A data-driven approach to estimate emissions for market-based power system test cases

Venkat Durvasulu | Timothy M. Hansen

Department of Electrical Engineering and Computer Science, South Dakota State University, Brookings, South Dakota, USA

Correspondence
Venkat Durvasulu, Department of Electrical Engineering and Computer Science, South Dakota State University, Brookings, South Dakota, USA. Email: venkat.durvasulu@jacks.sdstate.edu

Funding information
National Science Foundation; ECCS-1608722 The South Dakota State University Electrical Engineering; The South Dakota Board of Regents (SDBoR)

Abstract
A data-driven technique to determine greenhouse gas (GHG) and air-pollutant (AP) emissions from bulk-power system simulations is proposed. The proposed technique emulates the dispatch of a bulk-power system using open-source hourly fuel-energy data from an actual U.S. electricity market (i.e. PJM). Sixteen different fuel types were analysed from real generator data and dynamically assigned to power system test cases to statistically represent the real fuel-energy mix. Each test case generator is assigned a heat-rate function based on open-source real generator data to estimate realistic emissions from power system test case simulations. These augmented test cases can be used to simulate how changes in load and generation impact power system emissions to determine the environmental sustainability of new technologies (e.g. demand-side management). The proposed technique is implemented on three different power system test cases, and the simulated emissions are compared with the actual emissions of the PJM system. The results from the test systems are found to accurately emulate the time-series values of fuel-mix, emissions, and marginal costs of PJM.

1 | INTRODUCTION

In 2016, the electric power industry generated the largest share (28%) of greenhouse gas (GHG) emissions across all primary energy-consuming sectors in the United States. [1]. In the same year, approximately 68% of electric energy was generated from fossil fuel combustion—the majority of which is coal and natural gas [2]. In addition to GHG, fossil-fuelled power plants produce other harmful air-pollutants (AP), such as SO₂ and NOₓ, that reduce air quality and are the primary cause of acid rain. In addition to harmful environmental impacts, fossil-fuelled power plants emit particulates (PM10, PM2.5) that cause human health issues, including respiratory problems and increased risk of lung cancer [3].

There is a strong motivation across the world to create a sustainable and clean electric power system [4–6]. Power system generation resources have been globally undergoing rapid changes over the last decade. In the U.S. electricity sector, illustrated in Figure 1 as the fuel-mix since 2008, natural gas has overtaken coal as the primary source of electric energy [7]. Along with changes in generation, there have been increasing developments on the consumer side of the electric grid. Advancement in communication technologies in power systems has opened opportunities for the active participation of consumers in smart grid operations. Consumer participation in power systems, either by distributed generation or demand response, has been extensively studied for system economic, social, and environmental impacts [8, 9].

As real U.S. power system network data is considered Critical Energy/Electricity Infrastructure Information (CEII), which is a valuable asset to the security of the nation, research studies on economic impact [10–12] or environmental assessment of power systems [13–16] must publish their findings based on available data on standard power system test cases (e.g. IEEE 118-bus) as they do not have access to the real power system. The economic and environmental analysis of studies based on simulations depends on the accuracy of the generator data provided in the power system test cases. To accurately evaluate the impact of new generation or demand-side management techniques, the power system test cases used for simulation should represent the real power system. A data-driven technique is designed to augment existing power system...
test cases to accurately emulate the real-world generator fuel mix, along with realistic generator unit-types, to estimate emissions that reflect the real power system.

Most power system test cases were developed based on data from parts of a real power network. The most commonly used are IEEE power system test cases, which were synthetically developed based on portions of the 1970s U.S. power network, available in MATPOWER [17]. None of these test cases have the necessary data to estimate emissions. The reliability test system of 1996 (RTS-96) is provided with the necessary data to estimate emissions, but the fuel-mix on this test case is highly outdated [18], as illustrated in the changing U.S. fuel-mix in Figure 1.

Recently through the U.S. Department of Energy’s GRID DATA programme, large synthetic test cases were developed that are statistically similar to large geographical areas in the United States. [19]. These power system test cases have realistic lines, line limits, and load data based on demographically accurate data. The generation mix of these test cases is also based on the geographical location and are equipped with fuel-type data [20]. The generator cost functions were developed based on the heat-rate curves and fuel costs from open-source EIA data. These test cases, however, do not have the required data to perform accurate economic studies based on organised electricity markets [21]. Additionally, these test cases lack the required data to accurately estimate emissions from each fuel type.

The NREL 118-bus test system is a modified version of the IEEE 118-bus test case with extensive generator information [22]. The generator information is based on a possible future generation mix of the Southern California region. The source of the cost functions for the fuel types used is unknown, and there is no method to verify the accuracy of these prices compared with future market generator offers. The ISO-NE 8-bus test case is a deregulated market-based test case with generator cost functions derived from offer data of a real electricity market [23]. The cost functions in this work only represent the costs of generator offers for a single day, and a method has not been presented for updating the cost functions based on the changing market offers. The generator types are derived based on the installed capacity of the real system, but there is no data provided to estimate emissions from simulation studies.

The Federal Energy Regulatory Commission unit commitment test system (FERC-UC) was developed based on the generator information from PJM [24]. The test case has comprehensive turbine and fuel information with cost functions derived from the market-offer data, but the network information of this test case is not made public for security concerns. Based on the literature, there is no publicly available power system test case that is capable of accurately emulating the long-term dispatch of a real electricity market in terms of price and emissions.

Pacific Northwest National Laboratory (PNNL) developed a tool that can evaluate the impact on emissions for smart-grid projects [25]. The tool uses user specified load data before and after implementing the smart-grid technology and produces a comprehensive report on the impact of this technology based on the region it has been deployed. This tool uses the average emissions data published by Environmental Protection Agency (EPA). This tool is useful for post-power flow analyses. Many studies, however, such as assessment of new generation technology and emissions-constrained dispatch requires the time-series dispatch of generators, which can be performed only using power flow studies.

In our previous work, we proposed a technique to augment power system test cases to emulate marginal prices of an organised electricity market by replacing the fuel cost-based generator cost functions with market offer-based cost functions for economic simulation studies [21]. Day-ahead generator offers from a real electricity market were used to develop the new generator cost functions. In this work, we propose a technique to further augment the test cases by adding fuel and turbine types with the corresponding heat-rate curve to the test case generators to represent the fuel-mix and emissions of a real power system. The major contributions of this work are:

i. designed a data-driven technique to derive time-series fuel-mix data for power system test cases to represent the generation profile of a real electricity market; and
ii. developed a method to obtain generator heat-rate curve data for power system test cases based on the thermal efficiency of real generators to accurately estimate GHG and AP emissions.

To validate the proposed technique, three test cases were augmented and simulated with the hourly fuel-mix for 1 year, and the emissions from the simulation were compared with the fuel-energy mix and the emissions of a real system.

Here, Section 2 provides a brief discussion on emissions estimation in power systems. The proposed technique to augment existing test cases for economic and emission studies is presented in Section 3. The simulation setup and the data sources for developing the augmented test cases are presented in Section 4. The results are analysed and discussed in Section 5, followed by the conclusions in Section 6.
EMISSION ASSESSMENT IN POWER SYSTEMS

As of 2019, the majority of electricity generated in the U.S. is based on converting thermal energy to electricity (thermal generation units) and most is derived from fossil fuels [7]. Heat generated by burning fossil fuels is converted into kinetic energy to drive the turbine-generator shaft. During combustion of fossil fuels, GHG and AP are emitted. The heat required to generate a unit of electricity can be determined based on the thermal efficiency of the turbine generator. The total GHG and AP emitted can be determined based on the heat required by each generator and the emission factor of the fuel used to generate the total electric energy. The following data is required to estimate the quantity of GHG and other AP emitted from a generator and the total electric energy. The following data is required for each generator and the emission factor of the fuel used:

- Electric energy (MWh) produced by each generator,
- Efficiency (heat-rate curve given in MMBtu/MWh) of each turbine-generator (e.g. gas turbine),
- Fuel-type of each generator (e.g. natural gas),
- Heat value of each fuel-type (MMBtu/unit quantity fuel),
- Emission factor (e.g. emissions/unit quantity fuel) and emission control factor (i.e. percentage reduction of emission per-unit fuel) of each fuel-type.

