Engineering and Technology Journal e-ISSN: 2456-3358

Volume 09 Issue 03 March-2024, Page No.- 3619-3622

DOI: 10.47191/etj/v9i03.09, I.F. – 8.227

Estimation of Solar Power Generation Through the use of Effective Forecasting Algorithms

Parastou BEHGOUY [1], Osman YILDIRIM [2]

[1]MS Student, Electrical and Electronic Engineering, Istanbul Aydin University, Istanbul, Turkey, https://orcid.org/0000-0001-9653-2850

[2]Professor, Electrical and Electronic Engineering, Istanbul Aydin University, Istanbul, Turkey, https://orcid.org/0000-0002-8900-3050

ABSTRACT: The unpredictable nature of solar power generation poses significant challenges to energy management, particularly in power grids with high solar penetration, leading to potential imbalances between supply and demand. This has increased interest in developing accurate forecasting methods for photovoltaic power generation. This study focuses on short-term forecasting of solar power generation using MATLAB, based on historical generation trends through three different hybrid models. It uses a multilayer perceptron neural network as the forecasting tool, which is optimized through various algorithms to fine-tune the network's hyperparameters, such as the weights and biases of the input and hidden layers, for optimal performance. The Root Mean Square Error is used as the objective metric to minimize the discrepancy between actual and predicted values. The models use Particle Swarm Optimization (PSO), Imperialistic Competitive Algorithm, and Genetic Algorithm for network optimization. This research analyzes PV generation data from Belgium in July and August 2022, taken at 15-minute intervals, and evaluates model performance using six different error metrics, including Mean Absolute Error. The results show the accuracy and reliability in forecasting solar power generation time series.

KEYWORDS: Solar Energy, Neural Network, Energy Forecasting, Algorithms.

I. INTRODUCTION

In general, there are many algorithms for using big data for prediction. While multilayer perceptions (MLPs) are powerful for modeling and classification tasks due to their ability to learn from data, particle swarm optimization (PSO), the imperialist competitive algorithm (ICA), and genetic algorithm (GA) are optimization algorithms, each with unique strategies for exploring and exploiting a search space to find optimal solutions. PSO emphasizes exploration inspired by social behavior, ICA uses socio-political dynamics for competitive optimization, and GA relies on evolutionary processes. The choice between these methods depends on the specific problem, the nature of the search space, and the desired balance between exploration and exploitation.

MLP neural network is a class of feedforward artificial neural networks (ANNs) that consists of at least three layers of nodes: an input layer, one or more hidden layers, and an output layer. Unlike a single-layer perceptron, MLPs can capture nonlinear relationships between input and output data, making them suitable for solving complex problems where linear models fall short. Each node, except for the input nodes, is a neuron that uses a nonlinear activation function. MLP uses a supervised learning technique called backpropagation to train the network. During the training phase, the network learns by adjusting the weights of the connections between neurons based on the error between the predicted output and the actual output. MLP neural networks are widely used in various applications such as image recognition, speech recognition, and natural language processing due to their ability to learn and model non-linear and complex relationships.

PSO is a computational method that optimizes a problem by iteratively trying to improve a candidate solution with respect to a given quality measure. It simulates the social behavior of birds in a flock or fish in a school. Developed in 1995 by James Kennedy and Russell Eberhart, PSO is inspired by the social and cooperative aspects of biological populations. The algorithm initializes with a population of random solutions, called particles. Each particle represents a point in the multidimensional space of possible solutions to the optimization problem. These particles move through the solution space with velocities that are dynamically adjusted according to their own experience and that of their neighbors, with the goal of finding the optimal solution. Each particle keeps track of its coordinates in the solution space, which are associated with the best solution (fitness) it has achieved so far. This value is called pbest. Another "best" value tracked by the particle swarm optimizer is the global best (gbest), which is the best value obtained so far by any particle in the population. The potential solutions then move through the problem space by following the current optimal particles.

PSO does not require the gradient of the problem to be optimized, and therefore can be used on optimization problems that are irregular, noisy, changing over time, etc. Because of its simplicity and ability to converge quickly to a good solution (though not necessarily the global optimum), PSO has been successfully applied in various fields, including engineering, economics, and computer science, to solve complex optimization problems.

