Springer Proceedings in Energy

El Mehdi Elkhattabi · Mourad Boutahir ·
Konstantinos Termentzidis ·
Kohji Nakamura ·
Abdelhai Rahmani *Editors*

Advanced Materials for Sustainable Energy and Engineering

Selected Proceedings of the 2023 International Conference on Advanced Materials for Sustainable Energy and Engineering (ICAMSEE)

Springer

Springer Proceedings in Energy

Series Editors

Muhammad H. Rashid, *Department of Electrical and Computer Engineering, Florida Polytechnic University, Lakeland, FL, USA*
Mohan Lal Kolhe, *Faculty of Engineering and Science, University of Agder, Kristiansand, Norway*

The series Springer Proceedings in Energy covers a broad range of multidisciplinary subjects in those research fields closely related to present and future forms of energy as a resource for human societies. Typically based on material presented at conferences, workshops and similar scientific meetings, volumes published in this series will constitute comprehensive state-of-the-art references on energy-related science and technology studies. The subjects of these conferences will fall typically within these broad categories:

- Energy Efficiency
- Fossil Fuels
- Nuclear Energy
- Policy, Economics, Management & Transport
- Renewable and Green Energy
- Systems, Storage and Harvesting
- Materials for Energy

eBook Volumes in the Springer Proceedings in Energy will be available online in the world's most extensive eBook collection, as part of the Springer Energy eBook Collection. To submit a proposal or for further inquiries, please contact the Springer Editor in your region:

Kamiya Khatter (India)
Email: kamiya.khatter@springer.com
Loyola D'Silva (All other countries)
Email: loyola.dsilva@springer.com

El Mehdi Elkhattabi · Mourad Boutahir ·
Konstantinos Termentzidis · Kohji Nakamura ·
Abdelhai Rahmani
Editors

Advanced Materials for Sustainable Energy and Engineering

Selected Proceedings of the 2023 International Conference on Advanced Materials for Sustainable Energy and Engineering (ICAMSEE)

Springer

Editors
El Mehdi Elkhattabi
Ecole Normale Supérieure, Moulay Ismail University
Meknes, Morocco

Mourad Boutahir
Ecole Normale Supérieure, Moulay Ismail University
Meknes, Morocco

Konstantinos Termentzidis
CETHIL, INSA
Lyon, France

Kohji Nakamura
Department of Physics Engineering
Mie University
Tsu, Mie, Japan

Abdelhai Rahmani
Faculté des Sciences, Moulay Ismail University
Meknes, Morocco

ISSN 2352-2534 ISSN 2352-2542 (electronic)
Springer Proceedings in Energy
ISBN 978-3-031-57021-6 ISBN 978-3-031-57022-3 (eBook)
https://doi.org/10.1007/978-3-031-57022-3

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Ensemble Learning Method for Forecasting HVAC System Demand

Nihad Aghbalou[1](✉), Abdérafi Charki[2], Hanae Errousso[1], and Youssef Filali[1]

[1] EIGSI, Casablanca, Morocco
nihad.aghbalou@eigsica.ma
[2] LARIS, University of Angers, Polytech-Angers, Angers, France

Abstract. The efficiency of buildings' energy use is one of the energy transition's three fundamental pillars. It is a current issue that many countries are actively promoting as a means of lowering energy dependence by eliminating wasteful energy losses and environmental effect. Controlling energy flows within buildings is actually crucial for maximizing occupant comfort and timely use of consumer goods. In this brief study, ensemble machine learning techniques with different model structure, training process, interpretability, and performance, are developed to create a trustworthy tool for HL (Heat loading) and CL (Cool loading) estimation for future intelligent urban ecosystems. The results demonstrate that XGBoost and GBM with exhaustive feature selection provide accurate and more robust predictions. Moreover, the finding reveals that the proposed method handle better mixed Data.

Keyword: Building energy efficiency · HVAC load · Ensemble learning · Boosting algorithms

1 Introduction

Issues over supply shortages, resource depletion, and serious environmental effects (such as ozone layer destruction, global warming, and climate change) have been raised by the rapid increase in worldwide energy use. In developed countries, residential and commercial buildings now account for 20–40 % of total energy use, outpacing other significant industries and the transportation sector [1]. Because of the growing need for comfort and building services, the rising trend in energy consumption will continue. In most buildings, heating, ventilation and air conditioning (HVAC) systems use 20% of the energy [6]. As a consequence, energy efficiency in building design is one of the main focuses of regional, national, and international energy policy [3]. Building physics (dimensions, material conductivity, the form of the coefficient, length-to-width ratio, window-to-floor ratio, window-to-wall ratio (WWR), etc.), weather conditions, occupant behavior, and installed technology and equipment are several important variables that affect a building's energy consumption. All the influencing factors, must be considered in cost and scientific analysis for estimating Heating and Cooling load, given that

E. M. Elkhattabi et al. (Eds.): ICAMSEE 2023, SPE, pp. 1–9, 2024.
https://doi.org/10.1007/978-3-031-57022-3_1

they consume the most amount of energy and to formulate accurate energy consumption models. Particularly, early forecasts of HL and CL is useful for choosing the right heating and cooling equipment to maintain comfortable indoor air quality while designing sustainable buildings that focus more on energy efficiency. Recently, many studies have been conducted to deal with the building energy consumption forecasting. Building energy consumption forecasting models fall into two broad types in the literature: physical approaches and data-driven approaches [11]. The physical based models makes use of theoretical assumptions and equations and building's intrinsic physical properties and thermodynamic principles to develop equations that can be utilized to estimate its energy usage. This technique involves using algorithms that simulate building energy use [9]. The HVAC system, physical qualities, interior load, solar radiation, and other factors all affect how accurately the physical model predicts energy usage. This can be complicated to obtain these characteristics, and simplified simulation techniques might not be able to fully account for the intricate environmental aspects that influence energy use, leading to poor model performance[9]. Data driven approach also known as data-centric models, on the other hand, are constructed by analyzing and processing large datasets of building energy consumption and features to extract patterns, relationships, and insights and built directly from the those datasets[10]. Accurately estimating the energy consumption of buildings is a challenging issue due to the complicated and non-linear relationship between these influencing characteristics and the energy performance of buildings. Hence, this short paper aims to provide a practical data-driven model for the initial load forecasting at the early stages of residential building architectural design. The rest of the paper is organized as follows. Section 2 provides a literature review of the most relevant papers and recent works related to this topic. The materials and short explanation of the basic concepts underlying baseline models are enlightened in Sect. 3. Section 4 presents results produced by the framework used and discussions. Section 5 summarizes the paper findings.

