IET Journals

The Institution of Engineering and Technology

Market-based generator cost functions for power system test cases

Venkat Durvasulu[1] ✉, Timothy M. Hansen[1]

[1]*Department of Electrical Engineering and Computer Science, South Dakota State University, Brookings, USA*
✉ *E-mail: dvenkat@live.com*

Abstract: Post-restructuring, generators are dispatched using cleared offers in the day-ahead market in the cyber-physical power system. This study proposes a novel method to design generator cost functions to emulate existing market costs for power system test cases. Cost functions on existing test cases are based on fuel costs, which do not represent organised markets. In such markets, the marginal cost of energy is determined by generator *offers*, not fuel costs. In this work, the authors classify real market generator offers from an independent system operator organised electricity market into generator *types*. Generator offers are used to develop market-based generator cost functions for use in power system test cases to emulate electricity market behaviour. The proposed method is illustrated using PJM data on eight standard power system test cases from a six bus 240 MW generation case to 2000 buses with 95,000 MW of generation. The marginal price of the proposed market-based generator costs shows on average 280% improvement in accuracy of simulating the day-ahead PJM marginal energy price over existing fuel-cost-based test cases from 2014 to 2016. By using the new market-based generator cost functions, power system simulation studies will better represent actual economic impacts.

1 Introduction

Detailed power system test cases with realistic cost functions that represent actual electricity market wholesale prices are highly valuable to industry and academia. The power system in the USA has witnessed significant regulatory changes over the last two decades. The restructured power system has granted open access to the transmission system, which has encouraged competition between generators and distributors of electrical power [1]. To facilitate the trade of electricity between generators and distribution utilities, independent system operators (ISOs) were established as the central market-clearing entity to manage these organised markets [2]. In a restructured power system, each generator and load are independent entities participating in the ISO power market by submitting bids/offers through a cyber network, making the cyber-physical power system (CPPS) one of the largest cyber-physical systems in operation. A large part of the USA, Europe, Australia, and New Zealand are deregulated and have organised electricity markets.

Every generator/generation company participating in the bulk-power market is an independent entity trying to maximise their profit. Therefore, the fundamentals of economic analysis to determine the marginal cost of energy has to be changed from minimising the total sum of generator *fuel costs* to minimising the cost of discrete generator *offers (bids)* [3]. Fuel cost is just one of the deciding factor in determining the $/MWh offer price for a generator participating in an ISO-based electricity market [4]. Since the ISOs determine the marginal energy cost from minimising generator offers, to reflect these prices in simulations, test cases should be provided with market-based generator offers. Test cases that reflect market offers are required for performing analysis on power system studies such as demand response, congestion management, where the price of electricity is a deciding factor [5, 6].

Test cases with generator cost functions based on fuel costs were accurate to perform economic studies of power systems before restructuring. Simulation studies are an essential tool for power system engineers during planning, research, and operation. Simulation results for a new policy or operational changes should reflect real-world costs for deciding on whether or not to adopt the changes. Most power system test cases available for simulation

were developed based on the Mid Western American Electrical Power System in the 1960s and 1970s [7]. The generator cost functions used in these test cases are developed based on then fuel prices and heat curves of mostly thermal units. Studies such as contingency analysis and voltage stability can still be performed without issue on such systems, but accurate economic studies (e.g. demand response) cannot be performed without updates to the fuel-cost-based generator cost functions.

There have been multiple efforts to develop test cases for realistic economic studies. An approximate model of the European interconnection was developed based on the real network to study the impact of cross-border trades used for transmission pricing and congestion management [8]. The developed model was adequate to study a specific problem of cross-border trade, but due to unavailability of the latest data, the authors used thermal and nuclear plant pricing data published in 1995, which is more than two decades out-of-date [9]. Recently, there have been efforts to develop large-scale synthetic test cases that represent the complexity of today's electricity grid for energy economic studies. In [10, 11], two approaches were proposed to augment their developed synthetic test cases with generator cost data that represents the behaviour of an actual power system. Unit commitment (UC) and economic dispatch for 1 day were simulated on the modified test case to illustrate its capabilities to perform economic studies.

Most existing literature assigns fuel-cost-based cost functions to generators on test cases, whereas in the post-deregulation ISO-organised market, the market is cleared based on the generator-offer prices [12, 13]. There are multiple factors apart from the fuel costs that govern the offer strategy of a generator participating in a competitive electricity market [14]. There is sound research in bidding strategy for generator offers of various energy sources under different market scenarios, but there is a lacking methodology for using such bid strategies in test cases, in which the generator types are unknown [14, 15]. In [4], a method is presented to calculate the average bid price for groups of generators participating in the day-ahead market, but the work does not discuss a method to use these prices in power system test cases.

Detailed generator characteristics for a realistic test case based on the IEEE-118 bus system were developed in [16]. A detailed generation model with constraints such as ramping rates,

operational limits, and heat rates for the generators in the power system test case was presented. The generators in the modified IEEE-118 bus system use linear heat rates, and the generation capacity has been increased to 24,500 MW provided by 327 generators. This paper, however, does not describe results regarding marginal energy costs, which are vital to measuring impact in economic studies. Additionally, this work does not present a technique that can be used to keep up with ever-changing generator-offer prices. There exists a need for a general methodology to create *market-based* generator cost functions that represent ISO-organised market marginal energy costs.

In this paper, we present such a methodology to use publicly available market data to design market-based generator cost functions to augment existing power system test cases that accurately represent real marginal energy costs for use in economic studies. Hourly marginal energy costs and demand for most US electricity markets are publicly available. Additionally, daily generator-offer data is available with a 4 month delay [17]. Since the generator-offer data is sanitised of identifying information, we employ statistical pattern recognition (SPR) (*K*-means clustering in this paper) to cluster the market generators into *types*. The three generator types (*K* = 3 clusters) identified from the PJM market (base, intermediate, and peak generators) are used to fit market-based generator cost functions. These new cost functions can be used to augment existing power system test cases to more accurately represent market behaviour of marginal costs.

This methodology is used to augment eight power system test cases of different sizes to represent the generator mix of a real electricity market. Optimal power flow (OPF) was executed on these test cases, and the marginal energy costs produced by the simulation are compared with the real day-ahead market prices. The major contributions of this work are:

i. the design of a methodology to augment existing power system test cases with market-based generator cost functions to represent the marginal energy costs of real power markets for use in economic studies and
ii. a comparison of the proposed approach to the existing state-of-the-art fuel-cost-based power system test cases for multiple year-long OPF simulations with 1 h resolution.

The following section introduces the existing state-of-the-art fuel-cost-based power system test cases and illustrates the need for market-based cost functions for use in analysing economic impact in power system simulations. The methodology for augmenting power system test cases with market-based generator cost functions is described in Section 3. In Section 4, the eight power system test cases are presented and the simulation parameters are given. A comparative study of the marginal energy price of each augmented test case is conducted for 2014–2016 in a 1 h resolution in Section 5. The output of our method is compared with the actual PJM market and the existing fuel-cost-based method. Concluding remarks and directions for future work are presented in Section 6.

2 Economic analysis of electric power systems

2.1 Overview of existing techniques

In this section, we present a brief description of the calculation of the marginal cost of energy, and how key regulatory changes have caused existing fuel-based generator cost functions inadequate for simulation studies with an economical focus (such as the impact of demand response or distributed energy resources). Economic studies in power systems aim to reduce the cost of energy and to improve economic efficiency. In most of these studies, the deciding factor in adopting a method depends on the marginal price of energy, which can be obtained from the result of OPF [18]. In the following section, generator cost functions and the concept of OPF are introduced. An illustrative example comparing the marginal energy costs of PJM versus the existing state-of-the-art fuel-based generator cost functions highlights the need for market-based generator cost functions when trying to calculate energy costs in an ISO-organised market. The section concludes with changes to generator costs due to regulatory changes post-market restructure.

2.2 Optimal power flow

OPF is the core component of economic studies, introduced in the 1960s [19, 20]. In this paper, all simulations on power system test cases were carried out using the OPF formulated in MATPOWER [21]. To illustrate the cost minimisation problem of the power system, OPF can be formulated in its most simplified form without considering some of the constraints (e.g. bus voltage limits) in (1)–(5). For the generator *i*, let C_i be the cost function and P_i be the generation output

$$\min_{P_i} \sum_{i=1}^{N} C_i(P_i) \tag{1}$$

$$C_i(P_i) = \alpha_i P_i^2 + \beta_i P_i + \gamma_i \tag{2}$$

$$\sum_{i=1}^{N} P_i = \sum_{j=1}^{M} D_j + P_L \tag{3}$$

subject to

$$P_i^{\min} \leq P_i \leq P_i^{\max}, \quad \forall i \tag{4}$$

$$0 \leq P_{ij} \leq P_{ij}^{\max}, \quad \forall i, j \tag{5}$$

The objective function (1) is subject to constraints (3) and (5), where D_j is the demand at location *j* among *M* load points on the network. At any given instant, the total generation must equal the total demand plus system transmission losses (P_L), as represented in (3). Equation (4) describes the generator operational limits, and (5) describes the line flow limits for all lines/transformers in the network. Equation (2) represents the polynomial cost of operation of the generator *i* with coefficients α_i ($/MWh²), β_i ($/MWh), and γ_i ($/h). The Lagrangian multipliers of this optimisation problem provide the marginal energy cost of the system [19]. The marginal cost of energy is the cost incurred in the production of one additional unit (MWh). When line limits are reached, generators must be dispatched non-optimally (i.e. different marginal costs), and the marginal price of the entire system cannot be equal. The difference in marginal energy cost is called a congestion cost, and the sum of congestion cost and marginal energy cost is the location marginal price (LMP). Every generator is paid the LMP at the common point of coupling (POC) to the electric grid, and every load needs to pay the LMP at its POC. The last committed generator is called the marginal generator, and its incremental offer decides the cost of bulk electricity at every POC.

