The Creation and Validation of Load Time Series for Synthetic Electric Power Systems

Hanyue Li, *Student Member, IEEE,* Ju Hee Yeo, *Student Member, IEEE,*
Ashly L.Bornsheuer, *Student Member, IEEE,* Thomas J. Overbye, *Fellow, IEEE*

***Abstract*—Synthetic power systems that imitate functional and statistical characteristics of the actual grid have been developed to promote researchers' access to public system models. Developing time series to represent different operating conditions of these synthetic systems will expand the potential of synthetic power systems applications. This paper proposes a methodology to create synthetic time series of bus-level load using publicly available data. Comprehensive validation metrics are provided to assure that the quality of synthetic time series data is sufficiently realistic. This paper also includes an example application in which the methodology is used to construct load scenarios for a 10,000-bus synthetic case.**

***Index Terms*—Synthetic time series, residential commercial and industrial load, synthetic power systems**

I. Introduction

PBULIC access to real power system data is limited due to confidentiality concerns. Synthetic power system models and data are created to be functionally and statistically similar to real power systems. Synthetic systems are synthesized using public data of the actual grid, and they don't represent the actual system located on the same geographic footprint, or contain any confidential information about the actual grid.

Many efforts have been made on the creation of synthetic power system base cases, which contains systems topology and many of them have AC or DC power flow solutions. Early work of [1]-[2] came up with an approach to create transmission grid topologies based on the small world graph network. A methodology for generating large scale synthetic transmission systems with AC power flow solutions on the footprint of United States was proposed in [3]-[4], and several synthetic systems of different sizes and footprints were created. The work of [5] investigated the geographic and structural properties of North American and Mexican transmission grids, and large electric systems with synthetic nodes and node connections were created in [6]-[7]. European synthetic transmission grids with DC power flow solutions were also developed based on public information from utilities and regulatory agencies in the synchronous grid of Continental Europe (UCTE) [8].

Since synthetic power system base cases only reflect a one-time snapshot of the system, there is a natural need to expand the work and develop time series to represent changing system operating conditions over time. The combined data set of synthetic grid models and power system time series can be used as a benchmark for system scenario studies, a test bed for algorithms such as time sequence power flow, optimal power flow, and unit commitment.

Power system time series consist of data relevant to system status in a time sequence manner. It can span many aspects of the system such as the load, transmission line status, generator dispatch and electricity price. Real time series power system data are generally more publicly accessible compared to the data of actual grid topologies and models. For example, the Open Power System Data, for example, is a data depository that has load, wind, solar and price data in hourly resolution of 37 European countries [9]. In North America, for the transparent operation of electricity market, Independent System Operators (ISOs) often have hourly resolution load and price time series data publicly available as well [10]-[11].

Besides actual system-level data, the creations of synthetic time series for household-level load and renewable generation are also well-researched topics. The work of [12] developed a probabilistic mathematical model for residential load simulation. A top-down approach using domestic load patterns for household profile adaptation was implemented in [13]. A machine learning method to generate synthetic residential building load time series from smart meter data was proposed in [14]. The work of [15] developed an approach of generating renewable scenarios based on the machine learning concept of generative adversarial networks.

Developing time series based models from historic data is also a commonly used approach for synthetic data generation. For example, an autoaggressive integrated moving average (ARIMA) model was proposed in [16] to simulate the stochastic wind power generation time series while taking nonstationarity and physical limits into account. The work of [17] introduced a time-dependent, autoaggressive, Gaussian model to generate synthetic hourly solar generation time series. Periodic autoaggressive moving average (PARMA) models are also commonly used to construct the seasonal patterns of hydro power related time series [18], [19].

However, the techniques to develop end-use customer load and renewable generation time series are not easily transferable to create bus-level load time series. While time series data of household electricity consumption and renewable energy generation are widely available to the public, both customer-level load time series from other energy sectors such as commercial and industrial, as well as aggregated bus-level load time series are hard to obtain in large volume. Due to the lack

The authors are with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, 77840 USA (e-mail:hanyueli@tamu.edu, yeochee26@tamu.edu, abornsheuer@tamu.edu, overbye@tamu.edu).

Accepted for Publication in IEEE Transactions on Power Systems, 2020.

of training data, it is hard to directly apply machine learning or time series models for the creation of synthetic bus-level load time series.

This paper presents a bottom-up methodology for synthesizing bus-level time series load data. Building on the results of [20], the integration process is generalized and also improved by taking aggregation effects into consideration. The maximum value of load time series aims to match the corresponding load bus size determined in the above mentioned base case. The unique variation of each bus-level time series is a result of the heuristic aggregation of prototypical building and facility load time series. To ensure the quality and realism of the synthetic load time series, comprehensive validation metrics are established from the actual system-level load time series. It is important to note that the time series created in this paper is generalized and represents the bus-level load in a typical weather year. More customized scenarios incorporating other inputs such as temperature and irradiance can be created by conditioning the base load time series. An example that uses the load time series to create a high behind-the-meter solar installation scenario is shown in the application section.