2 Literature Review

Research on how to provide a safe and comfortable indoor environment and address energy-saving problems has been ongoing for a while. In a recent study [4], W. Cai et al. used the SVR supervised machine learning algorithm to predict and forecast the energy consumption of buildings and parameters were improved using six population-based (Slime Mold Algorithm, AEO, Sparrow Search Algorithm, Gray Wolf Optimizer, Artificial Bee Colony, and Arithmetic Optimization Algorithm). As a result, the SVR-Artificial Ecosystem-based Optimization hybrid model showed the best performance against the six available options for both heating and cooling analyses. Heap Optimization Based Generalized Intelligent Neural Fuzzy Control (HO-GINFC) was presented in [6], for predicting the cooling demand for air conditioning systems based on cold thermal storage and validated using data of a massive Saudi Arabian commercial structure. The suggested study demonstrated an improved performance on the the cold thermal storage-based air conditioning systems' forecasting and control capabilities against conventional methods such as Deep Reinforcement Learning, Elman neural network, Gradient boosting decision tree, and Mean Impact Value combined with Improved Gray

Wolf Optimizer-based SVR. Y. Huang and C. Li [7] improved Ant Colony Optimization and Wavelet Neural Network (I-ACO-WNN) model to forecast the HVAC system. The self-adaptive mutation operator was applied to avoid a premature local convergence of ants in order to optimize WNN learning rates. The model was trained based on the Spearman method findings indicating that the roof area and total height of a building have the biggest effects on load. N. Pachauri and C.W. Ahn [8] investigated three methods namely Gaussian process regression (GPR), least squared boosted regression trees (LSB), and marine predator optimization algorithm (MPO) to develop an ensemble predictive model called WGPRLSB. Moreover, a sensitivity analysis was carried out to evaluate how much the proposed predictive model expected output depends on its input variables. It was found that Roof Area, Surface Area, and Glazing Area are the most influential input variables for the forecasting of energy consumptions in case of HL. The level of dependency S_A attained by these variables were 0.9353, 0.9279, and 0.9020 for Heating Load. Whereas, in the case of Cooling Load, Surface Area, Roof Area, Wall Area, and Glazing Area are the important input variables with S_A equal to 0.9554, 0.9572, 0.9297, and 0.9356 respectively.

3 Methodology

The main purpose of this study is to leverage a well-known ensemble learning approach for HVAC load forecasting. A comparative evaluation of its main variants, i.e. Linear regression (LR), Decision Tree Regressor (DRT), Gradient Boosting Machine (GBM), Adaptive Boosting (Adaboost), eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LGB) and Categorical Boosting (CatBoost) was performed in order to purpose a reliable tool for HL (Heat loading) and CL (Cool loading) estimate for future intelligent urban ecosystems, ensemble machine learning techniques under mixed Data better. The flowchart shown in Fig. 1 outlines the proposed methodology with foor main steps: data pre-processing, training boosting models, testing models and then feature selection with the best models.

3.1 Ensemble Learning Methods [2]

Basically, the primary concept underlying ensemble learning is to combine numerous models, called learners, rather than utilizing a single model in order to improve machine learning performance. Linear regression dates back to the early $19^t h$ century and has been used for a very long period. It is among the most established and well-known statistical modeling methods that use one or more input features to make a continuous result variable forecasting. The target variable and the predictors are assumed to have a linear relationship. Regression Tree is a machine-learning model that employ a hierarchical network of leaf nodes and decision nodes frequently trained using techniques such as Recursive Partitioning. A stopping requirement, such as a maximum depth or a minimum number of samples per leaf, must be satisfied before the process can cease. Regression trees are vulnerable to overfitting since all the models are considered strong and particularly if they are allowed to grow excessively. One boosting approach that minimizes the likelihood of underfitting using the Bayesian strategy by merging numerous weak learners, called Adaboost. The the iterative AdaBoost was developed from this concept. In order to produce weak regressors with high bias error but low variance error, the AdaBoost regressor iteratively reweights training examples based on the prediction error. Initially, all data samples share equal weights, and in the next iteration round, the sample's weights are altered; each data sample is assigned a weight, signifying the data sample's importance to be selected as a training sample. By doing this, the next regressor highlights instances in which the predictions from the preceding phase were incorrect. In other words, its adaptability comes from the fact that the weights of the misclassified samples are increased, or the weights of the correctly classified samples are decreased. The new classifier emphasizes the misclassified samples, which might be adjacent to the classification margin, and finally reduce the error. The final prediction is made based on the weighted majority vote, the weak base classifiers are combined with being a strong classifier, which results in a mode with reduced low variance and bias errors. XGBoost is a well-known advanced implementation of an optimized Gradient Boosting treebased algorithm that can efficiently handle large-scale Machine Learning tasks. Merited by its performance superiority and affordable time and memory complexities, it has been widely applied to a variety of research fields. Its core idea is to train employ a sequential tree-building strategy with parallel implementations, and develop the tree to its maximum depth and then prunes backwards until the rise in regularized loss function falls below a predetermined threshold. LightGBM are very powerful computational algorithms that use tree-based learning methods. They were created to increase model training speed, reduce memory usage and improve predictability performance. Unlike other algorithmic trees, which grow horizontally, the methods grow vertically, which means they grow like leaves. Leaves with significant growth losses are chosen by LightGBM. They can reduce losses more than level-based strategies while expanding leaves. Model complexity can increase as trees grow in the direction of the leaves, which can lead to excessive data adjustments while using smaller data sets. LightGBM can also avoid over-fitting by limiting the depth of its tree structure. CatBoost solves the limitations of other decision tree methods, which generally require data preprocessing to convert categorical string variables into numerical values, perform one-hot encodings, etc. Without preprocessing, this method can directly use a combination of categorical and non-categorical explanatory variables. Without pre-processing, this method can directly use a combination of categorical and non-categorical explanatory variables. Preprocessing is performed by the algorithm. Ordered encoding is used by CatBoost to encode categorical entities. To calculate a value and replace the categorical entity, ordered encoding takes into account the target statistics of all rows before a data point. The use of symmetrical trees is another distinctive feature of the CatBoost algorithm. This means that all decision nodes use the same condition at each depth level. The training sample is split into right and left partitions by CatBoost using the same features, which results in a tree with exactly 2k leaves and a depth of k. A group of decision trees are sequentially built during training. In comparison to the preceding tree, each additional tree is constructed at a reduced loss. To avoid overfitting, the number of trees is regulated by the starting settings. If overfitting happens, CatBoost might end training earlier than what the training parameters call for.

