Combined Optimization and Regression Machine Learning for Solar Irradiation and Wind Speed Forecasting

Yahia Amoura[1,4][0000−0002−8811−0823] , Santiago Torres[4][0000−0002−3155−5039] , José Lima[1,3][0000−0001−7902−1207] , Ana I. Pereira[1,2][0000−0003−3803−2043]

[1] Research Centre in Digitalization and Intelligent Robotics (CeDRI), Instituto Politécnico de Bragança, Bragança, Portugal
[2] ALGORITMI Center, University of Minho, Braga, Portugal
[3] INESC TEC - INESC Technology and Science, Porto, Portugal
[4] University of Laguna, Spain
Email: {yahia,jllima,apereira}@ipb.pt, storres@ull.edu.es

Abstract. Prediction of solar irradiation and wind speed are essential for enhancing the renewable energy integration into the existing power system grids. However, the deficiencies caused to the network operations provided by their intermittent effects need to be investigated. Regarding reserves management, regulation, scheduling, and dispatching, the intermittency in power output become a challenge for the system operator. This had given the interest of researchers for developing techniques to predict wind speeds and solar irradiation over a large or short-range of temporal and spatial perspectives to accurately deal with the variable power output. Before, several statistical, and even physics, approaches have been applied for prediction. Nowadays, machine learning is widely applied to do it and especially regression models to assess them. Tuning these models is usually done following manual approaches by changing the minimum leaf size of a decision tree, or the box constraint of a support vector machine, for example, that can affect its performance. Instead of performing it manually, this paper proposes to combine optimization methods including the bayesian optimization, grid search, and random search with regression models to extract the best hyper parameters of the model. Finally, the results are compared with the manually tuned models. The Bayesian gives the best results in terms of extracting hyper-parameters by giving more accurate models.

Keywords: Renewable Energy· Forecasting · Machine Learning · Optimization · Wind Speed · Solar Irradiation.

1 Introduction

With significant population growth followed by rising demand, human energy consumption could become a major concern [1]. The main problem is the rapid increase in electricity consumption, which leads to higher electricity costs as well

as environmental pollution, especially when the energy comes from conventional combustion resources [2]. According to a study conducted by the International Energy Agency (IEA), fuel-based electricity production has been estimated to represent more than 64% of total world production, while the remaining 26% would have been shared between renewable resources, in particular only 5% wind and 2% solar [3]. Given their high world potential, this may seem strange, but several reasons have led to a reduction in their use, such as intermittent production, related to the presence of sun or wind, or as for example the solar production has a lower production in winter, while the consumption is higher, this makes their exploitation cost still expensive. For its regeneration, the states consecrated public aids until it becomes a competitive energy [4]. In particular, to be in line with the European Union (EU) objectives for tackling climate change, in the Framework for Action on climate and Energy 2021-2030, the following goals have to be assured: reduce greenhouse gas emissions by at least 40% (compared to 1990 levels), increase the contribution of renewable energy to final energy consumption by at least 32%, and increase energy efficiency by at least 32.5% [5]. Although renewable energy sources are inexhaustible and widely available, their randomness in terms of generation had led to provide innovative ways to get the most out of them. One of the methods is to predict their production allowing us to assess their performance and anticipate preventive actions. This allows, for instance, the size of the back-up systems to be reduced, or to adopt more efficient exploitation approaches. The inquiry of this resources, especially, wind speed and solar irradiation depends on large sizes of data and parameters availability [6]. Nowadays, Supervised Machine Learning (ML) methods are widely provided to achieve an accurate forecast model for wind speed and solar irradiation [7]. a model with inputs and their desired outputs is given to learn a general rule that maps inputs to outputs. These methods need expert intervention and the training data comprising a group of training examples. In supervised learning, each pattern may be a pair that includes an input object and the desired output value. The function of the supervised learning algorithm is intended to analyze the training data and produce an inferred function. As presented in the paper including the first part of this work, regression models are widely used in several types of research related to the prediction of weather parameters including wind speed and solar irradiation. Moreover, to predict wind speed in [16], the authors proposed Regression tree algorithms for the very short time gap. This method included the various types of regression trees in wind speed predictions to check the performances of different regression tree models. In [8] a new forecasting method based on bayesian theory and structure break model was used to predict wind speed parameters, by using prior information to improve the results of the time series model which are predicted as a set of values, different from other models. Gang et al in [9] proposed Bayesian combination algorithm and NN models for the prediction of wind speed in short term. The work includes the two-step methodology on Bayesian algorithms for wind speed forecasting and it also includes neural network models. Chang et al, in [14] used a hybrid approach of random forest regression and Bayesian model to predict