For a given generator $k$ using fuel-type $f$, let $E_k$ be the energy generated in MWh over a given time, $H_k$ be the thermal efficiency in MMBtu/MWh, and $M_f^Z$ be the emission factor in lbs./MMBtu for pollutant $Z$ (e.g. CO$_2$, NO$_x$, SO$_2$). The product of $E_k$ and $H_k$ results in the net thermal energy input (in MMBtu) and, based on the emission factor of a given emission $Z$, will result in the total emission by generator $k$. The total emissions of pollutant $Z$ in lbs. for all $N$ generators in a power system, $O^Z$, is evaluated using Equation (1) as the sum of emissions of each individual generator over a period of time. As the PJM Interconnection publishes emission details as monthly system average emissions in lbs./MWh, the monthly average emissions for this study are obtained by dividing the total emissions produced over a month by the total energy produced by all generators (renewable and non-renewable).

$$O^Z = \sum_{k=1}^{N} E_k H_k M_f^Z$$  \hspace{1cm} (1)

The U.S. fuel-mix has been dynamically changing in the long-term over the last decade, but also throughout the year based on climate, load, and dispatch, as shown in Figure 1. This seasonal and long-term change in the fuel-mix is not represented on existing power system test cases. The dispatch of generators in an organised electricity market is based on the offer price of each generator (i.e. the price per-unit energy bids into the daily energy market).

To accurately estimate the energy produced by each fuel-type from a simulation, the cost curves of the power system test case generators must be related to the offer price of generators in a real electricity market [21]. Herein, a statistical method is proposed to assign the generator-fuel information to the test case generators that have market-based cost functions. Along with the generator-fuel data, power system test cases are assigned the data mentioned in the list above (e.g. heat-rate curves, emission factors) that existing test cases are missing to accurately estimate the time-varying emissions of a U.S. electricity market using data-driven techniques.

2.1 Generator fuel types

The most critical data required to estimate the emissions from an electric power system is fuel quantity estimation. There are several political, economic, and environmental reasons for the change in the fuel-energy mix. Old coal-based generation is being replaced by cheaper and more efficient combined-cycle natural gas-fired units and renewable generation to improve economic and environmental sustainability [26]. In the PJM electricity market, natural gas units are quickly replacing coal units as both the base-load and the daily marginal generators (i.e. natural gas often sets the marginal cost of electric energy). Combined-cycle natural gas power plants are being installed as base-load generators in PJM, with the capacity factor rising from 50% in 2013% to 63% in 2016, during which the capacity factor of coal-based units reduced from 54% to 49% [27, 28]. Along with the fuel-type used, the turbine (prime mover) information is important for estimating accurate emissions. This information becomes more important to those fuels that are used with multiple turbine technologies (e.g. natural gas).

2.2 Generator heat-rate curves

Each turbine-generator combination has different thermal efficiencies represented by heat-rate curves. Generator heat-rate curve data is the ratio of total thermal energy input to the useful electric energy output that indicates the efficiency of a thermal power plant. The heat-rate of a unit varies based on the operating point; usually, the efficiency is lowest at minimum loading as most of the thermal energy is used to maintain the minimum temperature of the thermal system. Thermal power plants incorporate multiple heat recovery techniques, such as combined-cycle installation and air and water pre-heaters, to improve the efficiency of the system. Along with the technologies used, the age of the power plants also influences the heat-rate of the system.

Figure 2 is the distribution of average heat-rate (also known as thermal efficiency) of generators evaluated based on the annual heat input and electric energy produced by the major generators used in PJM Interconnection. The distribution is represented using modified box plots (letter-value plots) where the widest box represents the major middle quartile, and the black line is the median of the distribution. The advantage of using this letter-value plot over the box plot is the ability to
represent the distribution of data in addition to the quartiles and median. The average heat-rate $H_k$ in MMBtu/MWh of generator $k$ is evaluated using Equation (2), where $Q_k$ is the heat in MMBtu required to generate $E_k$ MWh of energy. In Figure 2, the thermal and electric energy data of the generators were estimated from *EIA, Power Plants Operations Report*, which publishes annual reports in a one-month time resolution [28]. Though the report has filings from all generators across the U.S., we used data of the generators operating under the PJM Interconnection for this work.

$$H_k = \frac{Q_k}{E_k}$$  \hspace{1cm} (2)

The important observations from Figure 2 are: (a) the distribution of heat-rates of a set of generators using a certain fuel and turbine-type has a wide range of efficiencies, and (b) the median efficiency of generators using a certain type of fuel is not similar across different turbine-types (e.g. natural gas generators). The median efficiency of natural gas generators using a combine-cycle steam unit is much greater than natural gas using a gas turbine, hence it is not enough to just assign a fuel type to a unit. The large variation in efficiencies among generators using the same fuel and turbine-type is caused by the different age of generators (newer generators are usually more efficient), and the generator's operating point on the non-linear heat-rate curve throughout the year. It is important that the test case generators represent this diverse thermal behaviour of generators to estimate emissions accurately from simulations.

### 2.3 Emission factors

Emission factors are a representative value that relates the physical quantity of GHG/AP emissions to the quantity of fuel consumed. These factors are often provided as a ratio of lbs. of emission output to produce one MMBtu of thermal energy. In the U.S., the EPA is the authority that evaluates the emission factors for GHG and AP for all fuel types used across all industries [29, 30]. These emission factors can be used in simulations to evaluate emissions from their respective industries. Realistic emission estimates can motivate authorities and researchers to develop control strategies to reduce the harmful impacts of emissions.

Emission factors of all fossil-fuelled generators must be estimated to evaluate the total emissions from the electric power industry. The emission factor of a power plant depends on the type of fuel used and the emission reduction techniques implemented. For a generator with fuel-type $f$, the emission factor $\mathcal{M}_f'(\text{lbs./MMBtu})$ is evaluated using Equation (3), where $\theta_f$ (lbs./MMBtu) is the uncontrolled emission factor of the fuel ($f$). An uncontrolled emission represents the actual emission when the fuel undergoes combustion prior to any emission reduction techniques. One of the major contributions of the work is deriving the controlled emission factors for test case generators, which are not provided on any other test case and are required to accurately emulate emissions.

Power plants in the U.S. are required to install pollution control techniques that reduce AP emissions. These control technique efficiencies change with development in new techniques. The emissions have to be scaled according to the emission reduction efficiency ($\eta_f^2$) of a pollutant $Z$ (NO$_x$, or SO$_2$), as shown in Equation (3), to evaluate the exact pollutants released into the environment. For this study, we use the generator data from the PJM area and the major fossil fuels used with their corresponding derived emission factors, presented in Table 1. The table contains emission factors for the major turbine types: steam turbine (ST), combined cycle (CC), gas turbine (GT), and internal combustion (IC). A detailed description of the uncontrolled emission factors $\theta_f$ and control efficiencies $\eta_f^2$ used for this study are provided in the appendix.

$$\mathcal{M}_f' = \theta_f' \left(1 - \frac{\eta_f^2}{100}\right)$$  \hspace{1cm} (3)
| Fuel class    | Actual fuel   | Turbine | $\lambda_i^{CO_2}$ (lb/MMBtu) | $\lambda_i^{SO_2}$ (lb/MMBtu) | $\lambda_i^{NO_x}$ (lb/MMBtu) |
|--------------|---------------|---------|-------------------------------|-------------------------------|-------------------------------|
| Coal         | Bituminous    | ST      | 210.6                         | 0.32                          | 0.2                           |
| Refined coal | ST            | 210.6   | 0.32                          | 0.2                           |                               |
| Sub-bituminous | ST          | 214.2   | 0.39                          | 0.16                          |                               |
| Waste coal   | ST            | 250.6   | 0.39                          | 0.16                          |                               |
| Gas          | Natural gas   | CC      | 117                           | 0.0006                        | 0.139                         |
| Natural gas  | ST            | 117     | 0.0006                        | 0.139                         |                               |
| Natural gas  | GT, IC        | 117     | 0.0006                        | 0.139                         |                               |
| Other gas    | Blast furnace gas | ST | 604.7                         | 0.006                         | 0.16                          |
| Landfill gas | IC            | 117     | 0                           | 0.04                          |                               |
| Biomass gas  | IC            | 117     | 0.028                         | 0.15                          |                               |
| Synthetic gas | CC          | 130.1   | 0.0009                        | 0.15                          |                               |
| Propane gas  | ST            | 135.5   | 0.0002                        | 0.01                          |                               |
| Multi-fuel   | Municipal waste | ST | 200                           | 0.005                         | 0.05                          |
| Waste wood   | ST            | 206.8   | 0.005                         | 0.088                         |                               |
| Tire-waste   | ST            | 189.5   | 5.34                          | 0.39                          |                               |
| Oil          | Residual fuel oil | IC, GT, ST | 163.1                        | 0.1                           | 0.24                          |
| Distilled-fuel | IC, GT    | 163.1   | 0.1                           | 0.06                          |                               |
| Kerosene     | IC            | 165.8   | 0.1                           | 0.06                          |                               |