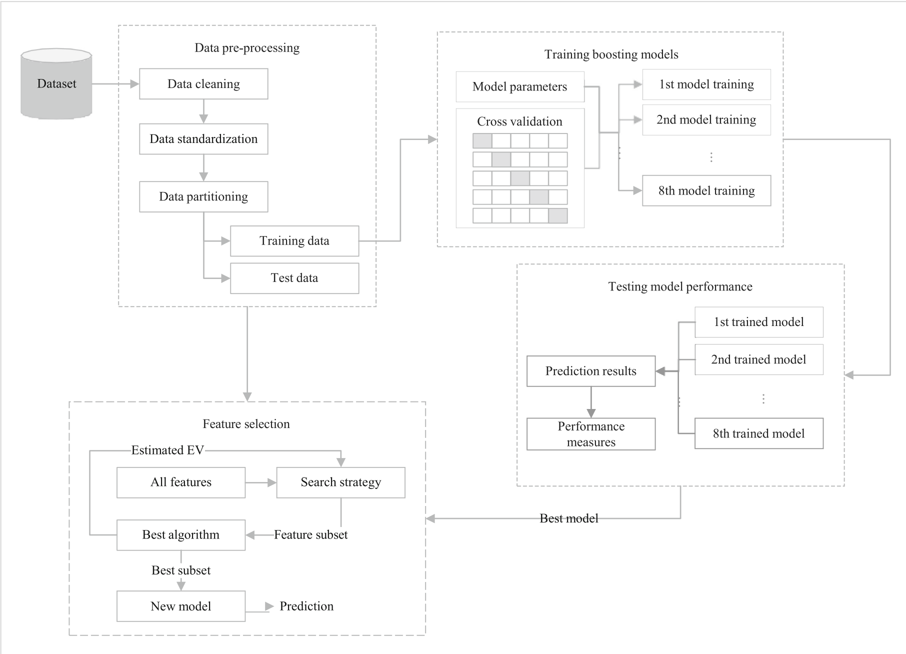

Fig. 1. Methodological framework.

3.2 Case Study and Data Set Information

This study makes use of the energy efficiency dataset produced by Angeliki Xifara and processed by the University of Oxford's Oxford Centre for Industrial and Applied Mathematics [5]. The dataset in Table 1 consists of eight input parameters used to predict two real outputs, HL and CL for 12 residential buildings of the same volume. Each structure is made of the same materials and has the same volume (771.75 m^3), but its relative compactness (RC), surface area (SA), wall area (WA), roof area (RA), overall height (OH), orientation, glazing area (GA) and glazing area distribution (GAD) are different. The statistical analysis of the input and output variables using the Spearman rank relations has been carried out. A significant correlation between output variables HL (Y1) and CL (Y2) and input variables X4 and X5 has been noticed. In fact, correlation coefficients between HL (Y1), CL (Y2) and input variables X5 are 0.89 and 0.9 respectively indicating strong positive monotonic. On the other hand, a strong negative monotonic relationship can be observed between HL (Y1), CL (Y2) and input variables X4 with a value of -0.86. However, a weak relationship with X6 is observed. Moreover, a few coupling and correlation relationships exist, e.g. X3 has a weak correlation with X1, X2, X4 and X5. Or, X1 has strong correlation with X2, X4 and X5 and X2 with X4 and X5 as X4 with X5.

Table 1. Description of the features of in the experimental dataset.

Variable	Parameters	Min	Max	Unit	Values
X1	Relative Compactness	0.62	0.98	N/A	12
X2	Surface Area	514.5	808.5	m^2	12
X3	Wall Area	416.5	318.50	m^2	7
X4	Roof Area	110.25	220.5	m^2	4
X5	Overall Height	3.50	7	m^2	2
X6	Orientation	2	5	N/A	4
X7	Glazing Area	0	0.4	m^2	4
X8	Glazing Area Distribution	0	5	N/A	6
HL (Y1)	Heating Load	6.01	43.1	kW	568
CL (Y2)	Cooling Load	10.9	48.03	kW	636

