



# Boosting algorithms in energy research: a systematic review

Hristos Tyralis<sup>1,2</sup> · Georgia Papacharalampous<sup>1</sup>

Received: 7 May 2020 / Accepted: 31 March 2021 / Published online: 19 April 2021

© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2021

## Abstract

Machine learning algorithms have been extensively exploited in energy research, due to their flexibility, automation and ability to handle big data. Among the most prominent machine learning algorithms are the boosting ones, which are known to be “*garnering wisdom from a council of fools*”, thereby transforming weak learners to strong learners. Boosting algorithms are characterized by both high flexibility and high interpretability. The latter property is the result of recent developments by the statistical community. In this work, we provide understanding on the properties of boosting algorithms to facilitate a better exploitation of their strengths in energy research. In this respect, (a) we summarize recent advances on boosting algorithms, (b) we review relevant applications in energy research with those focusing on renewable energy (in particular those focusing on wind energy and solar energy) consisting a significant portion of the total ones, and (c) we describe how boosting algorithms are implemented and how their use is related to their properties. We show that boosting has been underexploited so far, while great advances in the energy field are possible both in terms of explanation and interpretation, and in terms of predictive performance.

**Keywords** Artificial intelligence · Energy forecasting · Machine learning · Renewable energy

## 1 Introduction

Applications of machine learning algorithms to energy-related problems show an exponential increase, due to the rapid developments in the field of machine learning and the growing availability of big datasets [22]. Machine learning applications become more and more common in wind energy [31, 41, 61, 91], solar energy [88, 89, 101], water resources [98], energy in buildings [80, 86, 112], renewable energy in general [82, 102], fossil fuels [70], energy demand [100] and development of materials [24] to name a few disciplines.

The utility of machine learning algorithms in solving practical problems, including energy-related ones, has been highlighted in numerous studies. For instance, they are flexible, they have high predictive power, they can be fully automated (and, therefore, they can be used in operational settings), and they have been extensively tested exhibiting good performances in real-world cases. Several open-source implementations are available for these algorithms [74], while their combination with physics-based models [74, 97] is also possible. They have been extensively benchmarked in various settings [42, 73], and their properties are largely known. Although their primary focus is to generalize in previously unseen data after proper training [13], they can also be interpretable. Their strengths have been extensively exploited in energy research [22].

Boosting (an ensemble learning algorithm) holds a dominant position among the machine learning algorithms and has been described as the “*best-of-the-shelf classifier in the world*” [40 p. 340]. This latter description was specifically referred to AdaBoost (abbreviation for adaptive boosting, with trees algorithm), a boosting variant that does not need extensive preprocessing and hyperparameter tuning to perform well [40 p. 352, 12]. The concept behind boosting originated in [83]. In this latter work, it was shown that a weak learner (i.e. a learner that performs

✉ Hristos Tyralis  
hristos@itia.ntua.gr; montchrister@gmail.com

Georgia Papacharalampous  
papacharalampous.georgia@gmail.com;  
gpapacharalampous@hydro.ntua.gr

<sup>1</sup> Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Iroon Polytechniou 5, 157 80 Zografou, Greece

<sup>2</sup> Air Force Support Command, Hellenic Air Force, Elefsina Air Base, 192 00 Elefsina, Greece

slightly better than random guessing) can be transformed to (or “boosted” to become) a strong learner (worded as the “*strength of weak learnability*” at that time). This concept was later reworded as “*garnering wisdom from a council of fools*” [84 p. 1]. AdaBoost [32–35] was the most successful from the early boosting algorithms.

Early boosting algorithms focused on classification, while their procedures included the sequential training of a weak learner by using reweighted versions of the training data. The prediction would be a weighted majority vote of the sequence of the trained learners. Understanding the properties of boosting procedures was mostly a focus of the machine learning community. A statistical view of boosting was demonstrated in [38]. This latter work showed that AdaBoost could be seen as an additive logistic regression model. In subsequent developments, boosting was proposed as the estimation of a function by optimizing a loss criterion via steepest gradient descent in function space [36], albeit [67] question this view. The latter developments allowed the delivery of interpretable procedures in supervised learning, and extended boosting to the regression case using customized loss functions. On the other hand, some recent boosting variants (e.g. extreme gradient boosting [21], light gradient boosting [51], and categorical boosting [75]) focused on increased predictive performance and speed, thereby achieving remarkable results in real case applications and forecasting competitions. The history and the evolution of boosting algorithms can be found in [17, 20, 38, 65, 78], while reviews on the developments of boosting algorithms can be found in [9, 19, 66].

Considering the recent trends in energy research, manifested into both the growing number of the related statistical and machine learning applications and the success of boosting (due to its diverse properties) in several fields, in this work we aim to:

- (a) Increase understanding of boosting algorithms for energy researchers.
- (b) Provide a detailed review of boosting applications in energy research.
- (c) Review how particular strengths of the boosting algorithms were exploited in energy studies.

To this end, we present a brief theory of the boosting algorithms, as well as their strengths and weaknesses. We systematically review approximately 150 papers published in Journals specialized in energy research and relevant applications. We conclude with take-home remarks regarding potential applications of boosting in the energy sector.

The remainder of this paper is structured as follows. Section 2 presents the theory of boosting algorithms. Section 3 presents the properties of boosting algorithms, including their strengths and weaknesses. Section 4 follows with a systematic review on the implementation of boosting algorithms in energy research, while Sect. 5 complements Sect. 4 by focusing on specific applications which do not belong to the typical mode of “train and predict” nature of boosting applications in energy research. Section 6 concludes by summarizing results and by providing recommendations for better exploitation of the boosting algorithms.

## 2 Theory of boosting algorithms

Here we present a concise theory of boosting, while detailed descriptions can be found in books focusing on machine or statistical learning [28, 40, 46] or in books specialized in ensemble learning [114] in general, and in boosting in particular [84]. An overview of ensemble learning algorithms (among which boosting is a special case) can also be found in [81], while tutorials and introductory material for boosting can be found in [63, 64, 68].

Given the extensive use of machine learning algorithms in energy research, in what follows we assume that the reader is familiar with frequently used concepts appearing in the machine learning literature. Such concepts are the supervised and unsupervised learning, regression and classification, and training and testing, among others.

Let  $\mathbf{x} = (x_1, \dots, x_p)$  be the vector of predictor variables and  $y$  be the dependent variable. Let also  $\mathbf{x}_i$  be the  $i$ th observation,  $i = 1, \dots, n$  of  $\mathbf{x}$  and  $y_i$  be the respective  $i$ th observation of  $y$ . The focus is to estimate a function  $f: \mathbf{x} \rightarrow y$ , which approximates the unknown functional dependence between  $\mathbf{x}$  and  $y$  in a supervised setting.

### 2.1 Adaptive boosting

We begin by describing AdaBoost [32–35], the most popular boosting algorithm [40 p. 337]. AdaBoost is a classification algorithm using stumps (a two terminal-node decision tree) or decision trees as base learners (weak learners), albeit other options are also available [40 p. 339, 340]. In the two-class problem, where  $y \in \{-1, 1\}$ , the AdaBoost method is formulated according to Algorithm 1.

**Algorithm 1.** Formulation of the AdaBoost algorithm, adapted from [40 p.339]

Step 1: Initialize observation weights  $w_i = 1/n$ ,  $i = 1, \dots, n$ .

Step 2: For  $m = 1$  to  $M$ :

- a. Fit a classifier  $C_m(\mathbf{x})$  to the training data by using weights  $w_i$ .
- b. Compute  $\text{error}_m = \sum_{i=1}^n w_i I(y_i \neq C_m(\mathbf{x}_i)) / \sum_{i=1}^n w_i$ , where  $I(A)$  is the indicator function, which is equal to 1 when  $A$  materializes and 0 otherwise.
- c. Compute  $a_m = \log((1 - \text{error}_m) / \text{error}_m)$ .
- d. Set  $w_i$  equal to  $w_i \exp(a_m I(y_i \neq C_m(\mathbf{x}_i)))$ ,  $i = 1, \dots, n$ .

Step 3: Predict  $C(\mathbf{x}) = \text{sign}(\sum_{m=1}^M a_m \cdot c_m(\mathbf{x}))$  (i.e. by majority voting), where  $\text{sign}$  denotes the sign function.

The main idea behind AdaBoost is to iteratively train a weak learner (in this case  $C_m(\bullet)$ ) to reweighted versions of the data. In each iteration, larger weights are assigned to those data that were misclassified in the previous iteration (Algorithm 1, Step 2d). This reweighting aims at forcing the algorithm to correctly classify observations that are hard to predict. Therefore, AdaBoost focuses on manipulating the data rather than the weak learner. This procedure transforms a weak classifier to a strong one.

## 2.2 Gradient boosting

AdaBoost and other early boosting algorithms were developed by the machine learning community with a focus on predictive performance. The value of boosting for statistical inference was recognized by [38], while [36] derived the gradient-descent-based formulation of boosting methods (see Algorithm 2 for this formulation [16]). This formulation is directly linked to the statistical view of boosting.

**Algorithm 2.** Formulation of the gradient boosting algorithm, adapted from [16, 36, 68]

Step 1: Initialize  $f_0$  with a constant.