3 | AUGMENTED POWER SYSTEM TEST CASES

In an electricity market, participating generators submit a daily set of offers (i.e. blocks of $$/MWh), which can change the supply curve of the system. The daily generation fuel-mix profile not only depends on the changing load, but also with the change in generator offer costs. The cost functions of the test case generators should be dynamically updated with the latest generator offers from a real market to replicate the generator supply curve in simulation studies. Developing a test case with fixed cost functions will not result in a dispatch that emulates real electricity markets for simulations with longer time horizons. A fixed-cost test case is one that uses the same generator cost functions based on fuel costs of a certain date for the entire simulation duration.

This section describes (i) an improved method for augmenting the generator cost functions on power system test cases, and (ii) the design of new data-driven methods to augment power system test cases with emissions data. An outline of the steps involved in augmenting a test case is presented in Figure 3. The right part of the figure describes the layers of data augmented to the test case, and the left column blocks represent the flow of steps to obtain and augment data. The trapezoidal green boxes represent the approach to process the data, and each of the red, yellow, and blue boxes represent...
the augmented test case with different levels of augmentation (i.e. cost, fuel, and heat-rate data, respectively). Each of these processes are explained in detail in this section.

3.1 Augmented cost functions

To establish a power system test case that can represent the time-varying nature of generator offers in an organised electricity market, in prior work we introduced an augmentation technique for test cases based on pattern recognition [21]. Because the generator/fuel type of a unit is unknown from the offer data, an unsupervised pattern recognition technique (i.e. k-means clustering) was implemented to group generators with similar market offer characteristics. The generators were clustered into three groups (i.e. base, intermediate, peak) based on the capacity of the offer, weighted offer cost, minimum run time, and generator operating limits. In that study, the PJM market was chosen for augmenting the test cases.

Figure 4 represents two of the four cluster dimensions (i.e. capacity and weighted offer cost) for all generators participating in the PJM day-ahead market on 25 July 2016. Because the generators submit new offers every day, the clusters are recalculated daily. Generator cost functions are created using polynomial curve fitting from the market offer data, which is then used as the generator cost functions for the power system test case. Each cluster is sampled to ensure the augmented test case has the same percentage of base, intermediate, and peak as the real system.

As the number of generators in a power system test case is significantly fewer than the real power system market, undersampling issues can arise. For example, the generators in the peak-load cluster vary from ~10 MW offered at ~$900/MWh to ~700 MW offered at ~$10/MWh, shown in Figure 4 (note: this is due to the other generator offer features that are not shown on the 2D-projection). In Ref. [21], the polynomial cost functions were uniformly sampled from the real generator data and assigned to power system test case generators. This implies that the generator offer at ~$900/MWh has an equal probability of being assigned as the myriad near the cluster’s centroid.

To reduce the under-sampling issue, the generator cost functions should be selected based on the density of the generators in a cluster [31]. A well-established method for estimating the maximum likelihood or density of multi-dimensional data is principal component analysis (PCA) [32]. PCA is a dimensionality reduction technique that projects the multi-dimensional data on axes where the variance under projection is maximal. Using PCA, the multi-dimensional generator offer data was reduced to two dimensions along the major axes of principal components (PC). A bi-variate histogram is determined for each of the clusters along the two PC axes with the number of bins determined by the Scott formula presented in [33]. Figure 5 represents the probability distribution of the same peak-load generators from Figure 4 on 25 July 2016, along the PC1 and PC2 axes. As the clusters change daily, the PCA-based bi-variate histogram is evaluated each day to sample the generator cost functions.

We improve on the uniform sampling in [21] by sampling generator cost functions based on the probability of the bivariate distribution of generators in each cluster. Each bin has a different number of generators, and bin sampling is performed based on the density of each bin. The generator within a bin is uniformly selected once a bin is chosen. This process reduces the chance of outlier generators being selected for the power system test cases and increases the chances of a test case generator performing similarly to one that represents the cluster. For example, we would expect a large test case generator to behave as a base-load unit, and the cost function of such a unit must reflect the characteristics of a base-load unit consistently over the simulation time-series. This consistency of being assigned the cost functions from the most likely region of a cluster is vital to determining emissions, as a particular test case generator will be associated with the same generator-fuel type throughout a simulation.
3.2 Augmented fuel and turbine types for test cases

The data required to estimate emissions from a power system is presented in Section 2; in this section we present the technique to use real system data to augment test cases to estimate emissions from test case simulations. The cost data from Section 3.1 forms the first augmented data layer for the test case, which is required to perform accurate optimal power flow (OPF) to obtain generator dispatches. The test case is further augmented with fuel and turbine information to estimate emissions based on test case simulations.

One-to-one mapping of the real system generator data to the test case is not possible because the number of generators in the real system is much larger than most test cases. To overcome this issue, the proposed technique augments each test case generator with multiple fuel and turbine-types. These augmented units are then assigned heat-rate curves and emission factors based on the given fuel and turbine-types. The resulting test case will represent the real system in terms of the fuel-types and percentage of fuel-mix, and average system emissions.

Figure 6 is a representation of a test case generator that is augmented with all fuel and turbine types in proportion to the real system. The cost function $C$ in Figure 6 is a graphical representation of a cost function that represents the offer curve of an entire system on a single generator (i.e. the augmented cost function from the previous section). Each shaded segment of the test case generator represents an augmented generator that uses fuel type $f$ and turbine type $g$ from the valid set of fuel-turbine $(F-G)$ combinations used for this work (see Table 1).

For the sake of brevity we define the valid set of combinations of the fuel $F$ and turbine $G$ with each fuel-turbine-type as $i$ (see Table 1 with 16 types of fuel and four types of turbines) using a new single set of $I$ combinations with $i \in I$, from set $I$. The augmented generator 1 represents a generator with the least offer price and fuel-turbine type of a base-load generator, and similarly the augmented generator $I$ is the peak generator that has the highest offer price and uses a fuel-turbine combination of a peaking power plant. Each of these augmented generators share a segment of the cost function of a single test case generator. For the remainder, a generator segment refers to a specific delineated part of the test case generator cost function that is assigned to the augmented generator.

3.2.1 Fuel and turbine type assessment

The electricity market hourly fuel-mix data provides the energy produced from the primary sources to meet area demand plus exports. At the time of writing this work, the authors are aware of two markets (i.e. PJM and ISO-NE) that publicly host this data [34, 35]. This data only contains the hourly energy produced by fuel, but does not specify any other generator related parameters such as cost, size, or location. Without correlating parameters, the real hourly fuel-energy mix data cannot be directly attributed to power system test case generators.

There are no common attributes to relate the generator data from the two different data sources and the test case generators (i.e. energy market and EIA). In this work, we propose using the capacity factor of the fuel types in the real-system correlated with the capacity factors of the test case generators from the OPF solution. The capacity factor is the ratio of the total energy produced for a time period over the maximum energy that could be produced during the same time period. We chose this attribute because of the positive correlation of capacity factor and the generator type (i.e. fuel-turbine type). A generator with a higher capacity factor is generally related to that of a base-load unit, and a generator with a lower capacity factor is most likely a peak-load unit. For fossil fuel-based thermal units, the capacity factor can be related to the generator cost (high capacity factor-inexpensive offer price, low capacity factor-expensive offer price).

