



Research Article

Combined cycle gas turbine power output prediction and data mining with optimized data matching algorithm



David A. Wood¹ 

Received: 17 October 2019 / Accepted: 12 February 2020 / Published online: 18 February 2020
© Springer Nature Switzerland AG 2020

Abstract

Electrical power output (PE) for a combined cycle gas turbine (CCGT) consisting of 9568 data records collected over a 6-year period is evaluated by the transparent open box (TOB) machine-learning method to provide accurate PE predictions and insight to prediction errors. The PE predictions derived by applying the TOB optimized data matching technique are more accurate than published predictions for the dataset from fifteen correlation-based, machine-learning algorithms. TOB achieves this high-accuracy using a tuning subset of < 150 (~ 1.5%) data records. Its accuracy is confirmed by testing the optimized solutions against all dataset records in 15 runs spread across five shuffled datasets. The dataset has a few extreme outliers associated with its four independent variables and these negatively impact the prediction accuracy of machine-learning methods. Through its transparency and forensic-like auditability of its calculations for individual data records, the TOB algorithm is able to mine the dataset to provide useful insight to the interactions of the outliers with other data records. This enables a filtered dataset (9533 records), excluding 35 carefully selected data records, to be customized to deliver much improved prediction accuracy (RMSE = 2.89 MW). Mining the dataset also reveals significant differences in prediction accuracy achieved for different sectors of the PE distribution. This insight identifies that prediction accuracy could be further improved by dividing the dataset into separately optimized subsets, three along its main PE trend plus a fourth, small subset consisting of the outliers. The TOB algorithm demonstrates its value as a machine-learning tool capable of generating accurate predictions and easily auditable data mining. It is well suited for CCGT efficiency and performance optimization.

Keywords Gas-to-power predictions · Machine learning transparency · Memetic optimization · Outlier analysis · Data filtering · Segmental analysis

1 Introduction

Combined cycle gas turbines (CCGT) are widely exploited in modern power supply networks and their capacities and configurations as power blocks continue to grow, particularly in the United States due to low-priced shale gas [1]. CCGT power plant operators need to routinely predict in the medium term and forecast in the short-term the electrical power output (PE) at full load and the natural gas

consumption of their plants. They do this based on prevailing or expected conditions, in order operate efficiently, comply with contractual requirements and achieve maximum profitability. A CCGT's PE can be calculated based on the thermodynamic relationships governing the complex combination of processes involved. However, that is time-consuming and subject to a number of uncertainties and assumptions. Consequently, machine learning algorithms assessing historical performance data for specific plants,

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s42452-020-2249-7>) contains supplementary material, which is available to authorized users.

✉ David A. Wood, dwa@dwasolutions.com | ¹DWA Energy Limited, Lincoln, UK.



SN Applied Sciences (2020) 2:441 | <https://doi.org/10.1007/s42452-020-2249-7>

operating over a range of environmental conditions, are useful for providing rapid PE forecasts to operationally meaningful levels of prediction accuracy [2].

CCGTs generate their electricity from a gas turbine (GT) driven by high-temperature combustion gases coupled with a steam turbine (ST) fuelled by steam generated from a heat recovery system fuelled by the exhaust gases from the GT. The operational efficiency of CCGT is impacted by ambient conditions [3]. Ambient temperature continues to be widely studied in relation to the operational performance of CCGT [4–6], but other ambient factors such as atmospheric pressure, relative humidity, and wind speed also have an influence. The operating conditions of the equipment, e.g., turbine blade cooling, exhaust temperature and exhaust steam pressure also influence CCGT performance [7, 8].

A number of studies have applied various regression methods and other machine learning techniques to CCGT performance prediction [9] and fault detection [10]. The application of artificial neural networks (ANN) as reliable alternatives for various simulation control modelling and fault detection applications for CCGT is now well established [11–13]. Other machine learning methods including Least Squares Support Vector Machines (LSSVM) [14] and adaptive neuro-fuzzy inference system (ANFIS) [15] are also applied for modelling and controlling GT. Indeed, machine learning is now applied to many aspects of GT performance including the optimization of design for its components, asset operation and maintenance [16]. This typically involves data-driven and deep learning approaches [17], in some cases developing networks capable of unsupervised learning with input data that lacks a structured format. Machine learning regression approaches are widely applied to provide reliable performance predictions of compressor components and anomaly detection for gas turbine combustors [8, 18–20]. Whereas artificial neural networks remain the most commonly used for diagnosing fault conditions [21], gas turbine performance is now being more rigorously monitored by simulation, optimization [22] and hybrid networks such as a combination between ANN and high dimensional model representation [23]. However, all of these methods involve correlations regressions, correlations and statistical relationships between input parameters and dependent variables to underpin their machine learning capabilities. The novelty of this method and analysis applied to CCGT power output prediction is that it demonstrates that meaningful and accurate predictions can be achieved by data matching techniques that does not involve regression and/or correlations between its variables.

Here, transparent-open-box (TOB) machine learning [24–26] applies an optimized-data—matching technique to model CCGT performance based on a dataset published

by UCI [27]. That dataset is already extensively analysed to compare the prediction performance of several correlation and regression-based machine learning methods [9] including ANN [13]. The TOB method, which does not use correlations to make its predictions, offers several advantages over neural network and other correlation machine-learning methods. Comprehensive data mining and an ability to identify and avoid overfitting datasets are valuable TOB attributes.

The main objective and contribution of this study are to demonstrate that the recently developed optimized data matching algorithm employed (the TOB method) provides more accurate predictions of electrical power output for a large (nearly 10,000 data records) combined cycle gas turbine dataset of influencing variables than fifteen published correlation-based, machine-learning algorithms applied to the same dataset. Additional objectives and contributions are to describe how the transparency of the TOB calculations facilitates detailed data mining, prediction error analysis and segmental analysis of the dataset, providing substantial insight to the dataset that is difficult to generate with less-transparent machine learning methods.

The five remaining sections of this article are organized as follows. Section 2 describes the CCGT dataset evaluated. Section 3 explains the TOB method and its optimizers. Section 4 presents prediction results for 9568 data records and assesses their accuracy using standard statistical techniques on an individual and segmental basis. Further details of the methodology and cases run are provided in a supplementary file. Section 5 provides prediction outlier error analysis and considers data filtering possibilities to improve prediction accuracy. These findings support the final conclusions section.

2 Dataset

The dataset studied [27], with some analysis previously published [9, 28] contains 9568 data points collected from a CCGT operating in Turkey over a 6-year period (2006–2011). The plant (480-MW nominal generating capacity) involves *two* 160-MW ABB 13E2 gas turbines, *two* dual-pressure–heat-recovery steam generators, and *one* 160-MW ABB steam turbine. The operational data was recorded and filtered to provide information just for the periods when the CCGT was operating at full load.

Hourly average values of four independent variables are available for each data record. These are:

- Ambient temperature (AT)
- Atmospheric pressure (AP)
- Relative humidity (RH)

- Steam turbine’s exhaust pressure, referred to as vacuum (V)

Additionally, the CCGT’s net hourly electrical energy output (PE) in MW is provided for each data record. PE is the dependent variable for prediction purposes.

The data for these variables was recorded every second and averaged to provide hourly data. Sensors to measure variables AT, AP and RH are placed at the air inlet to the GT, whereas variable V is measured with a sensor at the exhaust on the low-pressure side of the ST. Five shuffled versions of the dataset (each with 9568 records) are provided as part of the UCI dataset that are useful for statistical and repeatability tests for the prediction models. In this study each of the five datasets is sampled independently three times for TOB analysis. The ranges, arithmetic means and units of the five dataset variables are listed in Table 1. See “Appendix” for accessing the complete dataset.

The independent variables in the CCGT dataset show distinctive correlations with dependent variable PE (Table 2). However, a few data records (< ~0.5%) do exist as outliers to the main trends as revealed in variable cross plots, e.g. AT versus PE (Fig. 1). These outliers have a significant impact on PE prediction accuracy, which is investigated in Sect. 5.

As well as generating accurate predictions of the dependent variable (PE) for the entire dataset, the few prediction outliers that are generated in the optimum prediction models (typically between 1 and 2% of all data records) are further assessed using a novel and simple data filtering and error analysis technique. This identifies the few outlying predictions that exceed specified squared-error (SE) limits. The SE limits applied in this study are 100 MW and 200 MW equating to PE prediction errors of plus or minus 10 MW and 14.1 MW for individual data records. Those data records that exceed this specified limits are identified as outliers and the data matching involved in their predictions forensically assessed. The details of this data filtering and outlier analysis applied to the CCGT dataset studied are provided in Sect. 5.

Table 2 Correlation coefficients (R) between variables in the 9568-UCI-CCGT dataset [27]. See Table 1 for variable abbreviation definitions

R	AT	V	AP	RH	PE
AT	1.0000	0.8441	-0.5075	-0.5425	-0.9481
V		1.0000	-0.4135	-0.3122	-0.8698
AP			1.0000	0.0996	0.5184
RH				1.0000	0.3898
PE					1.0000

3 Method

3.1 TOB method overview

The TOB learning network is a recently introduced machine-learning algorithm involving an optimized data matching approach [24]. It has been successfully applied to both small and large datasets to provide accurate predictions and transparent data mining [25, 26]. The TOB method is described in detail in the studies cited and is not repeated here. A flow diagram summarizing the calculation sequence adapted for application to the CCGT dataset summarizes the TOB method (Fig. 2). More details about the TOB method and its implementation sequence is provided in the supplementary file (Sects S.1 and S.2).

For small data sets (up to a few thousand data records) TOB Stage 2 can be efficiently conducted using Excel’s Solver optimizers. For mid-sized and large datasets, a fully coded memetic firefly optimizer is applied [29]. For the CCGT dataset both approaches are utilized. Excel’s Solver optimizers are both applied to verify the outcomes for the tuning subsets with up to 150 data records. For larger tuning subsets just the memetic firefly optimizer is applied.

An independent testing subset is used to verify that the optimum solutions derived by applying the two stages of the TOB algorithm to the tuning subset provide dependent-variable predictions that are statistically valid when applied to the dataset more generally. TOB derives two useful predictions of its dependent variable. The first (Stage 1) establishes and use the ten best data matches for each data records with other records in the dataset. The

Table 1 Variable characteristics of CCGT dataset [9, 27, 28] with 9568 data records

Variable	Symbol	Units	Min	Max	Mean
Temperature	AT	°C	1.81	37.11	19.65
Exhaust vacuum (steam turbine)	V	cm Hg	25.36	81.56	54.31
Ambient pressure	AP	Millibars	992.89	1033.30	1013.26
Relative humidity	RH	Percent	25.56	100.16	73.31
Net hourly electrical energy output	PE	MW	420.26	495.76	454.37