Step 2: For  $m = 1$  to  $M$ :

- a. Compute the negative gradient  $g_m(\mathbf{x}_i)$  of the loss function  $L$  at  $f_{m-1}(\mathbf{x}_i)$ ,  $i = 1, \dots, n$ .
- b. Fit a new base learner function  $h_m(\mathbf{x})$  to  $\{(\mathbf{x}_i, g_m(\mathbf{x}_i))\}$ ,  $i = 1, \dots, n$ .
- c. Update the function estimate  $f_m(\mathbf{x}) \leftarrow f_{m-1}(\mathbf{x}) + \rho h_m(\mathbf{x})$ .

Step 3: Predict  $f_M(\mathbf{x})$ .

Here one has to first decide which metric  $L(y, f(\mathbf{x}))$  should be optimized for estimating the function  $f$  (see Algorithm 2, step 2a). In the boosting terminology, this

metric is usually called “loss function”. Examples of loss functions can be found in Sect. 2.4. Second, one has to decide which base learner  $h$  to use (see Algorithm 2, step 2b). Examples of base learners can be found in Sect. 2.5. The selection of loss functions and base learners largely depends on the research question.

Boosting works by adding in each iteration, the newly fitted base learner  $h(\mathbf{x}, \theta_m)$  (see Algorithm 2, step 2c) to the previously estimated  $f_{m-1}$ . The base learner  $h(\mathbf{x}, \theta_m)$  has been fitted in step 2b (see Algorithm 2) to the negative gradient. The stopping iteration parameter  $M$  is important, since few iterations lead to unsatisfactory training (termed as “underfitting”), while too many iterations may result in overfitting (i.e. nice fitting to the training data but high probability of failing to provide good predictions of the test data). The step-length factor  $\rho$  ( $0 < \rho \leq 1$ ) is less important, while small  $\rho$  values (e.g.  $\rho = 0.1$ ) are usually satisfactory [16].

Some important remarks should be made on Algorithm 2. These are the following:

- (a) The concept of boosting is to train new base learners, aiming at correlating them with the negative gradient of the loss function [68]. To get an idea, the negative gradient of the loss function  $L_2$  is the error function; therefore, minimization of the loss function  $L_2$  is achieved by sequentially training the base learners to the errors of the model.
- (b) The term “stagewise additive expansion”, used by [36], does not imply that the model is additive in the predictor variables. Instead, it implies that the model is additive with respect to the sequentially trained functions [16].
- (c) Regarding the estimation of the optimal value of  $M$ , the most common approaches are based either on an early stopping rule within a cross-validation setting (by optimizing the empirical risk, i.e. the loss function averaged over the observations) or on the use of model selection criteria (e.g. the Akaike information criterion, Bayesian information criterion) [19]. On the other hand, [105] suggest that the boosting algorithms should be used without early stopping, and [65] suggest that overfitting resistance of the AdaBoost algorithm is an artefact of using different criteria for training and testing the algorithm.
- (d) Efforts on explaining the success of boosting algorithms in practical applications are their margin interpretation by [85], and their similarity to random forests as self-averaging interpolating algorithms [105].

- (e) The similarity between AdaBoost and gradient boosting has been clarified in [65]. The latter study notes that the primary idea of AdaBoost is to up-weight misclassified observations. On the contrary, gradient boosting focuses on identifying and modelling large errors resulting from the previous iterations.
- (f) Stochastic gradient boosting [37] is a development of gradient boosting. In each iteration of stochastic gradient boosting, a random subsample of the training data is used for training. This procedure increases robustness.
- (g) When implementing gradient boosting algorithms with decision trees, the relative importance of predictor variables in predicting the response variable can be estimated [36]. Relative importance allows the ranking of the predictor variables with respect to their importance in predicting the response variable. Partial dependence plots, which show the effect of a variable in predicting the response variable after considering the average effects of the other predictor variables, are also possible [36].

### 2.3 Important developments of statistical boosting

During the evolution of boosting algorithms from machine learning to statistical modelling [65], numerous boosting extensions were suggested and new possibilities appeared for improving modelling procedures. Important developments include component-wise gradient boosting [18]. This type of boosting allows the modelling of each predictor variable using diverse base learners. For instance, one could model the first predictor variable  $x_1$  using a decision tree, and the second predictor variable  $x_2$  using a linear model. Later developments of this approach include procedures for high-dimensional problems [14, 15, 44]. The generic framework was presented by [16].

The procedure of component-wise gradient boosting allows the modelling of different effects using different base learners, thereby resulting in interpretable models. Furthermore, the variable selection procedure is inherent to the framework, since at each iteration a single base learner (corresponding to a single predictor variable) is selected [64]. The framework can also be used for automatic model selection, in a similar way to that for variable selection [64]. Furthermore, using a single base learner for modelling blocks of predictor variables is also an option [64]. Related software implementing these developments can be found [43, 45].

Generalized additive models (GAM) assume an exponential family distribution for the response variable, while

the mean of the response variable is modelled as a function of the predictor variables. Likelihood-based boosting [96] was designed to circumvent problems appearing in generalized additive modelling (e.g. estimation in cases with a high number of predictor variables) by adopting stage-wise training of the base learners. This is done by maximizing an overall likelihood in each iteration [65].

### 2.4 Loss functions

Statistical boosting algorithms are flexible enough so that they can be optimized using diverse loss functions. The selection of a loss function depends on the problem that the researcher intends to solve. A non-exhaustive list of commonly used loss functions can be found in [43, 68]. This list includes the loss functions  $L_1$ ,  $L_2$ , the Huber loss function and the negative Gamma log-likelihood function with logarithmic link. These loss functions could be used when one aims to provide point predictions. The quantile and the expectile loss functions can be used when one aims to deliver probabilistic predictions. For categorical responses, one can use the binomial, the AdaBoost and the area under the receiver operating characteristic curve (AUC) loss function. Loss functions for survival analysis (the negative partial log-likelihood for Cox models) and counts data (the Poisson loss) also exist.

### 2.5 Base learners

The selection of the base learner depends on the nature of the problem that one wishes to solve. A non-exhaustive list of base learners can be found in [43, 68]. For instance, base learners can be linear models (ordinary linear regression, ridge penalized linear regression, random effects), smooth models (p-splines, radial basis functions), decision trees (tree stumps, decision trees in general), Markov random fields for spatial problems and wavelets, while relevant applications include time series models [87].

### 2.6 New techniques with high predictive performance

Recently, the interest in boosting algorithms from a predictive performance perspective has been revived. This revival has been made possible by the introduction of the extreme gradient boosting (XGBoost) algorithm [21]. This latter algorithm has been proved as a highly competitive one in machine learning competitions. The primary reason for its success is its scalability (i.e. its ability to handle cost-efficiently more and more workloads). Following the success of XGBoost, several new competitive algorithms that are based on similar ideas, were introduced. Such algorithms are light gradient boosting machine (LightGBM

[51]), categorical boosting (CatBoost [75]), and accelerated gradient boosting (AGB [8]), among others.

### 3 Properties of boosting algorithms

Boosting is accompanied by several strengths stemming from its own nature, while it also retains the strengths of its base learners. For its better exploitation, it is important to know and understand these strengths, as well as the weaknesses of the algorithm. The properties of the algorithm are largely known from theoretical studies. They are also empirically understood through the extensive use of the algorithm during the last two decades. A summary of these properties can be found in the following.

#### 3.1 Ease of use

Practitioners need algorithms that are easy to use. Ease of use may be simplicity of the algorithm, as well as its software implementation.

1. Boosting is an “*off the shelf*” algorithm, i.e. it does not need extensive preprocessing and hyperparameter tuning to perform well [12, 40 p. 352].
2. Several of its variants have been implemented in open-source software; therefore, they are freely available. Existing software is flexible enough so that customized models can be implemented [64, 65]. The modular nature of the boosting algorithms and their software implementations makes it even more appealing [16, 43, 45, 63, 66].

#### 3.2 Generalization ability

Algorithms should be accurate enough, so that for given data they provide the best possible results.

3. Boosting algorithms are more accurate than single decision trees, and very accurate in general [36, 63, 65, 68].
4. There are several regularization procedures to avoid overfitting, e.g. subsampling, shrinkage and early stopping (through cross-validation) [16, 63, 65, 66, 68], albeit boosting itself is also resistant to overfitting [16, 18, 65].
5. Boosting algorithms are robust against multi-collinearity issues [63]. This property is especially important in cases of high-dimensional settings.
6. Boosting improvements in predictive performance are rapid at the initial iterations, while they level off to slower increments. Therefore, boosting can be used for a preliminary approach to the problem, aiming at

understanding whether there is potential for further improvements, and later be implemented more intensively [36].

7. The first glance offered by boosting can be beneficial to understand the problem at hand and implement other algorithms later [36].

#### 3.3 Computational complexity and speed

Speed is another property required by algorithms, especially in operational settings in which big datasets have to be processed.

8. They are memory consuming [68].
9. They are time-consuming, due to the number of iterations needed for proper training [36, 63, 66].
10. They can be time-consuming when training, since sequential fitting is required in principle; therefore, benefits from parallelization are not always possible [66, 68].
11. They are time-consuming when they predict, because all base learners must be evaluated. That would be a problem, e.g. in online tasks [68].
12. Recent advances have merely addressed issues of parallelization [65, 66].
13. Boosting can be stopped, and the previous computations can be used when new data arrive, albeit this solution would be suboptimal compared to the case that the full dataset is used for training [36].
14. Recent advances of boosting in the machine learning community have managed to provide boosting algorithms (e.g. extreme gradient boosting, categorical boosting, light gradient boosting [21, 51, 75]) with unprecedented power and speed.