The creation and validation of load time series for 2,000 and 10,000- bus synthetic grids (i.e., the ACTIVSg2000 and ACTIVSg10K grids from [21]) are used as examples. The ACTIVSg2000 synthetic system shares the same footprint as the Electric Reliability Council of Texas (ERCOT), and the ACTIVSg10K has the same service region as the Western Interconnection (WI). However, the method of creating and validating synthetic time series is general enough to be applied to any system.

II. Creation of Bus-level Load Time Series

Electric load time series reflect electricity consumption patterns and provides insight on the absolute level and changing rate of load at different times. Having access to bus-level load time series is essential for the unit dispatch and commitment in power system operations since generation always needs to follow the time-varying system load. The load time series in synthetic power systems has hourly resolution with a duration of a year, and is created on the bus level so that every bus in the synthetic grid model has a unique profile. Each bus-level load time series is created using an iterative aggregation approach, where prototypical building load profiles are aggregated based on the size and composition of load buses.

A. Location and size of bus-level electric load

The location and size of the electric loads are determined during the creation of the synthetic base case discussed in [3]-[4]. The load buses are located based on the clustering of geographic coordinates associated with postal codes that are obtained from the public U.S. census database. The size of each load bus is then scaled according to the population of the corresponding postal code and the per-capita MW consumption. Based on the statistical analysis of the actual grid, a unique power factor is assigned to each bus in the base case to calculate the reactive load [4].

The base case is used as a reference to create load time series. The size of load buses in the base case are considered to be the peak value of each bus-level time series, and the geographic coordinates assigned to each load bus are then used to determine the unique location-dependent load features such as load composition ratio and prototypical building load time series.

B. Load bus composition ratio

The assignment of a composition ratio of residential, commercial and industrial load on each bus is helpful to realistically represent the uniqueness of load. It establishes the geographic and demographic dependence of electric load similar to reality.

U.S. utility companies' service territories as well as their residential, commercial and industrial megawatt-hours sales values from the Annual Electric Power Industry Report are used to determine the bus load composition ratios [22]. Each load bus is assigned to one utility company based off its geographic coordinates, and the company's sales ratio of the three load types is used as the average bus load composition ratio.

C. Prototypical building- and facility-level load time series

To bridge between the bus load composition ratio, and a unique hourly profile, prototypical end user level load time series under residential, commercial and industrial load types are synthesized from public data. Building- and facility-level time series gives the desired bus load a good base to incorporate both individual user load patterns and the aggregation effect. Different categories of buildings and facilities and their prototypical load time series are realistic approximations to represent the most common and important load features.

1) Prototypical residential and commercial building load: The prototypical building load time series synthesized in this paper are the same as the ones developed in [20], where open source data of simulated hourly residential and commercial building energy consumption are used [23].

The residential data contains buildings' hourly electricity usage value from space heating/cooling, High Voltage Alternating Current (HVAC) fan, interior/exterior lighting, as well as appliances and miscellaneous loads. Each data file covers one typical meteorological year 3 (TMY3) location in the United States, which represents geographic locations with different meteorology[24]. For commercial load, under each TMY3 location, 16 building electric load profiles are simulated using the Department of Energy (DOE) commercial reference building models[25], and the contents in each time series data are like that of residential data set.

Under the United States footprint, 1020 residential and 16,320 commercial building time series are calculated as a summation of all electricity consumption categories under each building type. They are created to host various features such as the unique profiles of residential and commercial buildings, and electricity consuming variations over time and geographic location.

Figure 1, for example, shows prototypical residential building load time series in a winter week and a summer week. The load shapes in two seasons are distinguishable, where winter profiles tend to have two peaks in one day due to winter heating, while summer profiles only have one peak per day. The magnitude of load can also be very different in each season, depending on geographic locations. In winter, regions with colder climate such as Helena, Montana, would have higher average load. While in summer, load within hot and arid climate zones, such as Phoenix,Arizona, will have much more electricity consumption.