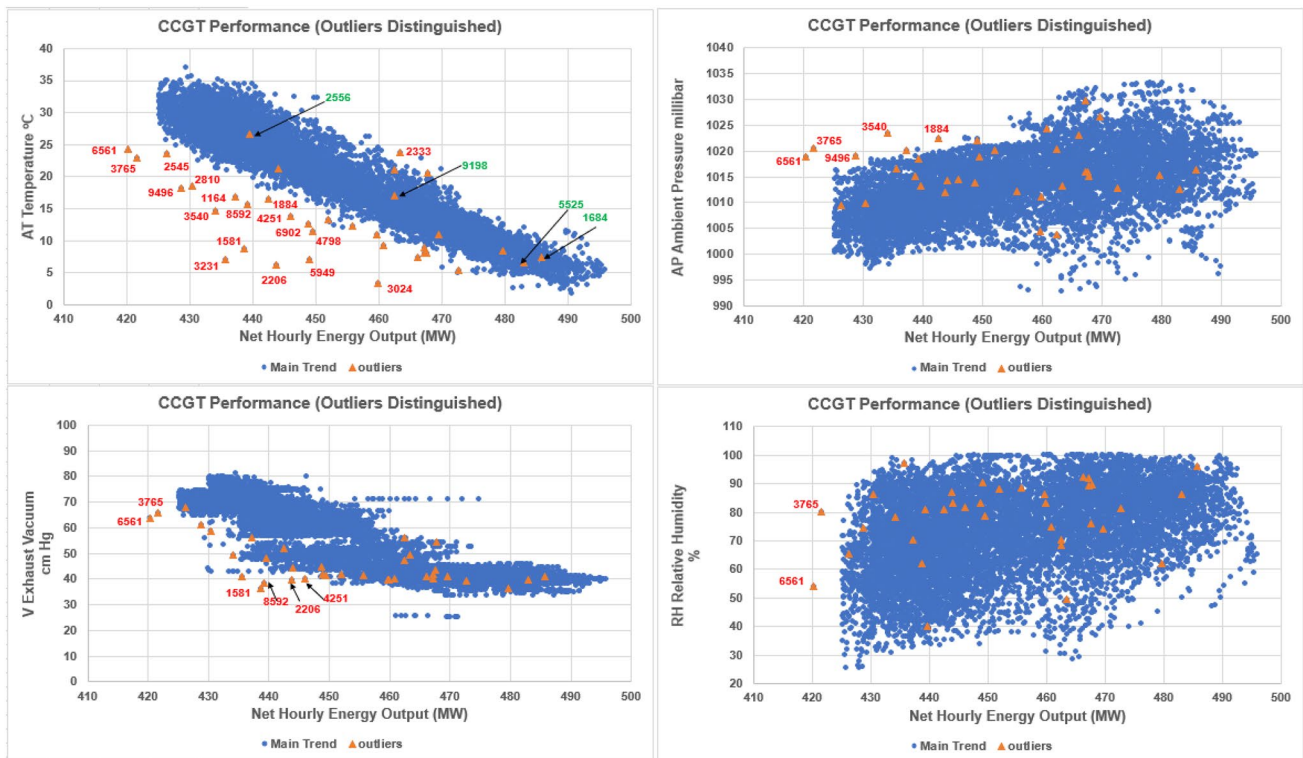


Fig. 1 Independent variables versus PE for the 9568-UCL-CCGT dataset [27]. Numbers identifying outliers refer to sequence positions of data records for shuffled set #5. These are addressed in Sect. 5.1

second prediction (Stage 2) minimizes the squared errors between the variables in the best matching data records by applying variable weights. Comparisons between these two predictions helps to reveal and prevent overfitting datasets.

3.2 TOB optimizers applied

The subsets used for tuning the TOB solutions are typically of modest size. If they involve about 150 data records or less, two of Microsoft Excel's "Solver" optimizers [30] can be usefully exploited for TOB Stage 2 optimization. These two optimizers ("Generalized Reduced Gradient" (GRG) and "evolutionary") are distinctive and the solutions they derive tend to complement each other. Together they help to find the most accurate or "best" optimized TOB Stage 2 solutions. These solvers can be usefully driven by "visual basic for application" (VBA) code but cannot be fully coded in Excel. They require the calculations of equations A3 to A6 (see supplementary file Sect. S2) setup as cell formulas in an Excel spreadsheet. For mid-sized and large datasets, such as the CCGT dataset evaluated here, with tuning subsets containing more than 150 data records it is more convenient to fully code a customized optimizer. This avoids the need to enter the calculations into a large number of Excel cell formulas. For the CCGT dataset a

customized memetic firefly optimizer [29] coded in VBA is applied and the results compared with Excel Solver (GRG and evolutionary) results for some of the smaller tuning subsets evaluated (Table 3).

The firefly algorithm was developed by Yang and his colleagues [31, 32] with computational analogies made to the behaviour of fireflies. Fireflies progressively move towards brighter individuals in a swarm. However, they do so in steps and incorporate a degree of randomness in the adjustment to their position in each step. The firefly algorithm adjusts the position of each member of a population based on specific rules related to a calculated "brightness" metric that is proportional to how close their objective function value is to the lowest value achieved so far [33]. The fireflies in the best positions are assigned the most intense brightness. On the other hand, the fireflies in the worst positions are assigned the least intense brightness [34, 35]. In addition to the number of fireflies (f), three key control variables are used to configure the firefly algorithm and control the scale of its solution adjustments from one iteration to the next. These control variables are: a randomized factor influencing firefly movements (α); an attractiveness factor (β) at distance $r=0$; and brightness coefficient (γ). The pseudocode listed here provides the values for these control variables used to configure the firefly algorithm used for the TOB analysis of the CCGT dataset.

Configuring the Transparent Open Box (TOB) Learning Network for Combined Cycle Gas Turbine Power Predictions, Forecasts and Data Mining from Diverse Input Variables

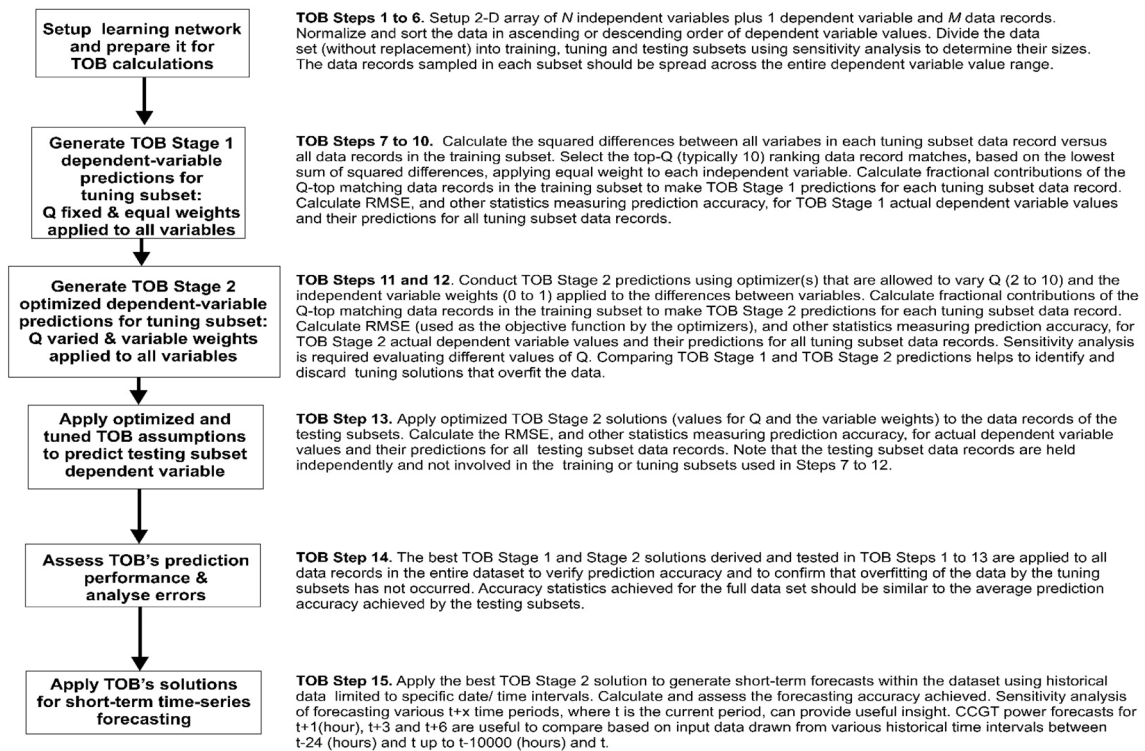


Fig. 2 Transparent open box algorithm workflow with steps configured for combined cycle gas turbine power prediction and data mining. See supplementary file for a step-by-step mathematical description of the TOB algorithm

The memetic firefly algorithm, customized specifically to the requirements of the TOB learning network, involves six integrated memes or metaheuristics (Mh) [29]. These memes improve its convergence speed to the global minimum (Fig. 3), compared to the standalone firefly optimizer. They enhance its ability to resist becoming trapped at sub-optimal local minima. Optimization algorithms exploring large non-linear solution spaces, such as the one generated by the CCGT dataset studied, have to explore that solution space both at coarse granularity for more global coverage and at finer granularity for detailed local investigation. There is a risk of algorithms becoming trapped at local optima and failing to find the global optima. Most high-performing optimization algorithms attempt to overcome this problem by involving metaheuristics that focus on global exploration and others that focus on local exploration of the solution space. The six memes of the memetic firefly optimizer are therefore carefully configured to have different focuses with respect to global and local solution space exploration. These memes operate on each of the multiple iterations applied to a progressively modified population of potential solutions:

Mh1: Creates a set of initial solutions randomly (good for global exploration of the solution space);
 Mh2: Creates some solutions applying the firefly algorithm (good for both global and local exploration of the solution space);
 Mh3: Creates some solutions by making minor adjustments to several selected variables (W_n or Q) of existing high-ranking solutions (good for local exploration of the solution space);
 Mh4: Creates some solutions by adjusting single variable weights to very low (~ 0) or very high values (~ 1) for existing high-ranking solutions (good for local exploration of the solution space but with some global exploration potential);
 Mh5: Creates some solutions by mutating one randomly selected variable (W_n or Q) in existing solutions of various rank; (good for global exploration of the solution space); and,
 Mh6: Creates some solutions by modifying one variable (W_n or Q) by alternating large and small amounts. It is selectively applied to only high-ranking solutions (good for local exploration of the solution space but with some global exploration potential).

Table 3 TOB accuracies recorded by the TOB tuning and testing subsets for the CCGT dataset (5 shuffled data sets each sampled three times, A, B and C)

Case identifier	Tuning subset size			Optimization results for tuning subset matched with the training subset			Testing subset size			Optimization results for independent testing subset matched with the training subset		
	Total data records	146	Q	RMSE	R ²	Total data records	197	Q	RMSE	R ²		
#1A	TOB stage 1	TOB stage 1	10	3.24697	0.96498	TOB stage 1	TOB stage 1	10	3.24991	0.96514		
	TOB stage 2:	Memetic firefly	9	2.85356	0.97281	TOB stage 2:	Memetic firefly	9	2.65903	0.97659		
	TOB stage 2:	Excel GRG	7	2.85356	0.97280	TOB stage 2:	Excel GRG	7	2.65900	0.97659		
#1B	TOB stage 2:	Excel evolutionary	7	2.85357	0.97281	TOB stage 2:	Excel evolutionary	7	2.65920	0.97659		
	TOB stage 1	TOB stage 1	10	3.32241	0.96327	TOB stage 1	TOB stage 1	10	3.40763	0.96155		
	TOB stage 2:	Memetic firefly	10	2.89522	0.97209	TOB stage 2:	Memetic firefly	10	2.68390	0.97621		
#1C	TOB stage 2:	Excel GRG	9	2.89851	0.97206	TOB stage 2:	Excel GRG	9	2.73116	0.97532		
	TOB stage 2:	Excel evolutionary	9	2.89851	0.97206	TOB stage 2:	Excel evolutionary	9	2.73118	0.97532		
	TOB stage 1	TOB stage 1	10	3.46045	0.96004	TOB stage 1	TOB stage 1	10	3.00790	0.97009		
#5A	TOB stage 2:	Memetic firefly	10	2.79048	0.97440	TOB stage 2:	Memetic firefly	10	3.18245	0.96642		
	TOB stage 2:	Excel GRG	6	2.77171	0.97478	TOB stage 2:	Excel GRG	6	3.73616	0.95422		
	TOB stage 2:	Excel evolutionary	10	2.79452	0.97434	TOB stage 2:	Excel evolutionary	10	3.15782	0.96694		
#5B	TOB stage 1	TOB stage 1	10	3.10777	0.96815	TOB stage 1	TOB stage 1	10	3.25655	0.96483		
	TOB stage 2:	Memetic firefly	10	2.74362	0.97519	TOB stage 2:	Memetic firefly	10	2.64944	0.97673		
	TOB stage 2:	Excel GRG	9	2.74897	0.97505	TOB stage 2:	Excel GRG	9	2.67193	0.97634		
#5C	TOB stage 2:	Excel evolutionary	9	2.74898	0.97505	TOB stage 2:	Excel evolutionary	9	2.67202	0.97634		
	TOB stage 1	TOB stage 1	10	2.88144	0.97230	TOB stage 1	TOB stage 1	10	3.63450	0.95613		
	TOB stage 2:	Memetic firefly	6	2.64891	0.97663	TOB stage 2:	Memetic firefly	6	3.18453	0.96633		
#5C	TOB stage 2:	Excel GRG	6	2.64891	0.98825	TOB stage 2:	Excel GRG	6	3.18447	0.96633		
	TOB stage 2:	Excel evolutionary	6	2.64905	0.98825	TOB stage 2:	Excel evolutionary	6	3.18783	0.96626		
	TOB stage 1	TOB stage 1	10	3.20608	0.96586	TOB stage 1	TOB stage 1	10	2.95508	0.97104		
#5C	TOB stage 2:	Memetic firefly	10	2.64582	0.97680	TOB stage 2:	Memetic firefly	10	2.85564	0.97308		
	TOB stage 2:	Excel GRG	10	2.64629	0.98833	TOB stage 2:	Excel GRG	10	2.83364	0.97350		
	TOB stage 2:	Excel evolutionary	10	2.65219	0.98829	TOB stage 2:	Excel evolutionary	10	2.86908	0.97280		

Results shown are for shuffled sets #1 and #5 only. Table S4 in the Supplementary file provides details of the prediction accuracies for all the shuffled cases evaluated

A pseudo-code for the memetic firefly optimizer configured for this study is as follows:

Objective function to minimize $f(X)$, $X = (X_1, \dots, X_m)^t$

Where X is the RMSE for the TOB tuning set dependent variable prediction versus its actual values for M data records in iteration t . Each data record m has N independent variables contributing to the value of X .