#### 3.4 Flexibility

Properties mentioned in Sects. 3.1–3.3 are important for practitioners. On the other hand, the flexibility of boosting algorithms has led to several theoretical developments, introducing variants that may also be of practical use.

15. Because of its flexibility, boosting can be considered as a framework, thereby giving the opportunity to the user to define his/her model for a specific task [16, 64–66, 68]. Several custom models are presented in the following.
16. Boosting algorithms can be nonparametric, semi-parametric or parametric (based on the choice of the base learner and the type of the dependent variable or parameter), depending on the specific requirements of the problem [16, 63–65, 68, 96].



17. Boosting algorithms can be combined with various base learners; thus, they can solve a variety of problems, ranging from typical machine learning problems (regression, classification) to less common ones (time series analysis, spatial analysis, density analysis, survival analysis, modelling of longitudinal data, functional regression) [16, 18, 43, 45, 63–66, 68, 87].
18. Diverse types of effects for each predictor variable can be modelled and estimated simultaneously (e.g. linear effect for one predictor variable and a decision tree for the second one) [43, 45, 63–65, 68].
19. Likelihood-based boosting can be used to estimate parameters in a likelihood setting [64, 65, 96].
20. Boosting can be combined with different loss functions (see Sect. 2.4 for a list of representative loss functions) [16, 18, 43, 45, 63–66, 68].

### 3.5 General properties

Besides being flexible, boosting algorithms inherit several other properties.

21. Boosting inherits the appealing properties of the base learners (usually decision trees), while it mitigates the negative ones [36].
22. It can handle missing values when using decision trees as base learners [36].
23. It is invariant under strictly monotone transformations of the predictor variables when using decision trees as base learners [36].
24. It is robust, not being sensitive to outliers and heavy-tailed distributions of the predictor and the dependent variables when using decision trees as base learners [36].
25. It can be regarded as an optimization tool [44, 66], e.g. by combining algorithms in stacking frameworks and optimizing the combination with respect to a given loss function.

### 3.6 Predictor variable selection

In high-dimensional problems, it is essential to screen the most important predictor variables, with the aim to obtain a parsimonious model. Boosting algorithms can perform variable selection.

26. Boosting algorithms can generalize well in the presence of redundant predictor variables [18, 36, 65].
27. They can be used in high-dimensional problems (in which predictor variables are far more than the observations). In these problems, a variable-selection

procedure is automatically implemented (in case that additive base learners are used) [16, 18, 43–45, 63–66, 68, 96].

28. In addition to their use in variable selection problems, they can also be used for model selection [16, 63–66]. For instance, predictor variables can be modelled by diverse models (see Sect. 2.3), while an automatic procedure retains the most useful ones.

### 3.7 Interpretability

An explanation of the concept of interpretability can be found in [46 p. 25]. For instance, a linear model is highly interpretable, since one can understand easily how the predictor variable is associated with the dependent variable. As a model becomes more flexible (i.e. with a wider range of possible shapes to estimate  $f$ , see Sect. 2), it becomes less interpretable (see Fig. 2.7 in [46]). Boosting algorithms are more flexible, e.g. compared to linear models; however, they can retain some degree of interpretability.

29. Boosting algorithms can be more interpretable compared to black-box machine learning models, due to their recent evolution resulting in statistical models [43, 45, 63–66] (see Sect. 2.3).
30. Depending on the base-learners used in additive modelling (linear effects, splines and more), they can be interpretable. This holds since the effect of each variable can be estimated independently [43–45, 63–66, 68].
31. Error estimates for resulting effects are not available in principle. Therefore, confidence intervals for effects cannot be estimated directly [63].
32. Boosting algorithms are less interpretable than single decision trees [36, 68].
33. Interactions of the predictor variables can be modelled [44, 68]. In regression settings, two predictor variables interact when the effect of the first predictor variable depends on the second predictor variable and vice versa [11].
34. Relative importance of the predictor variables can be computed; therefore, boosting algorithms can be interpretable [36, 68] (see Sect. 2.2, §(g) for description of relative importance).
35. Boosting can be used for computing partial dependence plots, thereby exploiting some of its potential for interpretability [36, 68] (see Sect. 2.2, §(g) for description of dependence plots).

### 3.8 Summary

To summarize, boosting algorithms have several appealing properties; however, there is no free lunch in algorithmic modelling, i.e. proposed boosting improvements are accompanied by some concessions. These mostly include computational complexity and speed (see properties (8)–(11)), and issues of interpretability compared to simpler models in general (see Sect. 3.7).

Depending on the specific application, the researcher could understand what would be the expected outcome of an implemented boosting algorithm, based on its properties. The next section summarizes typical problems in the field of energy that were addressed using boosting algorithms, as well as what class of boosting algorithms were used. As a general rule of thumb, our recommendation is to use the latest boosting algorithms (XGBoost, LightGBM, CatBoost; see Sect. 2.6) when the focus is predictive performance. However, for problems in which interpretability of the model is required for taking better decisions, one should switch to statistical boosting (see Sect. 2.3). A discussion on the typical problems met in energy research and possible new directions can be found in Sect. 6.

## 4 Quantitative analysis of boosting applications in energy research

Here we analyse articles that include applications of boosting algorithms in energy research. To this end, we conducted a literature survey using the Scopus database (including papers appearing until 2020-02-29). In particular, we searched for papers meeting three conditions simultaneously. Specifically, the papers should be (a) published in the subject area of Energy (also including papers from the traditional Energy Journal, which is not assigned in this subject area), (b) published until year 2019, and (c) citing at least one of the following 11 papers (the number of citations is in parentheses): [16] (0), [21] (34), [32] (5), [33] (4), [35] (28), [36] (98), [38] (11), [51] (6), [75] (1), [83] (11), [96] (0). We consider these latter papers as key papers, as they are related either to the machine learning view of boosting [32, 33, 35, 83], or the statistical view of boosting [16, 36, 38, 96], or the new generation of algorithms with high predictive performance [21, 51, 75]. The original dataset (composed by papers meeting the three conditions above) includes 202 papers. The final dataset includes 157 papers (see the list of publications uploaded as supplementary material), i.e. those that remained after excluding (from the original dataset) papers that do not include boosting applications. We consider that our sample is representative of energy-related-boosting studies.

Journals not belonging to the subject area of Energy of the Scopus database may occasionally publish papers related to Energy; however, we consider that including those Journals in our analysis would not result in significant improvements.

We thoroughly examined the 157 finally selected papers. In the remainder of Sect. 4, we present a quantitative analysis of our findings. Moreover, Sect. 5 complements Sect. 4 by presenting a qualitative analysis focusing on some special cases. The latter could not be revealed by our quantitative analysis.

### 4.1 General evaluation

To understand the main issues addressed when using boosting algorithms in energy research, we created the word cloud of Fig. 1. Figure 1 is a visual representation of the frequency of words appearing in the titles and keywords of the 157 examined papers; thus, it can serve as a basis for the rest of our analysis. The word cloud has been created by grouping similar words under a single term. For instance, “prediction”, “predictive” and “predicting” are represented by the term “prediction”.

One can conclude that the most frequently examined themes are related to “forecasting” (appearing 90 times), “prediction” (appearing 69 times), “regression” (appearing 18 times), “analysis” (appearing 25 times), “data mining” (appearing 30 times) and “modelling” (appearing 22 times) in general. Of particular importance is the frequent appearance of the term “probabilistic” (appearing 21 times), which is mostly connected to forecasting (e.g. probabilistic forecasting). We should also emphasize that most studies focus on methods and applications (with the word “using” (appearing 37 times) appearing frequently, while the terms “method” (appearing 36 times) and “approach” (appearing 18 times) are also frequently met). The frequent use of the term “machine learning” (appearing 79 times) implies that the authors perceive boosting as a machine learning algorithm, rather than as a statistical one.

Terms related to the kind of the modelled energy type also appear frequently (e.g. wind speed (appearing 20 times), wind turbine (appearing 22 times), electricity (appearing 20 times), power (appearing 40 times)). The kind of algorithms used is also emphasized with AdaBoost (appearing 18 times) and XGBoost (appearing 15 times) appearing frequently, while the term “neural networks” (appearing 37 times) is also frequently met. The presence of the latter term may be related to the fact that such applications require some kind of benchmarking or comparison, in which neural networks seem to be the most frequently used benchmark model. The term “ensemble” (appearing 30 times) also appears frequently, perhaps because boosting is an ensemble learner. Furthermore, its

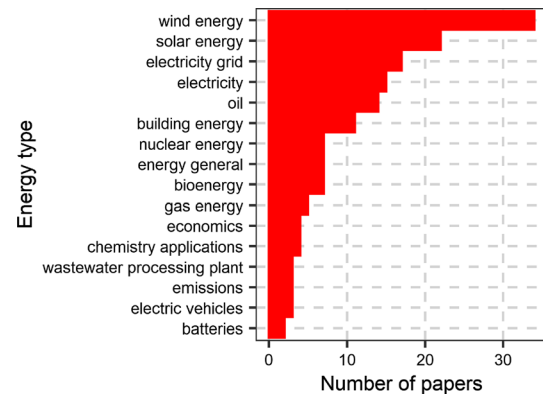




solar energy ones (see also discussion in Sect. 4.1). Electricity grids and electricity in general are also frequent themes, while almost every energy type is covered within our dataset. In the following visualizations (including Fig. 4), it may happen that the sum of the examined variables of interest may be higher than 157. For instance, in Fig. 4 some studies examine both wind energy and solar energy; therefore, they are counted in two energy types.

