact and resource depletion. Pollution and global warming, nuclear accidents, and the gap in recycling radioactive residuals—all in the context of new renewable technologies that are more reliable and effective than ever—offer a viable option to shift the paradigm toward a green and sustainable energy ecosystem. Economic performance and safe operation depend on the accuracy of the consumption forecast (Uniejewski *et al.*, 2019). Over the next 10 years, the European Commission aims to reduce greenhouse gas emissions by at least 40 % (from 1990 levels) to achieve at least a 32 % share of renewable energy and at least a 32.5 % improvement in energy efficiency (European Parliament, 2018). Taking into consideration the ongoing electrification process, the consumption pattern is likely to generate peaks in demand in the evening or at night. This reinforces the need to adopt the appropriate strategy for energy storage. In this context, the PV power output forecast becomes very important. With the same approach, but for a different purpose, the energy production forecast can be used to engage in energy trading (Carriere & Kariniotakis, 2019). The implementation of a forecasting system can reduce uncertainty over the energy price, leading to economic benefits (Alessandrini *et al.*, 2014; Barthelmie *et al.*, 2008; Kraas *et al.*, 2013).

Placing this article in a European and strategic context, according to SolarPower Europe, solar energy is generally cheaper than any other type of energy that we are capable of producing nowadays, and it is predicted to show the steepest cost reduction curve in the coming years (SolarPower Europe, 2019). In its Clean Energy for All Europeans Package and energy policy framework, the European Union encourages the use of solar-related energy production technologies through governance regulations and modernization of Europe's electricity market for greater flexibility with the aim of integrating a greater share of renewable energy. Currently, solar energy prices average from 0.03 EUR per kWh in Southern Europe to 0.05 EUR per kWh in Northern Euro and are expected to drop to 0.01 EUR per kWh by 2050. To achieve the clean energy goals, including the reduction of solar energy costs, the European Investment Bank has committed to finance clean energy innovation, renewable energy, and energy efficiency with up to EUR 1 billion over the next decade (SolarPower Europe, 2019). The present research has the potential to solve open points in any of the

three main interest areas identified. The aim is to propose a comprehensive solution that is ready to be deployed and that will increase the level of predictability, at least at production sites. There currently exists a gap in the literature dealing with solar energy forecasting through the implementation of recurrent neural-based solutions. This gap also includes the ways in the selected models are presented, and this is also addressed in the present paper. The sections that follow will focus on the theoretical and experimental considerations leading to a performant forecast.

Literature Review

PV power output forecasting models can be categorized based on their type (physical, statistical, machine learning, or hybrid) or time horizon (short-term, medium-term, or long-term). Physical models can deliver good accuracy if two conditions are respected: accurate weather data and deep understanding of the solar power installation to calculate the output based on the power curves. Statistical and machine learning models map the relationship between input and output variables without probing the profound details of how solar irradiance is transformed into electrical energy. In other words, the physical models' complexity and lack of flexibility can be eliminated. However, accurate weather data are still required. The hybrid approach combines the main categories listed before. Short-term forecasting can last from minutes to maximum days. Medium- and long-term forecasting are addressed more for strategic decisions, maintenance, and the need to develop and construct new infrastructure to fulfill power demands. Artificial neural networks (ANNs) are machine learning-based models that can address non-linear problems, such as solar energy forecasts. Recurrent neural networks (RNNs) are part of the learning networks that are suited to handling sequential data, such as time series or textual data. In this study, multivariate data and RNNs are used to forecast PV power output. Three types of RNN model were compared according to the same scope, under the same assumptions, and with the same data: standard recurrent neural networks (SRNN), long short-term memory (LSTM) neural networks, and gated recurrent units neural networks (GRU). Considering the dependence on weather parameters and the fact that weather forecasting remains an open topic, PV power output forecast has become a particularly complex problem. The compromise in this case is to reduce the forecast horizon to days or even hours for more accurate results. Forecasting weather in this research is out of scope, with weather data already available. ANNs are recognized as universal approximators that are able to find hidden patterns in the data. Even ANN, by its structure, can deal with noisy data; at some point, the noise and redundancies in data must be removed (Zhang *et al.*, 2018). De *et al.* (2018) performed various optimization scenarios to better understand the behavior of the model under the influence of shifting data set division, number of hidden nodes, number of input variables, or activation function. Adding more variables that seemed to impact PV power output did not add value in terms of model performance. Moreover, common sense leads to the conclusion that the higher the dimensionality, the more resources are needed for training. Another study proposed that univariate time series

forecasting highlights the superiority of the RNN over backpropagation neural networks, radial basis neural networks, support vector machine, and the persistence method (Li *et al.*, 2019). The same univariate approach has been tested for short-term forecasts in which LSTM outperformed the moving average (Poudel & Jang, 2017). The performance of hybrid models-including autoregressive integrated moving average coupled with support vector machine (Bouzerdoum *et al.*, 2013) or autoregressive model coupled with auto-regressive with exogenous input model in a time-step approach (Bacher *et al.*, 2009) -was reasonable for short-term forecasting results but not comparable with the results obtained using machine learning-based models. In the physical methods category, one interesting approach involves the use of satellites to monitor cloud movement and use it as input into solar radiation forecasting models (Hammer *et al.*, 1999; Lorenz *et al.*, 2009). However, although ANN and sky imaging methods can deliver excellent performance (Crisosto *et al.*, 2018), the overall cost is increased by deploying and operating expensive equipment, adding further complexity by working with images. Feed-forward neural networks (FFNNs) that are not designed for use with sequential data and are not as performant as RNN can deliver reasonable performance (Yona *et al.*, 2007; Gao *et al.*, 2019). Another potential approach is to use the radiation classification coordinate method to find and select similar sub-time series that are further used as input in learning networks-based models where LSTM was proven to perform better than backpropagation neural network, radial basis neural network, and Elman neural network (Chen *et al.*, 2020). A similar thought process can be applied to the weather data to categorize it into four main categories (rainy, cloudy, overcast, and sunny-ideal state), resulting in four different performance levels for four different scenarios (Lee & Kim,

2019). In the long-term forecast, LSTM can outperform multiple linear regression, bagged regression tree, and FFNNs (Abdel-Nasser & Mahmoud, 2017). RNN models exhibit better performance than statistical, physical, and even hybrid models. Statistical models are unable to capture all phenomena relating to solar power production, and physical models are too rigid and complex to resolve issues under the continuous uncertainty of weather data. Currently, in the literature on solar energy forecasting, the use of recurrent neural-based models is a gap. The most important recurrent neural-based models will also be examined in depth, as these will shape the models for a solar energy forecast use case.