2.3 Fuel-cost-based generator cost functions

Generator cost functions in the state-of-the-art synthetic test cases [11] are derived based on input–output characteristics, efficiency, and fuel costs of the major energy contributors such as natural gas, coal, nuclear, and hydro/renewable. The input–output characteristic of a thermal generator is the ability to convert thermal energy into electrical energy; these data may be obtained from design parameters of that generator. The cost function for a thermal generator *i* can be represented as (6), where a_{0i} ($/h) is its no-load cost to operate, and b_{1i} (MBtu/MWh) and b_{2i} (MBtu/MWh²) are the quadratic coefficients of the thermal input–output curve of that generator with fuel cost F_i, expressed as $/MBtu. To obtain the quadratic cost coefficients for a generator, (6) relates to (2), where $F_i b_{2i}$ equals α_i, $F_i b_{1i}$ equals β_i, and a_{0i} equals γ_i. The cost function of hydro/renewable generators comprises only no-load cost because they are non-thermal units and do not have fuel costs associated with generating electricity

$$C_i(P_i) = a_{0i} + F_i(b_{1i}P_i + b_{2i}P_i^2) \tag{6}$$

The South Carolina 500-bus system (SC500) is a synthetic test case statistically developed to represent a real network in a given
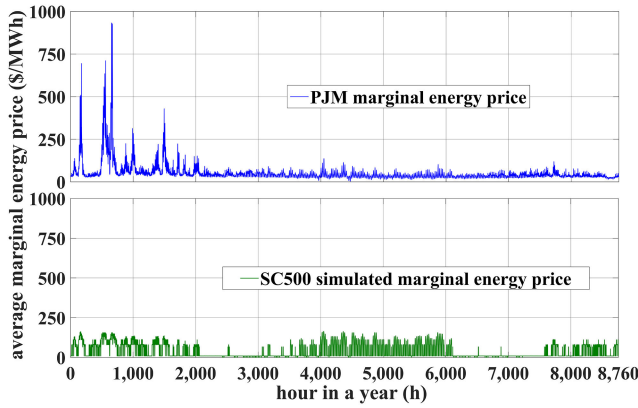
Fig. 1 *Comparison of the marginal cost of energy of PJM in 2014 (top curve in blue) to the average energy cost simulated on SC500 using a scaled PJM demand (bottom curve in green)*

Table 1 Monthly average fuel costs of coal and natural gas for the year 2014

Month	Coal, $/MBtu	Natural gas, $/MBtu
January	2.29	7.02
February	2.32	7.40
March	2.36	6.00
April	2.39	5.07
May	2.40	4.93
June	2.38	4.84
July	2.38	4.43
August	2.37	4.12
September	2.37	4.20
October	2.31	4.10
November	2.30	4.48
December	2.51	4.36

geographical area [11]; this test case is publicly available as ACTIVSg500 [22]. In this test case, one of the approaches to developing generator cost functions is to use the thermal model and fuel cost as described above. We chose this test case because the generation sources were augmented such that they represent the real-world generation mix of the PJM region for the year 2014. The cost functions are developed based on fuel costs of the major energy sources during 2014. The cost functions are developed using the no-load cost and the fuel costs of the thermal generators, and only the no-load costs for the non-thermal generators.

To illustrate the capabilities of the test case, OPF was simulated for 1 year using MATPOWER [21] with a demand curve with a 1 h resolution. The marginal energy cost of the test case was compared with the PJM marginal energy cost for 2014 using an appropriately scaled demand curve from PJM for the same time period. In this section, we will use the notation that lowercase variables describe quantities used for a power system *test case*, and capital variables describe quantities of the real power network. To simulate OPF at an hour t, the load $d_j(t)$ on any bus j is obtained using (7), where d'_j is the default demand on a bus j provided in the test case, $D_{mkt}(t)$ is the total demand on the real PJM network, and D_{mkt}^{max} is the annual peak load of PJM for the year in consideration (i.e. 2014 for this simulation). A scaling factor, ψ, is used to scale the default load such that the ratio of load to generation is similar to that of the real market. For this simulation, the scaling factor was set at $\psi = 1.25$. The scaling factor is described in detail in Section 4. The weighted average marginal price of all nodes in the test case is compared with the marginal price of the PJM market. The weighted average marginal price is defined using (8), where $\Lambda(t)$ is the weighted average marginal price of electricity (in $/MWh) of the test case for an hour t and $\lambda_j(t)$ is the LMP at bus j among the s buses of the power system test case

$$d_j(t) = \psi d'_j \frac{D_{mkt}(t)}{D_{mkt}^{max}}, \quad \forall j = 1, \ldots, s \quad (7)$$

$$\Lambda(t) = \frac{\sum_{j=1}^{s}(\lambda_j(t) \times d_j(t))}{\sum_{j=1}^{s} d_j(t)} \quad (8)$$

The simulation was performed on the SC500 test case using the 2014 PJM market data by scaling the annual demand curve using (7). Nuclear, coal, natural gas fired power, hydro, wind, and solar are the six types of generators defined in the test case. The thermal conversion rates for the nuclear, coal and natural gas fired power used in this simulation were based on [11]. The other three types of generations were non-thermal and had no associated fuel cost. Although the simulation is performed using 1 h resolution, the fuel-cost data obtained from Energy Information Administration [23] was the averaged fuel-cost price per month. Fuel cost of nuclear was almost constant, and 0.85 $/MBtu was used [10]. The fuel cost of the two other major energy resources of the test case, coal, and natural gas, are presented in Table 1. The results of OPF are

compared with the marginal energy cost of PJM in Fig. 1. The year 2014 was a very cold winter for North America, triggering a gas shortage for electricity, and resulted in a large spike in electricity prices. This effect did not appear in the simulated results, as the gas prices only varied by a factor of 1.7 throughout the year, compared with a factor of 100 in energy costs. These results establish that the cost of electricity in a market is not just based on the fuel-cost model of the marginal generator. Existing methods and test cases do not adequately represent real market costs, which may lead to large errors in analysing the economic impact of new power systems technologies as stated in research using existing test cases.

2.4 Generators in an electricity market

In this section, we describe the role of the generator in an organised market, as opposed to the fuel-cost method used prior to market restructuring. Generators that wish to sell electricity need to participate in an electricity market in their region. The cost of electricity is decided by the offer price of the marginal generator. Eligible generating entities need to submit the size of the generator and offer price to the ISO, along with their operational constraints (e.g. runtimes and economic/operational limits). The offer price of each generator is submitted in incremental blocks of their generators capacity ($/MW). Apart from the incremental costs, the generators must also submit their start-up and reserve costs which have to be paid to the generators if they are turned on or keep them in spinning reserve.

The price of bulk power is determined in a day-ahead market under each ISO across the USA The electricity markets perform UC and economic dispatch 1 day ahead of the supply day to schedule generation. The total cost of generation is minimised based on generator offers and demand bids submitted by generating entities and load-serving entities, respectively. Each generator submits a daily offer to the electricity market containing the following information:

i. incremental offer costs (energy cost per segment output range $/MWh versus MW);
ii. upper and lower limits of units economic operation in MW;
iii. start-up and no-load cost;
iv. time operational limits such as minimum runtime and maximum number of starts in a day; and
v. maximum and minimum economic, operational limits.

This information is published publicly on the electricity market websites as per the Federal Energy Regulatory Commission (FERC) ordered a lag period of 4 months [24]. The number of offer blocks can range from one to ten steps in increasing order of energy and cost. Some markets have a restriction on the maximum offer price, for example, PJM restricts their generators to bid a maximum of 1000 $/MW in their day-ahead energy market [25]. Some large generators who cannot shut down in a short note such as nuclear plants submit negative offer prices to ensure they are fully committed [18]. Each market has to maintain some

Input: day-ahead market generator offer data and test case generator data

1: extract the required features from the day-ahead market generator offer data
2: perform SPR with an optimal number of clusters (K)
3: determine the mean size of generator capacity in each cluster
4: arrange clusters in descending order of mean generator capacity
5: determine the percent share of capacity of each cluster ($y_1, ..., y_K$)
6: arrange test case generators in descending order of their capacity
7: initialize the index for test case generator $i = 1$
8: **for** $k = 1$ to K **do**
9:     initialize number of generators in *type k* ($n_k = 0$)
10:     determine the capacity of test case generation classified as *type k* ($\rho_k = y_k p^{max}$)
11:     **while** $\rho_k \geq 0$ **do**
12:         subtract the generator i from the required capacity $\rho_k = \rho_k - p_i$
13:         assign generator i to type k
14:         increase the generator index $i = i + 1$
15:         increase the generator count $n_k = n_k + 1$
16:     **end while**
17: **end for**

Output: test case with statistical classification of generator types

Fig. 2 *Algorithm 1: Algorithm to assign test case generators to a generator type obtained from SPR of market generator-offer data*

percentage of generation as a reserve, and the market needs to pay the generator the spinning reserve cost mentioned in the offer. The identity of the generators is protected so that this data is not used for market malpractices such as gaming the offer price to outbid other competitive generators. An electricity market uses all the information mentioned above in the offers to construct an optimisation problem to minimise the total cost of electricity for the required demand, resulting in hourly marginal energy prices for each location on their network.