4 Results and Discussion

The main focus of this section is on the experimental results of data processing, model building and model evaluation. To solve the time-consuming manual debugging of hyper-parameters and long search times in huge parameter spaces, A K-fold cross-validation technique was applied to find the best combination of hyperparameters that yield the optimal model performance. The Mean Absolute Error (MAE) is computed to measure the average magnitude of errors between predicted and actual values; a lower MAE indicates better model performance, as smaller errors are preferable. Then, appropriate performance indicators were chosen to evaluate the performance of the selected baseline models in the prediction and comparative experiments including MAE, MSE, R2, EXV, RAP, MPD and MGD. Results for LR hyper parameter tuning indicate that that the lowest value of MAE is obtained with the combination when no intercept is involved in calculations (fit_intercept = False) and when the input values are converted to a booleans (copy_X = True) for positive or negative coefficients. When training DRT models of predicting HL (Y1) and CL (Y2), the lowest value of MAE is found when the maximum depth of the tree is 30 and the minimum number of samples required to split an internal node is 5 to build the predictor of HL (Y1). However, lower values of MAE are obtained for the predictor of CL (Y2) when the maximum depth of the tree is 70 and the minimum number of samples required to split an internal node is 5. On the other side, the related MAE in the space of learning rate and estimators highlighted the Adaboost's resilience to the loss function. In both models for predicting HL (Y1) and CL (Y2), Adaboost with exponential loss function, 150 estimator and 0.1 for learning rater show the best performance. Similarly, the robustness of GBM model was demonstrated when tuning the the number of estimators selected by early stopping (n_estimators) and the learning rate whatever the function to measure the quality of a split Fridman mean squared error or squared error. However, a strong dependence on the data was underlined when training XGBoost. The ideal combination for training the HL (Y1) model is 30 for the maximum depth and a numberparallel tree adjusted to 10. However, the ideal combination for training the CL (Y2) model is 20 for the maximum depth and a (num_parallel_tree) adjusted to 5. To train models with LightGBM, hyperparameters learning rate and the maximum tree depth for base learners, as well as the type of booster were considered. It was concluded that optimal hyperparameters are 0.2 for learning rate and 15 for maximum depth for the tree. Moreover, the smaller MAE is observed when the type of the booster is tuned to the traditional Gradient Boosting Decision Tree. As far as as the Catboost algorithm, smaller MAE resulted in higher depth (depth = 30) and lower learning rate (0.1). Using the statistical measures MAE, MSE, R2, EXV, RAP, MPD and MGD, the performance of each trained model on the test dataset is assessed. The findings are shown in Tables 2 and 3. The best models that predict HL (Y1) and CL (Y2) are found after carefully evaluating the statistical measures of each learning algorithm. The XGBoost algorithm, followed by GBM, predicts HL (Y1) best with test errors (MAE = 0.56, MSE = 0.65, R2 = 0.98, EXV = 0.987, RAP = 0.020, MPD = 0.03 and MGD = 0.0009). Then. And for CL (Y2) Gradient boosting algorithms, followed by XGBoost, performs best with test errors of MAE = 0.60, MSE = 1.13, R2 = 0.992, EXV = 0.992, RAP = 0.022, MPD = 0.02 and MGD = 0.0008. It can be concluded that XGBoot and GBM are the best two models besides the experimented models. On the other side, XGBoost and GBM are sensitive to the dataset during hyperparameter process selection.

Table 2. Statistical validation metrics of the models for Y1 (HL).

Regressor	MAE	MSE	R2	EXV	RAP	MPD	MGD
LR	2.31	10.22	0.89	0.89	0.10	0.38	0.01
DRT	0.47	0.45	0.99	0.99	0.02	0.01	0.0008
Adaboost	1.70	4.68	0.9956	0.9535	0.08	0.1949	0.0093
Catboost	0.38	0.27	0.9972	0.9972	0.01	0.0127	0.0007
LightGBM	0.43	0.43	0.9956	0.9957	0.02	0.0208	0.0012
GBM	0.37	0.24	0.9975	0.9976	0.01	0.0092	0.0004
XGBoost	0.33	0.21	0.9978	0.9981	0.01	0.0078	0.0003

The experimental results of the adapted f-test statistic and the mutual information statistic exhibited that that X6 and X8 yields the minimum values for the both statistics. In contrast, the highest F-values are found for X5, X3, X2 and the highest MI value are found for X1, X2, X3, X4. To summarize, analyzing the F-values and the estimated MI indices for different dataset shows that there is a pronounced variability and significant significant amount of shared information between the features and variables, which could result in varying degrees of accuracy for various models. Hence, feature selection for the aim of reducing the number of input variables when developing a predictive model may be advised. The exhaustive search in the feature space with XGBoost for HL (Y1) and GBM for CL (Y2) has been carried out to find the potential subset of features from the given dataset. The goal is to determine which subset of features are most informative for the given predictive model. The advantage of exhaustive feature selection is that it ensures that all potential feature combinations are explored. It's quite clear that X1, X3, X6, X7, X8 best predict HL (Y1) and X1, X6, X7, X8 best predict CL (Y2). Table 4 presents the results of applying the above finding to predict HL (Y1) and CL (Y2) with XGBoost and GBM. It demonstrate in fact, important improvements in terms of metrics result when learning models with selected features.

Table 3. Statistical validation metrics of the models for Y2 (CL).

Regressor	MAE	MSE	R2	EXV	RAP	MPD	MGD
LR	2.64	13.33	0.85	0.85	0.0998	0.43	0.01
DRT	1.38	5.26	0.94	0.94	0.04	0.15	0.004
Adaboost	2.03	7.01	0.98	0.92	0.075	0.22	0.008
Catboost	0.96	2.05	0.97	0.978	0.034	0.06	0.002
LightGBM	0.86	1.57	0.981	0.983	0.030	0.04	0.0015
GBM	0.56	0.65	0.992	0.992	0.022	0.02	0.0008
XGBoost	0.60	1.13	0.98	0.987	0.020	0.03	0.0009

Table 4. Output metric after Feature Selection

Variable	Best subset	MAE	MSE	R2	EXV	MAPE	MPD	MGD
HL (Y1)	{X1, X3, X6, X7, X8}	0.278	0.145	0.998	0.998	0.0114	0.005	0.0002
CL (Y2)	{X1, X6, X7, X8}	0.49	0.48	0.99	0.994	0.0195	0.016	0.0006

5 Conclusion

In both cost analysis, building energy usage is a key parameter. High accuracy in the formulation of energy consumption models is essential since underestimating it can result in potential failures that have a detrimental impact on social and economic well-being, while overestimating it can result in waste and idle capacity. This article present a practical methodology for the initial load prediction at the early stages of residential building architectural design. An effort was made to model and forecast the cooling and heating load of buildings using supervised ensemble machine learning.The data used in this study includes eight input factors: relative compactness, area, wall, roof, overall height, orientation, glazing area, and glazing area distribution. The XGBoost and GBM were found as the best models in terms of correlation and error parameters, among the seven models for forecasting heating load and cooling load respectively. Moreover, combined with Exhaustive Feature Selection, those models generated a better performance in terms of correlation and error metrics.