Theoretical Fundamentals

An RNN is a type of neural network built to handle sequential data, such as time series, wherein a time step is predicted with respect to the previous time steps (Figure 1). Hidden layers store the information of the previous time steps, making it available as input when a new state is predicted. SRNN can be defined using the following equations (Pascanu *et al.*, 2014):

$$h_t = f_h(x_t, h_{t-1}) = \phi_h(\mathbf{W}^T h_{t-1} + \mathbf{U}^T x_t) \quad (1)$$

$$y_t = f_o(h_t, x_t) = \phi_o(\mathbf{V}^T h_t) \quad (2)$$

where f_h, f_o are state transition function and output function, further represented by element-wise non-linear functions, \mathbf{W}, \mathbf{U} , and \mathbf{V} , representing transition, input, and output matrices. Transition matrices are also known as recurrent weights matrices. The current state, t , is a function of the current input and previous state. The current state output y_t is a function of output synaptic weights and hidden vector h_t

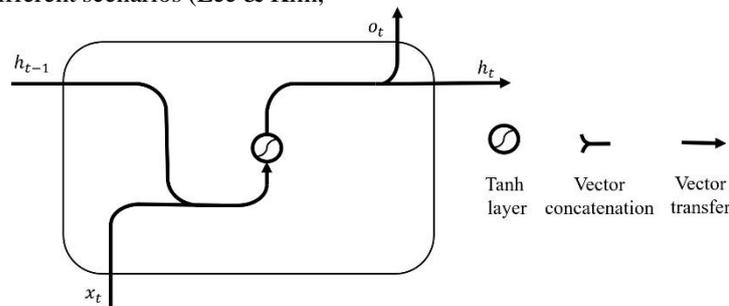


Figure 1. Standard RNN Cell

Network parameters can be estimated with gradient-descent-based algorithms, making the computations by backpropagation through time (BPTT) (Rumelhart *et al.*, 1986). BPTT, used for the RNN, is similar to the backpropagation of the FFNNs. For the RNN, the BPTT occurs from the current cell state back to the initial cell state. The gradient is calculated with respect to each individual time step. Since the weights are identical for all cells, the resulting gradients must be summed at the end. One of the main weaknesses of this design is represented by the vanishing or exploding gradient (Bengio *et al.*, 1994). The problem occurs for the synaptic weight that connects the hidden layers: if the weight value is small, it causes a vanishing gradient. If the weight is large, it causes an

exploding gradient. After multiplication operations, the gradient value, based on the weight's initial values, will become extremely small or large. For a lower gradient, the training time increases, and the weights update becomes harder as the change in the weights becomes insignificant. At the same time, the low gradient may cause the initial considered time steps to be improperly trained while the final time steps may have been trained well. However, not even their training cannot be considered to have been done well, since the input is coming from upstream. Another main weakness of the standard recurrent neural network is associated with the fact that in the case of applications where information of multiple earlier steps is needed, it fails to keep the correct long-term context data (i.e., important

information that happens to be at time step 2 may be lost until step 20—the final step—has been reached). This is called information morphing.

A new type of RNN incorporates three gates: the input, forget, and output gates (Figure 2). These three gates are represented by sigmoid-activated neural networks that bound the output between 0 and 1. The two interval ends mean that all the information is blocked when the value is 0 or that all the information is allowed when the value is 1. By convention, the different memory blocks that constitute the LSTM network are called cells. In the current state (denoted further by time step t), information from the previous cell ($t - 1$) is used—cell state and hidden state. The forget gate's task is to remove unnecessary information sent by the previous cell. This is described in (3), where W and U are the weights matrix of the previous cell hidden state (recurrent connection) and the weights matrix of the input of the current state (current hidden layer), respectively.

$$f_t = \sigma(x_t U^f + h_{t-1} W^f) \quad (3)$$

$$i_t = \sigma(x_t U^i + h_{t-1} W^i) \quad (4)$$

$$\hat{C}_t = \tanh(x_t U^g + h_{t-1} W^g) \quad (5)$$

$$C_t = \sigma(f_t \cdot C_{t-1} + i_t \cdot \hat{C}_t) \quad (6)$$

$$o_t = \sigma(x_t U^o + h_{t-1} W^o) \quad (7)$$

$$h_t = \tanh(C_t) \cdot o_t \quad (8)$$

For all three gates, the sigmoid function is used to decide which information is to be kept and which information is to be discarded. Through the input gate, the information is added to the cell state. By employing the same sigmoid-activated neural network (4), only the information that needs to be added to the cell state C is retained. Equation (5) creates a new vector of values that could be part of the state, \hat{C}_t . Current state, C_t , (6), depends on the information to be added via input gate and the information to be retained after passing the forget gate. Output gate (7) controls the information that leaves the LSTM cell (8); similarly, the forget gate and input gate

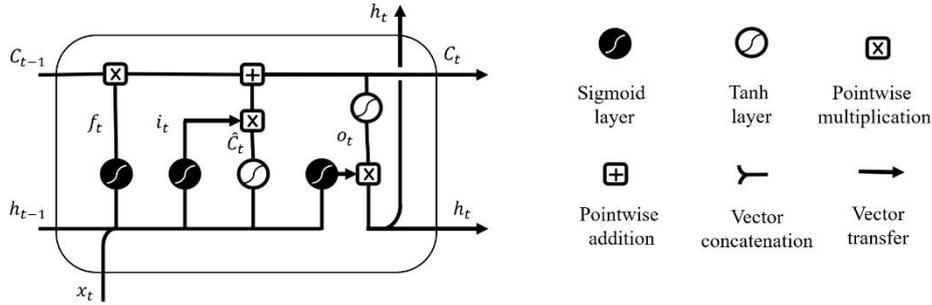


Figure 2. Long Short-Term Memory Cell

Gated recurrent unit (GRU) (Cho *et al.*, 2014), like LSTM, aims to solve the vanishing gradient problem encountered while using SRNN (Figure 3). Its structure allows long-term dependencies to be captured from long sequences of data, retaining information from earlier stages. Compared with LSTM, GRU does not pass two different

states between the cells; rather than cell state and hidden state, GRU passes only the hidden state. Owing to its design, GRU can keep both short-term and long-term dependencies through a similar gating mechanism, which regulates the flow of information but with two gates instead of three—the update gate and the reset gate.