ICA is an optimization algorithm inspired by the socio-political process of imperialism and colonialism. Introduced by Esmaeil Atashpaz-Gargari and Caro Lucas in 2007, ICA is a socio-politically motivated global search strategy that mimics the imperialistic competition among countries to effectively solve optimization problems. In ICA, potential solutions to the optimization problem are conceptualized as countries. These countries are divided into two categories: the imperialists (the best solutions) and colonies (the rest of the solutions). The algorithm begins with an initial population of countries, which are then divided into imperialists and colonies based on their strength or fitness, with the stronger countries becoming imperialists and the weaker ones becoming colonies. Each empire consists of an imperialist country and its colonies. The Imperialist Competitive Algorithm is notable for its ability to find global optima by effectively balancing exploration and exploitation of the search space. ICA has been successfully applied to various fields, including engineering optimization, scheduling, and classification problems, demonstrating its versatility and efficiency in finding solutions to complex optimization challenges (Atashpaz-Gargari and Lucas, 2007).

A Genetic Algorithm (GA) is developed by John Holland and his colleagues at the University of Michigan in the 1960s, GAs belong to the larger class of Evolutionary Algorithms (EA), which generate solutions to optimization and search problems using techniques inspired by natural evolution, such as inheritance, mutation, selection, and crossover. The process of a GA typically involves a population of individuals represented by chromosomes (i.e., strings of binary values, although other representations are common). Each individual in the population represents a possible solution to the problem at hand, and its fitness is evaluated by a fitness function designed specifically for the problem. GAs are widely used in optimization problems where the search space is large, complex, or poorly understood. They are particularly useful when traditional optimization techniques fail to produce a satisfactory solution due to the complexity of the problem or the presence of multiple local optima. Applications of GAs span diverse fields, including engineering design, economics and marketing, computer science, and artificial intelligence, demonstrating their versatility and effectiveness in solving a wide range of problems.

Delgadillo-Ruiz et.al. (2015) pointed out the development and application of a method for reconstructing missing temperature data in weather stations, specifically within the semi-desert Rio Queretaro basin. This method significantly reduces the uncertainty of temperature data, which is crucial for accurate evapotranspiration calculation. By utilizing a three-step process involving sinusoidal type regression, linear regression, and estimation of missing temperature values, the research not only enhances the precision of potential evapotranspiration estimates but also contributes to improving the accuracy of water balance calculations in the studied hydrological basin. This approach demonstrates a clear advancement in the field of hydrological modeling by providing a robust technique for filling gaps in climatological data, thereby ensuring more reliable water resource management and planning.

The research paper by Chicco, et.al. (2021) argues that the coefficient of determination (R-squared) is a more informative metric than other commonly used error metrics (SMAPE, MAE, MAPE, MSE, and RMSE) when evaluating regression analysis. It highlights the limitations of these traditional metrics, noting their inability to provide a clear indication of model performance without additional context. In contrast, R-squared provides a bounded and more interpretable measure that can indicate both the proportion of variance explained by a model and the model's performance relative to a baseline. Through a detailed comparison and application of these metrics in various scenarios, including synthetic scenarios, including synthetic use cases and real medical datasets, the paper demonstrates that R-squared not only provides a more nuanced understanding of model effectiveness, but also promotes a more standardized approach to regression evaluation across scientific domains. The authors suggest that R-squared's ability to reflect the majority of correct predictions and its sensitivity to the distribution of ground truth values make it the preferred choice for evaluating regression models in any scientific field. regression models in any scientific domain.