the solar irradiation. In [10] Lauret et al, had described the linear and non-linear statistical techniques auto-regressive (AR) and artificial neural network (ANN), while employing support vector machine (SVM) for predicting solar irradiation. However, employing a hybrid learning techniques combined with optimization methods to develop and optimize the construction of the learning algorithms is used in several works. For instance, in [11] the authors proposed optimized Least Squares Support Vector Machine (LSSVM) optimized by Particle Swarm Optimisation (PSO) algorithm to predict the speed of wind in a short time gap. In that paper, the accuracy rate is tried to improve by combining approaches like Ensemble Empirical Mode Decomposition (EEMD) and Sample Entropy (SE), also the optimization is done by PSO. In [12], an optimized model of wind speed forecasting with the combination of wavelet transform and Support Vector Machines is used. The proposed model includes wavelet transform, Genetic Algorithm (GA), and Support Vector Machines, as well as the consideration of two evolutionary algorithms, and the use of these together overcome the other models. In [13], Zameer et al employed a hybrid approach based on ANNs and Genetic Algorithm (GA) programming for short-term forecasting of solar irradiation.

The work presented in this paper is a continuation of an already performed approaches for prediction of wind speed and solar irradiation using regression machine learning models. It is proposed in that previous work to investigate further a method to increase the accuracy of tested models. One of the critical point observed were the tuning of the hyper parameters. As a first step, the internal parameters have been tuned manually affecting the performances of the models in certain cases. This paper proposes to combine optimization methods with machine learning to automate the selection of the best hyper parameters values.

This paper is divided into five sections. An overview of solar irradiation and wind speed prediction methods have been presented in Section 1. Section 2 presents, analyses and performs a pre-processing of the study data. The approach and use of optimization method to export the best hyper parameters for regression models is explained in Section 3. The results of the best optimizable models are presented and explained in Section 4. The last section concludes the study and proposes guideline for future works.

2 Data Pre-processing

The weather parameters including the temperature, wind speed, solar irradiation, and relative humidity are presented on Fig.1. They have been collected using a data monitoring system that can give 15 minutes range time samples. The system is composed of meteorological sensors in the station of Malviya National Institute of Technology in the Jaipur region in India (Latitude: 26.8622°N, Longitude:75.8156°E). A data set composed of 1922 features was imported from the platform of MNIT Weather Data - Live Feed.

The ranges for each weather parameters are presented on Table 1.