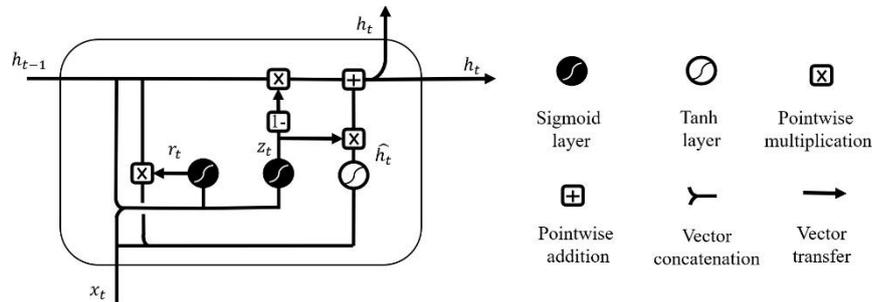


Figure 3. Gated Recurrent Unit Cell

$$z_t = \sigma(x_t U^z + h_{t-1} W^z) \quad (9)$$

$$r_t = \sigma(x_t U^r + h_{t-1} W^r) \quad (10)$$

$$\hat{h}_t = \tanh(x_t U^{\hat{h}} + (h_{t-1} \cdot r_t) W^{\hat{h}}) \quad (11)$$

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \hat{h}_t \quad (12)$$

Where " \cdot " represents element-wise multiplication.

Update gate (9) acts similarly to LSTM's input gate, deciding how much of the previous state information to keep. Multiplying the input and the hidden state with their associated synaptic weights and then summing and passing the result through the sigmoid function will bound the values between 0 and 1. This type of logic provides the

capability to filter the relevant data. Reset gate (10) is analogous to the LSTM’s forget gate and input gate, determining how much of the past knowledge to forget. The same formula is used for the update gate and reset gate vectors, but the weights vector, which is trainable, is different at the end. A sigmoid activation function is applied to the input x_t and previous hidden state h_{t-1} . If the values are close or equal to 1, then all the information coming from the previous state is relevant. Current memory content (11) is responsible for introducing non-linearity into the input and transforming it into a zero-mean. If the reset gate is equal to 0, then the previous hidden state information is filtered out, with only the current input, x_t , remaining. Finally, the vector h_t (12) containing the information of the current state that will be passed to the next state t+1 is calculated. The information passed to the next state is a

combination of both the current and the previous states. Owing to its design, containing fewer parameters than LSTM, GRU is more computationally efficient.

Method and Design

A key aspect of any data mining project is the selection of a data mining framework. In this way, the team members working on the project and the stakeholders have the appropriate levels of control and visibility. The cross-industry standard process for data mining (CRISP-DM) is one of these standards that are widely distributed across various industries. In this study, an instance of this standard proposed by IBM was used (IBM, 2015). The reason this standard was selected is that IBM made the support document based on their previous experience, being at the same time a living document (Figure 4).

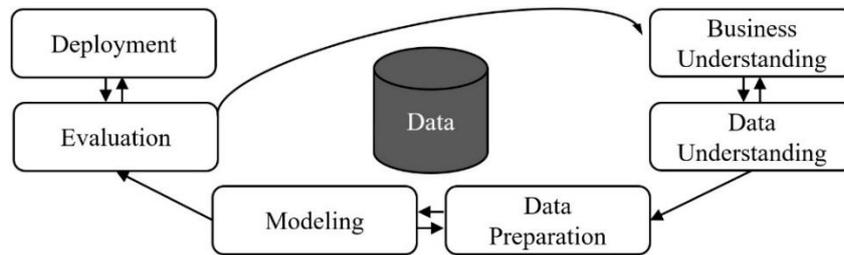


Figure 4. CRISP-DM Phases

The six-phase standard consists of business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The feedback between different phases ensures that the outcome aligns with expectations and the availability of resources. CRISP-DM’s first two phases-business understanding and data understanding-are crucial in determining whether the data mining project’s scope can be fulfilled using existing data and business understanding. Nevertheless, the probability of encountering blocking points during the other four phases remains. The data used for this study were made available by Open Power System Data (OPSD, 2020). OPSD is a data platform dedicated to electricity system research. The data are collected, checked, processed, and published free of

charge. OPSD’s goal is to avoid redundant work and to improve data quality with the aim of supporting researchers in their work. The initial raw data came from two different data sets. The first data set contains data relating to solar energy production while the second contains weather data. In the weather data set, the following variables were available: timestamp, v1 = wind speed measured at height h1; v2 = wind speed measured at height h2; v_50m = wind speed measured 50 meters above ground; h0 = roughness length; SWTDN = total top-of-the-atmosphere horizontal radiation; SWGDN = total ground horizontal radiation; T = temperature 2 meters above ground; rho = air density at surface; p = air pressure at surface

Table 1

Correlation Matrix

CM	v1	v2	v_50m	h1	h2	z0	SWTDN	SWGDN	T	rho	p
v1	1	0.99	0.90	0.37	0.37	0.38	0.20	0.15	0.17	0.15	0.23
v2	0.99	1	0.95	0.38	0.38	0.39	0.11	0.06	0.20	0.16	0.24
v_50m	0.90	0.95	1	0.40	0.40	0.40	0.16	0.18	0.27	0.18	0.24
h1	0.37	0.38	0.40	1	1	1	0.27	0.28	0.79	0.71	0.17
h2	0.37	0.38	0.40	1	1	1	0.27	0.28	0.79	0.71	0.17
z0	0.38	0.39	0.40	1	1	1	0.27	0.29	0.79	0.72	0.16
SWTDN	0.20	0.11	0.16	0.27	0.27	0.27	1	0.97	0.54	0.41	0.03
SWGDN	0.15	0.06	0.18	0.28	0.28	0.29	0.97	1	0.56	0.41	0.03
T	0.17	0.20	0.27	0.79	0.79	0.79	0.54	0.56	1	0.94	0.01
rho	0.15	0.16	0.18	0.71	0.71	0.72	0.41	0.41	0.94	1	0.29
p	0.23	0.24	0.24	0.17	0.17	0.16	0.03	0.03	0.01	0.29	1

During the data understanding step, as anticipated, a strong correlation was observed between several variables (Table 1). In the case of correlated variables, those with a correlation coefficient higher than 0.9 were removed. The timestamp has a standard YYYY-MM-DDTHH:MM:SSZ

format, from which information relating to year, month, day, hour, minute, and second can be extracted. Since the timestamp is created on an hourly basis, this is the maximum level of discretization. Categorical variables such as month, day, and hour were encoded.