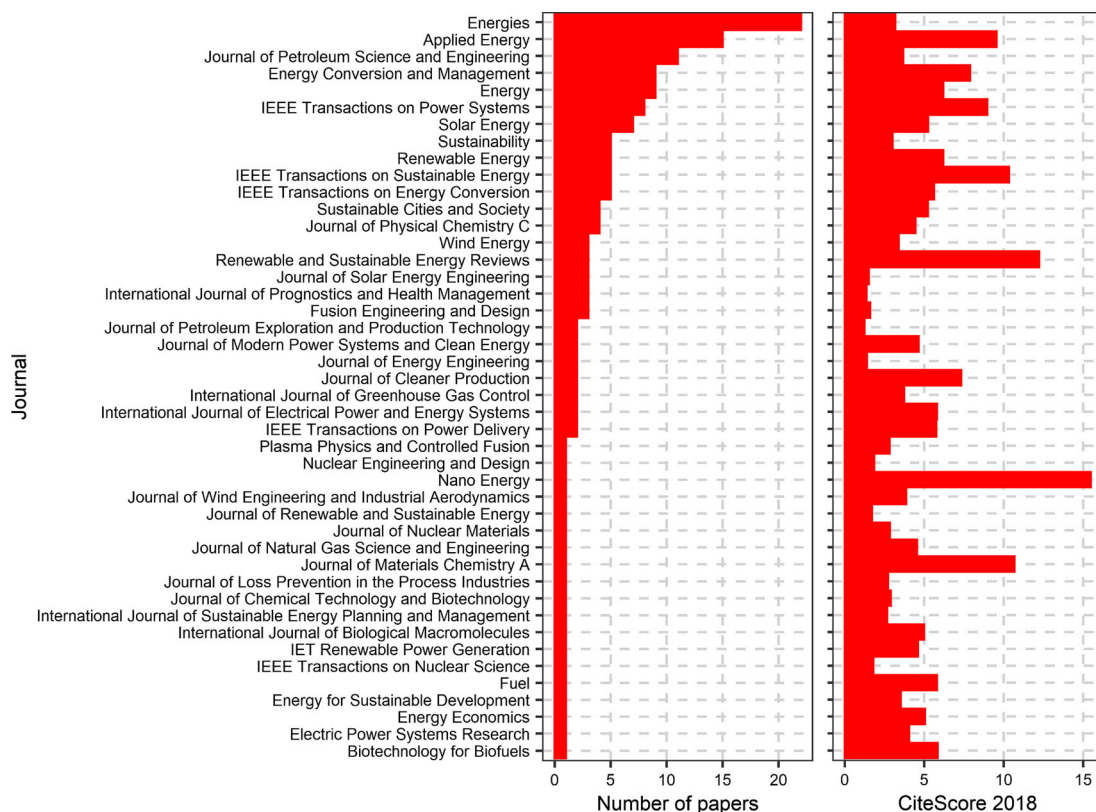
Most published papers use decision trees as base learners, followed by those using neural networks (Fig. 5). This is not surprising, given the fact that large part of the boosting literature is associated with decision trees.

Figure 6 presents the types of applications appearing in the examined articles. Most applications are related with prediction and forecasting (both applications are essentially regressions). The term “prediction” refers to the case of regression in general, while we use the term “forecasting” for studies which aim to predict future values in a regression (with time series) framework. Classification is met less frequently in the examined studies. On the other hand, relative importance is frequently computed; however, it is surprising that related concepts from statistical view of boosting (e.g. automatic variable selection and more) are not met.

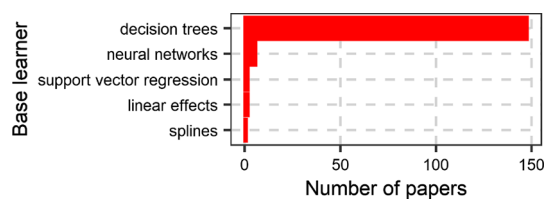


**Fig. 4** Number of published articles implementing boosting algorithms conditioned on the examined energy type

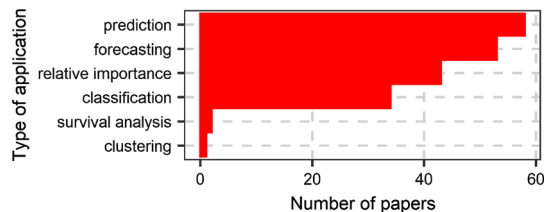
Classification applications are mainly conducted using the AdaBoost and the LogitBoost algorithms. In regression and forecasting applications, the interest is mostly on minimizing the  $L_2$  loss function, while the quantile (related to the estimation of conditional quantiles, e.g. probabilistic forecasting) and the  $L_1$  loss functions seem to be of less interest in existing applications. Gradient boosting decision trees are the main option in regression applications. Regarding the use of recent algorithms (see Sect. 2.6), XGBoost seems to dominate the field, perhaps due to its



**Fig. 3** Number of published articles implementing boosting algorithms conditioned on the Journal (on the left) and Journal CiteScores as of 2020-02-29 (on the right)



**Fig. 5** Number of published articles implementing boosting algorithms conditioned on the base learner



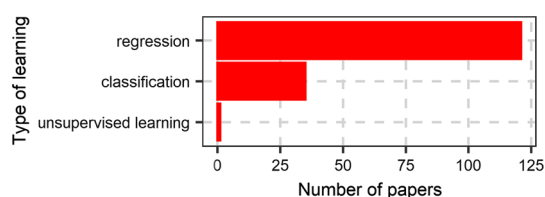
**Fig. 6** Number of published articles implementing boosting algorithms conditioned on the type of the application

hype gained by its high performance in machine learning competitions and to being introduced earlier. LightGBM and CatBoost are used less frequently, albeit we anticipate that this will change in the future.

Studies conducted in a regression context comprise the vast majority of applications followed by those conducted in a classification context. Unsupervised learning (clustering) is rarely met (Fig. 7).

The allocation of papers to Journals based on the examined energy type is presented in Fig. 8. Among these Journals, there exist some specialized in specific energy types (e.g. wind or solar energy, petroleum engineering). Multidisciplinary Journals (e.g. *Energies Journal*, *Applied Energy Journal*) are more balanced with respect to the examined energy type.

The type of application conditioned on the specific energy type is presented in Fig. 9. Studies on wind energy, solar energy and electricity focus on forecasting applications. Classification applications are mostly the focus of studies on electricity grids. Prediction applications are allocated evenly among studies examining the various energy types. Two applications of survival analysis can be found in studies examining nuclear energy.



**Fig. 7** Number of published articles implementing boosting algorithms conditioned on the type of the learning task

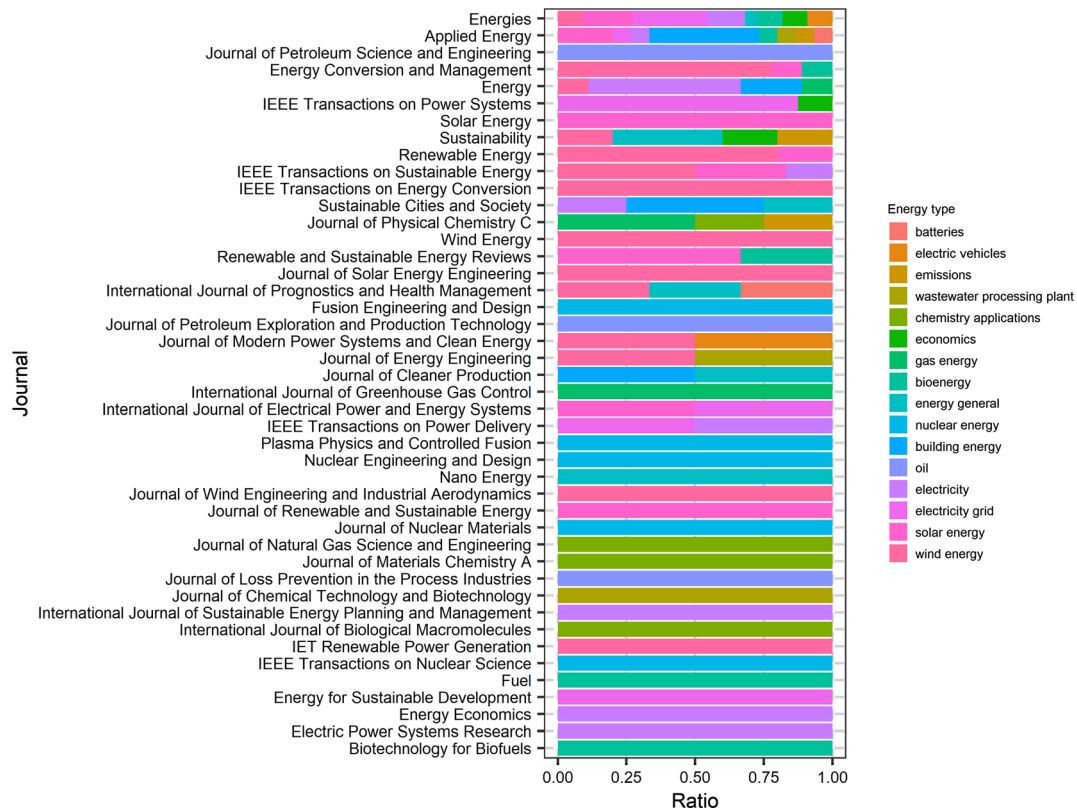
## 5 Qualitative analysis of boosting applications in energy research

The primary use of boosting algorithms is made within statistical frameworks, mostly forecasting (e.g. electricity forecasting, wind speed forecasting, solar irradiance forecasting) frameworks (see Sect. 4). Such frameworks are trivial and straightforward to apply. They necessarily involve the selection of predictor variables, which range from lagged time series variables to seasonal effects. Furthermore, forecasting frameworks usually involve a comparison of some statistical and machine learning algorithms. Such comparisons are common independently of the algorithm of focus (e.g. neural networks, support vector regression). In the following, we will refrain from analysing such frameworks since energy researchers are familiar with them. Instead, we will focus on some alternative aspects of the boosting applications.