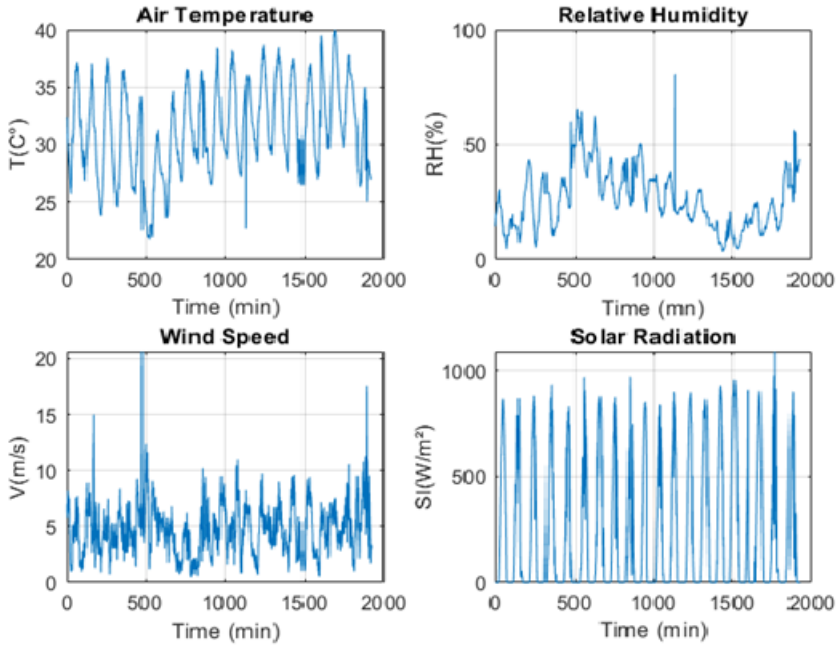


Fig. 1: Weather data

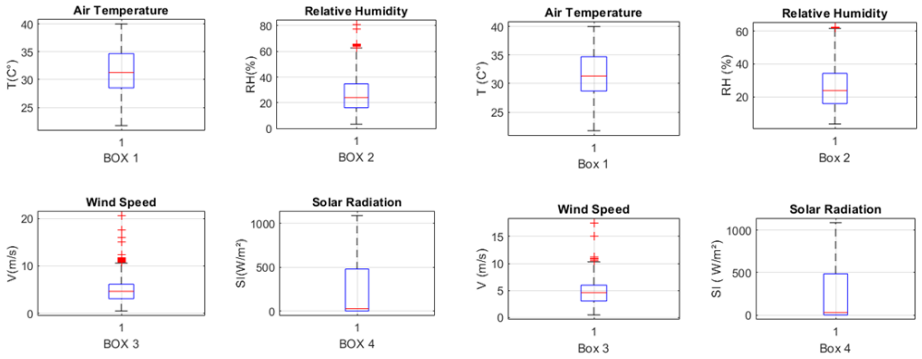
Table 1: Weather data informations.

Parameters	wind speed (m/s)	Relative Humidity (%)	Temperature (C°)	Solar Radiation (W/m^2)
Min	0.47	3.41	21.75	0
Max	20.62	80.72	39.97	1089
Average	4.71	26.14	31.39	233.79

2.1 Data Cleaning

Models' precision is highly impacted by the quality of the data employed. It is envisaged to carry out a data-cleaning procedure for improving the quality of the data by filtering it from any error or anomalies including outliers caused by the default in wind speed and solar irradiation measures. Several approaches have been applied in the literature to improve data quality by filtering errors such as the daily clearness index K [15]. In this paper, the inter quartile range (IQR) method is adopted for removing the outliers from wind speed and relative humidity data-set, after plotting the box and whiskers plot and getting the 25 percentile and 75 percentile, Q1, and Q3 respectively, the IQR is calculated considering the difference between Q1 and Q3. The IQR is used to Define the normal data range with a lower limit as $Q1–1.5 \times IQR$ and an upper limit as $Q3 + 1.5 \times IQR$. However, any data point outside this range is considered as

an outlier and should be removed for further analysis. Fig.2 is showing the data before applying the quarterlies method in Fig.2a and after Fig.2b.



(a) Box and Whiskers plot before data cleaning.

(b) Box and Whiskers plot after data cleaning.

Fig. 2: Data Box and Whiskers plots.

Fig 3 and Fig 4 represent the humidity and wind speed data cleaning, before and after removing the outliers using the quartiles method.