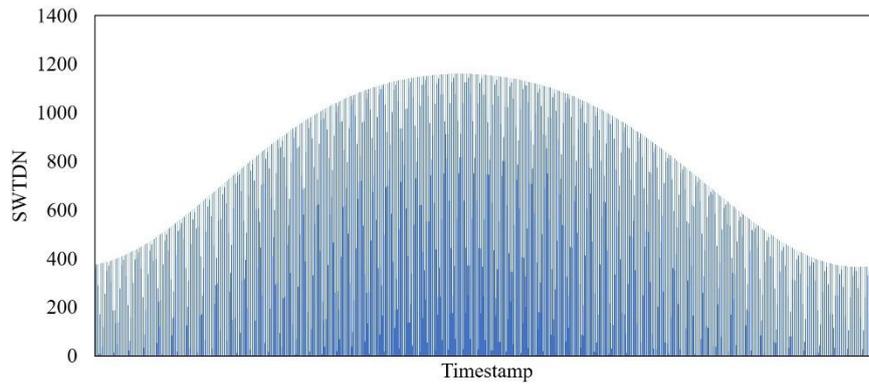


Figure 5. Total Top-of-the-Atmosphere Horizontal Radiation [W/m^2]

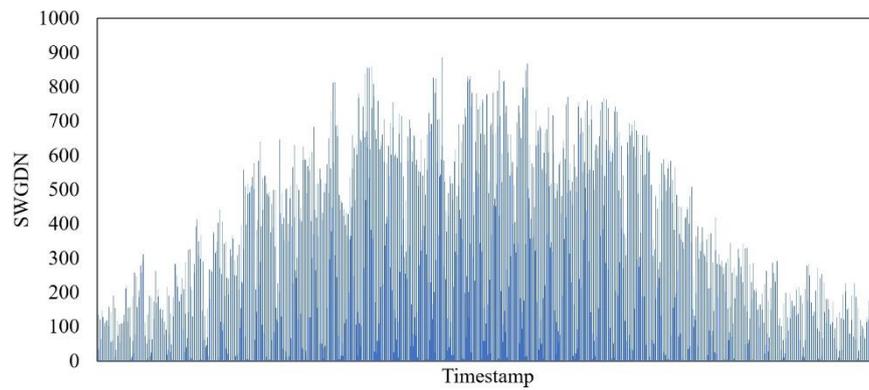


Figure 6. Total Ground Horizontal Radiation [W/m^2]

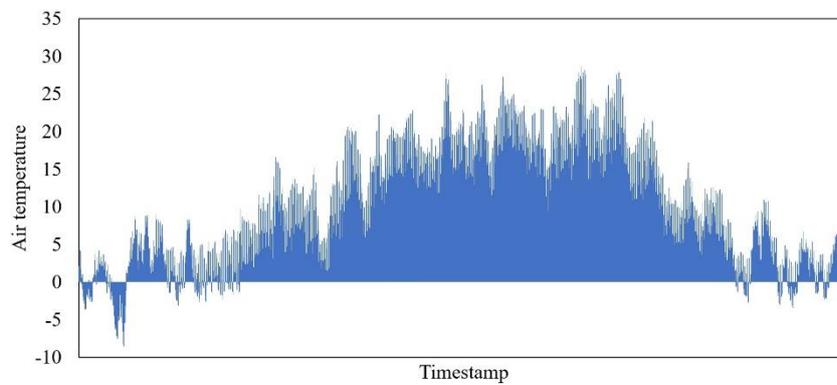


Figure 7. Air Temperature [C]

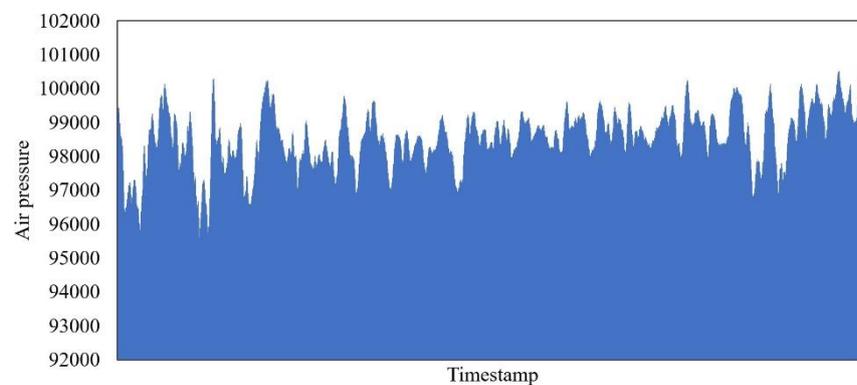


Figure 8. Air Pressure [Pa]

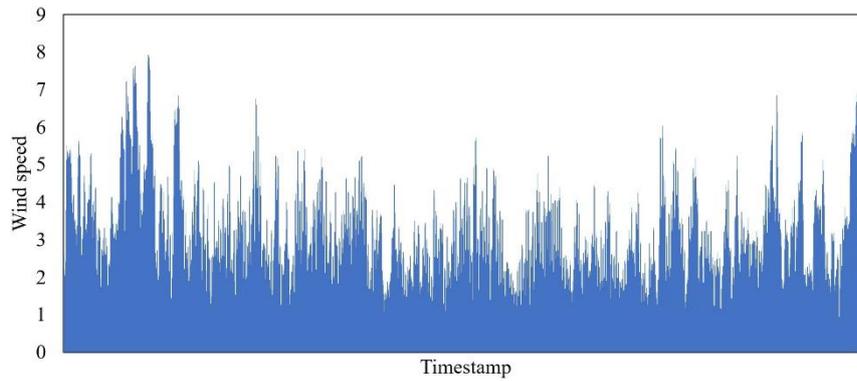


Figure 9. Wind Speed [m/s]