3 Market-based synthetic cost curves

3.1 Classification of generator types

Owing to cyber and physical security threats to the power system, the real grid data is not available in the public domain for simulation purposes. For economic studies, generator cost functions on available simulation test cases need to be similar to those in real bulk-power markets. There are a number of test cases that provide realistic line limits and line parameters (e.g. impedance and admittance) that physically represent real power networks, but they lack realistic cost functions. In this paper, we describe a general methodology to turn real generator market offer data into market-based generator cost curves and assign them to power system test cases, expanded from our prior work in [26]. The resulting augmented test cases emulate real market marginal energy prices in simulations.

The number of generators on real power systems is much higher than the number of generators in most test cases. Generators with similar fuel types and comparable sizes submit similar offers [27]. Even though the generation mix of a region is known, it is difficult to recognise the type of generator (e.g. coal, nuclear, and gas) by observing individual offers submitted to an electricity market as this data is not revealed (due to security concerns). An unsupervised pattern recognition technique is required to obtain groups of generators with similar bidding and operational attributes (i.e. generator *types*), but may not be of the same fuel type. Applying an SPR technique on the market offer data creates clusters of generators that are statistically similar, and these generator types can be used to serve as a synthetic representation of the generation mix of the electricity market. The idea of this work is to interpolate the cost functions of each generator type and apply these cost functions to the generators on the test cases in the same ratio as in the synthetic generation mix.

The generator-offer data published by an electricity market contains various information about the generator. The number of

dimensions (i.e. features) of the dataset is crucial for an accurate SPR. Among the various information that are published in the offer data of a generator, as described in the previous section, the following features were selected for SPR:

i. weighted average offer price ($/MWh);
ii. generator size (MW);
iii. minimum runtime per day (h), and
iv. ratio of minimum economic operation limit to the maximum operational limit.

Among the various features tested, these features gave consistent clusters for the selected SPR described in Section 4. A generator's offer can contain anywhere from one to ten incremental blocks, with the number of blocks represented by B_i. For the generator i, the selling price of energy is decided by generator output based on the offer curve defined by offer blocks $\mathscr{W}_i = \{w_{i1}, ..., w_{iB_i}\}$ (MW) versus the corresponding offer price $\mathscr{R}_i = \{r_{i1}, ..., r_{iB_i}\}$ ($/MWh). The weighted average offer of the unit i, $\mathscr{O}_i$ ($/MWh), is defined by (9), where b is the block index. Along with the average price of generation, another important feature for classifying generators is the size of the generating unit in MW (obtained as the quantity of the last offer block), as it separates the generators into clearly defined types (e.g. base versus peak) via the SPR. The minimum runtime is an operational constraint specifying the minimum number of hours a generator has to be committed in a day. The runtime information is used in classifying the generator types based on system loading conditions such as base load, peak load, and intermediate load [28]. The last feature of the dataset describes the ratio of minimum operation limit to maximum operation limit to represent the flexibility of the generator (e.g. a base-load generator will be inflexible with a ratio near 1 and a peak-load generator will be more flexible with a higher ratio)

$$\mathscr{O}_i = \frac{\sum_{b=1}^{B_i} (r_{ib} w_{ib})}{\sum_{b=1}^{B_i} w_{ib}} \qquad (9)$$

3.2 Assigning market generators to test case generators

Using the feature set described in the previous section and a suitable SPR, the *market* generator-offer dataset is classified into K clusters, which serves as statistical generator *types*. To create similar classification in the power system test case, the *test case* generators are also classified into K types, with each type having a similar percentage of generation capacity on the test case as the generator types from the market data. The process of grouping the test case into K types is described in Algorithm 1 (see Fig. 2). First, the mean generation size of each cluster in the market data is calculated. The K market clusters are arranged in descending order of their capacities, and their per cent share of the total generation is evaluated as $y_1, ..., y_K$, where y_1 is the per cent share of the cluster with the largest mean generator size.

A similar procedure is then carried out with the test case generators, by arranging the generators in descending order of their generation capacities. An index to point at a test case generator is initialised with one. The test case generation capacity of generators representing the generation type k is determined as ρ_k, which is $y_k\%$ of the total generation p^{max} as shown in step 3.2. Each generator's capacity (p_i) is subtracted from the generator *type's* capacity until the total capacity is met with n_k generators. This process starts from the largest test case generator, n_1 generators are chosen in descending order of their generation capacity to make a test case generator type that contains y_1 per cent of the total generation on the test case. This process is repeated to determine the K generator types on the test case from the $y_2, ..., y_K$ generation per cent from the market clusters.

For illustrative purposes, let us consider a 6-bus, 11-generator system [29] of 240 MW capacity with market offer data that is clustered into $K = 3$, with $y_1 = 59.5\%$, $y_2 = 19.2\%$, and $y_3 = 21.4\%$

Table 2 *K-means cluster summary for PJM bid offer data on 28 January 2014*

Unit type	Generation share, %	Mean offer quantity, MW	Mean offer price, $/MWh
peak load	21.4	83.1	687.9
intermediate	19.2	122.9	177.6
base load	59.5	311.8	180.8

Fig. 3 *Fitting a second-order polynomial as the generator market-offer-based cost function (green curve) using least-squared error onto an 800 MW generator's offer-rate curve with seven blocks (red curve)*

Input: market-based classified test case, clustered generator market data
1: convert every offer into offer-rate curve
2: fit a second-order polynomial to each offer-rate curve
3: **for** $k = 1$ to K **do**
4:     select n_k cost functions randomly from cluster k
5:     assign a cost function to every generator in the n_k sized test case cluster
6: **end for**
Output: updated test case generator cost functions

Fig. 4 *Algorithm 2: Algorithm to augment existing test cases with second-order polynomial market-offer-based cost functions*

(from Table 2). The four largest test case generators (three 40 MW and one 20 MW) representing 140 MW (i.e. 58.3% of the per cent share of the total test case generation) would be assigned to market generator type 1. The process would continue, assigning three generators with 20 MW capacity representing 25% to type 2, and assigning four generators (one with 20 MW, one with 10 MW, and two with 5 MW) to type 3 to represent the remaining percentages. The following section describes the process of assigning the market generator offers to the generator cost curves for each test case generator.

3.3 Offer-based generator cost curve fitting

3.3.1 Overview: Each generator in the test case is assigned a market-based cost function derived from the market generator-offer data of the generator *type* it is assigned. OPF tools such as MATPOWER [21] require generator costs represented as piecewise linear or polynomial functions. We investigate two techniques in this paper to design generator costs from the clustered market offer data: (a) second-order polynomial and (b) piecewise linear functions, described in the following sections.

3.3.2 Polynomial cost curve: Most existing test cases are provided with quadratic cost functions ($/h), as they were designed based on the quadratic heat rates of thermal units, i.e. (6). In this section, we replace the existing fuel-based cost functions with an equivalent market generator-offer-based cost functions ($/h). The offer-based cost function of the generator i is obtained by (i) multiplying the elements of the offer price, $r_b \in \mathcal{R}_i$, with the corresponding offer block, $w_b \in \mathcal{W}_i$, to obtain the offer rate, and then (ii) fitting a second-order polynomial to determine the coefficients α_i, β_i, and γ_i from (2). The resulting offer-based cost functions are used to augment the test case generator cost functions.

For all the generators with offers having three or more offer blocks (i.e. $B_i \geq 3$), a quadratic equation is fit to the offer-rate curve using the least-squared error method. For example, a generator offer $B_i = 7$ is converted into an offer-rate curve as shown in Fig. 3 in red and a second-order polynomial is fit to the offer-rate curve using least-squared error represented in the green curve. The resulting coefficients, $\alpha_i = 0.031$ $/MWh^2, $\beta_i = 8.75$ $/MWh, and $\gamma_i = 510.93$ $/h would represent the offer-based cost function for the generator i when placed onto a test case. The offer has a minimum limit of 270 MW at 5068 $/h. The fitted quadratic curve is plotted between the generators' operation limits between 270 and 800 MW.

For offers with two blocks, a linear curve is fit between the two operational points resulting in an $\alpha_i = 0$. Generators with offers of a single block have an γ_i equivalent to the offer rate, with $\alpha_i = \beta_i = 0$.