To evaluate the capacity factors of generators, data was analysed from EIA, Power Plants Operations Report (Form EIA-923) [28], which contains fuel, turbine, location, and thermal information. The capacity factor, defined as $\bar{C}_i$ for each fuel-turbine-type $i$ from set $I$, is evaluated as the ratio of the total energy produced for a time horizon $T$ over the maximum energy that could be produced, as shown in Equation (4). The energy produced for turbine-type and fuel type $i$ in time period $t$, defined as $E_{it}$, is obtained from EIA-923 [28]. To evaluate the maximum energy that could be produced, the installed capacity, defined as $P_{i,CAP}$, of the corresponding set of turbine generators and fuel types is multiplied by the number of hours in the time horizon $T$. The installed capacity of the major fuel types can be found from the annual report of the electricity market of interest [36].
used data from PJM. The efficiency rate of each generator (MWh) is estimated and the system input to the unit volume is provided. The EIA dataset contains monthly heat input (MMBtu) and the total electrical output (MWh) of each generator. This data is used to obtain the per-unit efficiency of each generator by dividing the heat input to the electrical output (MMBtu/MWh). The blue dots in Figure 7 represent the heat-rate of the steam-turbine portion of a combine cycle natural gas power plant operating in PJM. Each point represents the monthly performance of the unit during the year of 2016 as MMBtu required to produce a given MWh.

The heat-rate functions are non-linear and the operating point of the generator is important to estimate the accurate heat input from the heat-rate (H) in Equation (1). A heat-rate function developed by directly fitting the thermal performance may not be useful to most test cases because the size of the test case generator differs from the real generator. To overcome this issue, the energy output of the generators is converted to a relative capacity factor, and the heat-rate curve is fit between the monthly efficiency (monthly MMBtu/MWh) to the relative capacity of each generator. Because the actual installed capacity is not known from the EIA-923 data, the relative capacity $\hat{R}_k$ of each generator $k$ is estimated using Equation (6), which is the ratio of monthly electric output $E_k^t$ (i.e. each blue dot for the generator described in Figure 7) in MWh over the maximum energy $E_k$ (~72,000 MWh) for the generator in Figure 7 produced during any month in the same year.

\[
C_i = \frac{\sum_{j=1}^{T} E_{ij}}{P_i^{CAP} \times T} \quad (4)
\]

The efficiency of generating electricity from a certain fuel varies largely by the type of turbine used (e.g. natural gas plants using a steam turbine is more efficient than a gas turbine). It is important that the test case represents a similar share of each turbine type as in the real power system. The number of generators in the EIA-923 dataset is much larger than the number of power system test case generators (i.e. under-sampling issue as with the economic data). To statistically represent the real system, the percentage of each turbine-type using a particular fuel must be evaluated. A comprehensive dataset is derived to estimate the percentage share of each generator $i$, defined as $S_i$, using Equation (5), which represents the proportion of energy for a given turbine-type for each of the fuel types used in the bulk-power market. The percent share is evaluated by dividing the energy $E_i^t$ produced over a period $T$ by generators of type $i$ over the sum of energy produced by all generators that use the same fuel as generator type $i$ represented by $E_{ij}^t$ (e.g. the energy produced by natural gas combine-cycle plants over the energy produced by all natural gas plants). The period $T$ used in this study is one year, because we derive this data from the annual EIA-923 dataset [28].

\[
S_i = \frac{\sum_{j=1}^{T} E_{ij}}{\sum_{j=1}^{T} E_{ij}} \quad (5)
\]

3.2.2 | Heat-rate curves assessment

The thermal information of each generator is provided in EIA-923 data [28]. The generators that participate in a power system region are filtered by correlating the company/utility names and location with the generator member information of that region (i.e. PJM for this study [37]). The EIA-923 data contains monthly heat input (MMBtu) and the total electrical output (MWh) of each generator. This data is used to obtain the per-unit efficiency of each generator by dividing the heat input to the electrical output (MMBtu/MWh). The blue dots in Figure 7 represent the heat-rate of the steam-turbine portion of a combine cycle natural gas power plant operating in PJM. Each point represents the monthly performance of the unit during the year of 2016 as MMBtu required to produce a given MWh.

An example of a thermal efficiency curve is presented as the red squares in Figure 7, where each point corresponds to the thermal performance (in blue dots) in a month. The right
y-axis represents thermal efficiency (lower is better) as the ratio of thermal input (in MMBtu) per month over the electric energy produced (in MWh) in that month and the top x-axis is the relative capacity. The non-linear performance can be observed in the thermal efficiency curve where the maximum efficiency is achieved at the annual peak and stays relatively similar until the relative capacity reaches 0.3, after which there is a sharp decrease in efficiency. An augmented test case unit of natural gas and combine cycle can be assigned this thermal efficiency curve (heat-rate) irrespective of the size of the unit, and will represent the thermal characteristics of this real generator. Thermal efficiency curves of each generator in the set of functions based of generators using a fuel-turbine combination $i$.

### 3.2.3 Augmented generator model

A technique to assign the fuel and thermal data, as derived in the preceding sections, is presented in this section. The proposed Algorithm 1 is a statistical technique to augment test case generators with turbine and fuel data by correlating capacity factors of the real system to the test system. The algorithm is explained for a single generator for brevity, but can be extended to augment any test case. The data required to augment the test case are (a) the hourly fuel-mix percentage $U_{f_i}^{max}$ during the peak-hour of the day for each fuel $f_i$; (b) the detailed annual generator energy and capacity information from EIA-923 [28]; and (c) the OPF dispatch of each test case generator at each hour $p(t)$. The set of augmented generators for a test case are defined each day, similar to the cost functions that are valid for one market day due to changing market conditions and generator dispatches. The process can be repeated to simulate longer time periods.

Steps 1 and 2 of Algorithm 1 are evaluated using the yearly market performance information from the annual reports and data from EIA-923 [28, 36]. In Step 1, the capacity factors and share of each fuel are evaluated using Equations (4) and (5) for all turbines and fuel types $I$. The resulting data is sorted in descending order of capacity factor ($C_i$) to first assign the fuel-turbine with the highest capacity factor. This set of capacity factors and fuel-turbine-type energy share remains the same to augment any test case representing the same electricity market and time period.

The loads on the test case are augmented with real demand of an electricity market area to represent the market operation as described in [21]. For a test case with $m$ buses, the augmented demand $L_j(t)$ at hour $t$ on bus $j$ is obtained using Equation (7), where $p_j(t)$ is the default demand on bus $j$ provided with the test case, $L_{mk}(t)$ is the total demand of the real market (e.g. PJM, ISO-NE), and $L_{mk_i}$ is the annual peak-load of the market. A scaling factor, $\psi$, is used to scale the default load such that the ratio of peak demand to installed generation capacity is similar to that of the real market [21].

$$l_j(t) = \psi p_j(t) \frac{L_{mk}(t)}{L_{mk_i}} \quad \forall j = 1, \ldots, m \quad (7)$$

After evaluating the hourly demand for the day, the data of the peak hour is used to augment the test case generator. The capacity of augmented generator $a_i$ using fuel-type $f$ and turbine type $g$ is obtained as a percentage of the test case generator $p_{\text{max}}$ using Equation (8). The test case generator capacity is multiplied by the percentage of energy produced by a fuel ($U_f^{\max}$, obtained from the market data) and the share of a turbine-type $g$ using fuel $f$ as evaluated in Equation (5).

$$a_i = U_f^{\max} p_{\text{max}} S_i \quad (8)$$

The entire capacity of the test case generator is augmented with the capacity of each fuel and turbine type from the electricity market. A representation of the augmented test case generator is presented in Figure 6 where all the fuel turbine types are augmented over a test case generator with peak capacity $p_{\text{max}}$. The augmented generator units are assigned segments on the test case generator such that the rightmost segment $a_1$ corresponds to the fuel and turbine type $(f)$ that has the least capacity factor (peak-unit), and the augmented generator unit $a_1$ is assigned the left most segment with highest capacity factor (base-unit) as shown in Figure 6.

By assigning the leftmost segment of the test case generator to a fuel and turbine type with the maximum capacity factor, we are emulating the low offer price of the base load units (refer to Figure 4) on a single cost function. As the test case generator output increases to its maximum, the per-unit cost of generation increases (non-linear cost function) and the right most segment of the test case generator will represent the fuel and turbine type with the least capacity factor (peak unit). This step will establish a statistical relationship between the market offer data and the real fuel-turbine mix.