Figures 5 to 9 present a visualization of the main features used in the model. The source of the energy production volatility is clear, since the weather data impacting the output vary over time. The top-of-the-atmosphere horizontal radiation peaks in the summer, with values up to three times greater than those in winter. The ground horizontal radiation has peaks in the summer, but the amplitude is up to nine times greater than in the winter. This insight is important in that it is expected to show volatility across a 24-hour cycle but is less straightforward on a yearly cycle. Air pressure (Figure 8) and wind speed (Figure 9) do not show seasonality. Most of the study time was spent on data quality and quantity assessment, variable selection, and feature engineering. At the end of CRISP-DM’s data preparation step, 72 independent variables were considered for modeling. Prior to starting the modeling phase, the metrics used for model evaluation had to be decided. Root mean squared error (RMSE) and mean absolute error (MAE) facilitate the comparison of different models built on the same data set, both being measures of the accuracy of the forecast and scale-dependent. Coefficient of determination (R^2) is a measure of the variance in the dependent variable explained by the model.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{13}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \tag{14}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{15}$$

Results

Returning to the scope of this research, SRNN, LSTM, and GRU capabilities were assessed to determine which is more suited to the proposed use case (Table 2). The initial data set was divided into two sub-sets: a training data sub-set, representing 90 % of the initial data set, and a holdout data sub-set, representing the remaining 10 %. After several tests, the baseline was set.

Table 2

LSTM, GRU, and SRNN Baseline

Model	Parameters	RMSE	MAE	R2
SRNN_1	hl=2, nn=144, e=150, opt=Adam, batch=2048, wi=Xavier uniform, lr=0.001	453.2	276.3	0.9643
LSTM_1	hl=2, nn=144, e=50, opt=Adam, batch=2048, wi=Xavier uniform, lr=0.001	931.6	481.8	0.8490
GRU_1	hl=2, nn=144, e=100, opt= Adam, batch=2048, wi=Xavier uniform, lr=0.001	393.5	270.1	0.9731

Note: hl = number of hidden layers; nn = number of neurons on each hidden layer; e = number of epochs; opt = optimization algorithm, wi = weights initialization; lr = learning rate.

The third layer of neurons was added during the initial tests for all three main models, but no improvement in the forecast capability was observed. The same assessment was performed for the number of neurons on the hidden layers. Neither fewer nor more neurons on the hidden layers was associated with improvement. Moreover, it is known that increasing the network dimension will lead to an increase in the resources spent on training.

The next step in the search for the optimal hyperparameter combination was to determine the mini-batch size. For a smaller mini-batch size, LSTM showed an increase in R2 from 0.8490 to 0.9244, while GRU showed an increase from 0.9731 to 0.9826. No significant increase

was noted for the SRNN model. Initially, the Adam optimizer was used. To understand the impact of the optimizer in the model’s generalization performance, Adagrad (Lydia & Francis, 2019) and RMSProp (Hinton *et al.*, n.d.) were tested. RMSProp did not enhance any of the models’ performances, while Adagrad enhanced LSTM’s and GRU’s performances. Xavier uniform (Glorot & Bengio, 2010) remains the best method for weight initialization.

LSTM, GRU, and SRNN after Optimizer, Weight Initialization, and Batch Optimization

Model	Parameters	RMSE	MAE	R2
LSTM_2	hl=2, nn=144, e=50, opt= Adagrad, batch=128, wi=Xavier uniform, lr=0.001	396.3	222.9	0.9727
GRU_2	hl=2, nn=144, e=100, opt= Adagrad, batch=512, wi=Xavier uniform, lr=0.001	289.7	151.7	0.9854
SRNN_2	hl=2, nn=144, e=150, opt=Adam, batch=1024, wi=Xavier uniform, lr=0.001	450.1	275.2	0.9648

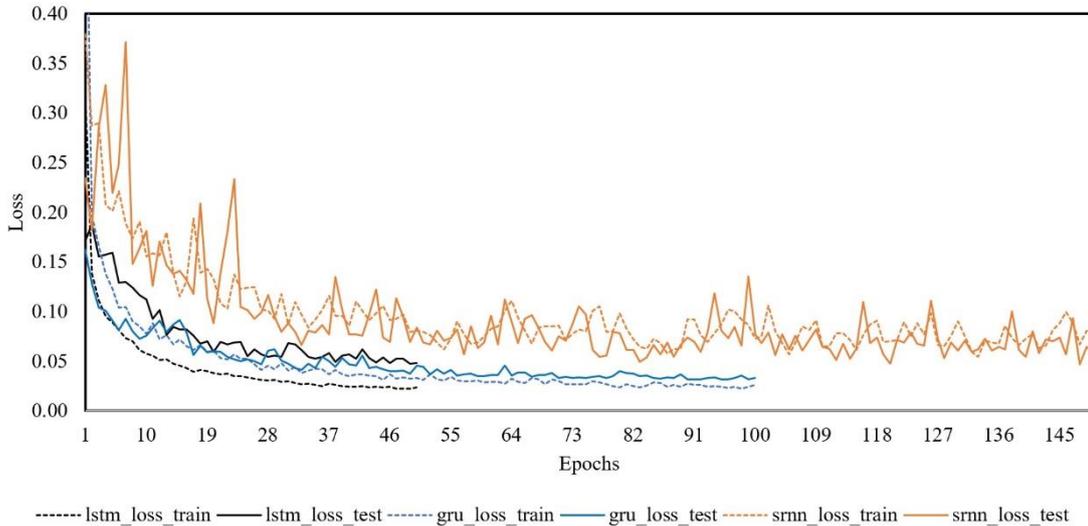


Figure 10. Learning Curves of LSTM, GRU and SRNN

Generally, learning curves represent a plot showing learning improvement over time (Figure 10). The dotted line represents the training learning curve, and the continuous line represents the validation learning curve: the smaller the loss, the better the forecast. The training learning curve evaluates how well the model generalizes the training data set, while the validation curve provides information about the generalization capability over holdout data set. Lowering the learning rate of the LSTM models has no impact on smoothing the learning curves. Both LSTM and SRNN learning curves show signs of possible overfitting or

underfitting. However, although these two unwanted behaviours may be present, the models’ overall performances are good. GRU not only performs better than LSTM and SRNN, but also has the smoothest learning curves and an almost equal capability for generalization over both training and holdout data, leading to the conclusion that is the best model for this use case. The first two days of production in the holdout data set were selected for visualization purposes (Figure 11). All three models have good forecast capabilities, and the predicted values are in sync with the actual values.