References

1. Mostafavi, S., Doddi, H., Kalyanam, K., Schwartz, D.: Nonlinear moving horizon estimation and model predictive control for buildings with unknown HVAC dynamics. IFAC-PapersOnLine **55**(41), 71–76 (2022)
2. Kunapuli, G.: Ensemble Methods for Machine Learning. Simon and Schuster (2023)
3. Abdelrazik, A.S., Shboul, B., Elwardany, M., Zohny, R.N., Osama, A.: The recent advancements in the building integrated photovoltaic/thermal (BIPV/T) systems: An updated review. Renew. Sustain. Energy Rev. **170**, 112988 (2022)
4. Cai, W., Wen, X., Li, C., Shao, J., Xu, J.: Predicting the energy consumption in buildings using the optimized support vector regression model. Energy **273**, 127188 (2023)
5. Tsanas, A., Xifara, A.: Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. Energy and buildings **49**, 560–567 (2012)
6. Irshad, K., Algarni, S.: Novel optimized hybrid neuro-fuzzy approach for analysis of cold thermal storage system-assisted air conditioning system performance. J. Build. Eng. **65**, 105729 (2023)
7. Huang, Y., Li, C.: Accurate heating, ventilation and air conditioning system load prediction for residential buildings using improved ant colony optimization and wavelet neural network. J. Build. Eng. **35**, 101972 (2021)
8. Pachauri, N., Ahn, C.W.: Weighted aggregated ensemble model for energy demand management of buildings. Energy **263**, 125853 (2023)
9. Wang, L., Xie, D., Zhou, L., Zhang, Z.: Application of the hybrid neural network model for energy consumption prediction of office buildings. J. Build. Eng. **72**, 106503 (2023)
10. Al-Shargabi, A.A., Almhafdy, A., Ibrahim, D.M., Alghieth, M., Chiclana, F.: Buildings' energy consumption prediction models based on buildings' characteristics: RESEARCH trends, taxonomy, and performance measures. J. Build. Eng. **54**, 104577 (2022)
11. Chen, Y., Guo, M., Chen, Z., Chen, Z., Ji, Y.: Physical energy and data-driven models in building energy prediction: a review. Energy Rep. **8**, 2656–2671 (2022)

RBDO Approach for Site-to-Wind Turbine Generator Pairing

Nihad Aghbalou[1](✉), Abdérafi Charki[2], Saida Rahali Elazzouzi[3], and Kamal Reklaoui[4]

[1] EIGSI, Casablanca, Morocco
nihad.aghbalou@eigsica.ma
[2] LARIS, University of Angers, Polytech-Angers, Angers, France
[3] Labtic, ENSAT, Abdelmalek Essaadi University, Tangier, Morocco
[4] Equipe Innovation et Ingénierie, FST, Abdelmalek Essaadi University, Tangier, Morocco

Abstract. The increasing focus on sustainable, clean, and abundant alternative energy sources has highlighted the need for robust and dependable methods to assess their potential. This is crucial in order to make informed investments in energy systems that can meet consumption and export requirements. When it comes to wind power, however, there is a higher level of uncertainty compared to conventional power sources regarding its contribution to meeting electricity demands. Reliabilitybased design optimization (RBDO) has been extensively employed in engineering design applications. This approach aims to strike a balance between safety and manufacturing costs while adhering to probabilistic constraints. In the context of wind turbine selection for a specific site, this paper seeks to introduce the RBDO method as a means to identify the optimal wind turbine. It takes into consideration the potential uncertainties associated with various parameters relevant to the turbine selection problem. Additionally, this study addresses a limitation in the conventional method by proposing an alternative approach that is less sensitive to approximated curves of wind turbine generators (WTGs). By introducing this novel method, the reliability of wind turbine processes can be improved.

Keyword: Probabilistic performance assessment · Random wind turbine conditions · Wind turbine-site matching

1 Introduction

In order to meet the growing energy needs, the scarcity of primary resources, the persistent volatility of oil prices and global warming, decision makers have opted for the massive use of renewable energy. There is now a strong desire to move towards completely clean energy, which is reflected in the growing share of renewable energy in the global energy mix. Among the energies that can be exploited in the context of sustainable development, wind energy is in the spotlight. Indeed, it is in a privileged position thanks to its technological progress and its relatively cheaper operating costs. Wind energy

E. M. Elkhattabi et al. (Eds.): ICAMSEE 2023, SPE, pp. 10–19, 2024.
https://doi.org/10.1007/978-3-031-57022-3_2

remains completely different from conventional resources. The fundamental difference is incorporated into its intermittent and uncertain character. Unlike predictable electricity generation systems, whose sources are storable such as fossil fuels and nuclear power and whose performance is only influenced by the maintenance and reliability of equipment, Wind power generation systems are highly weather dependent. Consequently, this type of production is difficult to control and "Not-dispatchable", faced with the variability of the resource on the one hand, and the effect of the operating environment on the reliability and availability of the equipment on the other hand. Thus, the inability to ensure regular supplies remains the major challenge for the integration of wind energy into electrical systems.

To reinforce these challenges, research has given rise to significant technologies and solutions in terms of design, predictive quality, power regulation, control and continuation of the optimal operating point, monitoring and maintenance, location and geographical dispersion, etc. However, fluctuations in production are often lower than estimated by modelling and forecasting studies, resulting in over-costs related to the reserve that must be insured to cover unforeseen losses. These errors in estimating production levels are, of course, dependent on the uncertainties associated with the measurements, the spatial and temporal wind extrapolation models, and the models used to convert these measurements into power predictions.

This work is an extension of a previous study [2] were the aim was providing a methodology for wind turbine-site matching by using a probabilistic approach taking in to account the random behavior of the wind speed climate and the unertainties of wind turbine characteristics, by considering the cost analysis. The paper is organized as follows: in Sect. 2 a literature review for wind power assessment is provided. Section 3 describes the RBDO approach. Section 4 is focused on the problem restatement. Conclusions and future work to be achieved are drawn in Sect. 5.

2 Literature Review

To develop a wind farm, several windy sites can be selected, the optimal choice is usually dictated by several criteria whether economic or technological. From a theoretical point of view, it is first necessary to ensure the profitability of the candidate site. Then, check the durability of the wind turbine or the candidate technology whith the characteristics of the site. Manufacturers have announced a 20-year lifespan. However, to ensure that a wind turbine is best suited to a site, reliable models representative of the turbine, the wind and the operating environment are needed to estimate its performance over time and space. According to Serrano et al. [3], the design of a wind farm is an optimization problem that must take into account several aspects: the study of wind behaviour, the analysis of the interactions between wind turbines, the wake effect, the design of auxiliary installations (access roads, electricity infrastructure), economic issues, component reliability and environmental impact assessment. In particular the geographical dispersion of wind turbines, because the presence of wind turbines in the vicinity can be beneficial for smoothing production but leads to an increase in turbulence in the airflow and fatigue stresses that affect the reliability of mechanical components.