---

**Algorithm 1 Algorithm to assign test case generators a fuel and turbine-type based on the market-hourly fuel data and capacity factor.**

**Input:** hourly market energy mix, hourly OPF dispatch data from market-based test case, EIA-923 data for the market region.
1: evaluate capacity factors ($C_i$) for all fuel and turbine type.
2: evaluate the share of each fuel and turbine type ($S_i$) as a percentage of energy by the fuel and sort them in descending order of ($C_i$)
3: evaluate piecewise step heat-rate function for all the generators from the EIA-923 generator data as $H_k(R_k)$
4: evaluate the average heat value of each generator in the EIA-923 data using Equation (2).
5: arrange the energy share of each fuel and turbine type in descending order of capacity factor.
6: evaluate the augmented demand $(l_j(t))$ for each hour of the day and obtain the generator output $(p_{\text{max}})$ during the peak hour of the day.
7: for $i = 1$ to $I$ do.
8: evaluate the size of each augmented generator unit for each fuel and turbine type during the peak hour using Equation (8) \((a_i)\).

9: a segment of the test case generators’ capacity is assigned to an augmented generator unit based on the capacity factor of the fuel and turbine type.

10: assign a heat-rate curve based on the density of average heat value for each augmented generator \(a_i\) from the corresponding fuel and turbine type piecewise step.

11: functions \(H_k(\lambda_i)\) assign the corresponding emission factor \(M_i\) of the fuel and turbine type for each augmented generator.

12: end for.

Output: augmented test case with fuel, turbine, thermal, and emission factor information.

Each augmented test case generator is assigned a heat-rate function \(H_k\) developed as described in Section 3.2.2. Each fuel-turbine combination \((i)\) has a set of heat-rate functions derived from the EIA-923 data. For assigning an augmented generator unit \(a_i\), a heat-rate function \(H_i\) is selected based on the likelihood value (similar to the cost curve selection in Section 3.1) of the average heat-rates of generators using the fuel and turbine type \(i\). The average heat value \(H_k\) in MMBtu/MWh of each generator is evaluated using Equation (2) as described in Section 2.2.

Electric energy is required for estimating emissions, as discussed in Section 2. OPF results are used from the augmented test case generator to estimate the electric energy produced. An augmented generator has a peak daily capacity of \(a_{i}\) MW, and at hour \(t\) the output of each augmented generator \(a_{i}^t\) depends on the test case generator output \(p(t)\) at hour \(t\). The heat-rate functions \(H_k\) derived in Section 3.2.2 are derived as a function of capacity factor. The capacity factor of the augmented generators is required to estimate the total heat input required to generate the electric energy. The capacity factor of an augmented generator \(\lambda_i\) is evaluated using Equation (9) as the sum of the output of augmented generator \(a_{i}^t\) at each hour of the day over the total energy that could be produced over 24 h. The derived capacity factor \(\lambda_i\) is then used to determine the heat-rate in MMBtu/MWWh input using the piecewise step heat rate function \(H_i(\lambda_i)\).

\[
\lambda_i = \frac{\sum_{t=1}^{24} a_{i}^t}{24a_{i}}
\]  

(9)

Emissions of pollutant \(Z\) in lbs. based on the OPF simulation using the augmented test case is evaluated using Equation (10), where the energy of each augmented generator is multiplied by the heat-rate function \(H_i(\lambda_i)\) and with the corresponding emission factor \(M_i^Z\). To estimate the emissions for any pollutant \(Z\), the emission factors from Table 1 are used. The system average emissions \(O^Z_\text{avg}\) in lbs./MWh are obtained by dividing the total emissions \(O^Z\) by the total energy generated over the day as shown in Equation (11).

\[
O^Z = \sum_{g=1}^{1} \sum_{b=1}^{34} a_{i}^t K_i(\lambda_i) M_i^Z
\]  

(10)

\[
O^Z_\text{avg} = \frac{O^Z}{\sum_{b=1}^{24} p(t)}
\]  

(11)

Algorithm 1 can be extended to any test case. To use the algorithm for a test case with multiple generators, \(p_{max}\) in Equation (8) must be replaced with the sum of capacities of all generators during the peak hour. Capacity factors of the test case generators should be used to assign the augmented capacities to multiple test case generators. The test case generator with the highest capacity factor will be assigned the fuel and turbine type with the highest capacity factor (base-load) and the test case generator with least capacity factor will be assigned the fuel and turbine type with the least capacity factor (peak-load). The complete process can be iterated daily through time to estimate emissions for a longer simulation time period.

4 | SIMULATION SETUP

The proposed augmentation algorithm is implemented using the PJM system fuel-mix data for 2016 in an hourly resolution [34]. Yearly energy output data from the EIA-923 generators report, and the summer installed capacity data from the annual PJM State of Market Report were used to calculate the capacity factors using Equation (4) for all major fuel-turbine types used in the PJM Interconnection [28, 36]. The capacity factors of all the fuel-turbine types are presented in detail in our provided open-source data repository [38]. PJM Interconnection contains a few fuel-turbine-types that have less than 0.1% of the total energy share and were not included for this study.

As there is no existing test case that represents the physical network of the PJM region, the proposed fuel and emission factor data augmentation technique is implemented on three of the best performing market-based augmented test cases [21]. Two of the test cases are IEEE reliability test systems (RTS) (a) RTS-79 [39], and (b) RTS-96 [40], and the third test case is the 500-bus synthetic test case developed based on the geographical area of South Carolina (SC-500) [41]. The test cases with augmented cost functions are further augmented with turbine, fuel, and heat-rate curves data using the proposed technique described in Section 3. All the augmented data for the test cases and data required to augment any test case are made available in an open-source repository [38].

All simulations were conducted in MATLAB 9.3 using MATPOWER for OPF [17]. RTS-96 obtained one of the highest goodness of fits (GoF) in representing the marginal price of PJM using the augmented cost functions and will be
the primary test case used to analyse all results in detail [21]. Although the GoF of two other test cases were higher (i.e. RTS-79 and SC-500), the number of generators in RTS-96 is greater which is important in this work to represent all the fuel turbine types on a single test case. Scaled demand of PJM using Equation (7) was used as load for the time-series simulation on the RTS-96 test case, where $D_{\text{load}}(t)$ is the total demand on the real PJM network, and $D_{\text{peak}}$ is the annual peak-demand of PJM for the year in consideration (i.e. 2016 for this simulation). A scaling factor, $y = 0.94$, is used to scale the default load such that the ratio of peak demand to installed generation capacity is similar to that of the real market [21].

Apart from the FERC RTO UC test case, there is no other test case that has real PJM data to compare against the proposed augmented test cases; the FERC RTO UC test case is unavailable to use for comparison purposes because the network information is not available to the general public which is the goal of this work. To establish a reference test case to compare to this study, the RTS-96 test case generators were modified to represent the fuel-mix of PJM. The RTS-96 test case was modified with the fuel-mix representing the installed capacity of PJM in summer 2016 (i.e. fixed cost functions and fuel-mix to represent the behaviour of existing test cases). In addition to the fuel data, the test case is also provided with generator heat-rate and emission factors derived from the FERC RTO UC test case data [24]. The cost functions used for this test case are derived using the fixed heat rate and the average fuel cost during 2016, and the same cost functions are used for the entire year. This test case will be referred as fixed-cost RTS-96 here and is provided in the MATPOWER format [38].

The OPF results of the proposed augmented RTS-96 (using the method proposed here) and the fixed-cost RTS-96 use the same hourly demand curves developed based on the 2016 PJM demand curve and Equation (7). In both test cases, renewable generation is considered non-dispatchable and is represented by subtracting the hourly renewable generation from the hourly demand. The emissions estimated from both techniques are compared to the emissions from the PJM system. PJM Interconnection publishes the system average emissions for CO₂ and air pollutants SO₂ and NOₓ. This information can be browsed from the PJM environmental information services website [42]. The performance of both augmented RTS-96 test cases is judged by comparing the annual fuel-mix and system average emissions from the simulation to the actual PJM system values.

5 | SIMULATION AND RESULTS

The fuel-mix is compared by estimating the percentage of annual energy produced by each fuel source. Figure 8 is the hourly fuel-mix composition of the PJM system compared to the two simulated fuel-mix test cases in an hourly resolution. The pie charts represent the annual percentage of each fuel type. The top plot represents the PJM system annual fuel-mix and hourly fuel-energy mix. The goal of this work was to obtain similar fuel-mix pattern from simulation. Some of the visually striking patterns are the lower coal consumption during the month of Mar.–June and again in Oct.–Nov. The proposed technique accurately represents all the major fuel types of the PJM system compared to the fixed-cost RTS-96. The proposed technique not only represents the annual percentage accurately, but also the hourly and seasonal variation in fuel, specifically the coal and gas mix variation throughout the year.