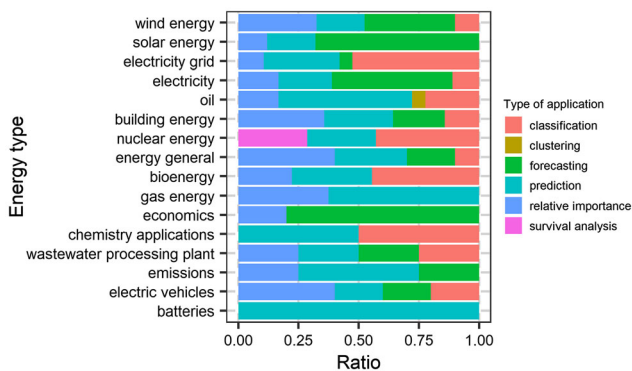
Based on the analysis of Sect. 4, we identified some special issues that need further consideration within the context of boosting applications in energy research. These themes are mostly related to the history of the first applications (Sect. 5.1), to the type of the modelled energy variable (Sect. 5.2), including the case of post-processing which differentiates in the sense that is based on a combination of physics-based and machine learning models (Sect. 5.3), as well as to several algorithmic issues (Sect. 5.4) and special cases (Sect. 5.5). In the following, some studies are analysed in more detail. In our opinion, these studies are representative for the examined special issues.

### 5.1 Early applications

The earliest applications of boosting algorithms in energy research trace back to 2009 (in spite of the fact that boosting was introduced much earlier). These applications focus on wind energy, with the group of Kusiak and co-authors being involved [53–56, 113]. (Note that Kusiak's group is involved in 17 published papers related to boosting.) These early studies have a strong engineering content, with boosting playing a secondary role, i.e. being used as an element of a larger framework. Therefore, they deviate from subsequent publications that have a strong statistical content. The latter category of studies involves e.g. the forecasting frameworks mentioned earlier in this section. The first application in which boosting plays a primary role is again conducted by the group of Kusiak and co-authors [52]. In this latter application, boosting is used for forecasting wind speed in comparison with multiple statistical and machine learning models. Neural networks outperformed other methods in this application. The



**Fig. 8** Ratio of the examined energy types conditioned on the Journal. The respective numbers of papers per Journal are shown in Fig. 3



**Fig. 9** Ratio of the types of the examined applications conditioned on the examined energy type. The respective numbers of papers are shown in Fig. 4

earliest boosting application that is not directly related to energy addresses the issue of wastewater plant energy consumption [110].

## 5.2 Applications related to a specific energy type

### 5.2.1 Electricity

We found some interesting electricity-related studies using boosting. An application to characterizing household

features based on electricity demand data highlights the potential of boosting implementations for solving energy—socio-economic issues [6]. Multidisciplinary issues have been addressed by boosting algorithms. This has been made by examining questionnaires in cross-sectional surveys. In such surveys, the number of predictor variables is high and the problems are more complex [23, 27].

Another interesting issue addressed by boosting is related to grounding systems, and in particular grounding topology [90]. An important—yet rarely met—theme in machine learning applications is handling imbalanced datasets, i.e. cases where we aim to predict a class that is under-represented in the dataset. Such cases can be addressed by boosting algorithms, e.g. by assigning higher costs to the minority class. A related interesting application can be found in [50]. This latter study attempts to predict thunderstorm-induced power outages.

### 5.2.2 Wind and solar energy

The most frequent themes addressed by boosting algorithms are wind energy and solar energy themes. Such are mostly related to forecasting issues (earliest applications include [52] and [109], respectively); however, alternative applications are also met. Among them, we found interesting the classification application on bird recognition

around a wind farm [107]. Another interesting application regards again an imbalanced dataset. In this latter application, random under-sampling combined with boosting was implemented for detection of non-technical losses in electricity grids [4]. An application on probabilistic solar power forecasting can be found in [7].

### 5.2.3 Other energy types

Besides electricity, wind energy and solar energy, other energy types have been examined using boosting, albeit to a lesser extent. An interesting study on probabilistic electricity price forecasting (economics of energy) can be found in [3]. Other studies include building energy consumption [79], survival analysis in nuclear energy [71, 72], heat load (in which XGBoost is used as benchmark) [94], applications in chemistry [95] and applications in nanotechnology [103].

An interesting study in which boosting is combined with reinforcement learning to optimize performance in energy consumption in a wastewater treatment plant can be found in [30]. Boosting is used to substitute costly experiments in a study on irradiated metals [49]. Furthermore, some multidisciplinary studies on water-electricity demand nexus using multiple sources of data and a socio-economic study on carbon dioxide emissions can be found in [69] and [92], respectively.

## 5.3 Post-processing applications

An alternative to directly forecasting weather variables (wind speed, solar irradiance) by only using machine learning algorithms and historical information is to additionally exploit information from numerical weather predictions (which are relatively accurate in short-term horizons) within post-processing frameworks. Post-processing usually improves forecasts, while it can also transform point forecasts to probabilistic forecasts. Given its high predictive accuracy and its ability to provide probabilistic predictions, boosting is suitable for post-processing applications. Such applications include [2] and [57], among others. The former application example combines numerical weather predictions and provides probabilistic forecasts, and the latter application example uses satellite data and post-processing of numerical weather forecasts for short-term forecasting of solar irradiance. The works by [5] and [10] are also of interest. These works compare multiple methods, including boosting, for probabilistic post-processing of solar radiation and energy consumption/production, respectively.

## 5.4 Algorithmic issues

### 5.4.1 XGBoost and related algorithms

An extended comparison of XGBoost with other models (including physics-based) in global solar radiation modelling can be found in [99]. XGBoost has also been combined with a time series model (to form a hybrid model) for energy supply security forecasting [58]. Finally, one study implements XGBoost, LightGBM and CatBoost to model non-technical loss detection of an energy system [76].

### 5.4.2 Stacking frameworks

Boosting can be regarded as an optimization tool; therefore, it can be effectively used in stacking frameworks by combining several algorithms and by optimizing their results with respect to a loss-function. Such studies include combinations based on AdaBoost for wind speed prediction [60], boosting for short-term electricity consumption forecasting [26], and XGBoost for gearbox fault prediction of wind turbines [108] and short-term photovoltaic power output prediction [115].

### 5.4.3 Large-scale studies and big datasets

Of particular importance are large-scale studies using multiple algorithms, studies using big datasets and forecasting competitions. Such empirical studies can provide understanding of the properties of the involved algorithms. They are rarely met in energy research, perhaps due to limited data availability. In [1], XGBoost is used for combining multiple physics-based models for solar irradiance modelling in 54 stations located worldwide. Multiple boosting algorithms are compared in forecasting electricity demand using multiple datasets [111]. Multiple boosting methods are also used in a building energy application [25]. Moreover, multiple algorithms are compared for long-term global solar radiation prediction in [39]. The largest study with respect to number of its compared algorithms is [106]. This latter study compares 68 methods in hourly solar forecasting. Finally, global datasets have been handled by boosting algorithms in bioenergy-related studies [47, 48].

### 5.4.4 Interpretability

Most of the examined studies focus on issues related to predictive performance, and only a few ones focus on interpretability and attempt to explain observed behaviours. The work by [29] applies model-agnostic methods to explain data-driven building energy performance models,



including XGBoost. However, the primary use of boosting algorithms is computation of relative importance (see, for example, the XGBoost application in [93]). An application involving many predictor variables can be found in [104]. This latter work focuses on understanding the properties of metal–organic frameworks.

### 5.5 Other uses of boosting algorithms

Model-based boosting has been used as a benchmark model, i.e. new models were compared against it for probabilistic electricity load forecasting [59]. Furthermore, the work by [62] applied boosting to classify counties in US states according to their involvement in green buildings projects. Finally, a new algorithm for classification, based on AdaBoost and using neural networks as base learners, has been proposed by [77] for nuclear energy systems.

## 6 Concluding remarks

Some important findings and take-home messages are presented in the following. These are related to: (a) properties of boosting algorithms, (b) a general overview of boosting algorithms in energy research, (c) the alternative views of boosting (machine learning and statistical) and their implications in energy research, and (d) the limitations of energy datasets, which affect the boosting applications and the potential of using boosting in energy applications.

Since their introduction in the early 1990s, boosting algorithms have largely progressed. The principal progress has been observed in the field of statistical boosting, thereby moving from the machine learning mind-set towards statistical thinking. On the other hand, new boosting algorithms (XGBoost, LightGBM, CatBoost, to name a few) oriented to issuing state-of-the-art predictions have appeared. The machine learning view of boosting mostly focuses on properties related to its high predictive performance and its good generalization ability. New boosting algorithms excel at such tasks. The statistical view of boosting focuses on properties related to its interpretability, and its potential to be applied by using a diverse set of base learners, and variable and model selection. On the other hand, boosting algorithms have some weaknesses. For instance, they can be time- and memory-consuming. Recent developments try to address these issues.