L. Soder et al. [1] present an overview of the adequacy challenge.The work of Chen et al. [4] study the effect of the different characteristics of the wind turbine, the start speed, the cut-off speed, the nominal speed and the height on the reliability of the electricity production systems in terms of adequacy based on the method *RBTS* (Roy Billinton Test System) to optimize the installation of wind turbines in relation to the candidate sites. This study showed that the starting speed is the most significant parameter, from 8 km/h to 18 km/h the *LOLE* (Loss of Load Expectation) from 0.6921 to 0.9209 h/year. On the other hand, we change the nominal speed from 32 to 42 km/h the *LOLE* increases from 0.7466 to 0.897 h/years or the cut-off speed from 40 to 60 km/h the *LOLE* increases from 0.7932 to 0.7895 h/years. Similarly, by varying the height of the wind turbine from 10 to 30 m the *LOLE* increases from 0.7895 to 0.7095 h/years, which shows that the change in wind speed with height has a slightly significant effect. In the same perspective, the author proposed two risk assessment indicators, the *LCCBR* (Load Carrying Capacity Benefit Ratio) and *ECR* (Equivalent Capacity Ratio), and showed that these indicators are better suited to select and classify sites and choose the optimal wind turbine.

During commissioning, the wind turbine and its subsystems are subjected to stresses or deformations over time. These variations lead to degradation and modification of material properties through stress accumulation and fatigue phenomena. To fully approximate the actual behaviour during export, designers must take into account the variability in the use of the wind system, namely the operating environment, the technological configuration, the different operating states, etc. In their work Staffell al. [5] indicated that wind turbines lose approximately 1, 6% of their production capacity each year. This is mainly due to the irrecoverable loss attributed by progressive deterioration, such as blade erosion and fouling, which causes a loss of power and a gradual reduction in the efficiency of parts such as blades, power modules, gearbox, bearings and generator due to aging.

3 Methodology Review

3.1 RBDO Approach

In the field of structural design in particular, the concern to design structures meeting a number of different criteria, such as the cost, safety, performance and durability, often leads to conflicting requirements that must be considered simultaneously [6]. This leads to the resolution of an optimization problem formulated as follows:

$$Min\, C_I(d)$$

$$\text{subject to} \begin{cases} G_i(d, p, X) \geq 0, \quad i = 1, ..., I \\ h_j(d) \geq 0, \quad j = 1, ..., J \end{cases} \tag{1}$$

This traditional deterministic optimization of the DDO (deterministic design design optimization) consists in the search of design parameters d that go minimize a C_I (d) objective function under performance constraints $G_i(.)$ and feasibility constraints h_j. Parameters p and variables d are considered deterministic. Uncertainty, especially for the most critical parameters in the problem, is introduced through the use of safety coefficients. These weight uncertain parameters to ensure an optimal margin of safety [8]. They are chosen on the basis of prior knowledge, for example determining by the test if it is a new product or comes from the standard sizing codes. This approach often leads to a safe and very expensive design (high reliability for high values of safety factors and considerably high structural cost) or vice versa [7].

$$Min\, C_I(d)$$

$$\text{subject to} \begin{cases} P_{rob}[G_i(d, p, X)] \leq P_{rob}^c, \quad i = 1, ..., I \\ h_j(d) \geq 0, \quad j = 1, ..., J \end{cases} \tag{2}$$

The objective of the design optimization method is to design structures both economical and reliable where the solution reduces the structural weight in the regions non-critical. This requires design optimization with a high level of confidence [8]. To explicitly account for parameter uncertainties, the optimization procedure can be coupled with the reliability analysis. This type of analysis requires a very high calculation time and leads to problems convergence. Indeed, solving the reliability problem involves a large number of calls of the performance function in the space of the random variables, or, the search for optimal design parameters requires at each iteration a limit state function assessment. The methods developed in Reliability-based design optimization involves nesting a reliability analysis in an optimization loop [9]. These methods are called double loop (double loop) or nested loop approaches while the outer loop allows to look for the optimal parameters and the inner loop performs the reliability analysis configuration selected. In this case, the reliability analysis is performed by the Performance measurement that verifies PMA (Performance Measure) target reliability Approach or the Reliability Index Approach (RIA) obtained by approximating the limit state around the most likely failure point MPFP (Most Probability Failure Point). Another alternative approach to solving this optimization problem is the decoupled sequential approach (sequential decoupled approaches) [8]. Each decoupling method uses a specific strategy to separate the optimization loops and then solve them sequentially until a certain convergence criterion is achieved. Among these methods, we can mention: TAM (traditional approximation method, SFA (Safety Factor Approach), SORA (Sequential Optimization and Reliability Assessment), SAP (Sequential Approximate Programming), etc. In addition, there are other so-called single-level approaches (Monolevel Approaches) which imply an entire reformulation of the RBDO problem into a problem Equivalent DDO for simple and efficient resolution using algorithms classical optimization [6]. Among the methods based on this approach: one method is to replace the reliability analysis by certain criteria of deterministic optimality KKT (Karush-Kuhn-Tucker) on the optimum (i.e., to impose it as a stress in the outer loop). Thus, they are executed (design optimization and reliability calculation) simultaneously and independently. A second method, the approximation of the statistical moments of the limit state function by a serial development of Taylor. This is the Approximate Moments Approach (AMA).

4 Problem Formulation

4.1 Wind Turbine Performance

As mentioned in [2], the power produced by a wind turbine depends mainly on the characteristics of the machine itself. These characteristics are established by the manufacturer under specific conditions according to the procedure prescribed in IEC 61400-12-1 for an external weather measurement mast or according to the procedure prescribed in IEC 61400-12-2 for wind turbines having a nacelle anemometer. However, different factors from one site to another such as the terrain topography and climatic conditions such as atmospheric stability and air density directly influence the wind shear profile and by the the performance of a wind turbine