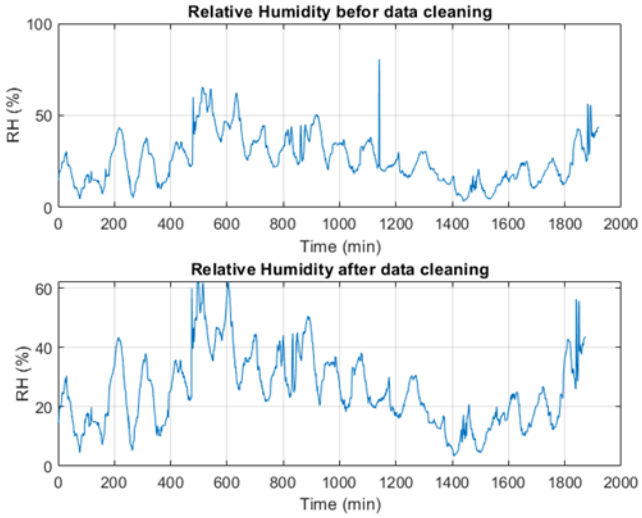


Fig. 3: Humidity data processing

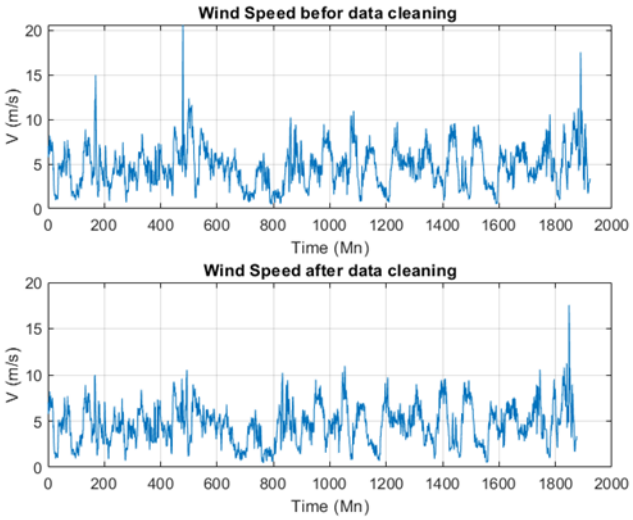


Fig. 4: Wind speed data processing

2.2 Data Correlation

After data analysis, having the relationship of variation between the metrological data is important to be used as predictors both for wind speed and solar irradiation, indeed, a correlation matrix will be calculated in this section. The data correlation is characterized by Pearson's correlation coefficients between all pairs of variables in the matrix of time series, including wind speed, solar irradiation, temperature, and relative humidity. The correlation coefficients are plotted on a matrix where the diagonal includes the distribution of a variable as a histogram. Each off-diagonal subplot contains a scatter-plot of a pair of variables with a least-squares reference line, the slope of which is equal to the correlation coefficient. From Fig. 7, a high negative correlation is remarked between the air temperature and relative humidity, physically, the relation between humidity and temperature formula simply says they are inversely proportional. If temperature increases it will lead to a decrease in relative humidity, thus the air will become drier whereas when temperature decreases, the air will become wet means the relative humidity will increase. Since air temperature is directly related to global solar radiation. So, an increase in solar radiation increases the air temperature and this justifies the positive correlation gated between the two items. Because of the solar energy (incoming short-wave radiation), the earth's surface heats up. Air near the ground tends to gain heat from the earth's surface. When air is heated, it expands and starts to rise. Solar energy causes wind due to its effect on air pressure and there is little correlation between them in the data analyzed. No correlation is detected between relative humidity and wind speed.

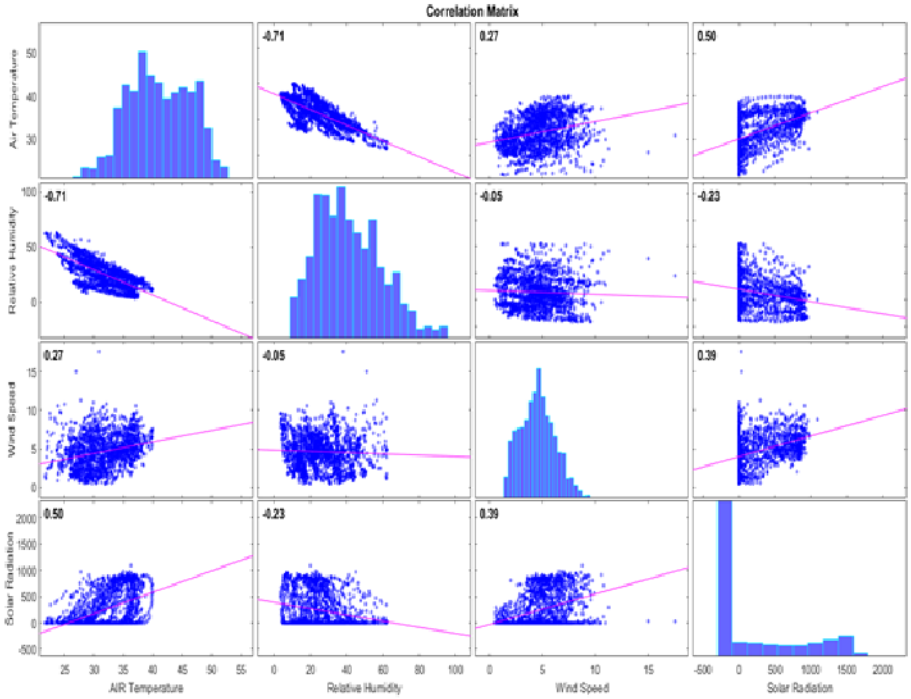


Fig. 5: Data correlation of considered weather parameters.

The calculation of the correlation allowed to determine the parameters used for the prediction of the solar irradiation and the wind speed. However to predict the solar irradiation the predictors used are: time sampling, relative humidity, temperature and wind speed. To forecast the future value of wind speed and as per of research experience made in [7] and actual data correlation, the used predictors were: time, solar irradiation, temperature and previous recorded wind speed.

3 Machine Learning

In the field of Artificial Intelligence, Machine Learning (ML) is widely used in several contexts among them forecasting. ML is composed of two main techniques including supervised and unsupervised learning. Among the approaches used in supervised learning to perform forecasting models is regression methods. Usually, regression models deal with real data mainly to predict continued responses as in the case of this paper regarding wind speed and solar irradiation.

The regression models employed in this paper are regression trees, Gaussian process regression models, support vector machines, and ensembles of regression

trees. As explained in Fig. 6, the data are first divided into two sets, one for training and the other for validation. The data preparation phase allows for reorganizing of the two sets by removing perturbations in the data distribution based on the quartile method.