Papers using boosting algorithms in energy research are devoted to a wide range of themes, including renewable energy, electricity, fossil fuels, nuclear energy and bioenergy themes, among others. Papers in energy research are mostly oriented in providing accurate predictions and forecasts. The vast majority of the examined papers are

related to some type of data-driven forecasting. Such applications have become trivial, since they require a procedure of selecting predictor variables and then algorithmic training. Such procedures have been automated in recent software developments.

Given the primary focus in energy research, adoption of the new boosting algorithms (XGBoost, LightGBM, CatBoost) has been speeded. On the other hand, applications with statistical orientation and focus on interpretation (as defined by the statistical view of boosting) have remained limited. The latter conclusion should be considered along with the fact that most of the applications use relative importance metrics associated with boosting algorithms with decision trees as base learners for interpreting fitted models. While relative importance metrics are well known in the machine learning community (e.g. similar metrics exist for random forests), interpretability-related properties originating from statistical boosting developments seem to be neglected.

Energy research is limited to the use of single time series, while it is a common practice to examine a few cases. Perhaps limited data availability prohibits larger applications (albeit a few exceptions have been noticed). Furthermore, data in multiple dimensions (e.g. spatiotemporal data) are even more limited. Consequently, boosting applications are limited to training a simple model, while the progress in combining multiple types of models (for instance time series and spatial models) has not been exploited. In addition, high-dimensional datasets (e.g. datasets with numerous variables) are also limited in energy, thereby further limiting the implementation of boosting models in energy research.

We anticipate that the increasing availability of diverse big datasets with multiple dimensions and a large number of features, combined with multidisciplinary approaches, will redefine the scope of boosting applications in energy, from a simple predictive algorithm to a more advanced statistical framework. Some exceptions from the observed general patterns can provide some new pathways for fully exploiting the properties of boosting algorithms. Such exceptions are, for example, post-processing applications in which boosting algorithms are combined with physics-based models to improve prediction accuracy. Since boosting can be regarded as an optimization tool, stacking frameworks (in which boosting is used to combine multiple models) have also been observed. Boosting applications using big datasets and multidisciplinary boosting applications (using datasets obtained from multiple sources and scientific disciplines) have also been conducted. These applications can serve as a guide for future research.

**Electronic supplementary material** The online version of this article (doi:<https://doi.org/10.1007/s00521-021-05995-8>) contains supplementary material, which is available to authorized users.

**Acknowledgements** We thank the Associate Editor and the Reviewers for their constructive comments and suggestions, which helped us to substantially improve the manuscript.

## Declarations

**Conflicts of interest** We declare no conflict of interest.

## References

- Aler R, Galván IM, Ruiz-Arias JA, Gueymard CA (2017) Improving the separation of direct and diffuse solar radiation components using machine learning by gradient boosting. *Sol Energy* 150:558–569. <https://doi.org/10.1016/j.solener.2017.05.018>
- Andrade JR, Bessa RJ (2017) Improving renewable energy forecasting with a grid of numerical weather predictions. *IEEE Trans Sustain Energy* 8(4):1571–1580. <https://doi.org/10.1109/TSTE.2017.2694340>
- Andrade JR, Filipe J, Reis M, Bessa RJ (2017) Probabilistic price forecasting for day-ahead and intraday markets: beyond the statistical model. *Sustainability* 9(11):1990. <https://doi.org/10.3390/su9111990>
- Avila NF, Figueroa G, Chu C-C (2018) NTL detection in electric distribution systems using the maximal overlap discrete wavelet-packet transform and random undersampling boosting. *IEEE Trans Power Syst* 33(6):7171–7180. <https://doi.org/10.1109/TPWRS.2018.2853162>
- Bakker K, Whan K, Knap W, Schmeits M (2019) Comparison of statistical post-processing methods for probabilistic NWP forecasts of solar radiation. *Sol Energy* 191:138–150. <https://doi.org/10.1016/j.solener.2019.08.044>
- Beckel C, Sadamori L, Staake T, Santini S (2014) Revealing household characteristics from smart meter data. *Energy* 78:397–410. <https://doi.org/10.1016/j.energy.2014.10.025>
- Bessa RJ, Trindade A, Silva CSP, Miranda V (2015) Probabilistic solar power forecasting in smart grids using distributed information. *Int J Electr Power Energy Syst* 72:16–23. <https://doi.org/10.1016/j.ijepes.2015.02.006>
- Biau G, Cadre B, Rouvière L (2019) Accelerated gradient boosting. *Mach Learn* 108(6):971–992. <https://doi.org/10.1007/s10994-019-05787-1>
- Bickel PJ, Ritov Y, Zakai A (2006) Some theory for generalized boosting algorithms. *J Mach Learn Res* 7:705–732
- Bogner K, Pappenberger F, Zappa M (2019) Machine learning techniques for predicting the energy consumption/production and its uncertainties driven by meteorological observations and forecasts. *Sustainability* 11(12):3328. <https://doi.org/10.3390/su10023328>
- Boulesteix AL, Janitza S, Hapfelmeier A, Van Steen K, Strobl C (2015) Letter to the Editor: on the term ‘interaction’ and related phrases in the literature on random forests. *Brief Bioinform* 16(2):338–345. <https://doi.org/10.1093/bib/bbu012>
- Breiman L (1998) Arcing classifiers. *Ann Stat* 26(3):801–849. <https://doi.org/10.1214/aos/1024691079>
- Breiman L (2001) Statistical modeling: the two cultures. *Stat Sci* 16(3):199–231. <https://doi.org/10.1214/ss/1009213726>
- Bühlmann P (2003) Boosting methods: why they can be useful for high-dimensional data. In: *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*
- Bühlmann P (2006) Boosting for high-dimensional linear models. *Ann Stat* 34(2):559–583. <https://doi.org/10.1214/009053606000000092>
- Bühlmann P, Hothorn T (2007) Boosting algorithms: regularization, prediction and model fitting. *Stat Sci* 22(4):477–505. <https://doi.org/10.1214/07-STS242>
- Bühlmann P, Hothorn T (2007) Rejoinder: boosting algorithms: regularization, prediction and model fitting. *Stat Sci* 22(4):516–522. <https://doi.org/10.1214/07-STS242REJ>
- Bühlmann P, Yu B (2003) Boosting with the  $L_2$  loss. *J Am Stat Assoc* 98(462):324–339. <https://doi.org/10.1198/0162145030000125>
- Bühlmann P, Yu B (2010) Boosting. *Wiley Interdiscip Rev Comput Stat* 2(1):69–74. <https://doi.org/10.1002/wics.55>
- Buja A, Mease D, Wyner AJ (2007) Comment: boosting algorithms: regularization, prediction and model fitting. *Stat Sci* 22(4):506–512. <https://doi.org/10.1214/07-STS242B>
- Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 doi: <https://doi.org/10.1145/2939672.2939785>
- Cheng L, Yu T (2019) A new generation of AI: a review and perspective on machine learning technologies applied to smart energy and electric power systems. *Int J Energy Res* 43(6):1928–1973. <https://doi.org/10.1002/er.4333>
- Christidis P, Focas C (2019) Factors affecting the uptake of hybrid and electric vehicles in the European Union. *Energies* 12(18):3414. <https://doi.org/10.3390/en12183414>
- Correa-Baena JP, Hippalgaonkar K, van Duren J, Jaffer S, Chandrasekhar VR, Stevanovic V, Wadia C, Guha S, Buonassisi T (2018) Accelerating materials development via automation, machine learning, and high-performance computing. *Joule* 2(8):1410–1420. <https://doi.org/10.1016/j.joule.2018.05.009>
- Cui B, Fan C, Munk J, Mao N, Xiao F, Dong J, Kuruganti T (2019) A hybrid building thermal modeling approach for predicting temperatures in typical, detached, two-story houses. *Appl Energy* 236:101–116. <https://doi.org/10.1016/j.apenergy.2018.11.077>
- Divina F, Gilson A, Gómez-Vela F, Torres MG, Torres JF (2018) Stacking ensemble learning for short-term electricity consumption forecasting. *Energies* 11(4):949. <https://doi.org/10.3390/en11040949>
- Dominguez C, Orehoung K, Carmeliet J (2019) Modelling of rural electrical appliances ownership in developing countries to project their electricity demand: a case study of sub-Saharan Africa. *Int J Sustain Energy Plan Manag* 22:5–16. <https://doi.org/10.5278/ijsepm.2564>
- Efron B, Hastie T (2016) *Computer age statistical inference*. Cambridge University Press, New York
- Fan C, Xiao F, Yan C, Liu C, Li Z, Wang J (2019) A novel methodology to explain and evaluate data-driven building energy performance models based on interpretable machine learning. *Appl Energy* 235:1551–1560. <https://doi.org/10.1016/j.apenergy.2018.11.081>
- Filipe J, Bessa RJ, Reis M, Alves R, Póvoa P (2019) Data-driven predictive energy optimization in a wastewater pumping station. *Appl Energy* 252:113423. <https://doi.org/10.1016/j.apenergy.2019.113423>
- Foley AM, Leahy PG, Marvuglia A, McKeogh EJ (2012) Current methods and advances in forecasting of wind power generation. *Renew Energy* 37(1):1–8. <https://doi.org/10.1016/j.renene.2011.05.033>