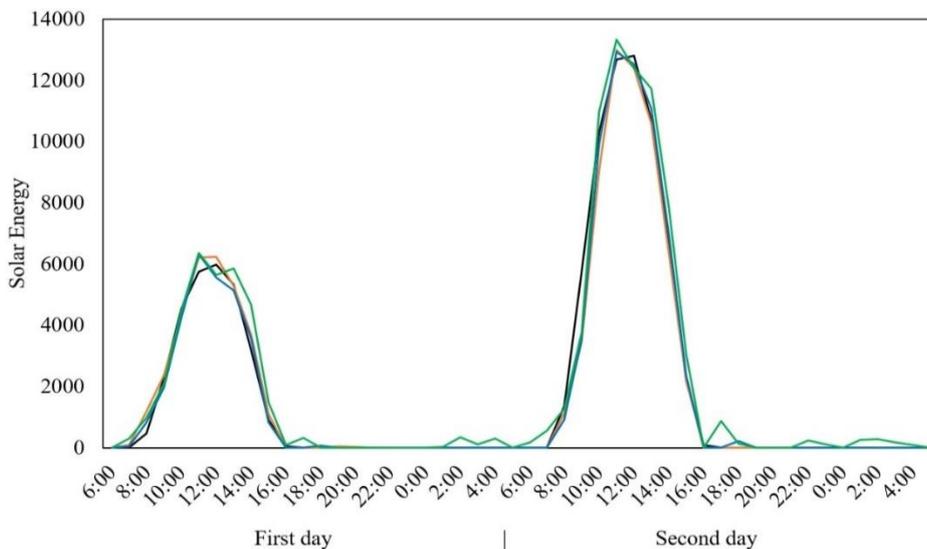


Figure 11. Solar Energy Forecast in the First 48 Hours [W]

In addition to the metrics mentioned above for the final model configurations, training time was analyzed. All models were run on a Dell Precision 7350 laptop equipped

with 2.5 GHz Intel Core i5-8400H CPU, Nvidia Quadro P2000, 32 GB RAM, Windows 10, and Python 3.6.10.

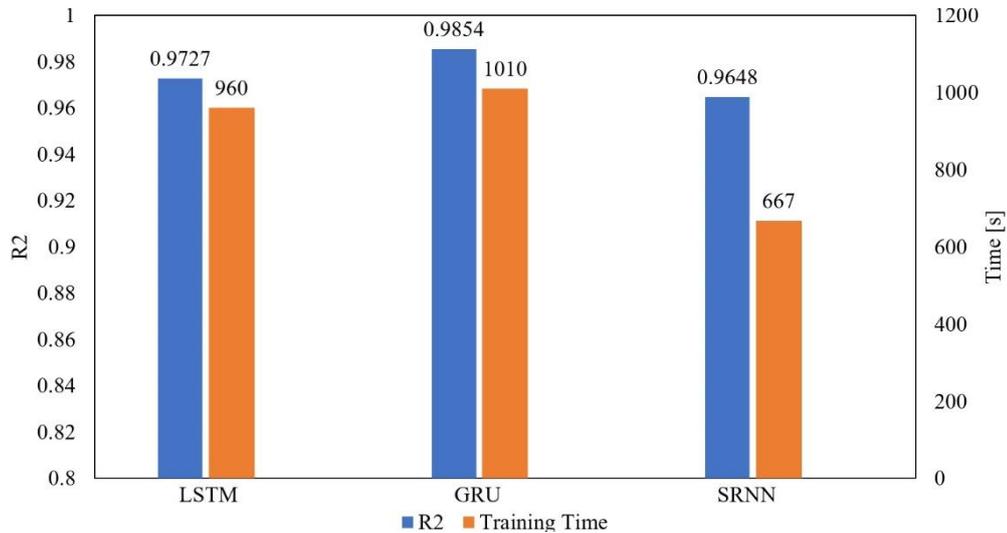


Figure 12. R2 and Ttraining Time Comparison

As of now, the training time was a matter of selected architecture and number of epochs. In terms of time per epoch, LSTM requires 19 seconds, GRU requires 10 seconds, and SRNN requires just 4 seconds. This was expected and is in line with RNN fundamentals. Owing to the different number of epochs needed for training, GRU’s training time is the largest while SRNN’s is the smallest. Although the R2 values are close, RMSE and MAE values show that the differences in terms of absolute values are important. For example, SRNN’s MAE is 80 % larger than GRU’s MAE, and LSTM’s MAE is 46 % larger than the same GRU’s MAE. Considering a pragmatic approach where training time is not a constraint, GRU is the best option for this use case. The parameters that added no improvement to the performance were not considered in the final discussion.

Conclusions

This paper proposed three different RNN types for PV energy production forecasts. Finally, the most powerful model of each considered model type has been fully described. SRNN, LSTM, and GRU were assessed in terms of complexity, performance, and training time. The baseline models proved from the beginning that these three architectures can provide reasonable performance. After hyperparameter tuning, both LSTM’s and GRU’s performances were improved, while SRNN’s performance remained the same. Accurate forecasting of energy production is among the key open points that need to be resolved to promote solar energy adoption. Moreover, obtaining close forecast performances for all three models is a means of validating that the results are meaningful rather than a fortunate coincidence. The proposed GRU architecture can model complex phenomena, such as solar energy production. Long-term dependencies were captured through the simple yet sophisticated gating system. The use of real-world data ensures that this study’s findings can be directly applied to installations in production, showing the