Keshtegar et.al. (2019) covers an innovative methodology that leverages polynomial chaos expansions (PCE) to predict dissolved oxygen (DO) concentration in rivers. This approach is distinct from traditional methods as it utilizes four specific water quality variables as predictors: water temperature, turbidity, pH, and specific conductance. By comparing this novel PCE model against multilayer perceptron neural network (MLPNN), MLPNN optimized with particle swarm optimization (MLPNN_PSO), and standard multiple linear regression (MLR) models, the study showcases the superior predicting performance of the PCE model. This is evident through various assessment metrics, including the coefficient of correlation (R), Nash–Sutcliffe efficiency (NSE), root mean squared error (RMSE), and mean absolute error (MAE), where the PCE model outperformed the other models across multiple stations operated by the United States Geological Survey (USGS). The paper demonstrates that employing PCE with these four water quality variables leads to the best model performances in predicting DO concentration at all stations, making it a valuable alternative to direct measurement methods. This contribution to the field of environmental earth

sciences represents a significant advancement in accurately modeling water quality in freshwater ecosystems, providing a robust tool for better water resource management and planning.

II. METHODOLOGY

This research consists in the application and refinement of hybrid models for short-term photovoltaic (PV) power generation forecasting using MATLAB. The research contributes to the field by providing a detailed comparison of the performance of these hybrid models in forecasting PV power generation, using a dataset of 15-minute intervals from July and August 2022 in Belgium. It successfully demonstrates that the MLP neural network optimized with PSO outperforms those optimized with ICA and GA in terms of accuracy, as shown by various evaluation metrics, including MSE, RMSE, MAE, MBE, R^2, and NRMSE. In particular, three hybrid models are presented that use Multilayer Perceptron (MLP) neural networks optimized with different algorithms: Particle Swarm Optimization (PSO), Imperialistic Competitive Algorithm (ICA), and Genetic Algorithm (GA). The novelty of this study lies in its comprehensive approach to optimizing the MLP neural network's hyperparameters - such as the input and hidden layer weight matrices and the bias vector - using these three different optimization algorithms, with the goal of minimizing the prediction errors represented by the Root Mean Square Error (RMSE).

MLP neural network can be expressed by the following equations:

$$f(X) = b_j + \sum_{i=1}^{n}(w_{ji}\, x_i) \quad (1)$$

$$\Phi = \frac{1}{1+exp(-f(X))} \quad (2)$$

$$y(X) = b + \sum_{j=1}^{M}(w_j\, \Phi_j) \quad (3)$$

Where bj is the hidden layer bias, n number of input layer neurons, wji input layer weights, xi input data, b output bias, Φ activation function, M number of hidden layer neurons, and wj hidden layer weights (Keshtegar, Heddam, Hosseinabadi, 2019).

The implemented models have been evaluated by the following errors: MSE, RMSE, MAE, R2 (Chicco, Warrens, Jurman, 2021) and MBE (Bellido-Jiménez, Estévez, García-Marín, 2021). The bellow formulas express these errors.

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(X_i - Y_i)^2 \quad (4)$$

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(X_i - Y_i)^2} \quad (5)$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|X_i - Y_i| \quad (6)$$

$$MBE = \frac{1}{m}\sum_{i=1}^{m}X_i - Y_i \quad (7)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(X_i - Y_i)^2}{\sum_{i=1}^{m}(\bar{Y} - Y_i)^2} \quad (8)$$

Xi, Yi, and m are the actual and forecast amount and test data's number respectively.

III.FINDING

Mean Squared Error (MSE) is the average of the squared differences between the actual and predicted values. It quantifies the variance of the prediction errors. A lower MSE value indicates a better fit of the model to the data. MSE gives more weight to larger errors due to the squaring part of the formula.

Root Mean Squared Error (RMSE) is the square root of the mean squared error. It measures the standard deviation of the prediction errors. Like MSE, a lower RMSE value indicates a better fit. RMSE is in the same units as the response variable and is more interpretable in the context of the original data.

Mean Absolute Error (MAE) is the average of the absolute differences between the predicted and actual values. It quantifies the average magnitude of the prediction errors. A lower MAE indicates a better fit. Unlike MSE and RMSE, MAE does not give extra weight to larger errors, making it more robust to outliers.