Initialise firefly control variables for metaheuristic (Mh)2

α (second iteration); (0.15 used for initial α in this study)

α (last iterations); (0.10 used for final α in this study)

β (second iteration); (0.4 used for initial β in this study)

β (last iteration); (0.2 used for final β in this study)

Sliding scales between initial and final α and β values are applied across the iterations

γ can vary between 2 and 10; (2 used for γ in this study)

Initialize population f of random solutions (Mh1) for iteration 1:

X_i (1,2, ..., f); ($f=100$ for this study)

For each solution: constraints applied are $2 \leq Q \leq 10$ and $0 \leq Wn \leq 1$

where Q is the number of top matching data records and Wn are the difference weights applied to each of the N variables

While ($2 \leq t \leq$ maximum number of iterations)

Ensure highest ranking solution from iteration $t-1$ is preserved as part of the population generated for iteration t

Evaluate and rank f solutions from iteration $t-1$

Modify top ranking f_2 of those solutions by Mh2; ($f_2=50$ for this study)

Modify f_3 of lower ranking solutions from iteration $t-1$ by Mh3 ($f_3=20$ for this study)

Modify f_4 of higher-ranking solutions from iteration $t-1$ by Mh4 ($f_4=10$ for this study)

Modify f_5 of higher-ranking solutions from iteration $t-1$ by Mh5 ($f_5=10$ for this study)

Modify f_6 of higher-ranking solutions from iteration $t-1$ by Mh6 ($f_6=10$ for this study)

Mh1 to Mh6 modifications collectively generate $f-1$ new solutions plus the best solution from iteration $t-1$

Rank and sort f new solutions generated for iteration t based on their X values and record them

End While

Graphically display best solution for each iteration

Mine solution data set to generate a metaheuristic profile of the optimization sequence

The effectiveness of each meme in contributing to a solution is monitored by a metaheuristic profiling technique [36, 37], a graphical representation of which is illustrated by Fig. 4 for an optimization solution derived for Case 5C.

Figure 4 indicates that the six memes described are able to contribute high-ranking solutions in the initial runs of the firefly optimizer. I.e., up to iteration #50, or so. On the other hand, random solutions only contribute to the ten best solutions during the first few passes. Beyond iteration 50 the top-ten solutions are dominated by Mh3 with Mh2

and Mh4 also making significant contributions. Such profiling verifies that all the memes in the memetic algorithm are contributing to the performance of the optimizer as a whole and facilitating efficient convergence without leading to entrapment at suboptimal minima. The optimizer collects the details (Wn and Q values) for all the solutions it evaluates. This is an advantage over the Solver optimizers that provide limited details on the interim best solutions found.

To summarize, three distinct optimizers: (1) Excel's Solver GRG—a gradient search optimizer; (2) Excel's Solver

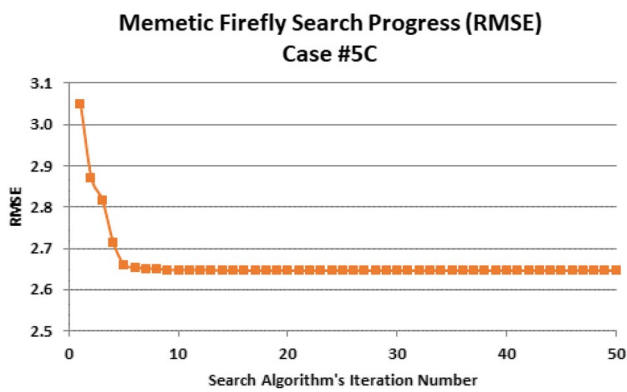


Fig. 3 Progress of the optimized solution towards a minimum RMSE value applying TOB's memetic optimizer

evolutionary optimizer; and, (3) the memetic firefly optimizer are used and their results displayed and compared for this study. Many other optimization algorithms (gradient search and evolutionary) with good global and local search capabilities could be successfully deployed for this purpose. It is not a necessity for the TOB method to use the memetic firefly optimizer but it is beneficial for it to do so. This is because the memetic firefly optimizer is specifically customized to maximize the transparency of the intermediate solutions generated, and to facilitate analysis exploiting its metaheuristic profiling capabilities.

3.3 Computational performance

For this CCGT power output prediction study, the TOB algorithm was executed on a computational platform

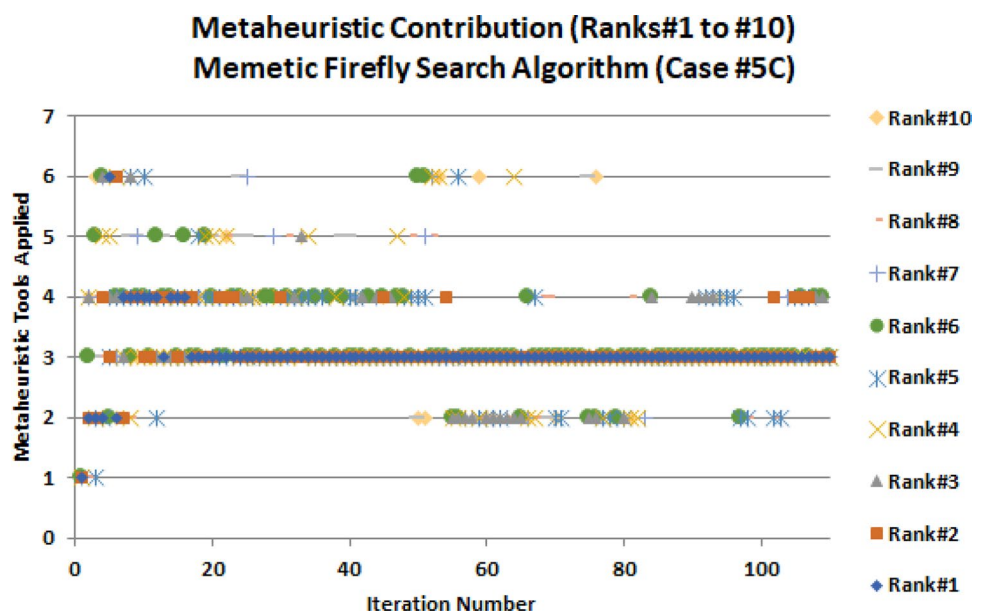
running Excel 365 using VBA code run on a laptop (Intel 5-core processor/CPU 1.80 GHz) with a Windows 10 operating system. Exact computation times vary depending upon the computational platform used as well as the number of variables evaluated and the number of data records constituting the datasets evaluated.

The approximate computational times, using the platform dataset and TOB configuration described, for the component TOB actions are:

- Allocating records to the tuning, testing and training subsets: ~ 3.5 s
- Identifying and ranking squared differences of each tuning and testing subset data record against all data records in the training subset (i.e., more than 9250 data records matched for each tuning and testing subset record) ~ 168 s
- Calculating TOB stage 1 predictions: ~ 0.5 s
- Executing TOB Stage 2 optimizers (Solver GRG, Solver evolutionary, memetic firefly) to train tuning subsets, consisting of ~ 150 data records, to reach convergence at an RMSE minimum (actual versus predicted PE for all data records in the tuning subsets); or be terminated after a specified running time: ~ <= 60 s
- Applying the optimized solution to calculate TOB Stage 2 predictions for the testing subsets consisting of about 200 data records: ~ 0.8 s

The TOB algorithm deployed for this study is configured to maximize the transparency of the intermediate calculations and to deliver insight to all the data records influencing those intermediate calculations rather than minimizing computation time. The data matching step to establish the

Fig. 4 A metaheuristic profile [36, 37] for the first 110 iterations of the optimizer (Case #5C). The top-ten solutions found by each meme are distinguished



top-10 best data matches from the large training subset is the most computationally expensive. This could be configured to run faster by reducing the data match information collected as part of its intermediate calculations, but that counters the objective of providing a fully transparent algorithm. Similarly, the memetic optimizer could be configured to run faster by restricting the intermediate solution data it collects but that would undermine the valuable metaheuristic profiling information it provides.

3.4 Prediction accuracy statistical measures exploited and monitored

The following statistical measures of accuracy are calculated as part of the TOB analysis:

- Mean square error (MSE)
- Root mean square error (RMSE)—TOBs objective function
- Percent deviation between measured and predicted values (PD)
- Average percent deviation (APD)
- Absolute average percent deviation (AAPD)
- Standard deviation (SD)
- Correlation coefficient (R) between measured and predicted values
- Coefficient of determination (R^2)

Detailed formulas to calculate these metrics are provide in Sect. 5.3 of the Supplementary File (equations A7 to A14).

4 Results

4.1 TOB applied through subset training to provide rapid and auditable predictions

The TOB learning algorithm evaluates five shuffled sets of the CCGT dataset [27] with three cases (A, B and C) run for each set. This means that each of the fifteen cases executed involves a different set of data records in both its tuning and testing subsets. Initial tests were performed to establish the appropriate number of data records for the small tuning and testing subsets to contain. Such tests are straightforward to execute and interpret. If a stage 2 solution does not lead to better prediction accuracy than a stage 1 solution applied to the testing subset, this indicates that the tuning subset used does not include sufficient data records to representatively sample the dataset. If that tuning subset solution shows high accuracy for the tuning subset records but poorer accuracy than the stage 1 solution when applied to the testing subset, it is likely to

be overfitting the limited data records it has at its disposal during the tuning process.

For the UCI CCGT datasets the sensitivity analysis suggested that the tuning subset needed to contain about 100 or more to provide meaningful improvements during the TOB Stage 2 optimization. On the other hand. The testing subset need to contain about 150 or more data records to meaningfully sample the distributions covered by the full dataset. Consequently, for this study, the sample sizes of the tuning subsets are all set at 146 data records and the testing subsets are all set at 197 data records. In both cases the data records are distributed across the entire data set in an arbitrary way to avoid potential clustering that may occur in random sampling. This means that the remainder of each dataset, 9225 data records, with tuning plus testing data records removed, becomes the training subset for the tuning process. In this study, the tuning subset size (i.e., number of data records) represents just 1.52% of the full dataset (9568) and the testing subset just 2.06% of each full dataset. Clearly, it is necessary to demonstrate that these small subset samples can provide robust and reliable predictions for the dataset as a whole. This results presented will confirm the repeatability and reliability of the method.

Table 3 displays the PE prediction accuracy of TOB analysis applied to tuning and testing subsets of selected cases for the dependent variable (PE). The results for all cases evaluated are listed in Table S4 in the supplementary file. RMSE (objective function) and R^2 are used to summarize prediction accuracy for each case and each optimizer applied. As should be expected, the stage 2 predictions provide higher tuning-subset accuracy than stage 1 in all cases evaluated. For all *tuning* subsets:

- RMSE varies from 2.573 to 3.510 MW
- R^2 varies from 0.95950 to 0.98833

The higher RMSE values (least accurate) derived for the tuning subset are Stage 1 predictions, plus cases where some of the tuning subsets have selected several data records that are outliers to the main input data trends (Fig. 1). The significance of outliers and prediction accuracy is addressed in Sects. 5.1 and 5.2.