Algorithm 2 (see Fig. 4) is used to augment the generators on the existing test cases with the market-offer-based cost functions. The proposed approach starts by converting the day-ahead market offer of each generator to a second-order polynomial market-offer-based cost function, as described above. The number of test case generators (n_k) for each generator *type* is known, and n_k offer-based cost functions are randomly selected from the corresponding cluster k. Each test case generator that represents a generator type is assigned a market-based cost polynomial that is randomly chosen from the cluster it represents. This approach ensures similar generation mix and costs in the test case as the market it represents.
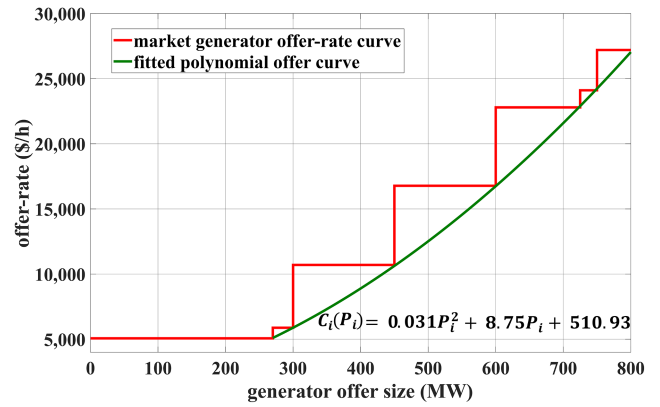
As shown in Fig. 1, the market LMP varies daily, because the generator-offer data changes daily (with hourly differences due to changing demand). This algorithm should be re-run for each day of interest to obtain accurate market simulations through time. Some generator offers do not have a good fit for a second-order polynomial. Such polynomials are eliminated from the dataset before starting the assignment process. The following section discusses an approach to overcome this issue.

3.3.3 Piecewise linear cost curve: Not all generators in the bulk-power market submit offers that can be assigned a quadratic cost curve with a good fit. Such offers that result in a poor fit may not accurately represent the actual generator-offer price. Most power system simulation software is capable of performing OPF using piecewise linear cost functions. For this approach, instead of fitting a second-order polynomial to the offer-rate curve, the market generator offers are directly used by scaling the offer quantity to fit the test case generator size. With this approach, the offer price of a generator remains the same on a test case as in the market, even though the offer has characteristics that make a second-order polynomial a poor fit (i.e. a higher-order polynomial would be required for a good fit).

The process of developing the clusters for the piecewise linear approach is similar to that of polynomial approach. Instead of developing offer-rate curves as in the previous approach, this approach takes the incremental offers $\mathcal{R}_i$ directly from the market data for fitting a piecewise linear cost function for the test cases. To utilise these offers to develop cost functions for test cases using Algorithm 3 (see Fig. 5), the market offers must be extrapolated from 0 MW to the generator's maximum operational limit. The market offer is partitioned into s equi-capacity blocks equal to $1/s$ of the generator's maximum output. These partitioned offer prices are used as offers at s equi-capacity blocks for the test case generator. In this paper, the lower limit of operation is neglected for all generators in the test cases because MATPOWER OPF does not consider UC, hence even in the partition of the market offer the lower operational limit of a generator is ignored. The cost function assignment process is similar to that of the polynomial approach, where the test case generators are grouped into generator *types* and

1: determine the s equi-capacity extrapolated market offer price intercepts for every generator
2: **for** $k = 1$ to K **do**
3:     divide each test case generator capacity into s equi-capacity incremental offers
4:     select n_k piecewise market offer intercepts randomly from cluster k
5:     assign the randomly selected incremental offer price to each test case generator
6: **end for**
Output: updated test case generator cost functions

Fig. 5 *Algorithm 3: Algorithm to augment existing test case with a market-based piecewise linear cost function*
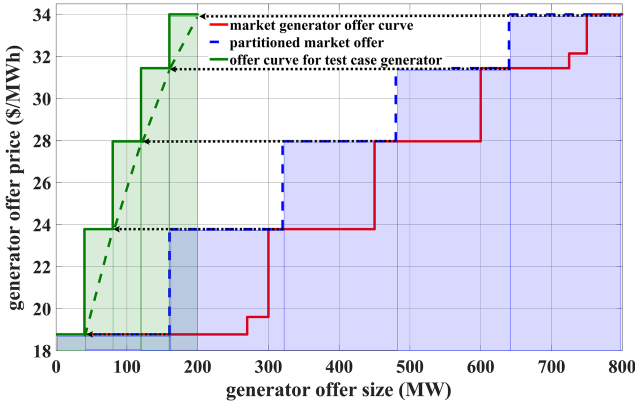

Fig. 6 *Developing generator-offer curve for a 200 MW test case generator (shown in green blocks) by partitioning an 800 MW generator offer submitted to the market (partitioned curve in blue dashed blocks and the offer curve submitted to the market in red). The piecewise linear curve of the developed generator offer is shown in green dashed line*

cost functions are chosen from their respective market generator-offer dataset clusters.

Fig. 6 illustrates an example of converting the same 800 MW generator market offer shown in Fig. 3 to a 200 MW test case generator's offer. The original offer submitted to the electricity market on 28 January 2014 had submitted an operational limit between 270 and 800 MW. If this generator offer is selected to augment a 200 MW test case generator with $s = 5$, the lower limit is higher than $1/s$ of the generator capacity, the lower limit is extrapolated with the first submitted offer price of 18.7 $/MWh. The market offer cost function shown in red in Fig. 6 is partitioned into five equi-capacity blocks shown as the dashed blue curve. The five offer prices of the partitioned curve are used as the five offer blocks for the test case offer curve shown in green. On the basis of the software that is used for the OPF, the offer blocks of the test case can be used as incremental step offers or piecewise linear offers as shown as the green dashed lines linking each incremental block in Fig. 6.

4 Simulation setup

4.1 Simulation overview

The methodology of assigning market-based generator cost functions is tested using the two different techniques to assign offer data to generator cost functions in eight standard test cases of varying sizes (6–2000 buses). The SPR from Section 3.1 was implemented using K-means clustering. Any unsupervised learning technique can be used to implement Algorithm 1 (Fig. 2). The choice of SPR is not explored in this work, as the main contribution of this paper is the design of the methodology to implement realistic market behaviour on power system test cases. Any electricity market that publishes the generator-offer data can be used for developing the cost functions. At the point of writing this paper, the authors are aware of at least two markets that

publish the generator-offer data publicly: (a) PJM [24] and (b) ISO-NE [30]. We chose PJM generator-offer data to develop generator cost functions for the eight test cases. Marginal energy prices obtained from OPF on these test cases are statistically compared with the real PJM market marginal energy price. Negative price offers were eliminated as these small number of offers (∼1%) did not impact the marginal cost of energy. The test cases were assigned scaled PJM hourly demand to each test case for the years 2014–2016. This section will describe in detail the K-means clustering method (including our choice of K), statistics of each cluster formed by K-means on PJM generator-offer data, PJM market demand, and generation statistics, and an introduction to the power system test cases used in this paper.

4.2 K-means clustering

In this work, we chose to implement K-means as an initial baseline implementation of the SPR. K-means is a commonly applied unsupervised learning technique for pattern recognition that is computationally efficient and produces well-separated clusters for well-defined data [31]. The 'K' value for the market generator-offer data is chosen such that the clusters produced have maximum separation and are well-defined. Well-defined clusters are clusters that have a maximum distance between cluster centroids, and least distance between each element within a cluster. We use the Calinski–Harabasz criterion (CHC) to determine the optimal number of clusters K for our offer dataset [32]. We chose CHC for this problem as CHC provides a K that gives maximum separation between cluster groups, so that the generator *types* can be properly identified. With N observations (i.e. generator offers) and K clusters, CHC can be determined by (10). CHC is a maximisation criterion that is directly proportional to the inter-cluster variance, defined in (11), and is inversely proportional to the intra-cluster variance, as defined in (12). For a cluster k, let n_k be the number of observations, m_k be the centroid, and x be a multi-dimensional data point. Additionally, let m be the centroid of the dataset

$$\text{CHC}_K = \frac{SS_B}{SS_W} \times \frac{(N - K)}{(K - 1)} \qquad (10)$$

$$SS_B = \sum_{k=1}^{K} n_k \parallel m_k - m \parallel^2 \qquad (11)$$

$$SS_W = \sum_{k=1}^{K} \sum_{x \in k} \parallel x - m_k \parallel^2 \qquad (12)$$

The CHC was determined for the electricity market data for each day of the year for the years 2014–2016 using the four-feature dataset. Fig. 7 is the histogram for the K value corresponding to the maximum CHC for each day's PJM generators offer data for the year 2014. The optimal number of clusters based on CHC was found to be three for 358 days out of 365 for the year 2014. Similar observations were found for the years 2015 and 2016, where all 365 daily datasets had maximum CHC with 3 clusters. For that reason, we use $K = 3$ for all further simulations.