32. Freund Y (1995) Boosting a weak learning algorithm by majority. *Inf Comput* 121(2):256–285. <https://doi.org/10.1006/inco.1995.1136>
33. Freund Y, Schapire RE (1995) A decision-theoretic generalization of on-line learning and an application to boosting. In: Vitányi P (ed) *Computational Learning Theory EuroCOLT Lecture Notes in Computer Science Lecture Notes in Artificial Intelligence*, vol 904. Springer, Berlin Heidelberg, pp 23–27
34. Freund Y, Schapire RE (1996) Experiments with a new boosting algorithm. In: *Proceedings of the Thirteenth International Conference on Machine Learning* 148–156
35. Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci* 55(1):119–139. <https://doi.org/10.1006/jcss.1997.1504>
36. Friedman JH (2001) Greedy function approximation: a gradient boosting machine. *Ann Stat* 29(5):1189–1232. <https://doi.org/10.1214/aos/1013203451>
37. Friedman JH (2002) Stochastic gradient boosting. *Comput Stat Data Anal* 38(4):367–378. [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2)
38. Friedman JH, Hastie T, Tibshirani R (2000) Additive logistic regression: a statistical view of boosting. *Ann Stat* 28(2):337–407. <https://doi.org/10.1214/aos/1016218223>
39. Ghimire S, Deo RC, Raj N, Mi J (2019) Deep learning neural networks trained with MODIS satellite-derived predictors for long-term global solar radiation prediction. *Energies* 12(12):2407. <https://doi.org/10.3390/en12122407>
40. Hastie T, Tibshirani R, Friedman J (2009) *The elements of statistical learning*. Springer, New York, NY
41. Helbing G, Ritter M (2018) Deep learning for fault detection in wind turbines. *Renew Sustain Energy Rev* 98:189–198. <https://doi.org/10.1016/j.rser.2018.09.012>
42. Hoffmann F, Bertram T, Mikut R, Reischl M, Nelles O (2019) Benchmarking in classification and regression. *Wiley Interdiscip Rev Data Min Knowl Discov* 9(5):e1318. <https://doi.org/10.1002/widm.1318>
43. Hofner B, Mayr A, Robinson N, Schmid M (2014) Model-based boosting in R: a hands-on tutorial using the R package mboost. *Comput Stat* 29(1–2):3–35. <https://doi.org/10.1007/s00180-012-0382-5>
44. Hothorn T, Bühlmann P (2006) Model-based boosting in high dimensions. *Bioinformatics* 22(22):2828–2829. <https://doi.org/10.1093/bioinformatics/btl462>
45. Hothorn T, Bühlmann P, Kneib T, Schmid M, Hofner B (2010) Model-based boosting 2.0. *J Mach Learn Res* 11:2109–2113
46. James G, Witten D, Hastie T, Tibshirani R (2013) *An introduction to statistical learning*. Springer, New York, NY
47. Jiang D, Ma T, Ding F, Fu J, Hao M, Wang Q, Chen S (2019) Mapping global environmental suitability for sorghum bicolor (L.) Moench. *Energies* 12(10):12101928. <https://doi.org/10.3390/en12101928>
48. Jiang D, Wang Q, Ding F, Fu J, Hao M (2019) Potential marginal land resources of cassava worldwide: a data-driven analysis. *Renew Sustain Energy Rev* 104:167–173. <https://doi.org/10.1016/j.rser.2019.01.024>
49. Jin M, Cao P, Short MP (2019) Predicting the onset of void swelling in irradiated metals with machine learning. *J Nucl Mater* 523:189–197. <https://doi.org/10.1016/j.jnucmat.2019.05.054>
50. Kabir E, Guikema SD, Quiring SM (2019) Predicting thunderstorm-induced power outages to support utility restoration. *IEEE Trans Power Syst* 34(6):4370–4381. <https://doi.org/10.1109/TPWRS.2019.2914214>
51. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu TY (2017) LightGBM: a highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst* 30:3146–3154
52. Kusiak A, Zhang Z (2010) Short-horizon prediction of wind power: a data-driven approach. *IEEE Trans Energy Convers* 25(4):1112–1122. <https://doi.org/10.1109/TEC.2010.2043436>
53. Kusiak A, Song Z, Zheng H (2009) Anticipatory control of wind turbines with data-driven predictive models. *IEEE Trans Energy Convers* 24(3):766–774. <https://doi.org/10.1109/TEC.2009.2025320>
54. Kusiak A, Zheng H, Song Z (2009) On-line monitoring of power curves. *Renew Energy* 34(6):1487–1493. <https://doi.org/10.1016/j.renene.2008.10.022>
55. Kusiak A, Zheng H, Song Z (2009) Short-term prediction of wind farm power: a data mining approach. *IEEE Trans Energy Convers* 24(1):125–136. <https://doi.org/10.1109/TEC.2008.2006552>
56. Kusiak A, Zheng H, Song Z (2009) Wind farm power prediction: a data-mining approach. *Wind Energy* 12(3):275–293. <https://doi.org/10.1002/we.295>
57. Lago J, De Brabandere K, De Ridder F, De Schutter B (2018) Short-term forecasting of solar irradiance without local telemetry: a generalized model using satellite data. *Sol Energy* 173:566–577. <https://doi.org/10.1016/j.solener.2018.07.050>
58. Li P, Zhang J-S (2018) A new hybrid method for China's energy supply security forecasting based on ARIMA and XGBoost. *Energies* 11(7):1687. <https://doi.org/10.3390/en11071687>
59. Li Z, Hurn AS, Clements AE (2017) Forecasting quantiles of day-ahead electricity load. *Energy Econ* 67:60–71. <https://doi.org/10.1016/j.eneco.2017.08.002>
60. Liu H, Tian H-Q, Li Y-F, Zhang L (2015) Comparison of four Adaboost algorithm based artificial neural networks in wind speed predictions. *Energy Convers Manage* 92:67–81. <https://doi.org/10.1016/j.enconman.2014.12.053>
61. Liu H, Chen C, Lv X, Wu X, Liu M (2019) Deterministic wind energy forecasting: a review of intelligent predictors and auxiliary methods. *Energy Convers Manage* 195:328–345. <https://doi.org/10.1016/j.enconman.2019.05.020>
62. Ma J, Cheng JCP (2017) Identification of the numerical patterns behind the leading counties in the U.S. local green building markets using data mining. *J Clean Prod* 151:406–418. <https://doi.org/10.1016/j.jclepro.2017.03.083>
63. Mayr A, Hofner B (2018) Boosting for statistical modelling: a non-technical introduction. *Stat Model* 18(3–4):365–384. <https://doi.org/10.1177/1471082X17748086>
64. Mayr A, Binder H, Gefeller O, Schmid M (2014) Extending statistical boosting. *Methods Inf Med* 53(06):428–435. <https://doi.org/10.3414/ME13-01-0123>
65. Mayr A, Binder H, Gefeller O, Schmid M (2014) The evolution of boosting algorithms. *Methods Inf Med* 53(06):419–427. <https://doi.org/10.3414/ME13-01-0122>
66. Mayr A, Hofner B, Waldmann E, Hepp T, Meyer S, Gefeller O (2017) An update on statistical boosting in biomedicine. *Comput Math Methods Med*. <https://doi.org/10.1155/2017/6083072>
67. Mease D, Wyner A (2008) Evidence contrary to the statistical view of boosting. *J Mach Learn Res* 9:131–156
68. Natekin A, Knoll A (2013) Gradient boosting machines, a tutorial. *Front Neurobot* 7:21. <https://doi.org/10.3389/fnbot.2013.00021>
69. Obringer R, Kumar R, Nateghi R (2019) Analyzing the climate sensitivity of the coupled water-electricity demand nexus in the Midwestern United States. *Appl Energy* 252:113466. <https://doi.org/10.1016/j.apenergy.2019.113466>
70. Okwu MO, Nwachukwu AN (2019) A review of fuzzy logic applications in petroleum exploration, production and