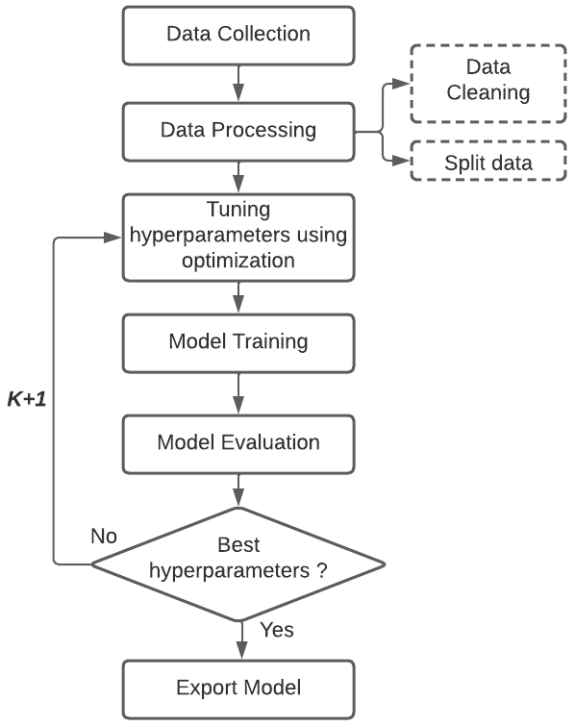


Fig. 6: Model processing flowchart.

Once the dataset is ready, it will be used for the model training inputs. A hyperparameter tuning is performed using optimization methods including bayesian, grid Search, and random Search. In general, the goal of optimization is to find a point that minimizes an objective function. In the context of hyperparameter tuning, a point is a set of hyperparameter values, and the objective function is the loss function. In each iteration the training model results are evaluated by the Root Mean Squared Error (RMSE) as shown in Equation 1.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(V,S)_{predicted} - (V,S)_{measured}}{n}} \tag{1}$$

where, V and S are wind speed and solar irradiation respectively.

If the error results are not sufficient, this will lead to going another step to export the best set hyperparameters for each model as described in Fig. 7.

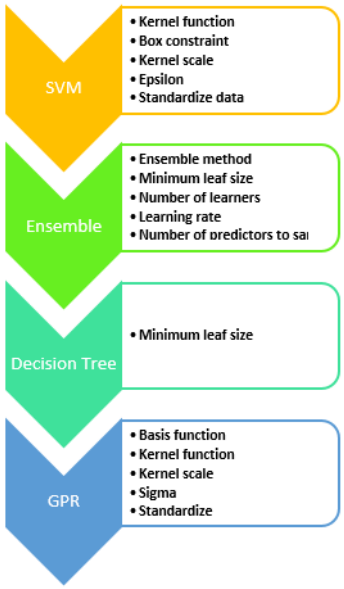


Fig. 7: Optimizable hyperparameters for each model.

4 Results and Discussions

In this paper, the aim is to predict future values of solar irradiation and wind speed from historical data, hence, the model will be implemented with regression techniques from supervised learning. These techniques are implemented in MATLAB, the discussion is divided into three parts, two dedicated to each predicted parameter, and the last part compares the results regarding the manual tuning and tuning by optimization.

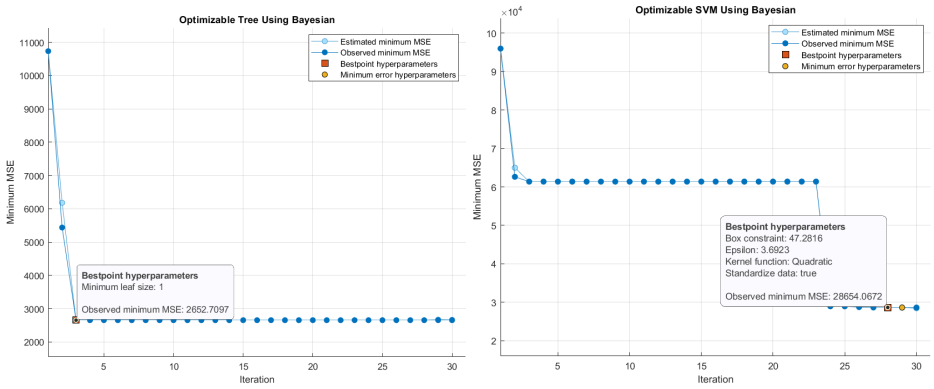
4.1 Optimized Models for Solar Irradiation Forecasting

For assessing and improving the regression learning models, tuning the model parameters by optimizing the hyperparameters is used to reduce the RMSE error values. The optimizer models used for hyperparameter tuning are the Bayesian optimization, Grid Search, and Random Search. Table 2 presents the RMSE obtained by tuning the models hyperparameters using the three optimization methods.