After clusters of generator types were determined via K-means, the generator types were labelled based on the analysis of each feature of the cluster data using power system domain knowledge. Base-load units, intermediate-load units, and peak-load units are the labels given to the three clusters. A base-load generator is defined as a generator that operates 24 h per day, intermediate units are those that meet the daily peaks (operating a few hours daily), and peak-load generators are operated only during annual peaks (a few hours annually) [28]. Similar minimum runtimes of generator types can be observed in Fig. 8. In this cumulative density plot of PJM generator-offer data on 28 January 2014, the cluster that has more than 90% of the generators with 24 h minimum runtime has been labelled as the base-load units, and the generators with lower minimum runtimes as peak load and intermediate load based on their cluster average weighted offer price.
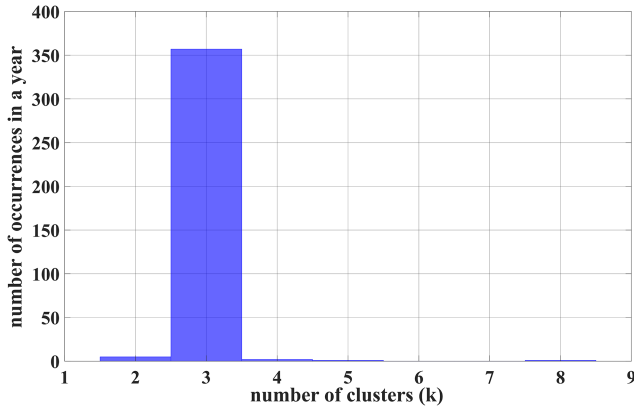
Fig. 7 *Histogram of the optimal number of clusters (K) for the PJM generator's offer data for the year 2014 using CHC*
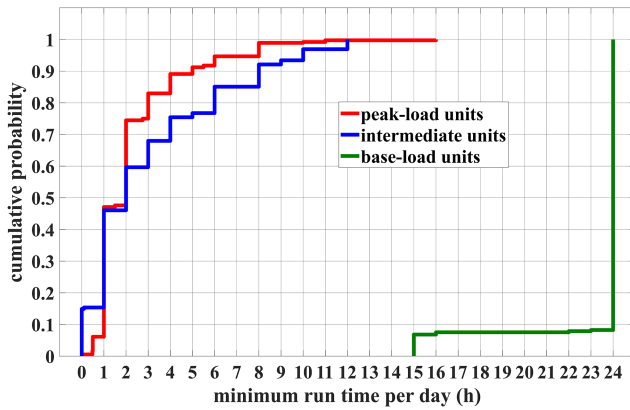

Fig. 8 *Cumulative density function of the minimum runtime of generators in each cluster for the market offer data of PJM on 28 January 2014*

Table 3 *K*-means cluster summary for PJM bid offer data on 6 July 2014

Unit type	Generation share, %	Mean offer quantity, MW	Mean offer price, $/MWh
peak load	18.3	60.0	213.6
intermediate	23.3	156.8	103.3
base-load	58.4	359.5	37.2

Fig. 9 shows the cluster of the generators based on the offer size and offer cost of the generator for the same day in January, as those are the features that are used to augment the test cases. The generators in each cluster appear to be overlapping as the data contains more than two plotted dimensions (i.e. four dimensions from the four features selected for the SPR). Table 2 presents the market offer data statistics based on the clusters shown in Fig. 9. The mean generator size of the base-load units is the largest, and the peak-load units are the smallest. The generation share of these three clusters would serve as the share of generators in test cases for any simulation based on this day (i.e. y_1, y_2, y_3).

28 January 2014 was the peak price day of the year with the marginal energy cost in the PJM region reaching 965 $/MWh at 16:00. The generator distribution on 6 July 2014, which was the peak *demand* day of the year of 139,571 MW, is presented in Fig. 10. Comparing the distributions of the two extreme days shows that the offer price of the generators does not depend only on the demand of the system. On a given day, the price of electricity follows the demand as the offer price of a generator is fixed for that day, but can change throughout the year as fuel prices and other expenses keep fluctuating. Table 3 shows the offer data statistics of 6 July 2014, where the share of each cluster does not change by a large margin, but the mean offer price is approximately six times less than that of 28 January. The lower generator-offer price of 6th July would result in lower electricity
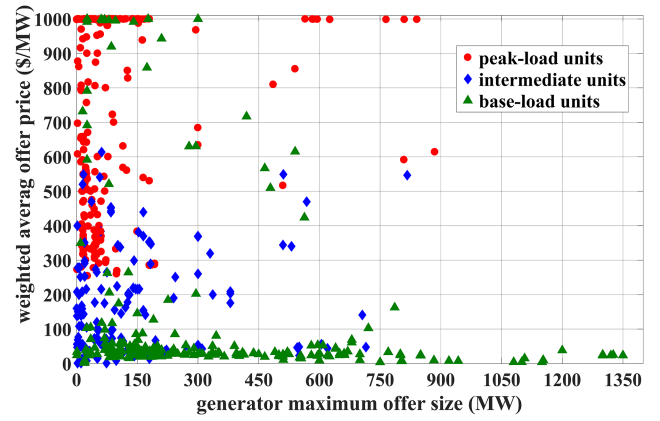

Fig. 9 *Distribution of all the generators by weighted average offer price to their offer size, clustered into three partitions using K-means for the generators participating in the PJM market on 28 January 2014*
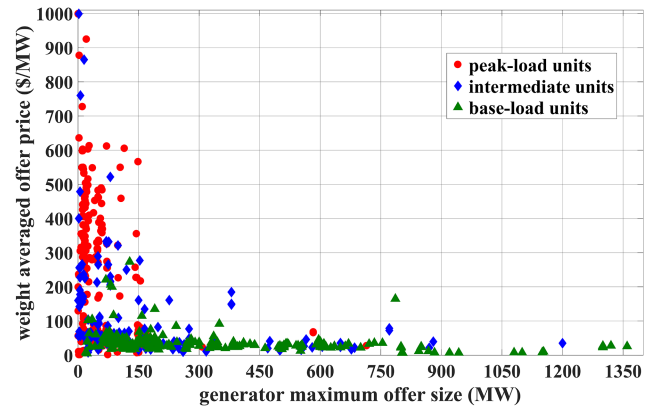

Fig. 10 *Distribution of all the generators by weighted average offer price to their offer size clustered into three partitions using K-means for the generators participating in the PJM market on peak demand day of the year, 6 July 2014*

prices, despite greater demand than that of 28 January, which would not be apparent in existing test case cost functions.

4.3 PJM market

This paper aims to compare the marginal energy prices from the OPF results of the augmented test cases to the marginal energy prices in the actual PJM electricity market annually with a 1 h resolution from 2014 to 2016. For a fair comparison, the test cases and the PJM market are statistically matched regarding demand and generation, described in Table 4. The average demand is calculated as the mean hourly demand in the calendar year. The peak demand is the largest hourly demand during the calendar year, and the generation capacity is the sum of the maximum operating limits of the generator offers on that particular day. The market demand factor (MDF) is obtained by the ratio of peak annual demand of the PJM market (D^{max}) to the generation capacity ($P^{D^{max}}$), given in (13). This ratio gives the maximum percentage of generation utilised during that year. This information is used to scale the PJM demand curve to match the test cases. For this paper, MDF = 80% was chosen for each of the 3 years

$$\text{MDF} = \frac{D^{max}}{P^{D^{max}}} \qquad (13)$$

4.4 Power system test cases for simulation

Eight power systems test cases ranging from 6 to 2000 buses were tested using the two proposed offer-based generator cost curve-fitting techniques. The eight test cases used are Roy Billinton test case (RBTS89) [29], IEEE 14-bus system (IEEE14) [33], IEEE

Reliability test case 1979 (RTS79) [34], IEEE 39-bus New England test case (NE-93) [35], IEEE Reliability test case 1996 (RTS96) [36], ACTIVSg200 Synthetic Illinois 200-bus power system model (IL200), ACTIVSg500 Synthetic South Carolina 500-bus power system model (SC500), and ACTIVSg2000 Synthetic Texas 2000-bus power system model (TX2000) [22]. Table 5 presents the details of the size and capacities of each test case considered for this paper. The test case demand factor (TCDF) is the ratio of the sum of all test case loads to the sum of the maximum generation of the test case generators. To match the PJM market MDF of 80%, the default loads on the test case were scaled using the scaling factor ψ, defined in (14), where p' is the total generation capacity of the test case and d' is the default load on the test case.

Owing to OPF convergence issues during some peak demand and demand valley times when using the scaled PJM load on some of the test cases, the scaling factor ψ was adjusted. In Table 5, the rows marked with a '*' have been adjusted to ensure convergence. The exact methodology for determining ψ for these test cases is described in the following sections. The minimum generation limit of all generators in the RTS79/96, IL200, SC500, and TX2000 test case was set to 0 MW, as described in Section 3.3.3 (all other test cases have default minimum limits of 0 MW). To maintain load model accuracy, the same scaling factor that was used for active power as described in (14) has been used to scale the reactive power demand at each bus

$$\psi = \frac{p'}{d'}\text{MDF} = \frac{\text{MDF}}{\text{TCDF}} \tag{14}$$

4.4.1 RBTS89: The RBTS89 test case is a small system with relatively high line flow limits that cannot be overloaded; this test case converges for all loading conditions. The largest generator of the RBTS89 system is 40 MW, which is smaller than the average peak-load generator of the PJM market. The cost function for a peak-load generator of the test case is developed based on PJM generator offers that are on an average ten times larger. The relative difference in size of generators creates cost functions that have steep variation due to the reduced size of the generator when compared with the market. There are 11 generators in this test case which represent each of the three generator types. Owing to a small number of generators, a small variation in demand can sometimes create a large jump in price, as the chance of the marginal generator changing from one type to another is high.

4.4.2 IEEE test cases: The IEEE14 test case has no line limits, and the only operational constraint is the generator maximum limits and voltage limits of the buses. This test case has only five generators varying between 100 and 334 MW. This small mix of generators and results in large step changes in hourly price over annual simulations because the smallest generator of the test case is

twice as the mean generator size of the peak-load generator of the PJM market. With just five generators, it is difficult to accurately group generators to resemble similar generation share as the market clusters. Even though the test case can converge for a scaling factor that would result in 80% loading, the resulting marginal prices are very high as the small changes in load (inter-day demand changes) does not result in marginal generator choice.