- distribution operations. *J Pet Explor Prod Technol* 9(2):1555–1568. <https://doi.org/10.1007/s13202-018-0560-2>
71. Olofsson KEJ, Humphreys DA, Haye RJL (2018) Event hazard function learning and survival analysis for tearing mode onset characterization. *Plasma Phys Controlled Fusion* 60(8):084002. <https://doi.org/10.1088/1361-6587/aac662>
  72. Olofsson KEJ, Sammulu BS, Humphreys DA (2019) Hazard function exploration of tokamak tearing mode stability boundaries. *Fusion Eng Des* 146:1476–1479. <https://doi.org/10.1016/j.fusengdes.2019.02.109>
  73. Papacharalampous G, Tyralis H, Koutsoyiannis D (2019) Comparison of stochastic and machine learning methods for multi-step ahead forecasting of hydrological processes. *Stoch Env Res Risk Assess* 33(2):481–514. <https://doi.org/10.1007/s00477-018-1638-6>
  74. Papacharalampous G, Tyralis H, Langousis A, Jayawardena AW, Sivakumar B, Mamassis N, Montanari A, Koutsoyiannis D (2019) Probabilistic hydrological post-processing at scale: why and how to apply machine-learning quantile regression algorithms. *Water* 11(10):2126. <https://doi.org/10.3390/w1102126>
  75. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A (2018) CatBoost: unbiased boosting with categorical features. *Adv Neural Inf Process Syst* 31:6638–6648
  76. Puig BC, Carmona J (2019) Bridging the gap between energy consumption and distribution through non-technical loss detection. *Energies* 12(9):1748. <https://doi.org/10.3390/en12091748>
  77. Razavi-Far R, Baraldi P, Zio E (2012) Dynamic weighting ensembles for incremental learning and diagnosing new concept class faults in nuclear power systems. *IEEE Trans Nucl Sci* 59(5):2520–2530. <https://doi.org/10.1109/TNS.2012.2209125>
  78. Ridgeway G (1999) The state of boosting. *Comput Sci Stat* 31:172–181
  79. Robinson C, Dilkina B, Hubbs J, Zhang W, Guhathakurta S, Brown MA, Pendyala RM (2017) Machine learning approaches for estimating commercial building energy consumption. *Appl Energy* 208:889–904. <https://doi.org/10.1016/j.apenergy.2017.09.060>
  80. Runge J, Zmeureanu R (2019) Forecasting energy use in buildings using artificial neural networks: a review. *Energies* 12(17):3254. <https://doi.org/10.3390/en12173254>
  81. Sagi O, Rokach L (2018) Ensemble learning: a survey. *Wiley Interdiscip Rev Data Min Knowl Discov* 8(4):e1249. <https://doi.org/10.1002/widm.1249>
  82. Salcedo-Sanz S, Cornejo-Bueno L, Prieto L, Paredes D, García-Herrera R (2018) Feature selection in machine learning prediction systems for renewable energy applications. *Renew Sustain Energy Rev* 90:728–741. <https://doi.org/10.1016/j.rser.2018.04.008>
  83. Schapire RE (1990) The strength of weak learnability. *Mach Learn* 5(2):197–227. <https://doi.org/10.1007/BF00116037>
  84. Schapire RE, Freund Y (2012) Boosting. The MIT Press, Cambridge, Massachusetts
  85. Schapire RE, Freund Y, Bartlett P, Lee WS (1998) Boosting the margin: a new explanation for the effectiveness of voting methods. *Ann Stat* 26(5):1651–1686. <https://doi.org/10.1214/aos/1024691352>
  86. Sha H, Xu P, Yang Z, Chen Y, Tang J (2019) Overview of computational intelligence for building energy system design. *Renew Sustain Energy Rev* 108:76–90. <https://doi.org/10.1016/j.rser.2019.03.018>
  87. Shafik N, Tutz G (2009) Boosting nonlinear additive autoregressive time series. *Comput Stat Data Anal* 53(7):2453–2464. <https://doi.org/10.1016/j.csda.2008.12.006>
  88. Sharma A, Kakkar A (2018) Forecasting daily global solar irradiance generation using machine learning. *Renew Sustain Energy Rev* 82(Part 3):2254–2269. <https://doi.org/10.1016/j.rser.2017.08.066>
  89. Sobri S, Koohi-Kamali S, Rahim NA (2018) Solar photovoltaic generation forecasting methods: a review. *Energy Convers Manage* 156:459–497. <https://doi.org/10.1016/j.enconman.2017.11.019>
  90. Souza FAA, Neto TRF, Magalhaes FRP, Silva FB, Pontes RST (2017) Predicting the grounding topology based on grounding impedance & the pattern recognition framework: a case study on one to four ground rods in straight line. *IEEE Trans Power Deliv* 32(4):1748–1757. <https://doi.org/10.1109/TPWRD.2016.2626339>
  91. Stetco A, Dinmohammadi F, Zhao X, Robu V, Flynn D, Barnes M, Keane J, Nenadic G (2019) Machine learning methods for wind turbine condition monitoring: a review. *Renew Energy* 133:620–635. <https://doi.org/10.1016/j.renene.2018.10.047>
  92. Sugiawan Y, Kurniawan R, Managi S (2019) Are carbon dioxide emission reductions compatible with sustainable well-being? *Appl Energy* 242:1–11. <https://doi.org/10.1016/j.apenergy.2019.03.113>
  93. Sun Y, Gao C, Li J, Wang R, Liu J (2019) Evaluating urban heat island intensity and its associated determinants of towns and cities continuum in the Yangtze River Delta urban agglomerations. *Sustain Urban Areas* 50:101659. <https://doi.org/10.1016/j.scs.2019.101659>
  94. Suryanarayana G, Lago J, Geysen D, Aleksiejuk P, Johansson C (2018) Thermal load forecasting in district heating networks using deep learning and advanced feature selection methods. *Energy* 157:141–149. <https://doi.org/10.1016/j.energy.2018.05.111>
  95. Toyao T, Suzuki K, Kikuchi S, Takakusagi S, Shimizu K-I, Takigawa I (2018) Toward effective utilization of methane: machine learning prediction of adsorption energies on metal alloys. *J Phys Chem C* 122(15):8315–8326. <https://doi.org/10.1021/acs.jpcc.7b12670>
  96. Tutz G, Binder H (2006) Generalized additive modelling with implicit variable selection by likelihood-based boosting. *Biometrics* 62(4):961–971. <https://doi.org/10.1111/j.1541-0420.2006.00578.x>
  97. Tyralis H, Papacharalampous G, Burnetas A, Langousis A (2019) Hydrological post-processing using stacked generalization of quantile regression algorithms: large-scale application over CONUS. *J Hydrol* 577:123957. <https://doi.org/10.1016/j.jhydrol.2019.123957>
  98. Tyralis H, Papacharalampous G, Langousis A (2019) A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water* 11(5):910. <https://doi.org/10.3390/w11050910>
  99. Urraca R, Martinez-de-Pison E, Sanz-Garcia A, Antonanzas J, Antonanzas-Torres F (2017) Estimation methods for global solar radiation: case study evaluation of five different approaches in central Spain. *Renew Sustain Energy Rev* 77:1098–1113. <https://doi.org/10.1016/j.rser.2016.11.222>
  100. Vázquez-Canteli JR, Nagy Z (2019) Reinforcement learning for demand response: a review of algorithms and modeling techniques. *Appl Energy* 235:1072–1089. <https://doi.org/10.1016/j.apenergy.2018.11.002>
  101. Voyant C, Notton G, Kalogirou S, Nivet ML, Paoli C, Motte F, Fouilloy A (2017) Machine learning methods for solar radiation forecasting: a review. *Renew Energy* 105:569–582. <https://doi.org/10.1016/j.renene.2016.12.095>
  102. Wang H, Lei Z, Zhang X, Zhou B, Peng J (2019) A review of deep learning for renewable energy forecasting. *Energy Convers Manage* 198:111799. <https://doi.org/10.1016/j.enconman.2019.111799>



103. Wu T, Wang J (2019) Global discovery of stable and non-toxic hybrid organic-inorganic perovskites for photovoltaic systems by combining machine learning method with first principle calculations. *Nano Energy* 66:104070. <https://doi.org/10.1016/j.nanoen.2019.104070>
104. Wu X, Xiang S, Su J, Cai W (2019) Understanding quantitative relationship between methane storage capacities and characteristic properties of metal-organic frameworks based on machine learning. *J Phys Chem C* 123(14):8550–8559. <https://doi.org/10.1021/acs.jpcc.8b11793>
105. Wyner AJ, Olson M, Bleich J, Mease D (2017) Explaining the success of AdaBoost and random forests as interpolating classifiers. *J Mach Learn Res* 18(48):1–33
106. Yagli GM, Yang D, Srinivasan D (2019) Automatic hourly solar forecasting using machine learning models. *Renew Sustain Energy Rev* 105:487–498. <https://doi.org/10.1016/j.rser.2019.02.006>
107. Yoshihashi R, Kawakami R, Iida M, Naemura T (2017) Bird detection and species classification with time-lapse images around a wind farm: dataset construction and evaluation. *Wind Energy* 20(12):1983–1995. <https://doi.org/10.1002/we.2135>
108. Yuan T, Sun Z, Ma S (2019) Gearbox fault prediction of wind turbines based on a stacking model and change-point detection. *Energies* 12(22):4224. <https://doi.org/10.3390/en12224224>
109. Zamo M, Mestre O, Arbogast P, Pannekoucke O (2014) A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production, part I: deterministic forecast of hourly production. *Sol Energy* 105:792–803. <https://doi.org/10.1016/j.solener.2013.12.006>
110. Zhang Z, Kusiak A (2011) Models for optimization of energy consumption of pumps in a wastewater processing plant. *J Energy Eng* 137(4):159–168. [https://doi.org/10.1061/\(ASCE\)JEY.1943-7897.0000051](https://doi.org/10.1061/(ASCE)JEY.1943-7897.0000051)
111. Zhang W, Quan H, Srinivasan D (2018) Parallel and reliable probabilistic load forecasting via quantile regression forest and quantile determination. *Energy* 160:810–819. <https://doi.org/10.1016/j.energy.2018.07.019>
112. Zhao Y, Li T, Zhang X, Zhang C (2019) Artificial intelligence-based fault detection and diagnosis methods for building energy systems: advantages, challenges and the future. *Renew Sustain Energy Rev* 109:85–101. <https://doi.org/10.1016/j.rser.2019.04.021>
113. Zheng H, Kusiak A (2009) Prediction of wind farm power ramp rates: a data-mining approach. *J Sol Energy Eng* 131(3):0310111. <https://doi.org/10.1115/1.3142727>
114. Zhou ZH (2012) Ensemble methods: foundations and algorithms. Chapman and Hall/CRC, Boca Raton, FL
115. Zhu R, Guo W, Gong X (2019) Short-term photovoltaic power output prediction based on  $k$ -fold cross-validation and an ensemble model. *Energies* 12(7):1220. <https://doi.org/10.3390/en12071220>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.