R^2, also known as the coefficient of determination, measures the proportion of the variance in the dependent variable that is predictable from the independent variable(s). R^2 ranges from 0 to 1, where 0 means that the model does not explain any variability of the response data around its mean, and 1 means it explains all the variability. Higher R^2 values indicate a model that fits the data better. Each of these metrics offers insights into the model's performance, but they have their own strengths and limitations and should be considered together when evaluating model accuracy and fit.

In this study, MLP neural network is implemented as a fast operation model (Ke and Huang, 2020). By using these formulas, the results are shown in Table 1.

Table 1. The Error Values of the Hybrid Models

Model	MSE	RMSE	MAE	MBE	R^2
Multilayer Perceptions-Particle Swarm Optimization	0.130	0.361	0.208	0.028	0.921
Multilayer Perceptions-Imperialist Competitive Algorithm	0.153	0.391	0.240	0.043	0.912
Multilayer Perceptions-Genetic Algorithm	0.144	0.380	0.267	0.091	0.905

IV. CONCLUSIONS

The estimation of solar power is a critical issue for several key reasons, such as improved grid stability and reliability, efficient use of resources, improved integration of renewable energy, economic benefits, policy and planning support, environmental impact, operational efficiency and economic benefits, and so on.

Accurate forecasting helps maintain grid stability by predicting the variability of solar power generation. This is

critical for grid operators to effectively manage the balance between supply and demand to ensure a reliable supply of electricity. By accurately predicting solar power output, utilities can optimize the use of backup power sources, reducing the need for expensive, fast-reacting fossil fuel plants. This results in more efficient use of resources and can help minimize operating costs.

Effective forecasting algorithms facilitate the integration of higher levels of solar (and other renewable) energy into the power system. The ability to accurately predict solar power generation can lead to significant economic benefits. For example, it can help reduce the costs associated with imbalance charges in electricity markets. These charges occur when actual power generation does not match the forecasted amount, and accurate forecasting can help minimize such discrepancies. Accurate forecasts of solar power generation are also essential for policymakers and planners. They provide valuable insights for designing energy policies, planning future power system expansions, and making investment decisions that take into account the increasing role of renewable energy sources.

By enabling greater penetration of solar energy into the grid, effective forecasting helps reduce greenhouse gas emissions and dependence on fossil fuels. This supports global efforts to combat climate change and promote sustainable development. Finally, forecasting plays a critical role in the operational efficiency of power systems. It enables better planning and dispatching of power plants, which can lead to economic benefits for power producers and consumers alike by optimizing production costs and market operations.

This research on optimizing neural networks for PV power forecasting represents a significant contribution to renewable energy management, especially for grids with high solar penetration, where accurate forecasting is critical for balancing load and generation. By comparing the effectiveness of different optimization algorithms within a neural network framework, the study provides valuable insights into improving forecasting accuracy and efficiency in renewable energy systems.

REFERENCES

1. Atashpaz-Gargari, E. and Lucas, C. (2007). Imperialist Competitive Algorithm: An Algorithm for Optimization Inspired by Imperialistic Competition, Evolutionary Computation IEEE Congress.
2. Bellido-Jiménez, J.A., Estévez, J. and García-Marín, A.P. (2021). New machine learning approaches to improve reference evapotranspiration estimates using intra-daily temperature-based variables in a semi-arid region of Spain. Agricultural Water Management, 245, 106558.
3. Chicco, D., Warrens, M.J. and Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science, 5, 7, e623.
4. Delgadillo-Ruiz, E., Ventura-Ramos, Jr. E., Trinidad, J.G., Júnez-Ferreira, H.E., Bautista-Capetillo, C.F. and Delgadillo-Ruiz, O. (2015). Estimating Potential Evapotranspiration by Missing Temperature Data Reconstruction, Journal of Applied Mathematics, 154864, 10 pages, http://dx.doi.org/10.1155/2015/154864
5. Ke, K.C. and Huang, M.S. (2020). Quality prediction for injection molding by using a multilayer perceptron neural network. Polymers, 12(8), 1812.
6. Keshtegar, B., Heddam, S. and Hosseinabadi, H. (2019). The employment of polynomial chaos expansion approach for modeling dissolved oxygen concentration in river. Environmental Earth Sciences, 78, 1-8.