This is similar to the NE39 test case, where there are ten generators varying between 508 and 1100 MW. The smallest generator on the test case is at least 15 times larger than the average peak-load generator of the PJM market. Just like the 14-bus system, an accurate percentage share of clusters cannot be formed. These two test cases represent a part of the power system which does not need to have the entire diversity of generation profile. The scaling factor on this test case has been lowered for the same reason as the IEEE14, where a higher scaling factor would result in a much higher median marginal energy price.

4.4.3 IEEE reliability test case: The RTS79 test case has realistic line limits, voltage limits, and generation limits. This test case has 32 generators varying between 12 and 400 MW. Even though this test case was proposed for performing reliability studies, it is still a good test case to conduct economic studies because of the realistic generation mix and line limits. The RTS96 test case is comprised of three RTS79 systems, interconnected by five lines, making a total of 96 generators and 73 buses. The generation and demand mix on each of the three subsystems of the RTS96 test case remains unchanged from RTS79.

4.4.4 Synthetic test cases: Three synthetic test cases (IL200, SC500, and TX2000) were developed using statistical techniques [10]. They are designed to represent the electrical grid of a geographical region by capacity. The IL200 test case represents a hypothetical grid in southern IL, USA, with 200 buses and 49 generators that range from 4 to 569 MW. This test case has a default load of 44% of its maximum generation, which was scaled to 70% using $\psi = 1.6$. This value is <10% would be calculated from (14), as beyond that percentage, the upper voltage limits on a few buses would be violated. On some buses, there is a little-to-no load which causes a voltage rise as the generation increases to meet the load on other buses. The SC500 test case represents a region of SC, USA, with 500 buses and 90 generators ranging between 1 and 772 MW. This test case was scaled by a factor of $\psi = 1.25$, which is equal to the value calculated from (14) because this case converges for all loading conditions. The TX2000 test case is a synthetic network that covers the entire Texas region with 2000 buses and 544 generators ranging between 1 and 1354 MW, which is very close to the original PJM market. Even though the upper limit of scaling is achievable, this was the only system that did not converge for any load that was scaled 84% below its default load. The simulation for this test case was set up such that the loads on this test case are scaled so that the minimum demand on the system is 84% of the default load of the test case, and the peak load is 110% of the default load. The scaling for the TX2000 is achieved using (15), to ensure convergence. With this scaling, the ratio of the valley load to the peak load over the annual simulation would be 77% (i.e. 0.84/1.1). The ratio of minimum demand to the maximum demand on the PJM network for the year 2014 was 39%

Table 4 PJM annual demand and generation summary

Year	Average demand, MW	Peak demand, MW	Generation capacity, MW	MDF, %
2014	89,317	139,571	176,126	79
2015	87,712	136,510	181,674	75
2016	87,793	142,928	178,064	80

Table 5 Power system test cases with their default setup and their scaling factors considered for this simulation

Test cases	Number of buses	Installed capacity, MW	Default load, MW	TCDF, %	ψ
RBTS89	6	240	185	77	1.04
IEEE14	14	772	259	34	2.20*
RTS79	24	3405	2850	84	0.96
NE39	39	7367	6254	84	0.88*
RTS96	73	10,215	8868	84	0.96
IL200	200	3379	1475	44	1.60*
SC500	500	12,189	7750	64	1.25
TX2000	2000	94,791	67,109	71	1.10*

IET Cyber-Phys. Syst., Theory Appl., 2018, Vol. 3 Iss. 4, pp. 194-205

201

$$d_j(t) = \psi \times d_j \times \frac{(1.1 - 0.84) \times (D_{\mathrm{mkt}}(t) - D_{\mathrm{mkt}}^{\min})}{D_{\mathrm{mkt}}^{\max} - D_{\mathrm{mkt}}^{\min}} + 0.84 \quad (15)$$

5 Simulation and results

5.1 Simulation overview

The capability of the proposed approaches is illustrated by comparing the energy prices from the OPF with the real PJM market marginal energy price. Simulations were performed on all eight test cases with the scaled PJM demand using (7) with the scaling factor ψ as mentioned in Table 5. All simulations were carried out using MATPOWER 6.0 in MATLAB (R2017a). All data from this work has been made publicly available in a GitHub repository with an open source licence at https://goo.gl/rSGeBX. While the simulation and analysis were conducted for 2014–2016, only the results from 2014 are presented and analysed in detail for brevity. The energy prices of 2014 are interesting as that year witnessed high variation in energy prices across the year due to its very cold winter temperatures in North America and unavailability of natural gas for electricity.

5.2 OPF simulation

The market-offer-based polynomial and piecewise linear approach simulations were performed using the generator cost functions derived using the proposed methods. The cost functions for the test cases were updated in 1 day resolution. The same load curve was used for simulating all three setups (default, polynomial, and piecewise linear) derived using (7) for each test case, except the TX2000 which uses (15). All the marginal energy prices presented in the results are the weighted average energy price using (8).

To compare the hourly energy prices over the 1 year time-series simulations with the PJM marginal energy prices, both visual and statistical techniques were investigated. One such statistical method is to compare the distribution of the hourly marginal energy price from the augmented test case OPF with that of PJM over the same year. In this case, the distribution is represented as violin plots, as shown in Fig. 11. Violin plots are similar to boxplots, but the probability density of the values can be observed, represented by the width of the violin plot. The dashed black lines in each of the violin plots represent the three quartiles of the distribution with the middle line representing the median energy price. The two ends of the violin plot represent the extremes of the observed data. In Fig. 11, the first violin plot is the energy price distribution of the PJM market during 2014 represented in the red plot. The distribution of marginal energy price of each test case using the polynomial approach is represented in the left half of each violin plot in blue, and on the right in green using the piecewise linear approach. The distribution of marginal energy price of each test case using the two proposed techniques is compared with the PJM market marginal energy distribution. Although the cost functions for all the test cases were drawn from the same pool of cluster data, the performance of each test case is different, as the physical structure and generator configuration are different.

From Fig. 11, we can observe that the test cases having realistic line limits, no convergence issues, realistic generation profile in terms of numbers and sizes (i.e. both the RTS system and SC500) produced energy price distribution similar to that of the PJM market. Although the TX2000 test case has all the characteristics of a real power system, due to its lower-bound convergence issues the energy prices have a higher median. However, the TX2000 price distribution shape closely resembles the PJM market distribution. The IL200 has the least matching distribution due to the upper limit convergence issues.

The marginal energy prices produced by OPF using the two approaches have comparable distributions. Regarding the density and mean of the marginal electricity price, both the RTS case and the SC500 test case are similar to the PJM market. One reason these test cases perform well is the availability of variously sized generators similar to that of the real market. The distribution of energy prices from the simulations of RBTS89, IEEE14, and NE39
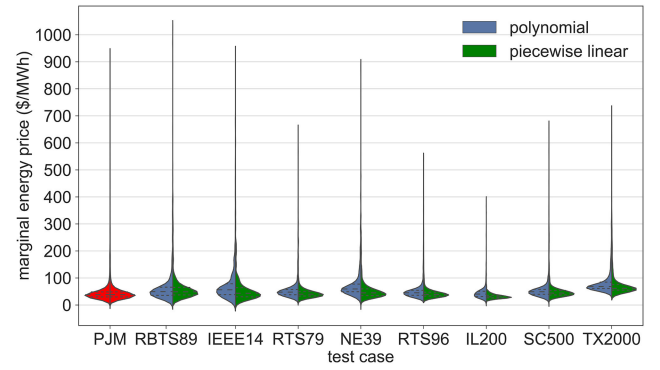

Fig. 11 *Violin plots representing the distribution of marginal energy price of the PJM market for the year 2014 in red and by simulation on test cases using polynomial approach presented in blue on the left half, and using the piecewise linear approach presented in green on the right-hand side of the violin plot*

are similar and have a narrow violin plot when compared with the PJM market. A narrow violin plot represents less density of observations around the mean, indicating more variation in marginal energy costs. The IL200 test case produced the lowest mean, as the upper scaling of this test case was reduced due to the voltage constraint convergence issues.

Another statistical metric to compare the performance of the proposed technique is the goodness of fit (GOF), given in (16), where Λ_{PJM} is the marginal energy cost of the PJM market and Λ is the marginal cost resulting from the augmented test case. GOF provides a metric between $-\infty$ and 1, where 1 is an exact fit and $-\infty$ indicates a very poor fit

$$\mathrm{GOF} = 1 - \frac{\parallel \Lambda_{\mathrm{PJM}} - \Lambda \parallel}{\parallel \Lambda_{\mathrm{PJM}} - \Lambda_{\mathrm{PJM}}^{\mathrm{mean}} \parallel} \quad (16)$$

Table 6 presents the GOF, and the distribution of marginal prices produced by each test case using the two proposed approaches compared with the original PJM market for 2014. The table compares the OPF simulation results using the mean and standard deviation of marginal energy cost for 1 year with a 1 h resolution. The first row shows the results from 1 year OPF using the single set of default cost functions that are provided with the test cases, except for results of SC500 and TX2000 which are simulated using the heat rates of the generators and monthly fuel costs (as described in Section 2.3). The distribution of the SC500 marginal price for default cost curve is derived from the plot presented in Fig. 1.