starting point of the optimization and beyond to ensure the best result. Several intermediate steps remain to be taken before the final state can be reached. Renewable energy production systems must become integrated into existing energy production and transport systems. Owing to its unpredictable nature, wind energy cannot adequately fulfill end users’ demands. Concerns relating to renewable energy in general and solar energy in particular will intensify in the coming years, at least in the European Union context. The progress made worldwide is transforming this into a global initiative, despite the fact that the efforts have been made in different workstreams. The European Green Deal creates opportunities for proposing, financing, and implementing courageous innovations for the common good on both the environmental and economic levels. It is clear that predictability leads to optimization and optimization leads to cost reduction. The cost reduction is made initially where the energy is produced but is further propagated downstream to the customer, with all parties involved taking advantage of it. Another particularly salient aspect is the rapid decrease in production costs, which will drive the increase in the installed capacity, owing to its attractiveness to business developers. The key to integrating this and other green technologies together with classical technologies is to put in place a predictive system: by anticipating when the gaps in energy delivery will appear, a proactive approach on the part of production and distribution companies will lead them to take the necessary actions to fill such gaps using classical energy resources. By integrating a larger share of cheap energy, the price for the end customer will be reduced proportionally with the share. However, several limitations and unresolved issues must be considered in real-world scenarios. First, the model’s performance relies on the input data, namely, weather data, which already include a degree of uncertainty. This uncertainty is further propagated into the model, thereby impacting the performance. Second, equipment wear is not considered in the current setup. After deployment, the performance of assets will deteriorate along

with the capability of converting solar irradiance into electricity. Third, no fully end-to-end implementation accounts have been published, and so unexpected challenges may yet emerge.

Author Contributions

Adrian-Nicolae Buturache conducted the research to identify the real-world use case and the state of the art in addition to data gathering, data modeling, and results interpretation. Stelian Stancu was responsible for the article structure, supervision, and advising during the research.

Disclosure Statement

The authors claim that there is no conflict of interest, any competing financial, professional, or personal interests from others. None of the research in the paper is under consideration for publication elsewhere or has been published.

References

- Abdel-Nasser, M., & Mahmoud, K. (2017). Accurate photovoltaic power forecasting models using deep LSTM-RNN. *Neural Computing and Applications*, 31, 2727–2740. <https://doi.org/10.1007/s00521-017-3225-z>.
- Alessandrini, S., Davo, F., Sperati, S., Benini, M., & Delle Monache, L. (2014). Comparison of the economic impact of different wind power forecast systems for producers. *Advances in Science and Research*, 11, 49–53. <https://doi.org/10.5194/asr-11-49-2014>.
- Bacher, P., Madsen, H., & Nilsen, H. A. (2009). Online short-term solar power forecasting. *Solar Energy*, 83(10), 1772–1783. <https://doi.org/10.1016/j.solener.2009.05.016>.
- Barthelmie, R. J., Murray, F., & Pryor, S. C. (2008). The economic benefit of short-term forecasting for wind energy in the UK electricity market. *Energy Policy*, 36(5), 1687–1696. <https://doi.org/10.1016/j.enpol.2008.01.027>
- Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2), 157–166. <https://doi.org/10.1109/72.279181>.
- Bouzerdoum, M., Mellit, A., & Pavan, A. M. (2013). A hybrid model (SARIMA-SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant. *Solar Energy*, 98(Part C), 226–235. <https://doi.org/10.1016/j.solener.2013.10.002>.
- Botterud, A., Zhou, Z., Wang, J., Bessa, R. J., Keko, H., Sumaili, J., Miranda, V. (2012). Wind Power Trading Under Uncertainty in LMP Markets. *IEEE Transactions on Power Systems*, 27(2), 894–903. <https://doi.org/10.1109/TPWRS.2011.2170442>.
- Carriere, T., & Kariniotakis, G. (2019). An Integrated Approach for Value-Oriented Energy Forecasting and Data-Driven Decision-Making Application to Renewable Energy Trading. *IEEE Transactions on Smart Grid*, 10(6), 6933–6944. <https://doi.org/10.1109/TSG.2019.2914379>.
- Chen, B., Lin, P., Lai, Y., Cheng, S., Chen, Z., & Wu, L. (2020). Very-Short-Term Power Prediction for PV Power Plants Using a Simple and Effective RCC-LSTM Model Based on Short Term Multivariate Historical Datasets. *Electronics*, 9(2), 289. <https://doi.org/10.3390/electronics9020289>.
- Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* 1724–1734. Association for Computational Linguistics. <https://doi.org/10.3115/v1/D14-1179>.
- Crisosto, C., Hofmann, M., Mubarak, R., & Seckmeyer, G. (2018). One-Hour Prediction of the Global Solar Irradiance from All-Sky Images Using Artificial Neural Networks. *Energies*, 11(11), 2906. <https://doi.org/10.3390/en11112906>.
- De, V., Teo, T. T., Woo, W. L., & Logenthiran, T. (2018). Photovoltaic Power Forecasting using LSTM on Limited Dataset. *2018 IEEE Innovative Smart Grid Technologies - Asia* 710–715. IEEE. <https://doi.org/10.1109/ISGT-Asia.2018.8467934>.
- European Commission. (2017). *Clean energy for all Europeans package*. https://ec.europa.eu/energy/topics/energy-strategy/clean-energy-all-europeans_en
- European Parliament. (2018, December 11). *Directive (EU) 2018/2001 of the European Parliament and of the Council of 11 December 2018 on the promotion of the use of energy from renewable sources*. <http://data.europa.eu/eli/dir/2018/2001/oj>
- Gao, M., Li, J., Hong, F., & Long, D. (2019). Short-Term Forecasting of Power Production in a Large-Scale Photovoltaic Plant Based on LSTM. *Applied Sciences*, 9(15), 3192. <https://doi.org/10.3390/app9153192>.
- Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Journal of Machine Learning Research*, 9, 249–256.
- Granger, C. W. J., & Pesaran, M. H. (2000). A Decision Theoretic Approach to Forecast Evaluation. *Statistics and Finance*, 261–278. https://doi.org/10.1142/9781848160156_0015.