The middle curve of Figure 8 represents the fuel-mix from the simulation using the fixed-cost RTS-96. These simulation results do not represent the PJM fuel-mix accurately, because the fixed-cost functions do not exhibit the same price sensitivity to commit generators of a real energy market with daily submitted offer prices. Energy from coal remains almost same throughout the year, and does not exhibit the seasonal reduction during the months described earlier. The lower percentage of natural gas in the fixed-cost RTS-96 is because the cost functions used in this test case are based on the FERC RTO UC test case, which was developed in 2010 when natural
The system emissions because of the accurate fuel-mix and the heat-rate curves that are developed from actual bulk-power system data. Goodness of fit is a statistical metric used in estimating the accuracy of a series $x$ to a reference series $x_{ref}$ as shown in Equation (12), where $\| \cdot \|$ indicates the L2 norm of a vector. The goodness of fit is a value between $-\infty$ (bad fit) and 1 (perfect fit). A goodness of fit of 0.93 was achieved with the proposed augmented RTS-96, and -0.08 with the fixed-cost RTS-96.

$$GOF = 1 - \frac{\|x_{ref} - x\|}{\|x_{ref} - \text{mean}(x_{ref})\|}$$

As mentioned in Section 1, $SO_2$ and $NO_x$ are the main contributors for AP and acid rain. The proposed data-driven augmentation method for estimating emissions on test cases can also estimate AP emissions using the emission factors $\langle M_i^2 \rangle$ for $SO_2$ and $NO_x$ presented in Table 1 and Equation (3). A detailed description of the emission factors of the AP and their respective reduction efficiencies is presented in the appendix. The system average AP from simulation is compared to the PJM system average AP in Figure 10. The goodness of fit for $SO_2$ was found to be at 0.68, and the goodness of fit for $NO_x$ was found to be 0.56. We could not establish the AP emissions from the fixed-cost RTS-96 because the controlled emission factor for AP were not available in any existing data repository. To the best of the authors’ knowledge, this is the only available power system test case that is capable of estimating the controlled emissions of these air pollutants. Controlled emission factors are important so that any analysis done based on these simulations reflect the real power system. Overestimating these emissions might result in pessimistic analysis and conclusions.

There is some randomness in Algorithm 1 derived from random sampling of generator cost functions and heat rates within the bins. Figure 11 presents the system average $CO_2$ emissions from 50 trials of a Monte Carlo simulation performed on the three standard power system test cases mentioned in Section 4. In each trial, the heat-rate curves for each fuel and

**Figure 9** Bar graph representing the monthly system average $CO_2$ emissions of PJM during 2016 compared to the simulated emissions on the fixed-cost RTS-96 and augmented RTS-96 test cases

**Figure 10** Comparison of simulated system average $SO_2$ and $NO_x$ emissions on the augmented RTS-96 test case in green to system average emissions from the PJM interconnection in blue for the year 2016.

Gas prices were significantly higher compared to the simulated year of 2016. Typically the use of natural gas for electricity generation peaks in summers and lowers in winter months because of the heating needs; this trend cannot be observed when simulating using the fixed-costs test case.

The bottom curve in Figure 8 represents the fuel-mix from the simulation using the proposed augmented test case that not only accurately represents the energy prices of a real market, but also the fuel-mix of the real system. By updating the cost functions and fuel-type of the test case generators daily based on the data from a real electricity market, the simulation results are closer to the PJM system than any other test case in literature; an accurate representation of the time-varying fuel-mix is necessary to accurately estimate emissions. This result provides supporting evidence to show the augmented test case represents the annual fuel-energy mix and temporal variation of the fuel-mix of the PJM system.

Next, the PJM system average emission information was used to compare the emissions estimated from the simulation. The system average $CO_2$ emission is evaluated as a ratio of the total emissions over a time period in lbs. to the total energy produced during the same time period in MWh. PJM Interconnection publishes the system average emissions in a monthly interval [42]. The plots in Figure 9 are a comparison of system average $CO_2$ emissions from PJM represented in blue with the emissions estimated from simulation using the fixed-cost test case (in red) and the proposed augmented RTS-96 test case (in green).

Coal has one of the highest emission factors of all fossil fuels (210 lbs./MMBtu), and natural gas has one of the least (117 lbs./MMBtu) as previously discussed in Table 1. Because the fuel-mix estimations from the simulations using fixed-cost RTS-96 estimated a higher proportion of coal-based energy and a lower share of natural gas, the resulting estimated $CO_2$ emissions were almost twice that of the real system. In Figure 9 this is noticeably observed in March, where the actual system average emission reduced with a reduction in coal-based energy, but the fixed-cost RTS test case estimated a rise in emissions.

The green bars in Figure 9 represent the emissions from the augmented RTS-96, which closely represents the PJM
turbine type were randomly selected using PCA. Each bar, except for the one representing PJM, represents the mean of emissions from the 50 trials simulated on the corresponding test case. The black lines over the bars represent the standard deviation for the 50 trials. The number of Monte Carlo trials depends on the probability of the bin and number of generators with a bin. Because the variance of the function within the bin is low and inter-bin variance is high, the number of scenarios depends mainly on selecting the bin. A bin is selected for a generator based on its density (25%–35% chance of peak bin), and we found that a Monte Carlo simulation using 50 trials to select 96 generators (RTS-96), each from one of 54 bins (see Figure 5), was adequate to be confident in the algorithm output.

During the summer months, the mean emissions of all the power system test cases are very close to the actual PJM system, and in many months fall within one standard deviation. This is because it is easier to accurately model the heat-rate curves for coal and natural gas that are the major share of fuel-energy for these months. January and February do not perform as accurately (although the effect is amplified by the zoomed in y-axis in Figure 11), as the majority of gas generation during these months is classified as other gas by PJM, and the exact composition (i.e. percentage of each fuel-type in the class of other gases) of these gasses is unknown. The heat-rate curves of all fuel types under other gas are considered as one subset for Step 4 in Algorithm 1. Sampling from this subset to assign heat-rate curves for the test case generator for Step 6 in the algorithm introduces this wide variation, causing less accurate estimates of emissions. Nevertheless, to make an accurate test case only one set of accurate heat-rate curves for a test case is required, and that set can be saved for multiple other simulations when the test case is used.

6 | CONCLUSIONS

Herein, a data-driven approach was proposed to augment open-source power system test case generators with real generator data to represent the dispatch of a real electricity market in terms of emissions and marginal costs. The proposed technique was implemented on three test cases of different sizes using generation data from PJM Interconnection, real generator data published in EIA, and emissions data from EPA [28, 30, 34]. Simulation results using the augmented test cases were shown to produce fuel-mix and emissions that accurately represent the dispatch of the PJM Interconnection. Additionally, the results showed the need of using market-offer data as generator cost functions along with the daily market fuel-energy mix to properly estimate emissions (as opposed to fixed-cost and fixed-fuel test cases found in literature). These augmented test cases can be used for simulating transmission level dispatch based on a deregulated power system, and evaluate the long-term economic and environmental impact of changing load, generation, and distributed energy resources.

The proposed data-driven technique can be used to augment any power system test case with market offer-based cost functions and heat-rate functions to analyse emissions. Any bulk-power market that provides access to certain generator data (e.g. offer price, fuel-energy mix, installed capacity per fuel-type) can be used. For the case of using the power system test cases from this work, we have made available the augmented RTS-96 test case in MATPOWER format, and all data required to augment other test cases based on PJM market data for 2016 in an open source repository [38]. The primary motivation is to enable researchers and authorities to simulate new initiatives and methods on power systems while accurately analysing the impact on GHG/AP emissions and energy costs.

ACKNOWLEDGEMENT

This work was supported by the National Science Foundation under grant number ECCS-1608722, the South Dakota State University Electrical Engineering Ph.D. fund, and the South Dakota Board of Regents (SDBoR).