All three optimizers found similar optimum solutions for each case evaluated. For almost all cases evaluated, the memetic firefly and Excel's Solver GRG optimizer find the same minimum RMSE. In several cases those two optimizers slightly outperform Excel's evolutionary algorithm by finding optimum solutions with better accuracy. It is reassuring that all three optimizers consistently agree with each other. In all cases, two out of the three optimizers agree on the Q value for the optimum solution. However, as tuning and testing subsets have their data records

selected independently, the relative differences between the PE prediction accuracies they achieve do not vary systematically. This is one of the reasons why it is worthwhile using more than one optimizer and running several cases with different independent testing subsets. It is also worthwhile testing more than one optimum solution, as there are a large number of good but sub-optimal solutions to choose from.

Of more significance than the tuning set prediction accuracies, in terms of TOBs prediction performance, is the accuracies achieved for the **testing** subset (right side Table 3). RMSE for the testing subset varies from 2.620 to 3.736 MW. Whereas, R^2 ranges from 0.95422 to 0.97742 for the testing subset. Stage 1 predictions for the testing subset are less accurate (higher RMSE). Also, testing subsets that sample more outlier data records (Fig. 1) tend to generate less accurate predictions. However, notice (Table 3 and S4 in the supplementary file) that most of the tuned solutions perform well achieving high prediction accuracies when applied to testing subsets. This is so despite some of the tuning subsets sampling several outlier data records.

These results are extremely encouraging. To place this prediction performance in context, the best of fifteen regression-based machine learning methods evaluated for the same dataset by Tufekci [9], i.e., the Bagging algorithm with REPTree, achieved an RMSE of 3.787 MW. Moreover, the ANN method applied to the dataset by Elfaki & Hassan [13] only achieved an RMSE of 4.32 MW. The remaining algorithms evaluated by Tufekci [9] achieved RMSE's of between 3.861 MW (K-lazy -learning) and 8.487 MW (RBF function). The TOB Stage 1 testing subset predictions outperform all of those methods when applied to this dataset. Moreover, the TOB Stage 2 predictions do much better. For

the independent testing subset for all five datasets shuffles they achieve $RMSE < 3.3$ MW for the best optimized solution found by one or other of the three optimizers applied. The best performance for the testing subset is an RMSE of 2.62 MW for Case 2C (see Table S4 in the supplementary file).

More details for selected solutions are provided in Tables 4 and 5. Full solution and prediction accuracy details for all 15 cases evaluated are provided in Table S5 in the supplementary file. Table 4 lists the Q values and variable weights selected by the optimized TOB Stage 2 solutions for Cases 1B, 4B and 5C. Table 5 shows the full range of statistical accuracy metrics for these three cases applied to the testing subsets; highlighting that Cases 1B and 5C achieve higher levels of prediction accuracy than Case 4B (as that subset selected several outlying data records).

The optimum solutions for these selected cases have distinctive values of optimization variable Q and for the weights applied to the squared errors between the variables in the closely matching records. These are values selected by the optimizers to derive their optimum solutions (Table 4). For Cases 1B and 5C the Q value is high (9 or 10) and the highest weight is applied to variable V and the weights are arranged $V \gg AP > AT > RH$, with the RH weight approaching zero. For the sub-optimum testing subset Case 4C, Q is much lower (6) and the highest weight is applied to variable AP and the weights are arranged $AP \gg V > AT > RH$. The significance of these distinct solutions will become apparent when they are applied to the full datasets.

RMSE (TOB's objective function) is treated as the most important statistical accuracy metric to monitor. However, the other statistical accuracy metrics displayed in Table 5) usefully express different aspects of the degree of accuracy

Table 4 Q values (number of best matching records included) and weights for variables for three representative shuffles of the CCGT dataset evaluated

Selected cases run for shuffled full datasets			Tuning subset (146 data records) optimum solutions				
	Case solutions	Q	AT	V	AP	RH	
Case 1B	TOB stage 1	10	0.5	0.5	0.5	0.5	
	TOB stage 2	Memetic firefly	10	1.96E-02	1.00E+00	3.73E-01	1.25E-04
	TOB stage 2	Excel GRG	9	2.42E-02	1.00E+00	6.82E-01	2.02E-04
	TOB stage 2	Excel evolutionary	9	1.09E-02	1.00E+00	9.44E-02	3.72E-04
Case 4B	TOB stage 1	10	0.5	0.5	0.5	0.5	
	TOB stage 2	Memetic firefly	6	1.02E-01	8.01E-01	1.00E+00	3.65E-02
	TOB stage 2	Excel GRG	6	1.02E-01	8.02E-01	1.00E+00	3.65E-02
	TOB stage 2	Excel evolutionary	6	1.02E-01	7.99E-01	1.00E+00	3.65E-02
Case 5C	TOB stage 1	10	0.5	0.5	0.5	0.5	
	TOB stage 2	Memetic firefly	10	1.17E-02	1.00E+00	2.87E-02	0.00E+00
	TOB stage 2	Excel GRG	10	1.08E-02	1.00E+00	3.44E-02	5.40E-05
	TOB stage 2	Excel evolutionary	10	1.58E-02	9.66E-01	2.49E-02	1.36E-04

For the results of all 15 cases (45 optimizer runs) evaluated see Table S5 in the supplementary file

Table 5 Detailed prediction accuracy measures for the three representative case optimum solutions listed in Table 4

Selected cases run for shuffled full datasets		Applying optimum TOB tuning subset solutions applied to independent testing subset (197 data records)						
Case solutions		RMSE	MSE	APD %	AAPD %	SD	R	R ²
Case 1B	TOB stage 1	3.4076	11.6119	-0.0358	0.5387	3.4134	0.9806	0.9615
TOB stage 2	Memetic firefly	2.6839	7.2033	-0.0446	0.4220	2.6841	0.9880	0.9762
TOB stage 2	Excel GRG	2.7312	7.4593	-0.0407	0.4300	2.7329	0.9876	0.9753
TOB stage 2	Excel evolutionary	2.7312	7.4593	-0.0407	0.4300	2.7329	0.9876	0.9753
Case 4B	TOB stage 1	3.5856	12.8566	-0.0731	0.5417	3.5815	0.9786	0.9576
TOB stage 2	Memetic firefly	3.2688	10.6852	-0.0530	0.4940	3.2700	0.9822	0.9647
TOB stage 2	Excel GRG	3.2687	10.6845	-0.0530	0.4939	3.2699	0.9822	0.9647
TOB stage 2	Excel evolutionary	3.2690	10.6865	-0.0530	0.4940	3.2702	0.9822	0.9647
Case 5C	TOB stage 1	2.9551	8.7325	0.0133	0.4776	2.9617	0.9854	0.9710
TOB stage 2	Memetic firefly	2.8556	8.1547	-0.0093	0.4506	2.8627	0.9865	0.9731
TOB stage 2	Excel GRG	2.8336	8.0295	-0.0094	0.4484	2.8407	0.9867	0.9735
TOB stage 2	Excel evolutionary	2.8691	8.2316	-0.0075	0.4522	2.8763	0.9863	0.9728

For results of all 15 cases evaluated see Table S5 in the supplementary file

achieved by the optimum solutions. It would be possible to configure TOB to use AAPD % or SD as its objective to minimize. Such configurations do lead to similar (particularly using SD) but not identical solutions and prediction accuracies.

4.2 Applying optimum solutions to all 9568 data records in the CCGT data set

The robustness of the stage 1 and 2 solutions derived for the tuning subsets in terms of the Q and squared error weights derived need to be verified. This is done by applying those optimum tuning subset solutions to the full datasets. Computationally that is quite a significant task, as each record needs to be tested against every other data record in its dataset and the sum of its squared errors calculated. However, such an analysis provides the best means of verifying that the prediction accuracy suggested by the testing subset evaluations is repeatable for the entire CCGT UCI dataset. Tables 6 and 7, in the same format as Tables 4 and 5, summarise the results. Table 6 lists the optimum solutions derived by the memetic firefly optimizer. Optimum solutions derived from the other optimizers are not displayed but are very close to those listed for the memetic firefly optimizer.

It is noticeable in Table 6 that most of the solutions displayed apply variable squared-error weights in the magnitude order $V \gg AP > AT > RH$ and select Q values at or close to 10. Solutions for Cases 4B and 5B (the poor performers with the testing subsets, Table 3) stand out in Table 6, because they have lower Q values and the variable weights derived for their optimum solutions are in the magnitude order $AP \gg V > AT > RH$. It is interesting to evaluate how these two quite different type of "optimum"

solutions derived perform with respect to all records in the CCGT dataset.

The low correlations with PE (Table 2) and the data dispersion (Fig. 1) explain why the variable RH is assigned such low weights by the TOB Stage 2 optimizers. It is interesting to consider why the TOB's optimizers impose higher squared-error weights on variables V and AP, in most cases, than to AT. Both AT and V have strong negative correlations with PE (Table 2). However, V shows less dispersion due to outliers than AT. By applying small weights to AT this potentially minimizes the impact of some that dispersion in the ranking of the matches. Also, the moderate positive correlation between AP and PE versus the negative correlation between V and PE work effectively together to exploit contrasting and complex relationships between the data records. It is less productive for the optimizers to assign high weights to both V and AT that display more similar relationships with PE. Assigning a small, but non-zero weight to a variable can often be an effective way for TOB Stage 2 to distinguish between quite similar data records. Lower weight, therefore, does not necessarily equate to lower significance in a variable's contribution to the prediction accuracy.

Table 7 reveals that the prediction performance results fall within quite a narrow range for all the datasets and cases evaluated. The TOB Stage 1 performance is almost identical for all five shuffled datasets, as should be expected (RMSE = 3.53; AAPD % = 0.55; SD = 3.53), significantly outperforming the best of the regression algorithms evaluated by Tufekci [9] and ANN analysis of Elfaki and Hassan [13]. All cases, except Case 5B, show improvements on that accuracy by applying the optimized solutions. Average prediction accuracy achieved for all data records (including all the outliers, Fig. 1) of all fifteen

Table 6 Values of Q and variable squared-error weights derived for all fifteen cases applied to the five shuffled CCGT datasets

TOB applied to five shuffled datasets			Optimum solution weights applied			
9568 records	Applied to testing subset	Q	AT	V	AP	RH
Set 1	TOB stage 1	10	0.5	0.5	0.5	0.5
Case 1A	Memetic firefly	7	1.10E-02	1.00E+00	9.43E-02	3.77E-04
Case 1B	Memetic firefly	10	1.96E-02	1.00E+00	3.73E-01	1.25E-04
Case 1C	Memetic firefly	10	8.69E-02	1.00E+00	6.25E-02	0.00E+00
Set 2	TOB stage 1	10	0.5	0.5	0.5	0.5
Case 2A	Memetic firefly	9	3.28E-02	1.00E+00	7.58E-03	1.21E-03
Case 2B	Memetic firefly	8	9.54E-03	1.00E+00	3.19E-01	2.37E-03
Case 2C	Memetic firefly	10	4.53E-02	1.00E+00	1.64E-01	0.00E+00
Set 3	TOB stage 1	10	0.5	0.5	0.5	0.5
Case 3A	Memetic firefly	10	4.55E-02	1.00E+00	1.53E-02	8.13E-03
Case 3B	Memetic firefly	8	1.92E-05	1.00E+00	0.00E+00	0.00E+00
Case 3C	Memetic firefly	10	2.03E-02	1.00E+00	1.78E-01	8.14E-02
Set 4	TOB stage 1	10	0.5	0.5	0.5	0.5
Case 4A	Memetic firefly	5	2.16E-02	1.00E+00	2.28E-03	2.42E-03
Case 4B	Memetic firefly	6	1.02E-01	8.01E-01	1.00E+00	3.65E-02
Case 4C	Memetic firefly	10	1.28E-02	1.00E+00	8.93E-01	6.34E-02
Set 5	TOB stage 1	10	0.5	0.5	0.5	0.5
Case 5A	Memetic firefly	10	6.60E-03	1.00E+00	5.93E-02	4.05E-04
Case 5B	Memetic firefly	6	1.14E-01	2.45E-01	1.00E+00	2.66E-04
Case 5C	Memetic firefly	10	1.17E-02	1.00E+00	2.87E-02	0.00E+00