The mean and standard deviation of the SC500 and the two RTS systems are closest to the real PJM distribution. Even though the TX2000 system has physical dimensions comparable with the real system, due to the simulation limitations in scaling the load, a much higher mean is observed. The simulation results using the scaled 2015 and 2016 showed similar trends, where the RTS and SC500 have the best marginal energy price distribution and GOF among the other test cases. The SC500 had obtained a GOF of 0.54 using the polynomial approach for the year 2015 and the least performed test case was IL200. The year 2016 had low fluctuation in marginal price, and a low peak, because of this the TX2000 could only achieve a 0.04 GOF for the year 2016 as the median price from the simulation was higher than the first quartile price of the PJM market.

None of the results from simulations using default cost functions could generate a positive GOF for the year 2014, but managed to get a positive GOF of 0.11 for SC500 and 0.04 for RTS96 using 2016 demand. The RTS79 using the piecewise linear cost functions produced the closest marginal price distribution in terms of mean and standard deviation to the PJM marginal price for 2014. The GOF of RTS79 is also the highest among all the test cases with 0.45, closely followed by polynomial cost function-based SC500 with 0.44. The highest GOF of 0.49 was obtained among all the simulations for RTS96 during 2015 with market-based polynomial cost functions. The IL200 test case did not

202

Table 6 Statistical comparison of marginal energy price by simulation with the PJM market

| Cost function	Statistical parameter	PJM	RBTS89	IEEE14	RTS79	NE39	RTS96	IL200	SC500	TX2000
default cost curve	mean energy price, $/MWh	49.16	0.70	41.43	17.31	7.94	24.46	18.51	37.8	158.42
	standard deviation	51.87	0.07	0.99	7.74	1.36	13.76	2.62	43.29	6.0
	GOF	N/A	−1084.36	−51.60	−6.45	−47.21	−2.87	−21.87	−0.13	−18.99
market-based polynomial	mean energy price, $/MWh	N/A	65.53	73.77	55.65	79.77	64.78	46.30	57.66	87.21
	standard deviation	N/A	71.35	71.58	43.74	72.16	47.93	32.66	46.32	57.16
	GOF	N/A	0.36	0.17	0.36	0.23	0.36	0.15	0.44	0.16
market-based piecewise linear	mean energy price, $/MWh	N/A	66.59	69.57	49.76	58.96	48.29	45.85	53.49	81.54
	standard deviation	N/A	71.17	88.59	43.77	49.68	40.29	35.04	38.09	61.41
	GOF	N/A	0.39	0.21	0.45	0.29	0.41	0.14	0.15	0.24
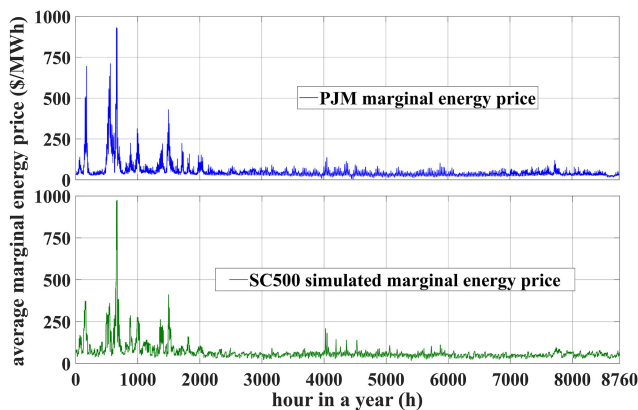
Fig. 12 *Annual marginal energy price of PJM for 2014 with 1 h resolution (top blue curve) and the simulated marginal energy price using the market-offer-based polynomial cost functions on the SC500 test case with scaled PJM demand (bottom green curve)*

perform as well as the other test cases as the scaling of this particular system was 10% lower than that of the system that is being compared. Similarly, the case with TX2000 which suffers from lower scaling produced higher rate when compared with other test cases. It is to be noted that these values of GOF are far from the best fit value because the comparison is between OPF results from a test case that is physically different from the real market in terms of size and complexity. Second, the results presented in this paper are marginal energy prices of test cases from OPF, which only resembles the economic dispatch problem of an electricity market, but not the UC. Considering these two approximations, the positive GOF obtained are significant improvements when compared with the existing default test case results.

Fig. 12 presents the weighted average marginal energy cost from OPF simulation on SC500 (green) using the market-based polynomial generator cost function for the scaled PJM demand of 2014. The marginal energy cost has been compared with the marginal energy of PJM during the year 2014. The hours during winter (first 2000 h) show a higher marginal energy cost similar to that of the market. The hours during the summer (4000–6000), where the system demand is higher compared with other seasons of the year, the price was relatively low. When the marginal energy cost time-series of the proposed technique is compared with the fuel-cost-based approach in Fig. 1, the time-series in the proposed technique could represent most of the valleys and peaks similar to that of the real PJM market.

5.3 Monte Carlo simulation

There is randomness in the selection process of cost functions from the PJM market cluster data to generators on the test cases, as the number of generators in the PJM system is much greater than the number of generators on any test case (i.e. 853 in PJM versus 544 in TX2000, the largest test case available in this paper). The proposed algorithm is a viable solution to create test case cost functions only if its solution is stable in the random selection of generators. Monte Carlo simulation was performed to show the consistency and stability of the proposed technique. OPF was
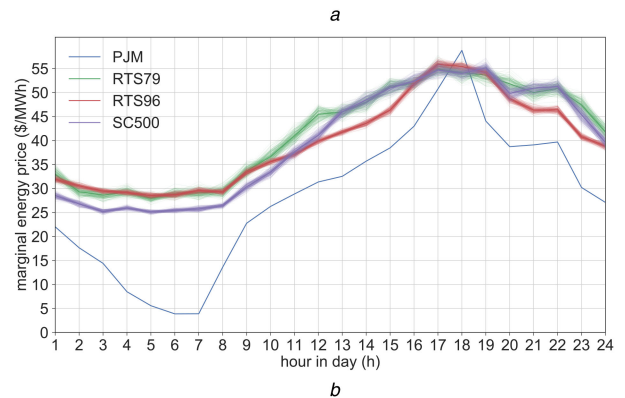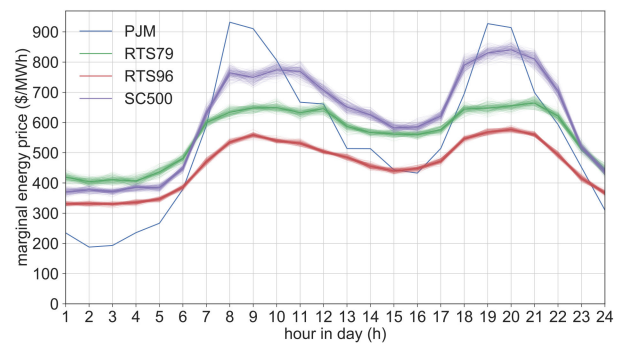
Fig. 13 *Monte Carlo simulation of OPF with a 1 h resolution of RTS79, RTS96, and SC500 for 100 trials*
(a) Simulation for the peak price day of the year in PJM 28 January 2014, *(b)* Least price day of the year 18 June 2014

performed for 1 day with a 1 h resolution for 100 Monte Carlo trials. Each trial, the selection of generators to the test case from the PJM market was resampled. Although the Monte Carlo simulation was performed on all test cases for both curve-fitting techniques, only results of RTS79, RTS96, and SC500 using market-based polynomial curves were presented for brevity, as they had the highest GOF. The other test cases showed similar stability.

Fig. 13 compares the stability of the three test cases versus the PJM market using bootstrap plots of the weighted average marginal cost. Each thin coloured curve represents one Monte Carlo trial of the test case, with the dark curve representing the mean. Two days were chosen for the Monte Carlo simulation, as they are the two extreme priced days (the peak and the valley) of the year.

Fig. 13*a* presents the Monte Carlo simulation of the peak priced day of the year (28 January 2014). The SC500 test case performed the best among the three for this day and could reproduce the peak hour of the day at 8:00. Both the peak and valley of that day in the simulation performed on the test cases appear at the same hour they occur on the PJM system for most of the trials. The valley on the simulations performed on RTS96 has a 1 h offset at 15:00, instead of at 16:00 for the actual market valley. Fig. 13*b* represents the Monte Carlo simulation results for the lowest price day of the year 18 June 2014. All three systems performed equally, but none could reproduce the valley effect of PJM, even on a single trial. The inability of the proposed technique to match the valley can be

attributed to the approximation of the test case simulations explained previously in Section 3.3.3. All test cases could reproduce the peak of the day and the sudden change in slope between 20:00 and 22:00. Simulations are expected to reflect such variations, as services such as demand response and battery charging strategies are most likely to respond to similar pricing signals.

6 Conclusions

This paper presented a general methodology for augmenting power system test cases with generator cost functions that represent real electricity market energy prices. This technique is compared with the current state-of-the-art fuel-cost and heat-rate-based generator cost functions. Eight test cases were tested using the two proposed techniques by comparing their OPF results with the real PJM market day-ahead marginal energy prices. The OPF was executed using a scaled annual demand curve of PJM with 1 h resolution and compared with the marginal energy cost of the PJM market using a GOF metric. Simulation with our proposed technique on all the test cases resulted in better energy price estimates than the default cost functions of the system. A Monte Carlo simulation was presented to show the stability of the proposed techniques, even with the randomness in the algorithms. The proposed technique is intended to be used to augment test cases based on the ISO-based market which publishes their bid/offer data. This technique should not be used for economic analysis of other market models and on non-deregulated power systems.