ORCID

Venkat Durvasulu https://orcid.org/0000-0002-0644-9648
Timothy M. Hansen https://orcid.org/0000-0001-8096-1255

REFERENCES

1. U. S. Environmental Protection Agency: Sources of greenhouse gas. https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions. Accessed 30 August 2019
2. U. S. Energy Information Administration: Annual energy Outlook 2017, with projections to 2050, US Department of Energy,pp. 1–127. (2017) https://www.eia.gov/outlooks/archive/aeo17/. Accessed 30 August 2019
3. Kan, H., Chen, R., Tong, S.: Ambient air pollution, climate change, and population health in China. Environ. Int., 42, 10–19. (2012)
4. Cullen, J.A., Mansur, E.T.: Inferring carbon abatement costs in electricity markets: a revealed preference approach using the shale revolution. Am. Econ. J. Econ. Pol. 9(3), 106–133 (2017)
5. Lund, H.: Renewable energy strategies for sustainable development. Energy. 32(6), 912–919 (2007)
6. Lund, H., et al.: 4th Generation District Heating (4GDH): Integrating smart thermal grids into future sustainable energy systems. Energy. 68, 1–11 (2014)
7. EIA: U.S. Energy Information Administration (EIA), Electricity Data Browser. https://www.eia.gov/electricity/data/browser/. Accessed 30 August 2019
8. Nguyen, D.T., Le, L.B.: Risk-constrained profit maximisation for microgrid aggregators with demand response. IEEE Trans. Smart Grid. 6(1), 135–146 (2015)
9. Allcott, H.: Rethinking real-time electricity pricing. Resour Energy Econ. 33(4) 820–842 (2011). Special Section: Sustainable Resource Use and Economic Dynamics

10. Pineda, S., Bock, A.: Renewable-based generation expansion under a green certificate market. Renew. Energy. 91, 53–63 (2016)

11. Gkatzikis, L., Koutsopioulou, I., Salomidou, T.: The role of aggregators in smart grid demand response markets. IEEE J. Sel. Area Commun. 31(7), 1247–1257 (2013)

12. Sharma, S., et al.: Metrics-based assessment of sustainability in demand response. In: 15th IEEE International Conference on Smart City, Bangkok, pp. 150–157 (2017)

13. Hayashi, Y., et al.: Versatile modelling platform for cooperative energy management systems in smart cities. Proc. IEEE. 106(4), 594–612 (2018)

14. Raviprabhakaran, V., Subramanian, R.C.: Enhanced ant colony optimization to solve the optimal power flow with ecological intrusion. Int. J. Syst. Assur. Eng. Manag. 9(1), 58–65 (2018)

15. Parvania, M., Fotuhi-Firuzabad, M., Shahidehpour, M.: Assessing impact of demand response in emission-constrained environments. In: 2011 IEEE Power and Energy Society General Meeting. Detroit (2011)

16. Kang, C., et al.: Carbon emission flow from generation to demand: a network-based model. IEEE Trans. Smart Grid. 6(5), 2386–2394 (2015)

17. Zimmerman, R.D. et al.: MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education. IEEE Trans. Power Syst., 26(1), 12–19 (2011)

18. Fong, C., Haddad, S., Patton, D.: The IEEE reliability test system - 1996. IEEE Trans. Power Syst. 14(3) (1999)

19. Birchfield, A.B., et al.: Grid structural characteristics as validation criteria for synthetic networks. IEEE Trans. Power Syst. 32(4), 3258–3265 (2017)

20. Anido, C.B., Hodge, B.: An extended IEEE 118 bus test case generator for synthetic networks. IEEE Trans. Power Syst. Theor. Appl. 3(4), 429–444 (2021).

21. Durvasula, V., Hansen, T.M.: Market-based generator cost functions for power system test cases. IET Cyber-Phys. Syst. Theory Appl. 3(4), 194–205 (2018)

22. Peña, I., Martínez-Anido, C.B., Hodge, B.: An extended ieee 118-bus test system with high renewable penetration. IEEE Trans. Power Syst. 33(1), 281–289 (2018)

23. Krishnamurthy, D., Li, W., Tesfatsion, L.: An 8 zone test system based on iso new england data: development and application. IEEE Trans. Power Syst. 31(1), 234–246 (2016)

24. Krall, E., Higgins, M., O’Neill, R.P.: RTO Unit commitment test system. (Washington D.C.: Federal energy and regulatory Commission (FERC), https://www.ferc.gov/industries-data/electric/power-sales-and-markets/increasing-efficiency-through-improved-software-1 (2012) Accessed 30 August 2019

25. Studarus, M., et al.: Quantifying the emissions impacts of smart grid projects with a publicly available web calculator. 2016 IEEE Power and Energy Society General Meeting (PESGM), Boston, MA (2016)

26. United Nations Organization. Geneva (2015). The role of gas in sustainable energy. United Nations Economic Commission for Europe

27. U. S. Energy Information and Administration (EIA): EIA, Today in Energy: natural gas-fired power plants are being added and used more in PJM Interconnection. Washington. https://www.eia.gov/todayinenergy/detail.php?tid=37293. Accessed 30 August 2019

28. U. S. Energy Information and Administration (EIA): EIA, electricity: Form EIA-923. https://www.eia.gov/electricity/data/eia923/. Accessed 30 August 2019

29. U. S. Environmental Protection Agency (EPA): EPA emission factors for greenhouse gas Inventories. https://www.epa.gov/sites/production/files/2018-03/documents/emission-factors_mar_2018_0.pdf. Accessed 30 August 2019

30. U. S. Environmental Protection Agency (EPA). EPA, AP-42: Compilation of air emissions factors. https://www.epa.gov/air-emissions-factors-and-quantification/ap-42-compilation-air-emissions-factors. Accessed 30 Aug 2019

31. Fraley, C., Raftery, A.E.: Model-based clustering, discriminant analysis, and density estimation. J. Am. Stat. Assoc. 97(458), 611–631 (2002)

32. Tipping, M.E., Bishop, C.M.: Probabilistic principal component analysis. J. Roy. Stat. Soc. B. 61(3), 611–622 (1999)

33. He, K., Meeden, G.: Selecting the number of bins in a histogram: a decision theoretic approach. J. Stat. Plann. Infer. 61(1), 49–59 (1997)

34. Interconnection, PJM: Data miner 2, generation by fuel type. http://dataminer2.pjm.com/list. Accessed 30 August 2019

35. ISO New England: ISO-NE operations report: dispatch fuel mix. https://www.iso-ne.com/isosexpress/web/reports/operations/-/tree/gen-fuel-mix. Accessed 30 August 2019

36. State of the market report for PJM. Monitoring Analytics, LLC. (2016). https://www.monitoringanalytics.com/reports/PJM_State_of_the_Market/2016.shtml. Accessed 30 August 2019

37. Interconnection, PJM: PJM member list. https://www.pjm.com/about-pjm/member-services/member-list.aspx. Accessed 30 August 2019

38. Durvasula, V., Hansen, T.M.: “Market-Based Power System Test Cases Data”. Mendeley Data (2021), https://doi.org/10.17632/829dr326s.1

39. IEEE RTS Task Force of APM Subcommittee: IEEE reliability test system. IEEE PAS, 98(6), 2047–2054 (1979)

40. Grigg, C., 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. 14(3), 1010–1020 (1999)

41. Texas A and M University: Electric grid test case repository. https://electricgrids.engr.tamu.edu/electric-grid-test-cases/activeg500/. Accessed 30 August 2019

42. Interconnection, PJM: PJM environmental information services: public reports. https://www.pjm-eis.com/reports-and-events/public-reports.aspx. Accessed 30 August 2019

43. Overview of Greenhouse Gases (Washington, DC). https://www.epa.gov/ghgemissions/overview-greenhouse-gases. Accessed 30 August 2019

44. Power plant Carbon Dioxide Capture and Storage projects (Cambridge, MA). http://sequestration.mit.edu/tools/projects/index.capture.html. Accessed 22 November 2020

45. U. S. Energy Information and Administration, Coal (Washington, DC). https://www.eia.gov/coal/markets/. Accessed 30 August 2019

46. Table A1. Sulphur Dioxide Uncontrolled Emission Factors (Washington, DC). https://www.eia.gov/coal/electricity/annual/html/eapa_a_01.html. Accessed 30 August 2019

47. Table A2. Nitrogen Oxides Uncontrolled Emission Factors (Washington, DC). https://www.eia.gov/coal/electricity/annual/html/eapa_a_02.html. Accessed 30 August 2019