Only the optimized solutions obtained with the memetic firefly optimizer are displayed and evaluated with the full datasets

Table 7 Detailed prediction accuracy achieved applying TOB's optimum solutions (Table 6) to the entire CCGT UCI datasets (9568 records)

TOB applied to five shuffled datasets			Statistical measures of prediction accuracy						
9568 records	Applied to full dataset	Q	RMSE	MSE	APD %	AAPD %	SD	R	R ²
Set 1	TOB stage 1	10	3.5288	12.4525	-0.0084	0.5450	3.5290	0.9784	0.9572
Case 1A	Memetic firefly	7	3.3465	11.1993	-0.0124	0.4850	3.3465	0.9806	0.9616
Case 1B	Memetic firefly	10	3.3261	11.0630	-0.0141	0.4874	3.3260	0.9809	0.9621
Case 1C	Memetic firefly	10	3.3204	11.0251	-0.0110	0.4884	3.3204	0.9809	0.9622
Set 2	TOB stage 1	10	3.5287	12.4520	-0.0083	0.5450	3.5289	0.9784	0.9572
Case 2A	Memetic firefly	9	3.4867	12.1569	-0.0206	0.5339	3.4861	0.9789	0.9583
Case 2B	Memetic firefly	8	3.5105	12.3233	-0.0200	0.5342	3.5100	0.9786	0.9578
Case 2C	Memetic firefly	10	3.4990	12.2427	-0.0187	0.5339	3.4986	0.9788	0.9580
Set 3	TOB stage 1	10	3.5288	12.4523	-0.0084	0.5450	3.5289	0.9784	0.9572
Case 3A	Memetic firefly	10	3.2864	10.8007	-0.0105	0.4862	3.2865	0.9813	0.9629
Case 3B	Memetic firefly	8	3.3918	11.5042	-0.0110	0.4948	3.3918	0.9802	0.9608
Case 3C	Memetic firefly	10	3.3759	11.3970	-0.0124	0.5087	3.3760	0.9802	0.9609
Set 4	TOB stage 1	10	3.5282	12.4483	-0.0085	0.5449	3.5284	0.9784	0.9573
Case 4A	Memetic firefly	5	3.4705	12.0445	-0.0106	0.5083	3.4706	0.9792	0.9588
Case 4B	Memetic firefly	6	3.4528	11.9218	-0.0117	0.5129	3.4528	0.9794	0.9591
Case 4C	Memetic firefly	10	3.4002	11.5613	-0.0133	0.5107	3.4002	0.9800	0.9603
Set 5	TOB stage 1	10	3.5284	12.4493	-0.0085	0.5449	3.5285	0.9784	0.9573
Case 5A	Memetic firefly	10	3.2703	10.6951	-0.0131	0.4735	3.2703	0.9815	0.9633
Case 5B	Memetic firefly	6	3.5425	12.5495	-0.0110	0.5261	3.5426	0.9783	0.9570
Case 5C	Memetic firefly	10	3.2678	10.6784	-0.0114	0.4732	3.2678	0.9815	0.9634

cases is RMSE = 3.40; AAPD % = 0.5038 and SD = 3.39. For those performance-accuracy metrics, the five datasets are ranked as follows (best to worst performance): for RMSE and SD, Case 1 < Case 3 < Case 5 < Case 4 < Case 2; for AAPD % Case 1 < Case 5 < Case 3 < Case 4 < Case 2. So, for those metrics, Case 1 provides slightly better accuracy than the rest, whereas Case 2 clearly provides slightly worse accuracy than the rest. Interestingly datasets 4 and 5, with the somewhat problematic results with the testing subsets associated with Case 4B and Case 5B, lie in the middle and do not underperform the other cases.

It is not surprising that the performance accuracy metrics for the full dataset applications are slightly poorer than for the small, 197-data-record, testing subsets, because all the outliers in the datasets are included in this full-dataset analysis. The possibility to significantly improve upon this accuracy, by careful analysis and selective filtering of the outliers, is addressed in Sect. 5.1. Firstly, in order to conduct that analysis, it is necessary to consider the wealth of

additional information the calculated TOB learning network cases provide on each data record. This makes it possible to gain significant insight to the datasets as a whole; something that is not possible with neural network methods. The availability of such information makes it possible to conduct various meaningful data mining tasks on such datasets that have the potential to assist CCGT operators in improving plant performance.

4.3 Segmental analysis of the CCGT dataset

It is generally useful to know how machine-learning-prediction algorithms are performing over different segments of a datasets. The CCGT dataset is evaluated in twelve segments of incremental ranges of electrical power output (PE). Table 8 displays prediction accuracy by segment in terms of RMSE, AAPD and SD for two representative cases (Case 2C and Case 5C).

Table 8 Segmental analysis of the datasets for two selected case evaluated: Case 2C and Case 5C

Dependent variable segment	Data records in segment	Min PE (MW)	Max PE (MW)	RMSE	AAPD %	SD
<i>From case 2C TOB stage 2 solution</i>						
1	800	420.26	432.62	3.9576	0.6786	2.9507
2	800	432.64	436.37	3.4531	0.5118	3.3726
3	800	436.37	439.78	3.7963	0.6026	3.7986
4	800	439.78	443.06	3.2299	0.5553	3.2319
5	800	443.06	446.92	3.7172	0.5483	3.6752
6	800	446.93	451.71	3.5875	0.5468	3.5383
7	800	451.74	457.67	3.6070	0.5794	3.6092
8	800	457.68	463.86	3.7348	0.5687	3.6901
9	800	463.86	468.58	3.1643	0.4745	3.1655
10	800	468.60	474.16	3.2191	0.5033	3.1956
11	800	474.17	480.54	2.8805	0.4117	2.8522
12	768	432.64	495.76	3.4870	0.4212	3.2253
Total	9568		Full set:	3.4990	0.5339	3.4986
<i>From case 5C TOB stage 2 solution</i>						
1	800	420.26	432.62	3.6544	0.5919	3.1370
2	800	432.64	436.37	3.4042	0.4884	3.3477
3	800	436.37	439.78	3.5446	0.5271	3.5466
4	800	439.78	443.06	3.1780	0.5230	3.1789
5	800	443.06	446.92	3.5459	0.5183	3.5293
6	800	446.93	451.71	3.4162	0.4844	3.3902
7	800	451.74	457.67	2.9775	0.4613	2.9777
8	800	457.68	463.86	3.4435	0.4681	3.4268
9	800	463.86	468.58	2.7706	0.4027	2.7723
10	800	468.60	474.16	2.8952	0.4109	2.8935
11	800	474.17	480.54	2.7350	0.3883	2.7184
12	768	432.64	495.76	3.4762	0.4113	3.2569
Total	9568		Full set:	3.2678	0.4732	3.2678

Such information is routinely generated as part of the TOB algorithm configured for this study

It is quite clear for the cases shown in Table 8 that segment 1, at the lower end of the PE scale achieves lower prediction accuracy than the other segments. Indeed, the lowest three segments, collectively, typically underperform the other segments in terms of RMSE and AAPD %. However, in terms of SD the lowest segment generally displays a lower SD than the other segments in the lower half of the distribution. At the other end of the distribution, segments 9–11 show the lowest RMSE, AAPD % and SD compared to other segments. These relationships are observed for all 15 cases evaluated. Careful scrutiny of the spreads of the distributions for the AT and V variables in Fig. 1 go some way to explaining these relationships. From a plant operator's perspective, a case could be made for reconfiguring the predictions for these datasets into at least three groupings: (1) from PE = ~425 to ~440 MW; (2) from PE ~440 to ~465 MW; and (3) from ~465 to ~495 MW. A case can also be made for separating out a fourth, smaller group of outliers (see Sects. 5.1 and 5.2). By approaching predictions in this way greater accuracy would likely be achieved. It would be a straightforward exercise for TOB to conduct the analysis in that way.

The level of detail generated by each TOB prediction is available for more in-depth data mining to evaluate in more detail all outlying predictions. Indeed, all interim calculation details for each data record and its predictions are recorded. In this study, such details are recorded for the full dataset as part of the Sect. 4.2 analysis. These details are made available for complete inspection and forensic analysis with each execution of the TOB algorithm. Section 5.4 of the Supplementary File (Tables S1 to S3) provides a detailed example of TOB's forensic auditing capabilities using this information to gain insight to the prediction made for an arbitrarily selected data record (#4370). The independent variables of the selected best data record matches make different relative contributions to the PE predictions. Understanding how these contributions are calculated opens up data-mining possibilities, which is addressed next in Sect. 5.1 for outlier analysis.

5 Discussion

5.1 Outlier analysis of the CCGT dataset

The UCI dataset of long-term CCGT performance data includes a number of outlier data records. These few outliers limit the prediction accuracy that can be achieved by optimized data matching (Table 7), regression and other machine-learning algorithms. The extent of that impact is illustrated by Fig. 5 for Case 5C showing the predicted versus measured PE predictions applying the memetic firefly optimized TOB Stage 2 solution.

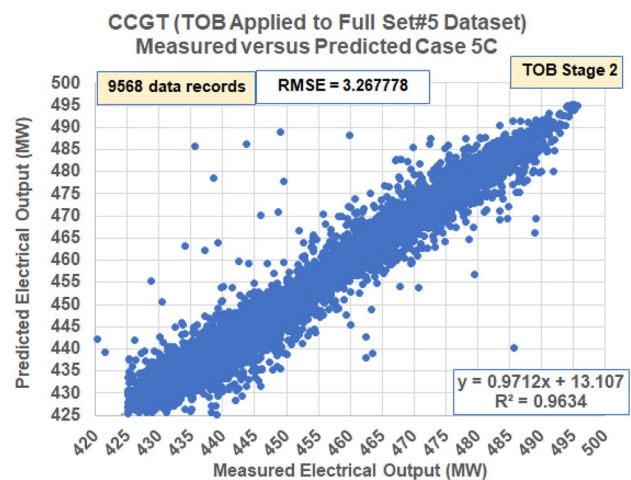


Fig. 5 Predicted (TOB Stage 2) versus measured PE for Case 5C for full dataset #5 including all outliers

A simple filter can be applied to identify and remove data records from the dataset that exceed a specified squared-error (SE) magnitude between predicted and measured PE values. For Case 5C, a filter of SE = 200 MW (maximum SE accepted for Stage 1 predictions). Applying this SE filter excludes between 35 and 38 data records, depending on the dataset. That represents < 0.4% of the entire dataset. For a lower filter threshold, SE = 100 MW, more data records breach the filter's threshold. About 126 data records (~ 1.3% of the CCGT dataset) are identified and excluded applying SE = 100 MW. The 36 outliers identified for dataset 5 with the filter SE = 200 MW are shown in Fig. 6.