- Gyalai-Korpos, M., Zentko, L., Hegyfalvi, C., Detzky, G., Tildy, P., Hegedusne Baranyai, N., Pinter, G., & Zsiboracs, H. (2020). The Role of Electricity Balancing and Storage: Developing Input Parameters for the European Calculator for Concept Modeling. *Sustainability*, *12*, 811. <https://doi.org/10.3390/su12030811>
- Halldorsson, A., Gremyr, I., Winter, A., & Taghahvi, N. (2018). Lean Energy: Turning Sustainable Development into Organizational Renewal. *Sustainability*, *10*, 4464. <https://doi.org/10.3390/su10124464>
- Hammer, A., Heinemann, D., Lorenz, E., & Luckehe, B. (1999). Short-term forecasting of solar radiation: a statistical approach using satellite data. *Solar Energy*, *67*(1), 139–150. [https://doi.org/10.1016/S0038-092X\(00\)00038-4](https://doi.org/10.1016/S0038-092X(00)00038-4)
- Hanna, R., Kleissl, J., Nottrott, A., & Ferry, M. (2014). Energy dispatch schedule optimization for demand charge reduction using a photovoltaic-battery storage system with solar forecasting. *Solar Energy*, *103*, 269–287 <https://doi.org/10.1016/j.solener.2014.02.020>.
- Hinton, G., Srivastava, N., & Swersky, K. (n.d.). *Lecture 6a: Overview of mini-batch gradient descent*. https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf.
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-term Memory. *Neural Computation*, *9*(8), 1735–80. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- IBM. (2015). *IBM SPSS Modeler CRISP-DM Guide*. <ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/18.0/en/ModelerCRISPDM.pdf> accessed at 25.06. 2020
- Jarvis, A., Leedal, D., & Hewitt, N. C. (2012). Climate-society feedbacks and the avoidance of dangerous climate change. *Nature Climate Change*, *2*, 668–671. <https://doi.org/10.1038/nclimate1586>.
- Kahan, A. (2019, September 24). *EIA projects nearly 50% increase in world energy usage by 2050, led by growth in Asia*. U.S. Energy Information Administration <https://www.eia.gov/todayinenergy/detail.php?id=41433>
- Kingma, D. P., & Ba, J. (2014). *Adam: A Method for Stochastic Optimization*. Paper presented at the International Conference on Learning Representations, San Diego, CA.
- Kraas, B., Schroedter-Homscheidt, M., & Madlener, R. (2013). Economic merits of a state-of-the-art concentrating solar power forecasting system for participation in the Spanish electricity market. *Solar Energy*, *93*, 244–255. <https://doi.org/10.1016/j.solener.2013.04.012>.
- Lee, D., & Kim, K. (2019). Recurrent Neural Network-Based Hourly Prediction of Photovoltaic Power Output Using Meteorological Information. *Energies*, *12*(2), 215. <https://doi.org/10.3390/en12020215>.
- Lee D. H., Kim, M. S., Roh, J. H., Yang, J. P., & Park, J. B. (2019). Forecasting of Electric Vehicles Charging Pattern Using Bayesians method with the Convolution. *IFAC-PapersOnLine*, *52*(4), 413–418. <https://doi.org/10.1016/j.ifacol.2019.08.245>.
- Li, G., Wang, H., Zhang, S., Xin, J., & Liu, H. (2019). Recurrent Neural Networks Based Photovoltaic Power Forecasting Approach. *Energies*, *12*(13), 2538. <https://doi.org/10.3390/en12132538>.
- Lorenz, E., Hurka, J., Heinemann, D., & Beyer, H. (2009). Irradiance forecasting for the power prediction of grid-connected photovoltaic systems, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *2*(1), 2–10. <https://doi.org/10.1109/JSTARS.2009.2020300>.
- Lydia, A. A., & Francis, F. S. (2019). Adagrad-An Optimizer for Stochastic Gradient Descent. *International Journal of Information and Computing Science*, *6*(5), 566–568.
- Open Power System Data. (2020). <https://open-power-system-data.org/>
- Pascanu, R., Gulcehre, C., Cho, K., & Bengio, Y. (2014). *How to construct deep recurrent neural networks*. Paper presented at Proceedings of the Second International Conference on Learning Representations (ICLR 2014), Banff, Canada.
- Poudel, P., & Jang, B. (2017). *Solar Power Prediction Using Deep Learning Technique*. Paper presented at Advance in Future Generation Communication and Networking, Jeju Island, South Korea. <https://doi.org/10.14257/astl.2017.146.26>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagation errors, *Nature*, *322*, 533–536. <https://doi.org/10.1038/323533a0>
- SolarPower Europe. (2019). *EU Market Outlook for Solar Power / 2019–2023*. https://www.solarpowereurope.org/wp-content/uploads/2019/12/SolarPower-Europe_EU-Market-Outlook-for-Solar-Power-2019-2023_.pdf?cf_id=5387
- Uniejewski, B., Marcjasz, G., & Weron, R. (2019). Understanding intraday electricity markets: Variable selection and very short-term price forecasting using LASSO. *International Journal of Forecasting*, *35*(4), 1533–1547. <https://doi.org/10.1016/j.ijforecast.2019.02.001>.
- 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*(4), 38–46. <https://doi.org/10.17775/CSEEJPES.2015.00046>.
- Yona, A., Senjyu, T., & Funabashi, T. (2007). Application of Recurrent Neural Network to Short-Term-Ahead Generating Power Forecasting for Photovoltaic System. *2007 IEEE Power Engineering Society General Meeting*, 1–6. IEEE. <https://doi.org/10.1109/PES.2007.386072>.

Zhang, J., Chi, Y., & Xiao, L. (2018). Solar Power Generation Forecast Based on LSTM. *2018 IEEE 9th International Conference on Software Engineering and Service Science*, 869–872. IEEE. <https://doi.org/10.1109/ICSESS.2018.8663788>.

Authors' biographies

Adrian-Nicolae Buturache is an aerospace engineering graduate with experience in aerospace propulsion systems, automotive advanced manufacturing engineering, and renewable energy machine learning-based predictive analytics. Currently Ph.D. student at Economic Cybernetics and Statistics Doctoral School of Bucharest University of Economic Studies with a keen interest in artificial neural networks and renewable energy.

Stelian Stancu is a professor with an academic career spanning almost 30 years. He received his Ph.D. in Economic Cybernetics and Statistics. He is a member and Vice Head of the Informatics and Economic Cybernetics Department, Bucharest University of Economic Studies, teaching courses as Processes and Business Modelling, Artificial Neural Networks Quantitative Economics.

The article has been reviewed.
Received in March 2021; accepted in October 2021.



This article is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 (CC BY 4.0) License (<http://creativecommons.org/licenses/by/4.0/>)