Table 2: Optimized model results for solar irradiation prediction (RMSE).

Model \ Optimizer	Iterations	Bayesian	Grid Search	Random Search
Optimizable TREE	30	51.50	53.70	51.504
Optimizable SVM	30	169.38	250.72	107.82
Optimizable GPR	30	0.09	0.19	0.14
Optimizable Ensembles	30	0.65	48.18	36.52

The results obtained from all different regression models show that the bayesian method outperforms the two others optimization methods. Bayesian optimization is an approach that uses the bayes theorem to drive the search in order to find the minimum of the objective function. It is an approach that is most useful for objective functions that are noisy as in the case of data presented in this paper. Applied to hyperparameter optimization, Bayesian optimization builds a probabilistic model of the function mapping from hyperparameter values to the objective evaluated on a validation set. Bayesian Optimization differs from Random Search and Grid Search in that it improves the search speed using past performances, whereas the other two methods are uniform or independent of past evaluations.

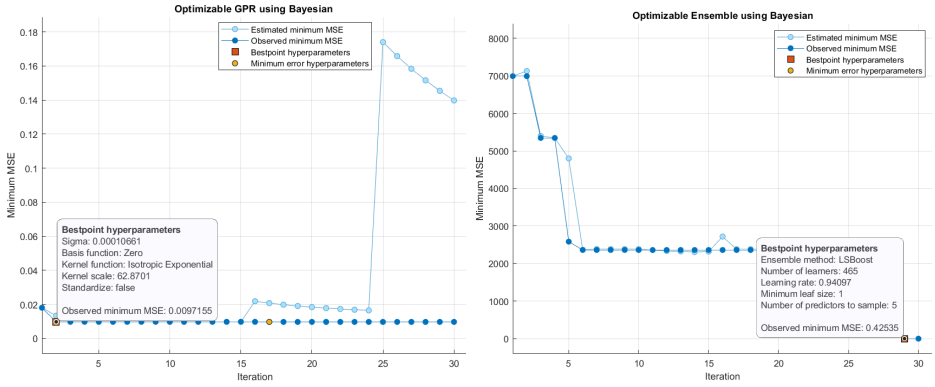


(a) Optimizable Tree. (b) Optimizable SVM.

Fig. 8: Optimizable models results for solar irradiation forecasting

Fig. 8 and 9 shows the optimization results obtained for the best hyperparameters regarding the four regression models using Bayesian optimization.

The plots are showing the estimated minimum MSE. Each light blue point corresponds to an estimate of the minimum MSE computed by the optimization process when considering all the sets of hyperparameter values tried, including the current iteration. Each dark blue point corresponds to the observed mini-



(a) Optimizable GPR. (b) Optimizable Ensembles.

Fig. 9: Optimizable models results for solar irradiation forecasting.

mum MSE computed by the optimization process. Best point hyperparameters. The red square indicates the iteration that corresponds to the optimized hyperparameters. Minimum error hyperparameters. The yellow point indicates the iteration that corresponds to the hyperparameters that yield the observed minimum MSE.

4.2 Optimizable Models for Wind Speed Forecasting

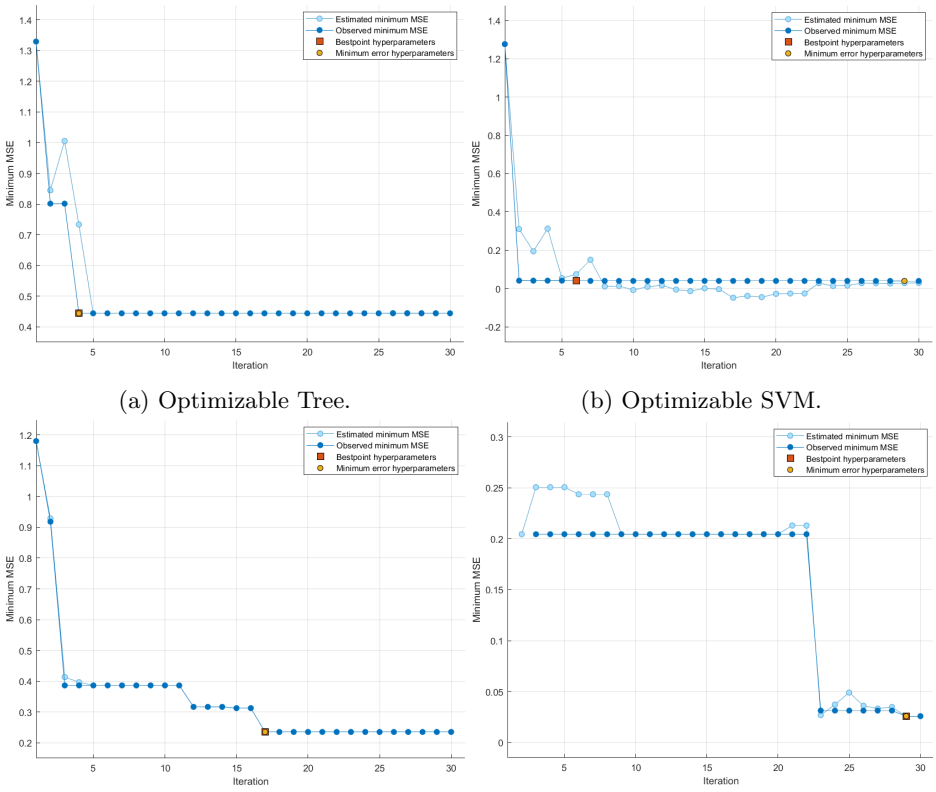
It was applied the same approach of solar irradiation, to forecast the future values of wind speed four regression methods including, Support Vector Machine (SVM), Gaussian Process Regression, Ensembles, and Regression Tree. The hyperparameters were optimized based on the three optimization methods: Bayesian, Grid Search, and Random Search.