Even though all test cases produced energy price closer to the realistic costs than their results using the default cost functions, not all test cases performed the same when compared with the real market price. The SC500 and both the RTS test cases had their marginal energy price close to PJM marginal energy price. These test cases performed better when compared with the other test cases as the SC500 and both the RTS test cases as they had no scaling issues and had a wide range of generator sizes (i.e. range of generators from a large units in 100s of MW to small units in range of 10 MW) as they closely represent a real power system generation profile of a large area. Such a test case achieved a GOF as high as 0.5, and the test cases that had generation profile that only represent a part of a power system produced lower GOF. The proposed technique performed better than the existing cost functions and also the state-of-the-art synthetic test cases that use generator heat models and fuel costs to develop the cost functions. These test cases would allow the research community to perform more accurate economical analysis of CPPS with more realistic energy prices.

To produce more accurate energy prices, the proposed augmented test cases can be used along with UC. The formulation of the UC problem does not change, but the OPF following the UC using the proposed market-based generator cost functions will result in realistic energy prices when compared with the fuel-cost-based generator cost functions. There are other SPR techniques that may result in different clusters and different distributions. In this paper, we only used one such unsupervised technique, as the main contribution of this work is to present the method of using real market data to augment test cases for economic studies. Exploring various pattern recognition technique for this problem is not in the scope of this paper, but could be explored in future studies. As the proposed technique has resulted in realistic energy prices in OPF simulation than the default cost functions on all eight test cases, it is fair to conclude this technique can be applied to any other power system test case for economic studies. Economic studies performed on such augmented test cases would result in realistic energy costs, and conclusions drawn from such studies would better represent the behaviour of new technologies on the real-world power system.

7 Acknowledgments

8 References

[1] Federal Energy Regulatory Commission: 'FERC order 888'. Available at https://www.ferc.gov/legal/maj-ord-reg/land-docs/order888.asp, accessed 10 December 2017

[2] Federal Energy Regulatory Commission: 'FERC order 2000'. Available at https://www.ferc.gov/legal/maj-ord-reg/land-docs/RM99-2A.pdf, accessed 10 December 2017

[3] Wang, H., Murillo-Sánchez, C.E., Zimmerman, R.D., et al.: 'On computational issues of market-based optimal power flow', IEEE Trans. Power Syst., 2007, 22, (3), pp. 1185–1193

[4] Zhang, N.: 'Generators' bidding behavior in the NYISO day-ahead wholesale electricity market', Energy Econ., 2009, 31, (6), pp. 897–913

[5] Su, C.L., Kirschen, D.: 'Quantifying the effect of demand response on electricity markets', IEEE Trans. Power Syst., 2009, 24, (3), pp. 1199–1207

[6] Fang, R.S., David, A.K.: 'Transmission congestion management in an electricity market', IEEE Trans. Power Syst., 1999, 14, (3), pp. 877–883

[7] 'Power systems test case archive'. Available at https://www2.ee.washington.edu/research/pstca/, accessed 10 December 2017

[8] Zhou, Q., Bialek, J.W.: 'Approximate model of European interconnected system as a benchmark system to study effects of cross-border trades', IEEE Trans. Power Syst., 2005, 20, (2), pp. 782–788

[9] SEI: 'The European fossil-fueled power station database used in the SEI CASM model' (Stockholm Environment Institute, York, England, 1996)

[10] Xu, T., Birchfield, A.B., Gegner, K.M., et al.: 'Application of large-scale synthetic power system models for energy economic studies'. Proc. 50th Hawaii Int. Conf. System Sciences, Waikoloa Village, HI, USA, 2017, pp. 3123–3129

[11] Birchfield, A.B., Xu, T., Gegner, K.M., et al.: 'Grid structural characteristics as validation criteria for synthetic networks', IEEE Trans. Power Syst., 2017, 32, (4), pp. 3258–3265

[12] Ren, Y., Galiana, F.D.: 'Pay-as-bid versus marginal pricing – part I: strategic generator offers', IEEE Trans. Power Syst., 2004, 19, (4), pp. 1771–1776

[13] Ren, Y., Galiana, F.D.: 'Pay-as-bid versus marginal pricing – part II: market behavior under strategic generator offers', IEEE Trans. Power Syst., 2004, 19, (4), pp. 1777–1783

[14] Wen, F., David, A.K.: 'Optimal bidding strategies and modeling of imperfect information among competitive generators', IEEE Trans. Power Syst., 2001, 16, (1), pp. 15–21

[15] Fan, L., Wang, R.J.J., Guan, Y.: 'Min-max regret bidding strategy for thermal generator considering price uncertainty', IEEE Trans. Power Syst., 2014, 29, (5), pp. 2169–2179

[16] Peña, I., Martinez-Anido, C.B., Hodge, B.M.: 'An extended IEEE 118-bus test system with high renewable penetration', IEEE Trans. Power Syst., 2018, 33, (1), pp. 281–289

[17] PJM Interconnection: 'Daily day-ahead LMP'. Available at http://www.pjm.com/markets-and-operations/energy/day-ahead/lmpda.aspx, accessed 10th December 2017

[18] Zhao, Z., Wu, L.: 'Impacts of high penetration wind generation and demand response on LMPs in day-ahead market', IEEE Trans. Smart Grid, 2014, 5, (1), pp. 220–229

[19] Dommel, H.W., Tinney, W.F.: 'Optimal power flow solutions', IEEE Trans. Power Appar. Syst., 1968, PAS-87, (10), pp. 1866–1876

[20] Lee, K.Y., Park, Y.M., Ortiz, J.L.: 'A united approach to optimal real and reactive power dispatch', IEEE Trans. Power Appar. Syst., 1985, PAS-104, (5), pp. 1147–1153

[21] Zimmerman, R.D., Murillo-Sánchez, C.E., Thomas, R.J.: 'MATPOWER: steady state operations, planning, and analysis tools for power systems research and education', IEEE Trans. Power Syst., 2011, 26, (1), pp. 12–19

[22] Texas A and M University: 'Electric grid test case repository'. Available at https://electricgrids.engr.tamu.edu/electric-grid-test-cases/activsg500/, accessed 20 January 2018

[23] U.S. Energy Information Administration (EIA): 'U.S. Natural Gas Electric Power Price'. Available at https://www.eia.gov/dnav/ng/hist/n3045us3m.htm, accessed 20 January 2018

[24] PJM Interconnection: 'Daily energy market offer data'. Available at http://www.pjm.com/markets-and-operations/energy/real-time/historical-bid-data/unit-bid.aspx, accessed 10th December 2017

[25] PJM, Cost Development Subcommittee: 'PJM manual 15: cost development guidelines'. Available at http://www.pjm.com/~/media/documents/manuals/m15.ashx, accessed 20 January 2018

[26] Durvasulu, V., Hansen, T.M., Tonkoski, R.: 'Classification of generators participating in the bulk-power market'. 2017 IEEE Int. Conf. Industrial Technology ICIT, Toronto, ON, Canada, 2017, pp. 1–5

[27] Bezerra, B., Barroso, L.A., Pereira, M.V.: 'Bidding strategies with fuel supply uncertainty in auctions of long-term energy call options', IEEE Trans. Power Syst., 2011, 26, (2), pp. 653–660

[28] Nag, P.K.: 'Power plant engineering' (Tata McGraw-Hill Publication Company Limited, New Delhi, 2008, 3rd edn.), p. 7

[29] Billinton, R., Kumar, S., Chowdhury, N., et al.: 'A reliability test system for educational purposes-basic data', IEEE Trans. Power Syst., 1989, 4, (3), pp. 1238–1244

[30] ISO-New England: 'ISO-NE pricing reports: day-ahead energy offer data'. Available at https://www.iso-ne.com/isoexpress/web/reports/pricing/-/tree/day-ahead-energy-offer-data, accessed 20 January 2018

[31] Jain, A.K., Duin, R.P.W., Mao, J.: 'Statistical pattern recognition: a review', IEEE Trans. Pattern Anal. Mach. Intell., 2000, 22, (1), pp. 4–37

204

IET Cyber-Phys. Syst., Theory Appl., 2018, Vol. 3 Iss. 4, pp. 194-205

[32] Caliński, T., Harabasz, J.: 'A dendrite method for cluster analysis', *Commun. Stat.*, 1974, **3**, (1), pp. 1–27

[33] Reid, G.F., Hasdorff, L.: 'Economic dispatch using quadratic programming', *IEEE Trans. Power Appar. Syst.*, 1973, **PAS-92**, (6), pp. 2015–2023

[34] IEEE RTS Task Force of APM Subcommittee: 'IEEE reliability test system', *IEEE Trans. Power Appar. Syst.*, 1979, **98**, (6), pp. 2047–2054

[35] Athay, T., Podmore, R., Virmani, S.: 'A practical method for the direct analysis of transient stability', *IEEE Trans. Power Appar. Syst.*, 1979, **PAS-98**, (2), pp. 573–584

[36] Grigg, C., Wong, P., Albrecht, P*., et al.*: 'The IEEE reliability test system-1996. A report prepared by the reliability test system task force of the application of probability methods subcommittee', *IEEE Trans. Power Syst.*, 1999, **14**, (3), pp. 1010–1020

IET Cyber-Phys. Syst., Theory Appl., 2018, Vol. 3 Iss. 4, pp. 194-205

205