The general form of the Weibull probability density function is given by:

$$f(u) = \frac{k}{c}\left(\frac{v}{c}\right)^{k-1} \exp\left[-\left(\frac{v}{c}\right)^{k}\right] \tag{3}$$

where c is the scale parameter, k is the shape parameter and v is the wind speed. The accuracy of wind energy estimation is influenced by a number of issues, including wake effect brought on by interactions in the wind farm, component degradation, and mechanical/electrical losses that alter the predicted performance of wind turbines. More focus should to be given attention to the variables used to measure the wind climate, including average wind speed, wind shear, air density, and turbulence [2]. Generally, one or more meteorological masts are used to measure wind speed. A number of extrapolation and long-term correction methods have been put out in the literature to determine the variation of wind speed with height at the target site from a reference site.The law of power is an empirical law that is not based on any physical law. The formula was proposed by Hellman as follows:

$$V(z) = V(z_{ref})\left(\frac{z}{z_{ref}}\right)^{\alpha} \tag{4}$$

z_{ref} is the height of the mast hub, $V\ (z_{ref})$ is wind speed at the reference hub of the studied site, α is the Hellman shear coefficient that can be deduced from measurements made at two different heights that can be calculated as follows:

$$\alpha = \frac{1}{\log(z * z_{ref}/z_0)} - \frac{0.0881 * \log(V(z_{ref})/6)}{(1 - 0.881 * \log(z_{ref}/10))} \tag{5}$$

where z_0 is the ground roughness.

To depict relationship between wind power turbine and hub height wind speed without any other details about WTG dynamics. The power curve is expressed as follows:

$$P_{elec}(v) = \mathrm{P}_{rd}\begin{cases} g(v) & u_c < v < u_r \\ 1 & u_r < v < u_o \\ 0 & otherwise \end{cases} \tag{6}$$

where P_r is the rated power, $g(v)$ is the non-linear function that represents the wind turbine output between the cut-in and rated speed; cut-in speed u_c, cut-out speed (or furling) u_o and rated speed u_r. Overall, power curves $g(v)$ of most of the available wind turbine generators from different manufactures are expressed with u_c, u_o, u_r and v the wind speed. Inspiring by the work of Jangamshetti S. H. et al. [10] and M. EL-Shimy [11], a normalized power curve model in rated power and wind speed can be fixed as depicted in Fig. 1. This result is based on the fact that most of models for wind power curve underestimate the rated wind speed and the power in the partial charge between cut in and rated wind speed. To overcome this, an hypothetical rated wind speed can be defined corresponding to 99% of the rated output power of a commercial WTG and is about 90% of the measured rated wind speed. Moreover, the cut-in and the cut-out wind speed can be related to the normalized rated wind speed by the relations $u_c = q.u'_r$ and $u_o = p.u'_r$ with p the ratio of cut-in wind speed and rated wind speed < 1 and q the ratio of cut-out wind speed and rated wind speed > 1 as. Furthermore, this normalization allows investigating and select the optimal values based on wide range of commercial WTGs instead of making several calculations.

$$\begin{cases} P_{rd} = 0.99 * P'_{rd} \\ u'_r = 0.90 * u_r \\ u_i = q * u'_r \\ u_o = p * u'_r \end{cases} \tag{7}$$

As underlined by N. Aghbalou et al. in [2], the characteristics of the wind turbine itself largely determine how much electricity can produce. These characteristics are determined by the manufacturer under certain settings in accordance with the method for external meteorological data or nacelle anemometer i.e. the mast. Climate conditions, terrain, and atmospheric stability are additional contributing elements that affect the wind shear profile and air density. The probability of functioning p_o, or the likelihood of generating electrical power, under the assumption that the two operating modes of the wind turbine are independent, can be expressed as follows:

$$p_o = 1 - (1 - p_{fc})(1 - p_{pc}) \tag{8}$$

where p_{fc} the probability of operating at full charge and p_{pc} the probability of operating at partial charge.

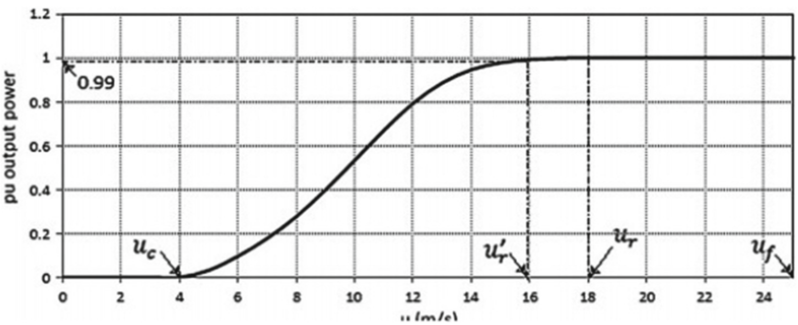

Fig. 1. Normalized power Curve.

According to a failure criterion in structural engineering, a structure is safe if the applied stress stays below the component's strength. The difference between the component's strength and the applied stress can be viewed as the performance function at any given time:. In our study, one can place the normalized produced power $P'_{elec}(.)$ of the *WTG* of Eq. 6 in the strain side and the critical power $P_c(.)$ in the strength side. The performance function may be expressed as:

$$G(X_{turbine}, X_{wind}) = P'_{elec}(X_{turbine}, X_{wind}) - P_c(.) \quad (9)$$

where $X_{turbine} = (H, p, u_r^{'}, q, P_r^{'})$ is vector of random variables related to the *WTG* specification and $X_{wind} = (V_m, c, k, \alpha, z_0)$ is vector of random variables related to the wind climate. Then, one can define the following probabilities:

$$p_{fc} = 1 - \Pr(G_1(X_{turbine}, X_{wind}) < 0) \quad (10)$$

and

$$p_o = 1 - \Pr(G_2(X_{turbine}, X_{wind}) < 0) \quad (11)$$

where $G_1(X_{turbine}, X_{wind})$ is the performance function such that the produced power $P_e^{'}{}_{lec}(.)$ of the *WTG* less than or eqals to zero $P_c = 0$. And $G_2(X_{turbine}, X_{wind})$ is the performance function such that when the produced power $P_e^{'}{}_{lec}(.)$ is less than the rated power $P_c = P_r$. The p_{pc} can be deduced from the equations equpo, Eqs. 10 and 11.

$$p_{pc} = \frac{p_o - p_{fc}}{1 - p_{fc}} \quad (12)$$

4.2 Cost Function of the Wind Power Generation

In wind power projects it is more meaningful and practical to minimize the turbine cost of wind turbine and maximizing the the amount of annual energy production (AEP). To this aim, the annual profit (ANP) [16] is an objective function takes into account both energy production and costs. Additionally, an objective of profit can consider more refined measures of the value of energy, such as time-of-day pricing where the price of electricity varies depending on the time of day it is produced. Because a primary interest of most businesses is to make money, this objective would likely be of more interest to wind power plant developers. It is calculated by converting annual energy production into annual electricity sale revenue C_e, and the annual cost is calculated by adding annual operating and maintenance costs $C_{o\&M}$ and initial capital costs *ICC*, converted into 1-year units [12]. The $C_{o\&M}$ is proportional to the AEP and can be expressed considering the P_{rd} of the wind turbine. The major assumptions when calculating ANP are as follows [12]:

- All electricity from a wind turbine is sold.
- Operating and maintenance costs are proportional to the amount of energy production.
- ICC increases linearly with hub height.
- ICC per unit capacity decreases linearly with the rated power of a wind turbine.