The comparative Table 3 is showing the comparison of the four regression models based on the three optimization methods. The Bayesian outperforms the optimization methods in terms of model improvement.

Table 3: Optimizable model results for wind speed predection (RMSE).

Optimizer / Model	Iterations	Bayesian	Grid Search	Random Search
Optimizable Tree	30	0.66	0.69	0.68
Optimizable SVM	30	0.19	0.58	0.34
Optimizable GPR	30	0.48	0.79	0.62
Optimizable Ensembles	30	0.16	0.73	0.64

Fig. 10 shows the optimization results search for the best hyperparameters regarding the four regression models using Bayesian optimization for wind speed prediction.



(a) Optimizable Tree.

(b) Optimizable SVM.

(c) Optimizable GPR.

(d) Optimizable Ensembles.

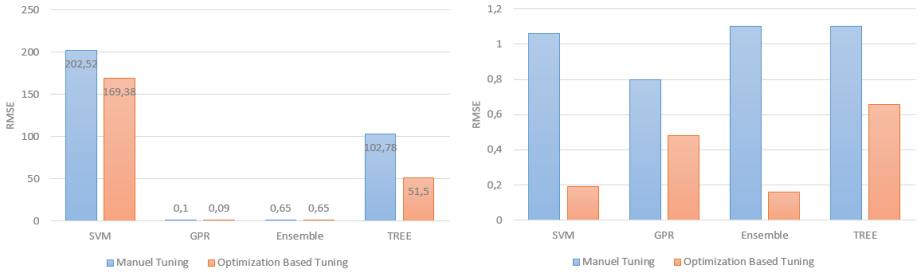
Fig. 10: Optimizable models results for wind speed forecasting.

4.3 Comparison of Manual and Optimized Tuning of Hyperparameters

The Fig. 11 presents the comparison of models in term of accuracy between manually and automated tuned model. The optimization helps to export the best hyper parameters combination. In fact, the accuracy of the models have remarkably been increased.

Automatic Tuning exclude the undifferentiated heavy lifting required to search the hyperparameter space for more accurate models.

This feature allowv saving significant time and effort in training and tuning of machine learning models. In general, Automated hyperparameter tuning of



(a) Solar irradiation (b) Wind speed.

Fig. 11: Manual and optimized tuning comparison.

machine learning models can be accomplished using Bayesian optimization. In contrast to random search, Bayesian optimization chooses the next hyperparameters in an informed method to spend more time evaluating promising values.

5 Conclusion and Future works

In this article, a combined optimization with machine learning has been performed to predict the future values of wind speed and solar irradiation. It is made by performing a tuning of hyper-parameters using optimization methods to build optimized models, four regression models including Regression Tree, Support Vector Machine (SVM), Gaussian Process Regression (GPR) and optimized ensembles, combined with three optimization approaches for each model including, Bayesian, Random Search and Grid Search. The three methods have been compared on a scale of 30 iterations. The Bayesian optimization has shown the best automated hyper-parameter tuning for the four regression models. The results have been compared to the manual tuned models presented in the first part of this work, the optimized models gave a higher accuracy. As a continuation of this work, it is proposed to perform a time series prediction for long term time horizons.