TOB's PE predictions for each of these outlying data records, when analysed in detail, reveal that there are different factors influencing the anomalous predictions versus measured values for these data-record outliers. Moreover, there are interesting interactions between some of these outliers. Considering the two outliers #3231 and #1684 (Fig. 6), an analysis of the TOB data records, summarized in Table 9, reveals that #3231 is indeed anomalous, but record #1684 is only anomalous because it is closely matched by ΣwSE with #3231. Figure 1a reveals that #3231 is an extreme outlier in terms of AT with respect to the on-trend #1684. The PE values of the top-ten matching records for #3231 are all quite different from the anomalous PE value for #3231 (column 4 top part of Table 9). Unfortunately for #1684, it has the closest ΣwSE match with #3231, meaning that the AT, V, AP and RH values for these two records are very similar. Consequently, #1684 contributes 92.07% to the PE prediction for #3231. A more undesirable consequence of the close match between these records is that #3231 contributes 90.59% to the PE prediction of #1684. Note that the other nine matching records for

Fig. 6 36 outlying data records identified with filter SE maximum limit of 200 MW applied to dataset 5. Data record numbers identify the specific outliers

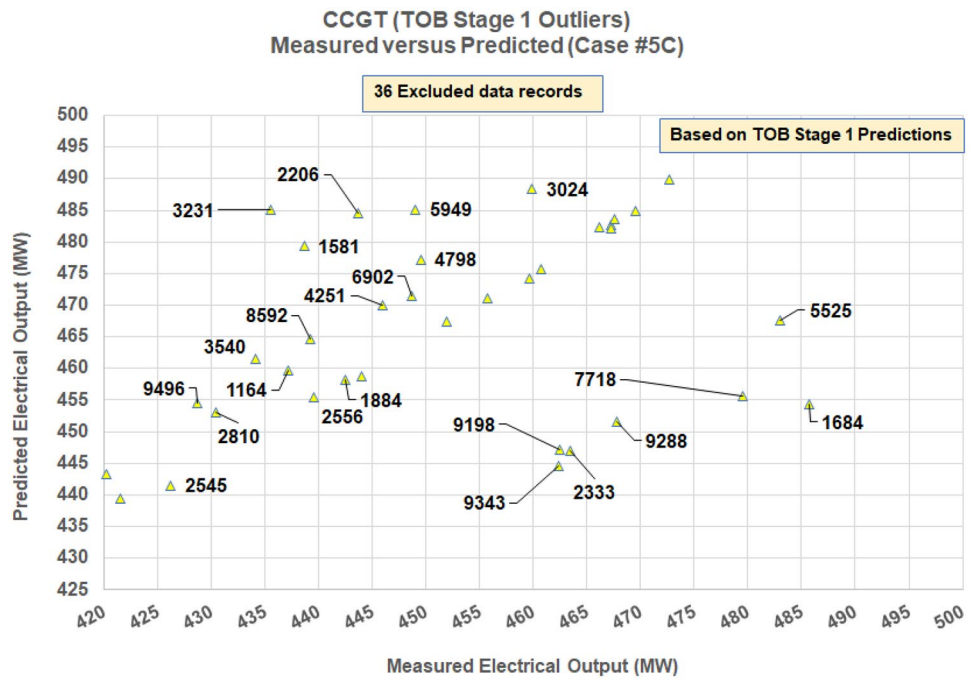


Table 9 TOB calculation data records reveal that outlying data records do sometimes interfere with and disrupt the predictions for on-trend data records

		ΣwSE	Measured normalized PE	Contributions to PE prediction	Predicted normalized PE
Rank of match to #3231	3231		-0.5942		0.7318
1	1684	0.00034	0.7343	0.9207	0.6761
2	8803	0.00177	0.6940	0.0311	0.0216
3	709	0.00264	0.6702	0.0233	0.0156
4	5037	0.00392	0.8143	0.0144	0.0117
5	9455	0.01003	0.6109	0.0016	0.0010
6	5308	0.0105	0.5597	0.0006	0.0004
7	6706	0.01168	0.6975	0.0006	0.0004
8	7027	0.01172	0.4135	0.0022	0.0009
9	4252	0.01215	0.7502	0.0049	0.0037
10	1671	0.01242	0.7160	0.0006	0.0005
Rank of match to #1684	1684		0.7343		-0.4721
1	3231	0.00034	-0.5942	0.9059	-0.5382
2	8803	0.00169	0.6940	0.0390	0.0271
3	709	0.00302	0.6702	0.0278	0.0186
4	5037	0.00404	0.8143	0.0166	0.0135
5	9455	0.00707	0.6109	0.0016	0.0010
6	7027	0.00845	0.4135	0.0022	0.0009
7	5308	0.00847	0.5597	0.0006	0.0004
8	6706	0.00857	0.6975	0.0006	0.0004
9	4252	0.00861	0.7502	0.0051	0.0038
10	1671	0.01051	0.7160	0.0006	0.0005

The prediction of #1684 (on-trend data record) is disrupted because it closely matches (ΣwSE values) with the true outlier #3231

#1684 have normalized PE values closer to that measured for #1684. If record #3231 is excluded from the dataset the PE prediction for #1684 would fall on trend with a low squared error from its measured value. On the other hand, if record #1684 is excluded from the dataset record #3231 would still plot as an extreme outlier. In this case applying the SE error filter correctly removes #3231 from the dataset but also unfortunately removes the on-trend record #1684.

Three other pairs of data records in the 36 excluded behave in a similar way. The outlier #2206 negatively impacts on-trend record #5525. The outlier #1581 negatively impacts on-trend record #7718. The outlier #1164 negatively impacts on-trend record #9198. If records #3231, #2206, #1581 and #1164 are excluded from the dataset, when the analysis is re-run (Stage 1 and Stage 2), records #1684, #5525, #7718 and #9198 plot on-trend with low prediction errors. This analysis highlights that there are interactions between outliers and data records that they match closely with in terms of their independent variable values. For every true outlier that is removed a beneficial outcome should be observed for those other data records that it potentially impacted. Such impacts are likely as the outliers can be selected as one of the top-ten highest matches for some other data records.

5.2 Impacts achieved by mildly filtering the CCGT dataset

Figure 7 plots the PE prediction analysis that excludes 35 of the outliers associated with dataset 5 (Case 5C) and returns it to become Case 8C. Note that the RMSE for the 9533-dataset for the Case 8C solution (retuned) is 2.945 compared to the full dataset (9568 data records) displayed in Fig. 4 for Case 5C with a RMSE of 3.268. Note that the

RMSE of this filtered dataset is quite close to the RMSE of 2.856 achieved for the testing subset of Case 5C (Tables 3 and 5). This is extremely encouraging as it confirms the validity of the TOB methodology. I.e., training small and representative tuning subsets with large training subsets leads to optimum solutions that is not unduly impacted by outliers. The average of three or so such tuning/testing subset runs can be considered as providing realistic predictions for all data records.

Table 10 displays the prediction accuracy statistics for all 15 cases evaluated with a 200 maximum SE filter applied, removing some 36 records consisting of genuine outliers and some data records influenced by outliers. Clearly, there is further scope for more detailed outlier analysis, adjustment and filtering to fine-tune and clean up the dataset for operational usage. The TOB calculated dataset provides the key information to do this in an easy-to-access and interpret format.

Table 10 reveals that the prediction performance results fall within quite a narrow range for all the filtered datasets and cases evaluated. The TOB Stage 1 performance is almost identical for all five shuffled datasets, as should be expected (RMSE = 3.23; AAPD % = 0.53; SD = 3.33), showing significant improvements from the results for the full dataset shown in Table 7. All cases, including Case 5B, show improvements on that accuracy by applying optimized solutions originally tuned and evaluated using the complete dataset. The average prediction accuracy achieved for filtered datasets by re-evaluating all fifteen cases is RMSE = 3.07; AAPD % = 0.4876 and SD = 3.07. For those performance-accuracy metrics, the five datasets are ranked as follows (best to worst performance) for AAPD %, RMSE and SD, Case 1 < Case 5 < Case 3 < Case 4 < Case 2.

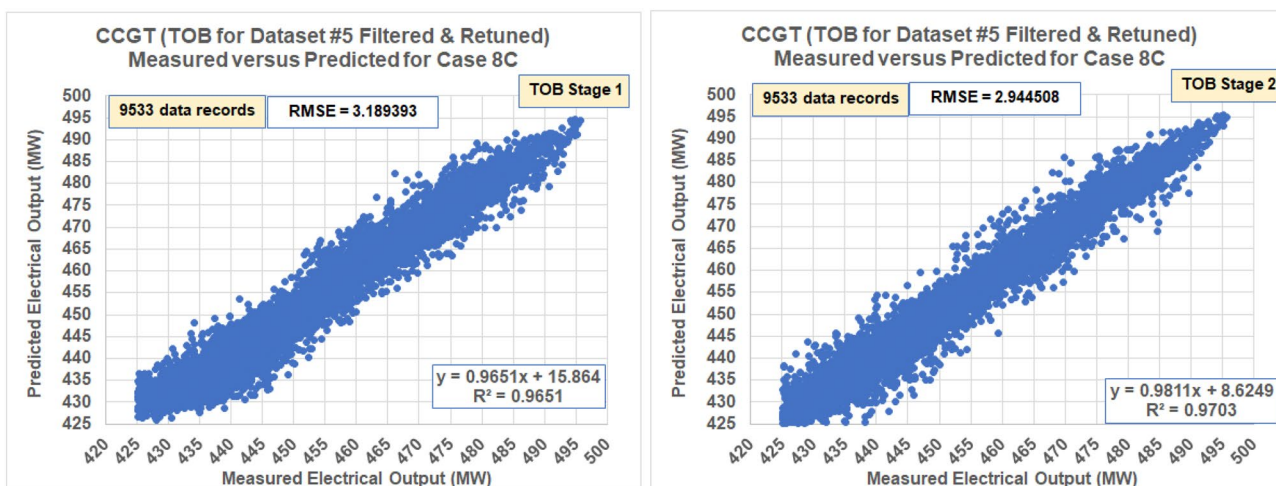


Fig. 7 Predicted (TOB Stage 1 left; TOB Stage 2 right) versus measured PE for Case 8C for dataset #5 excluding 35 of the outliers identified in Fig. 5 and for the tuning subset for Case 5C retuned by sampling only the filtered dataset (9533 records)

Table 10 TOB applied to filtered datasets (9532 data records; just a few records removed) leads to improved prediction accuracy in the five shuffled CCGT UCI datasets

TOB applied to five shuffled datasets		Statistical measures of prediction accuracy							
9532 or 9533 records	Applied to full dataset	Q	RMSE	MSE	APD %	AAPD %	SD	R	R ²
Set 1	TOB stage 1	10	3.2258	10.4059	0.0040	0.5287	3.2257	0.9820	0.9643
Case 1A	Memetic firefly	7	3.0056	9.0339	-0.0008	0.4687	3.0058	0.9844	0.9690
Case 1B	Memetic firefly	10	2.9870	8.9221	-0.0024	0.4712	2.9871	0.9846	0.9694
Case 1C	Memetic firefly	10	2.9869	8.9215	0.0011	0.4721	2.9870	0.9846	0.9694
Set 2	TOB stage 1	10	3.2257	10.4054	0.0040	0.5286	3.2256	0.9820	0.9643
Case 2A	Memetic firefly	9	3.1893	10.1718	-0.0080	0.5179	3.1895	0.9824	0.9651
Case 2B	Memetic firefly	8	3.1889	10.1692	-0.0077	0.5178	3.1890	0.9824	0.9651
Case 2C	Memetic firefly	10	3.1803	10.1145	-0.0064	0.5175	3.1805	0.9825	0.9653
Set 3	TOB stage 1	10	3.2258	10.4057	0.0040	0.5287	3.2257	0.9820	0.9643
Case 3A	Memetic firefly	10	2.9632	8.7808	0.0014	0.4700	2.9633	0.9848	0.9699
Case 3B	Memetic firefly	8	3.1182	9.7230	0.0018	0.4806	3.1183	0.9832	0.9668
Case 3C	Memetic firefly	10	3.0488	9.2953	-0.0007	0.4923	3.0489	0.9839	0.9681
Set 4	TOB stage 1	10	3.2252	10.4017	0.0038	0.5285	3.2251	0.9820	0.9643
Case 4A	Memetic firefly	5	3.1615	9.9953	0.0019	0.4918	3.1616	0.9827	0.9657
Case 4B	Memetic firefly	6	3.1222	9.7479	0.0003	0.4962	3.1223	0.9831	0.9665
Case 4C	Memetic firefly	10	3.0799	9.4858	-0.0015	0.4946	3.0800	0.9836	0.9674
Set 5	TOB stage 1	10	3.2253	10.4027	0.0038	0.5285	3.2252	0.9820	0.9643
Case 5A	Memetic firefly	10	2.9225	8.5410	-0.0016	0.4573	2.9226	0.9852	0.9707
Case 5B	Memetic firefly	6	3.2190	10.3619	0.0013	0.5095	3.2190	0.9821	0.9644
Case 5C	Memetic firefly	10	2.9207	8.5308	0.0002	0.4569	2.9208	0.9853	0.9707

For the filtered dataset metrics, Case 1 provides slightly better accuracy than the rest, whereas Case 2 clearly provides slightly worse accuracy than the rest. Case 5 consolidates its position in second place despite the problematic results applying the optimum solution found for the Case 5b tuning subset to the testing subset.