Then Annual Net Profit (ANP) to be maximized is expressed based on these assumptions as follows [12]:

$$ANP = (C_e - C_{o\&m})AEP - FCR * ICC \quad (13)$$

where FCR is the annual fixed charge rate. It is a factor by which the ICC is multiplied to convert the initial capital cost into an annual cost, determined by a comprehensive consideration that includes the wind turbine lifespan, interest rate, and state of capital [12]. The AEP is determined by the power characteristics and the wind distribution [12] and can be expressed using Eq. 7 as follows

$$AEP = \frac{0.99 * P'_{rd}T}{0.9^2 - q^2}\left\{\frac{e^{Cq^2V_{rn}^2} - e^{CV_{rn}^2}}{CV_{rn}^2} - (0.9^2 - q^2)e^{-Cq^2V_{rn}^2}\right\} \tag{14}$$

where C is a parameter used to simply the equation, given by:

$$C = \frac{\pi}{4}\left(\frac{Z_{ref}^{\alpha}}{v_{ref}Z^{\alpha}}\right)^2 \tag{15}$$

On the other hand, the cost of energy (COE) to be minimized is considered as the averaged turbine output energy cost and expressed as follows [13]:

$$COE = \frac{FCR * ICC + AOE}{AEP} \tag{16}$$

where the AOE is the annual operation energy that include land-lease costs, operation and maintenance (OM) wages and material, and levelized replacement costs [13]. It is expressed based on the AEP and P_{rd} as reported in [13] and [14]:

$$AOE = f(AEP, P'_{rd}) \tag{17}$$

Then based on Eqs. 13, 16 and 17, the equation to be minimized is given by:

$$COE = (C_e - C_{o\&m}) - \frac{ANP}{AEP} + f(AEP, P'_{rd}) \tag{18}$$

In the other hand, the annual net profit ANP is a standard objective in wind power energy optimization stated in [16] as follows:

$$ANP = (AEP(1 - L)) * PPA - FCR * CapEx - C_{o\&m} \tag{19}$$

In this equation, L are any losses experienced within the wind turbine implementation assumed tu be a constant value. The power purchase agreement (PPA) in this case establishes the monetary value of the generated energy. Instead of applying time of day pricing, seasonal or annual PPA modifications, or adding PPA incentives or penalties for power quality, it is assumed constant. CapEx (Capital expenditures) is the total of the depreciable expenses, which are further broken down into direct costs (system costs) and indirect costs; expenditures associated with site preparation, engineering and design, project contingencies, and upfront permitting fees [15], at the scale of a wind turbine in this case.

Based on the above proposal, the reliability based design optimization for wind turbine-site matching is formulated as follows:

$$\text{Minimise} \quad \text{COE}(\text{X}_{wind}, X_{turbine})$$

$$s.t. \begin{cases} p_o < P_{target} \\ q_l < q < q_u \\ p_l < p < p_u \\ v_r < vr_u \\ H < h_u \end{cases} \tag{20}$$

where the upper and lower values of p, q and h_u are taken from the available data base. P_{target} is a probability target to be fixed. The vectors of design $X_{turbine}$ and X_{wind} of random and deterministic variables can be fixed based on the selected WTG and measured wind speed in the candidate location.

5 Conclusion and Future Work

The purpose of this brief study is to modify the RBDO method to determine the best wind turbine for a given location while taking uncertainties into consideration. Additionally, an enhanced method that reduces sensitivity to estimated WTG curves in comparison to traditional methods is proposed. The uncertainties pertaining to wind turbine generator (WTG) features and models used to approximate the WTG power curve are also taken into consideration in this wind turbine-site matching process. These uncertainties stem from the unpredictable behavior of wind speed climate. Additionally, the energy cost was stated as a function of various costs, random and deterministic variables, and design. The approach can be used to investigate a wind power plant project and can incorporate time-varying wind speed for extended. The purpose of a feature work is to test the technique with an actual case study.

References

1. Söder, L., et al.: Review of wind generation within adequacy calculations and capacity markets for different power systems. Renew. Sustain. Energy Rev. **119**, 109540 (2020)
2. Aghbalou, N., Charki, A., Elazzouzi, S.R., Reklaoui, K.: A probabilistic assessment approach for wind turbine-site matching. Int. J. Electr. Pow. Ener. Sys. **103**, 497–510 (2018)
3. Serrano González, J., Burgos Payán, M., Santos, J.M.R., González-Longatt, F.: A review and recent developments in the optimal wind-turbine micro-siting problem. Renewa. Sustain. Ener. Rev. **30**, 133–144 (2014). ISSN: 1364–0321. visited on 20 May 2014
4. Chen, H.: Generating system reliability optimization, PhD thesis. University of Saskatchewan Saskatoon (2000)
5. Staffell, I., Green, R.: How does wind farm performance decline with age?. Renewable Energy **66**, 775–786 (2014). ISSN: 0960-1481
6. Tsompanakis, Y., Lagaros, N.D., Papadrakakis, M.: Structural Design Optimization Considering Uncertainties: Structures Infrastructures Book, Vol. 1, Series. In: Frangopol, Dan M. (ed.) Google-Books-ID: ThIfsVSdekcC. CRC Press (2008). ISBN: 978-0-203-93852-2
7. Kharmanda, G., Mohsine, A., Makloufi, A., El-Hami, A.: Recent methodologies for reliability-based design optimization. Int. J. Simul. Multidiscipl. Desi. Optimiz. **2**(1), 11–23 (2008). ISSN: 1779627X, 1779–6288