48. Reisman, J.L.: Air Emissions From Scrap Tire Combustion (Washington, DC). https://cfpub.epa.gov/si/si_public_record_Report.cfm?Lab=NRMRL&dirEntryID=115147. Accessed 22 November 2020

How to cite this article: Durvasulu, V., Hansen, T.M.: A data-driven approach to estimate emissions for market-based power system test cases. IET Smart Grid. 4(4), 429–444 (2021). https://doi.org/10.1049/stg2.12015

APPENDIX

Table 2 provides a detailed overview of the uncontrolled emission factors and control efficiencies of the corresponding fuel-turbine types. Generators used in the real system have different ages and different construction technologies, which changes the control efficiencies of each fuel and turbine type. In this appendix, we describe the particularities of each fuel type, and the assumptions made in determining the control efficiencies. For comparison purposes, the heat values of all

non-solid fuels have been converted to Btu/lb using the fuel mass densities under standard 1 bar pressure and 32°F temperature for gaseous fuels, and 1 bar/60°F for liquids [30]. The AP-42 data provides a detailed description of AP emission factors of most fuels [29, 30]. Because CO₂ is the single largest contributor (82%) to GHG emissions in the U.S., this study only considered CO₂ emissions for GHG [43]. There are control techniques to reduce CO₂, but they are not widely adopted (e.g. sequestration installed in only three power plants in the U.S. as of 2017), and the uncontrolled emissions were used in this study [44]. For a given fuel-type, the uncontrolled emissions are fixed as it depends on the chemical composition and combustion conditions of the fuel. The control efficiencies change as new control techniques are implemented to meet governmental regulations.

The chemical composition of coal in the U.S. differs depending on which coal basin was mined. This study considers generating plants from the PJM region, hence it is assumed that the coal is mined from the Appalachian and Illinois region. The heat value of the coal used for this study assumes 12,000 Btu/lb. for bituminous coal, and 9500 Btu/lb. for sub-bituminous coal based on the mean heat value of coal available from the Appalachian and Illinois Basin [45]. Even though the sulphur content of sub-bituminous is lower than bituminous coal, the uncontrolled emission of sub-bituminous and waste coal is higher than bituminous because of the lower heat content. In the EPA AP-42 data, the wet scrubber technology (80% efficiency) is mentioned as the typical technique for high thermal applications such as a power plants. Overfire air and low NOₓ burners are the most popular technologies (efficiency of about 60%) implemented to control nitrogen oxides.

The average heating value of natural gas in the U.S. varies between 950 and 1050 Btu/scf. This study uses the approximate heat value of 1020 Btu/scf, as provided in the EPA AP-42 data. Because natural gas has very low sulphur content, there is no dedicated SO₂ emissions reduction technique installed. Low NOₓ burners with an efficiency of about 26% are used to control nitrogen oxides emissions. As there was no turbine specific information available in the EPA AP-42 data, the same emission factors were assumed for all turbine types.

Blast furnace byproduct gas, which is collected from the steel/iron furnace top, contains CO and other particulates. Blast furnace gas is a low heating value gas, with just 92 Btu/scf [29]. Because of the high CO content and low heating value, the CO₂ emission factor is very high. The SO₂ and NOₓ depend on the steel/iron furnace from which the gas is extracted as there are a number of techniques in the metallurgy industry to reduce emissions. The AP emission factors provided in Table 2 are extracted from EIA uncontrolled emission factor database for the electricity sector [46, 47]. To the best of

| Table 2 | Detailed emissions factors and emission control efficiencies of all the major fossil fuels used under PJM |
|---|---|
| Fuel Class | Actual Fuel | Turbine | Heat Value (Btu/lb.) | θ CO₂ (lb./MMBtu) | θ SO₂ (lb./MMBtu) | η SO₂ % | θ NOₓ (lb./MMBtu) | η NOₓ % |
| Coal | Bituminous | ST | 12,000 | 210.6 | 1.62 | 80 | 0.5 | 60 |
| Refined coal | ST | 12,000 | 210.6 | 1.62 | 80 | 0.5 | 60 |
| Sub-bituminous | ST | 9500 | 214.2 | 1.94 | 80 | 0.41 | 60 |
| Waste coal | ST | 7300 | 250.6 | 1.94 | 80 | 0.41 | 60 |
| Gas | Natural gas | CC | 22,000 | 117 | 0.0006* | n.a | 0.186 | 25 |
| Natural gas | ST | 22,000 | 117 | 0.0006* | n.a | 0.186 | 25 |
| Natural gas | GT, IC | 22,000 | 117 | 0.0006* | n.a | 0.186 | 25 |
| Other gas | Blast furnace gas | ST | n.a | 604.7 | 0.006* | n.a | 0.16* | n.a |
| Landfill gas | IC | n.a | 117 | 0 | n.a | 0.04* | n.a |
| Biomass gas | IC | n.a | 117 | 0.028* | n.a | 0.15* | n.a |
| Propane gas | ST | 21,600 | 135.5 | 0.0002 | n.a | 0.01 | 60 |
| Multi-fuel | Municipal-waste (solid) | ST | 5100 | 200 | 0.025 | 80 | 0.49 | 90 |
| Waste wood (dry) | ST | 6900 | 206.8 | 0.025 | 80 | 0.22 | 60 |
| TDF | ST | 14,000 | 189.5 | 5.34 | n.a | 0.39 | n.a |
| Oil | Residual fuel oil | IC, GT, ST | 19,600 | 162.8 | 1 | 90 | 0.37 | 35 |
| Distilled-fuel | IC, GT | 18,900 | 163.1 | 1 | 90 | 0.15 | 60 |
| Kerosene | IC | 18,900 | 165.8 | 1 | 90 | 0.15 | 60 |

Please note that values with an asterisk are uncontrolled emission factors with either low emission factors where control techniques were not required, or there is not sufficient data to provide control efficiency.
the authors' knowledge, there are no emission control techniques information for this fuel.

Methane and CO are the primary constituents of landfill gas, and are produced by micro-organisms within the landfill under anaerobic conditions. The average heat content of landfill gas is about 485 Btu/scf. Only the controlled emission factor of the oxides of nitrogen was obtained from the EPA AP-42 data [30]. Biomass fuels are used with or without a fossil fuel. Biomass fuel can be produced by many sources, and the chemical composition depends on the source of biomass. For this study, plant-based biomass was used. The AP-42 data only provides a controlled emission factors for biomass fuel. Propane is a high energy density fuel with 2500 Btu/scf. The only controls developed for liquid propane gas combustion are to reduce NOx emissions using low NOx burners and flue gas recirculation.

Waste wood/residual wood heat values highly depend on the moisture content. The EPA AP-42 data provides the uncontrolled emissions in lb./MMBtu, so there was no need for converting the emission factors. The nitrogen oxides control efficiency depends on the ammonia to NOx ratio, and varies from 35% to 75% [30]. For this study, the mean value was used as the control efficiency. Liquid fuel oil has a very efficient SO2 reduction technique for all fuel and turbine types, hence for this study they have been approximated to be the same.

Tire-derived fuel (TDF) can be achieved in multiple combinations of percentage of tires and fossil fuel. The distribution of the emission factors for TDF is wide, and a single number cannot be used unless the actual details of all the generating units using this fuel are known. The AP emission factors in Table 2 for TDF are controlled emission factors from a real power plant. This data corresponds to facility 'D' in the EPA document on TDF [48]. We used the values from this facility because, with the best of the authors' knowledge, there was no other source to derive an industry standard emission factor for TDF as the composition of this type of fuel changes from plant to plant.

Oil-based fuels were the choice of peaking power plant until the 1973 oil crisis. Though not the major choice of fuel for larger plants, oil is still a popular choice of distributed generation in remote areas and some grid-connected generators. Oil has a high energy density of 150 MMBtu/10^3 gallons for residual fuel, and 140 MMBtu/10^3 gallons for distilled fuel and kerosene. Distilled fuel and kerosene are obtained by distillation and are volatile with lower ash and particulate pollutants compared to residual fuel oil. The CO2 emission is similar among all oil-based fuels. Wet scrubbing is cited as the most common control technique of reducing SO2, which is in the range of 80% to 98%. For this study, the mean of the range of efficiency is considered as 90%. There are different technologies for NOx reduction specified in the EPA AP-42 data [30]. The control efficiencies specified in Table 2 are based on the mean of flue gas recirculation and staged combustion efficiencies specified in the AP-42 data sheet.