The question arises as to what should be done about the excluded outliers. Although plotting in anomalous positions in Fig. 1, and resulting in poor PE predictions, they are genuine cases accepted by the data cleansing process applied by Tufekci [9]. A case can be made that at least some of the 36 outliers excluded by the filtering process constitute one or more data sets that should be treated separately as a special case(s). For instance, data records 1581, 1884, 2206, 3540, 3765, 6561, 8592, 9496 and some others plot in a particular region of the PE distribution < 435 MW. These records have anomalously low AT and AP values and, for the most part, anomalously high V values (Fig. 1). This suggests that these records are related to a specific set of operating and/or climatic conditions. Although those conditions rarely occur, there is a chance that they will occur from time to time. This suggests that it is appropriate, from an operational standpoint, to identify and define these specific conditions. Such defined conditions can then be treated as limited special cases. These small subsets can then be evaluated separately from the

main dataset using their own TOB feature selection to derive more appropriate tuned solutions focused on the outliers.

5.3 Exploiting ratios of input variables to enhance data characterization

In datasets with relatively few independent variables, the discrimination between closely matching data records can sometimes be enhanced by considering the inclusion of ratios between those variables as potential additional data input variables for the TOB Stage 2 optimizers to exploit. Prediction accuracy benefits are most likely to be achieved in this way for datasets where the independent variables show much poorer correlations with the dependent variables than their ratios.

Table 11 Correlation coefficients (R) among ratios of variables and PE for the 9568-UCI- CCGT dataset [19]

R	AP/AT	V/AT	RH/AT	PE
AP/AT	1.000	0.945	0.971	0.832
V/AT		1.000	0.931	0.676
RH/AT			1.000	0.796
PE				1.000

Table 12 Solution-selected values for Q and squared-error weights applied to variables for optimized solutions applied to Cases 6A, 6B, 6C, 6D and 7 (retuned)

197 records	Applied to testing subset	Q	AT	V	AP	RH	AP/AT	RH/AT
<i>Optimum solutions for tuning subset applied to training subset (both including outliers)</i>								
Case 6A	TOB stage 1	10	0.5	0.5	0.5	0.5	0.5	0.5
TOB stage 2	Memetic firefly	10	0.00E+00	1.00E+00	7.64E-02	1.36E-04	3.41E-01	6.79E-05
TOB stage 2	Excel GRG	10	0.00E+00	1.00E+00	7.61E-02	1.35E-04	3.39E-01	0.00E+00
TOB stage 2	Excel evolutionary	10	0.00E+00	9.62E-01	9.91E-02	1.38E-04	4.11E-01	3.25E-03
Case 6B	TOB stage 1	10	0.5	0.5	0.5	0.5	0.5	0.5
TOB stage 2	Memetic firefly	10	1.47E-02	1.00E+00	3.74E-01	1.47E-04	7.61E-01	0.00E+00
TOB stage 2	Excel GRG	10	1.35E-02	1.00E+00	3.10E-01	1.18E-04	5.47E-01	0.00E+00
TOB stage 2	Excel evolutionary	7	1.84E-01	7.50E-01	1.00E+00	1.81E-03	1.09E-01	2.46E-06
Case 6C	TOB stage 1	10	0.5	0.5	0.5	0.5	0.5	0.5
TOB stage 2	Memetic firefly	6	1.79E-02	1.00E+00	5.48E-02	0.00E+00	0.00E+00	0.00E+00
TOB stage 2	Excel GRG	6	1.81E-02	1.00E+00	5.51E-02	0.00E+00	0.00E+00	0.00E+00
TOB stage 2	Excel evolutionary	6	1.85E-02	9.84E-01	5.84E-02	0.00E+00	0.00E+00	3.66E-05
<i>Optimum solutions from tuning subset applied to both full (9568 records) and filtered data sets (9532 records)</i>								
Dataset #5	TOB stage 1	10	0.5	0.5	0.5	0.5	0.5	0.5
Case 6A	Memetic firefly	10	0.00E+00	1.00E+00	7.64E-02	1.36E-04	3.41E-01	6.79E-05
Case 6B	Memetic firefly	10	1.47E-02	1.00E+00	3.74E-01	1.47E-04	7.61E-01	0.00E+00
Case 6C	Memetic firefly	6	1.79E-02	1.00E+00	5.48E-02	0.00E+00	0.00E+00	0.00E+00
<i>Optimum solutions from tuning subset applied to re-tuned filtered data sets (9533 records)</i>								
Dataset #5	TOB stage 1	10	0.5	0.5	0.5	0.5	0.5	0.5
Case 7A	Memetic firefly	9	1.14E-03	1.00E+00	6.94E-03	3.12E-05	1.24E-05	0.00E+00
Case 7B	Memetic firefly	9	0.00E+00	1.00E+00	1.22E-03	8.26E-05	4.73E-03	4.86E-05
Case 7C	Memetic firefly	10	9.92E-05	1.00E+00	0.00E+00	1.44E-04	3.25E-05	0.00E+00
Dataset #5	TOB stage 1	10	0.5	0.5	0.5	0.5	No ratios used	
Case 8A	Memetic firefly	9	7.68E-04	1.00E+00	4.11E-03	1.36E-04		
Case 8B	Memetic firefly	9	4.78E-04	1.00E+00	8.00E-04	7.88E-05		
Case 8C	Memetic firefly	9	1.98E-04	9.45E-01	0.00E+00	5.52E-05		

These cases use dataset 5 but with six input variables including two ratios. Case 5C (four input variables) is retuned for the filtered dataset is shown for comparison and labelled Case 8

Correlation coefficients (R) for the ratio variables with AT as the denominator for the CCGT datasets are shown in Table 11. AP/AT and RH/AT show better correlations with PE than AP and RH (Table 2). For this reason, dataset #5 is analysed using six input variables (AT, V, AP, RH, AP/AT and RH/AT) instead of the four independent variables used in the other cases evaluated.

Tables 12 and 13 show the results of that analysis. Table 12 considers prediction accuracy when the solutions are applied to the testing subsets. Table 13 focuses on accuracy when those solutions are applied to the entire dataset. Filtered and filtered plus retuned datasets for Cases 6A, 6B, 6C and 6D are also evaluated. These cases, involving six input variables, are using the same data records as Cases 5A, 5B, 5C, 5D but with which they are compared (Tables 6 and 7).

The optimum solutions derived for Case 6 still assign the highest variable weight to variable V (Table 12). For Case 6A (weights $V \gg AP/AT > AP > RH > RH/AT$) the

optimizer assigns the second highest weight to ratio AP/AT and assigns no weight to variable T. For Case 6 B (weights $V \gg AP/AT > AP > AT > RH > RH/AT$) the optimizer assigns higher weights to AP/AT than AT. For Case 6 C (weights $V > AP > AT$) no weights are assigned to the ratio or RH. Of course, for all cases the ratios are assigned equal weight to the other variables in selecting the top-ten record matches for TOB Stage 1.

The prediction performance results for Cases 6A, 6B and 6C displayed in Table 13, when compared with Cases 5A, 5B, 5C (Tables 7 and 10) show that the Case 6 prediction accuracy is very similar but slightly poorer than for Case 5 based on just 4 independent variables. For the full dataset the average RMSE for Cases 5 and 6 are both 3.360, whereas, for the filtered dataset the average RMSE for Case 6 is 3.025 versus 3.021 for Case 5. For the filtered and retuned datasets, the average RMSE for Case 7 (6 input variables) is 2.99 versus 2.89 for Case 8 (4 input variables).

Table 13 Detailed prediction accuracy measures established for case 6 applied to CCGT UCI dataset 5 utilizing 6 input variables including ratios AP/AT and RH/AT

197 records	Applied to testing subset	Q	RMSE	MSE	APD %	AAPD %	SD	R	R ²
<i>Performance accuracy for optimum tuning subset solution applied to testing subset (including and excluding outliers)</i>									
Case 6A	TOB stage 1	10	3.2631	10.6477	-0.0210	0.5384	3.2707	0.9822	0.9647
TOB stage 2	Memetic Firefly	10	2.8115	7.9044	-0.0105	0.4659	2.8185	0.9868	0.9738
TOB stage 2	Excel GRG	10	2.8115	7.9046	-0.0105	0.4659	2.8185	0.9868	0.9738
TOB Stage 2	Excel evolutionary	10	2.8140	7.9186	-0.0115	0.4665	2.8210	0.9868	0.9737
Case 6B	TOB stage 1	10	3.6711	13.4767	-0.0011	0.5600	3.6803	0.9774	0.9552
TOB stage 2	Memetic firefly	10	3.3885	11.4816	0.0116	0.4855	3.3963	0.9808	0.9620
TOB stage 2	Excel GRG	10	3.3662	11.3316	0.0111	0.4806	3.3741	0.9811	0.9625
TOB stage 2	Excel evolutionary	7	3.5789	12.8082	-0.0044	0.5070	3.5880	0.9786	0.9576
Case 6C	TOB stage 1	10	3.0050	9.0302	0.0150	0.4833	3.0115	0.9849	0.9700
TOB stage 2	Memetic firefly	6	2.9174	8.5110	0.0193	0.4613	2.9233	0.9860	0.9722
TOB stage 2	Excel GRG	6	2.9172	8.5100	0.0194	0.4613	2.9232	0.9860	0.9722
TOB stage 2	Excel evolutionary	6	2.9144	8.4939	0.0193	0.4608	2.9204	0.9860	0.9723
<i>Optimum solutions applied to the full dataset (9568 records) including outliers</i>									
Dataset #5	TOB stage 1	10	3.5363	12.5055	-0.0068	0.5467	3.5365	0.9783	0.9571
Case 6A	Memetic firefly	10	3.3229	11.0419	-0.0054	0.5017	3.3231	0.9809	0.9621
Case 6B	Memetic firefly	10	3.3391	11.1494	-0.0077	0.4936	3.3392	0.9807	0.9617
Case 6C	Memetic firefly	6	3.4168	11.6743	-0.0109	0.4953	3.4168	0.9798	0.9600
<i>Optimum solutions applied to the filtered dataset (9530 records) excluding outliers (38 data records)</i>									
Dataset #5	TOB stage 1	10	3.2369	10.4777	0.0057	0.5301	3.2367	0.9818	0.9640
Case 6A	Memetic firefly	10	3.0120	9.0723	0.0069	0.4855	3.0117	0.9843	0.9688
Case 6B	Memetic firefly	10	3.0034	9.0206	0.0043	0.4765	3.0033	0.9844	0.9690
Case 6C	Memetic firefly	6	3.0588	9.3560	0.0006	0.4775	3.0589	0.9838	0.9679
<i>Optimum solutions applied to the filtered & re-tuned dataset (9533 records) excluding outliers (35 data records)</i>									
Dataset #5	TOB stage 1	10	3.2035	10.2622	-0.0052	0.5253	3.2036	0.9822	0.9648
Case 7A	Memetic firefly	9	2.8594	8.1762	-0.0077	0.4498	2.8595	0.9859	0.9720
Case 7B	Memetic firefly	9	3.1478	9.9087	-0.0018	0.5122	3.1480	0.9829	0.9660
Case 7C	Memetic firefly	10	2.9749	8.8497	-0.0087	0.4687	2.9749	0.9848	0.9697
Dataset #5	TOB stage 1	10	3.1894	10.1722	-0.0071	0.5232	3.1895	0.9824	0.9651
Case 8A	Memetic firefly	9	2.8520	8.1336	-0.0119	0.4481	2.8518	0.9860	0.9721
Case 8B	Memetic firefly	9	2.8696	8.2349	-0.0118	0.4501	2.8695	0.9858	0.9718
Case 8C	Memetic firefly	9	2.9445	8.6701	-0.0095	0.4637	2.9445	0.9851	0.9703

The optimized solutions applied for these cases are displayed in Table 12

Although prediction accuracy improvements were not achieved by adding the variable ratios for this dataset, it is worth conducting sensitivity analysis to establish their impact. Frequently, data mining and data characterization benefits are achieved by including ratio variables, particularly for datasets where the dependent variable to be predicted is poorly correlated with the input variables. The challenge with the CCGT dataset is for TOB Stage 1 calculations. As stage 1 predictions assigns equal squared-error weightings to all input variables, the choice for the ten-best matches will be quite different using six variables rather than four variables. Even if the subsequent optimized solutions choose not to apply weights to the squared errors of certain variables, those variables have

already influenced the selection of the best-matching records to be used in the stage 2 prediction analysis. Additional sensitivity analysis could investigate replacing some of the independent variable with ratios to avoid increasing the total number of variables above four or five.