References

1. U. Akpan, G. Friday, G.E. Akpan. "The contribution of energy consumption to climate change: a feasible policy direction." International Journal of Energy Economics and Policy 2.1 (2012): 21-33. https://doi.org/10.1201/9781003126171-4.
2. A. Shahsavari, M. Akbari. "Potential of solar energy in developing countries for reducing energy-related emissions." Renewable and Sustainable Energy Reviews 90 (2018): 275-291. https://doi.org/10.1016/j.rser.2018.03.065.
3. J. F. A. Lopez, A. Granados, A. P. Gonzalez-Trevizo, M. E. Luna-Leon, A. G. Bojorquez-Morales. "Energy payback time and greenhouse gas emissions: studying the international energy agency guidelines architecture." Journal of Cleaner Production 196 (2018): 1566-1575. https://doi.org/10.1016/j.jclepro.2018.06.134.

4. K. Engeland, M. Borga, J. D. Creutin, B. François, M. H. Ramos, J. P. Vidal. "Space-time variability of climate variables and intermittent renewable electricity production–A review." Renewable and Sustainable Energy Reviews 79 (2017): 600-617. https://doi.org/10.1016/j.rser.2017.05.046.
5. Y. Amoura, Â.P. Ferreira, J. Lima, A. I. Pereira. "Optimal Sizing of a Hybrid Energy System Based on Renewable Energy Using Evolutionary Optimization Algorithms." International Conference on Optimization, Learning Algorithms and Applications. Springer, Cham (2021): 153-168. https://doi.org/10.1007/978-3-030-91885-912.
6. Y. Amoura, A.I. Pereira, J. Lima. "Optimization methods for energy management in a microgrid system considering wind uncertainty data." Proceedings of International Conference on Communication and Computational Technologies. Springer, Singapore (2021): 117-141. https://doi.org/10.1007/978-981-16-3246-410 .
7. Y. Amoura, A. I. Pereira, J. Lima. "A Short Term Wind Speed Forecasting Model Using Artificial Neural Network and Adaptive Neuro-Fuzzy Inference System Models." International Conference on Sustainable Energy for Smart Cities. Springer, Cham, (2021): 189-204. https://doi.org/10.1007/978-3-030-97027-712.
8. J.Wang, S. Qin, Q.Zhou, H.Jiang. "Medium-term wind speeds forecasting utilizing hybrid models for three different sites in Xinjiang, China." Renewable Energy 76 (2015): 91-101. https://doi.org/10.1016/j.renene.2014.11.011.
9. E.Cadenas, R. Wilfrido. "Short term wind speed forecasting in La Venta, Oaxaca, México, using artificial neural networks." Renewable Energy 34.1 (2009): 274-278. https://doi.org/10.1016/j.renene.2008.03.014.
10. P. Lauret, C. Voyant, T. Soubdhan, M. David , P. Poggi. "A benchmarking of machine learning techniques for solar radiation forecasting in an insular context." Solar Energy 112 (2015): 446-457. https://doi.org/10.1016/j.solener.2014.12.014.
11. L. Yang, L. Wang, Z. Zhang. "Generative Wind Power Curve Modeling Via Machine Vision: A Deep Convolutional Network Method with Data-Synthesis-Informed-Training." IEEE Transactions on Power Systems (2022). https://doi.org/10.1109/tpwrs.2022.3172508.
12. D. Liu, D. Niu, H. Wang, L. Fan. "Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm." Renewable energy 62 (2014): 592-597. https://doi.org/10.1016/j.renene.2013.08.011.
13. A. Zameer, J. Arshad, A. Khan, M. A. Z Raja . "Intelligent and robust prediction of short term wind power using genetic programming based ensemble of neural networks." Energy conversion and management 134 (2017): 361-372. https://doi.org/10.1016/j.enconman.2016.12.032.
14. J. F. Chang, N. Dong, K. L. Yung. "An ensemble learning model based on Bayesian model combination for solar energy prediction." Journal of Renewable and Sustainable Energy 11.4 (2019): 043702. https://doi.org/doi.org/10.1063/1.5094534.
15. K. Ferkous, F. Chellali, A. Kouzou, B. Bekkar. "Wavelet-Gaussian process regression model for forecasting daily solar radiation in the Saharan climate." Clean Energy 5.2 (2021): 316–328. https://doi.org/10.1093/ce/zkab012.
16. A. Troncoso, S. Salcedo-Sanz, C. Casanova-Mateo, J.C. Riquelme, L. Prieto. "Local models-based regression trees for very short-term wind speed prediction." Renewable Energy, 81 (2015), 589-598. https://doi.org/10.1016/j.renene.2015.03.071.