6 Conclusions

The transparent open box (TOB) machine-learning method is successfully applied to a published CCGT dataset of 9568 data records. It is used to predict the hourly full-load electrical power output (PE). It does this utilizing four independent variables: ambient temperature

(AT), atmospheric pressure (AP), relative humidity (RH) and steam turbine exhaust pressure or vacuum (V). The TOB Stage 2 calculations achieved PE prediction with average accuracy for 15 executions tested against all the data points in five shuffled versions with an RMSE of 3.4 MW. This compares with the best RMSE achieved by fifteen machine learning regression methods and by artificial neural network, published for the same dataset, of 3.79 MW and 4.32 MW, respectively. Indeed, even the TOB Stage 1 preliminary average prediction for PE achieved an RMSE of 3.53 MW. This is additionally impressive because the TOB established its optimum solutions using a tuning subset of just 146 data points trained against a training set of 9225 data records. The verification testing showed that these optimum solutions (found by the small tuning subsets) were also effective over the full datasets. Moreover, their prediction accuracy was not excessively distorted by data outliers. A memetic firefly optimization algorithm is effective at generating the TOB stage 2 solutions. However, Microsoft Excel's Solver optimizers, run in parallel for this study, are also effective in generating high-performing TOB optimized solutions.

In addition to demonstrating its highly-accurate prediction capabilities the CCGT dataset is analysed to demonstrate its detailed data mining and auditing capabilities. The data set has significant outliers in its variable distributions. About 0.35–0.5% of the total data records can be considered as outliers. These outliers are shown by detailed analysis of the TOB prediction calculations to negatively impact the accuracy of the PE predictions. The outlier analysis identifies different types of outliers and leads to effective filtering of the dataset. The majority are outliers in the prediction-error distribution because of their anomalous independent variable values. However, some data records generate outlying predictions because their prediction calculations are influenced by the former group of outliers. Filtering the datasets by removing just 36 outlying data records achieves notable PE prediction accuracy improvements: TOB Stage 1 (RMSE = 3.23 MW) and TOB Stage 2 (RMSE = 3.07 MW). By removing 35 outliers and retuning the TOB analysis with the filtered datasets, prediction accuracy is improved further (Case 8) TOB Stage 1 (RMSE = 3.19 MW) and TOB Stage 2 (RMSE = 2.89 MW).

Data mining of the CCGT dataset also reveals that certain sectors of the PE distribution achieve different levels of prediction accuracy. This makes it possible to refine the TOB optimum solutions for specific segments of that distribution if required. Adding two ratios AP/AT and RH/AT as input variables (6 input variables in total) and applying the TOB model to a filtered dataset of 9533 data records achieved further improvements to prediction accuracy and repeatability. The detailed analysis, accuracy and reliability of the machine-learning predictions achieved

by the TOB algorithm demonstrated with the CCGT dataset are possible because of its transparency and forensic auditing capabilities. This provides TOB with a significant data mining advantage over correlation-based, machine-learning methods.

Compliance with ethical standards

Conflicts of interest The author has no conflicts of interest with respect to this study

Appendix

A supplementary file provides more details regarding the mathematical calculations required for the transparent open box method, examples of its forensic auditing capabilities and statistical measures of its performance accuracy.

An additional supplementary Excel data file is also available to download for the filtered CCGT dataset used for TOB analysis in this study. The file include the 9533 data records of the filtered dataset (35 outliers excluded) divided between TOB-selected training, tuning and testing subsets.

References

1. EIA (2019) Power blocks in natural gas-fired combined-cycle plants are getting bigger. US energy information administration today in energy 12 February 2019 report. <https://www.eia.gov/todayinenergy/detail.php?id=38312>
2. Kesgin U, Heperkan H (2005) Simulation of thermodynamic systems using soft computing techniques. *Int J Energy Res* 29:581–611
3. Arrieta FRP, Lora EES (2005) Influence of ambient temperature on combined-cycle power-plant performance. *Appl Energy* 80(3):261–272
4. De Sa A, Zubaidy AS (2011) Gas turbine performance at varying ambient temperature. *Appl Therm Eng* 31(14–15):2735–2739
5. Gonzalez-Díaz A, Alcaraz-Calderon AM, Gonzalez-Díaz MO, Menendez-Aranda A, Lucquiaud M, Gonzalez-Santalo JM (2017) Effect of the ambient conditions on gas turbine combined cycle power plants with post-combustion CO₂ capture. *Energy* 134:221–233. <https://doi.org/10.1016/j.energy.2017.05.020>
6. Şen G, Nil M, Mamura H, Doğan H, Karamolla M, Karaçor M, Kuyucuoğlu F, Yörükeren N, Ruhul M, Bhuiyan A (2018) The effect of ambient temperature on electric power generation in natural gas combined cycle power plant—a case study. *Energy Rep* 4:682–690. <https://doi.org/10.1016/j.egy.2018.10.009>
7. Lee JJ, Kang DW, Kim TS (2011) Development of a gas turbine performance analysis program and its application. *Energy* 36(8):5274–5285
8. Lee JH, Kim TS, Kim E-H (2017) Prediction of power generation capacity of a gas turbine combined cycle cogeneration plant. *Energy* 124:187–197. <https://doi.org/10.1016/j.energy.2017.02.032>

9. Tüfekci P (2014) Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods. *Int J Electr Power Energy Syst* 60:126–140. <https://doi.org/10.1016/j.ijepes.2014.02.027>
10. Tayarani-Bathaie SS, Sadough Vanini ZN, Khorasan K (2014) Dynamic neural network-based fault diagnosis of gas turbine engines. *Neurocomputing* 125(11):153–165
11. Arriagada J, Genrup M, Loberg A, Assadi, M (2003) Fault diagnosis system for an industrial gas turbine by means of neural networks. *Proceedings of the International Gas Turbine Congress 2003 Tokyo November 2–7, 2003*. p 6
12. Asgari H, Venturini M, Chen XQ, Sainudiin R (2014) Modelling and simulation of the transient behaviour of an industrial power plant gas turbine. *ASME J Eng Gas Turb Power* 136(6):061601–0616010. <https://doi.org/10.1115/1.4026215>
13. Elfaki E, Hassan AHA (2018) Prediction of electrical output power of combined cycle power plant using regression ANN model. *Int J Comput Sci Control Eng* 6(2):9–21
14. Du D, Jia X, Hao C. (2016) A new least squares support vector machines ensemble model for aero engine performance parameter chaotic prediction. *Math Probl Eng* p 8, Article ID 4615903. <http://dx.doi.org/10.1155/2016/4615903>
15. Elhosseini MA, Sehlemey RA, Salah AH, Abido MA (2017) Modeling and control of an interconnected combined cycle gas turbine using fuzzy and ANFIS controllers. *Electr Eng* 1:1. <https://doi.org/10.1007/s00202-017-0547-x>
16. Michelassi V, Allegorico, C, Cioncolini, S, Graziano A, Tognarelli L, Sepe M (2018) Machine learning in gas turbines. *Mech Eng*. 140(09): 54–55. Paper No: ME-2018-SEP5 <https://doi.org/10.1115/1.2018-SEP5>
17. Allegorico C, Mantini V (2014) A data-driven approach for on-line gas turbine combustion monitoring using classification models. In: *European conference of the prognostics and health management society*, July 8–10, Nantes, France 2014, p 9
18. Yan W, Yu L (2019) On accurate and reliable anomaly detection for gas turbine combustors: a deep learning approach, p 8. [arXiv:1908.09238](https://arxiv.org/abs/1908.09238)
19. Batayev N, Onbayev A (2018) Prediction of gas turbine parameters based on machine learning regression methods. In: *The 6th International virtual conference on advanced scientific results* June 25–29. *Electrical and Electronic Engineering*, pp 217–221. <https://doi.org/10.18638/scieconf.2018.6.1.495>
20. Kim JH, Kim TS (2019) A new approach to generate turbine map data in the sub-idle operation regime of gas turbines. *Energy* 173:772–784
21. Perez-Ruiz JL, Loboda I, Miro-Zarate LA, Toledo-Velazquez M, Polupan G (2017) Evaluation of gas turbine diagnostic techniques under variable fault conditions. *Adv Mech Eng* 9(10):1–16. <https://doi.org/10.1177/1687814017727471>
22. Liu Z, Karimi IA (2018) Simulation and optimization of a combined cycle gas turbine power plant for part-load operation. *Chem Eng Res Des* 131:29–40
23. Liu Z, Karimi I (2020) Gas turbine performance prediction via machine learning. *Energy* 192(February):116627. <https://doi.org/10.1016/j.energy.2019.116627>
24. Wood DA (2018) A transparent open-box learning network provides insight to complex systems and a performance benchmark for more-opaque machine learning algorithms. *Adv Geo-Energy Res* 2(2):148–162
25. Wood DA (2018) Transparent open-box learning network provides auditable predictions for coal gross calorific value. *Model Earth Syst Environ*. <https://doi.org/10.1007/s40808-018-0543-9> (published online 16 November, 2018)
26. Wood DA (2019) German solar power generation data mining and prediction with transparent open box learning network integrating weather, environmental and market variables. *Energy Convers Manag* 196:354–369. <https://doi.org/10.1016/j.enconman.2019.05.114> (published online 12 June 2019)
27. UCI Machine Learning Repository (2014) Combined cycle power plant data set (9568 data records) deposited by Pinar Tüfekci. <https://archive.ics.uci.edu/ml/datasets/Combined+Cycle+Power+Plant#>. Accessed 2 August 2019
28. Kaya H, Tüfekci P, Gürgeç SF (2012) Local and global learning methods for predicting power of a combined gas & steam turbine. In: *Proceedings of the international conference on emerging trends in computer and electronics engineering ICETCEE (Mar. 2012, Dubai)*, pp 13–18
29. Wood DA (2018) Thermal maturity and burial history modeling of shale is enhanced by use of Arrhenius time-temperature index and memetic optimizer. *Petroleum* 4:25–42. <https://doi.org/10.1016/j.petlm.2017.10.004>
30. Frontline Solvers (2019) Standard excel solver—limitations of nonlinear optimization. <https://www.solver.com/standard-excel-solver-limitations-nonlinear-optimization>. Accessed: 19 August 2019
31. Yang X-S (2009) Firefly algorithms for multimodal optimization. In: *Stochastic algorithms: foundations and applications, SAGA*, lecture notes in computer sciences, 5792: 169–178
32. Yang X-S, He X (2013) Firefly algorithm: recent advances and applications. *Int J Swarm Intell* 1(1):36–50
33. Pal SK, Raj CS, Singh AP (2012) Comparative study of firefly algorithm and particle swarm optimization for noisy non-linear optimization problems. *IJ Intell Syst Appl* 10:50–57
34. Arora S, Singh S (2013) The firefly optimization algorithm: convergence analysis and parameter selection. *Int J Comput Appl* 69(3):48–52
35. Arora S, Singh S (2014) Performance research on firefly optimization algorithm with mutation. In: *International conference on communication, computing & systems*, pp 168–172
36. Wood DA (2016) Metaheuristic profiling to assess performance of hybrid evolutionary optimization algorithms applied to complex wellbore trajectories. *J Nat Gas Sci Eng* 33:751–768. <https://doi.org/10.1016/j.jngse.2016.05.041>
37. Wood DA (2016) Evolutionary memetic algorithms supported by metaheuristic profiling effectively applied to the optimization of discrete routing problems. *J Nat Gas Sci Eng* 35:997–1014. <https://doi.org/10.1016/j.jngse.2016.09.031>

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