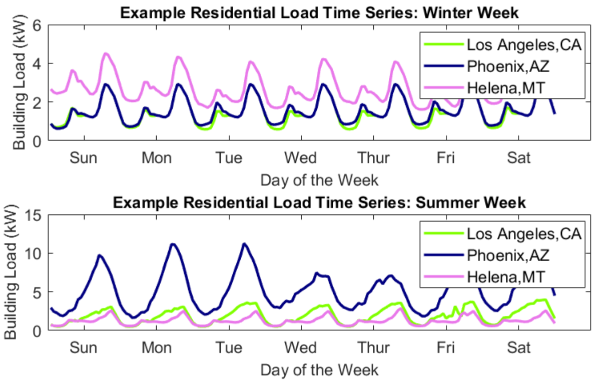

Fig. 1. Prototypical residential building load time series examples by location

Similarly, figure 2 shows prototypical commercial load time series for the large office building type. The load shape of a specific building type is generally consistent regardless of the location, and the load level is slightly higher in summer season compared to that in winter. In figure 3, weekly load profiles of three commercial building types (full-service restaurant, small office, and strip mall) in Los Angeles, California are shown. The load shape and size under each building type is unique. Small offices have steady load during weekdays and low load during weekends. For full-service restaurants and trip malls, load levels are constant through out the week, while full-service restaurants observe two peaks near lunch and dinner time, strip malls only peak once every day.

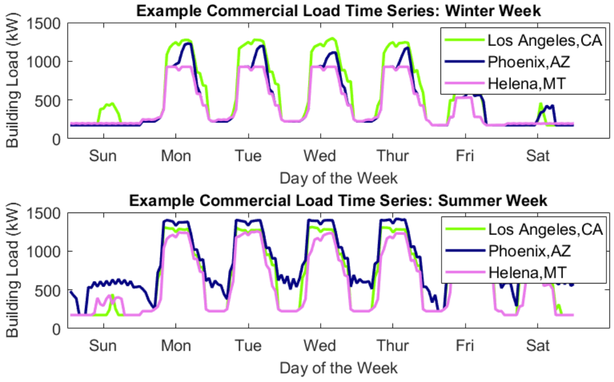

Fig. 2. Prototypical commercial building load time series examples by location

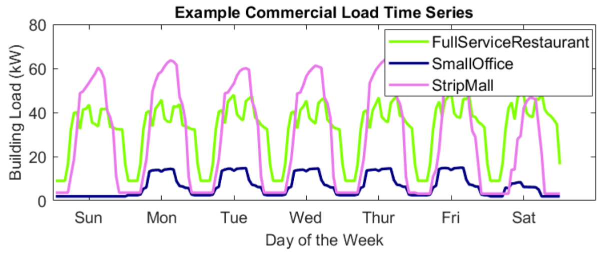

Fig. 3. Prototypical commercial building load time series examples by type

2) Prototypical industrial facility load: Prototypical industrial facility load time series are created based on publicly available per-unit industrial load curves from Oak Ridge National Laboratory (ORNL) [26] and the industrial assessment data in Industrial Assessment Centers (IAC) Database [27].

The ORNL per-unit curves provide daily profiles of different industrial sectors, presented by different Standard Industrial Classification (SIC) codes with their unique load factor. The IAC Database contains information on the industry SIC code, total electricity usage and yearly operating hours of over 14,000 facilities in the United States, which are used to modify the ORNL curves into facility-specific load time series for a year.

For each industrial facility, the yearly operating hour is first used to determine the total number of operating days. The ORNL daily curves of the same SIC code is then expanded to a yearly load curve, with small white noise imposed and a random selection of starting day of the year. The synthesized yearly curve is then scaled so that the integral value of the curve matches the total electricity usage.

Figure 4 presents prototypical industrial load time series for four facilities from food, petroleum and refining, primary metal, as well as electronic and electrical equipment industries. As those load curves are adopted from the ORNL per-unit daily curves, they have similar daily variations and weekly shapes, with different load levels and load factors.

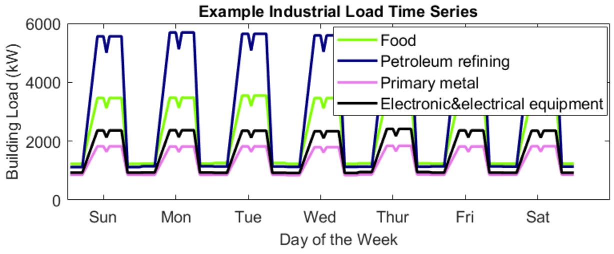

Fig. 4. Prototypical industrial facility load time series examples by type

D. Aggregation of load

The bus-level load time series is created by iteratively aggregating prototypical building and facility load time series of each load type. This aggregation process has three main aspects: integrating realistic amounts of end users under each load type, selecting representative prototypical time series, and mimicking the effect of load aggregation described in [28].

A flow chart of this aggregation process is shown in figure 5. The reference peak values of residential, commercial and industrial bus load type of each bus are first determined by the multiplication of bus load size and the load composition ratio. This is used to integrate realistic amount of end users under each load type, where the peak component values are the stopping criterion for the iterative aggregation process.

A pool of representative prototypical building load time series are then selected for each load bus. For residential and commercial load, the selection used the top five shortest distances between the load bus geographic coordinates and TMY3 locations. All industrial facility load time series that have smaller maximum value than the calculated peak industrial load component are included in this pool since industrial loads are less correlated with geographic locations.

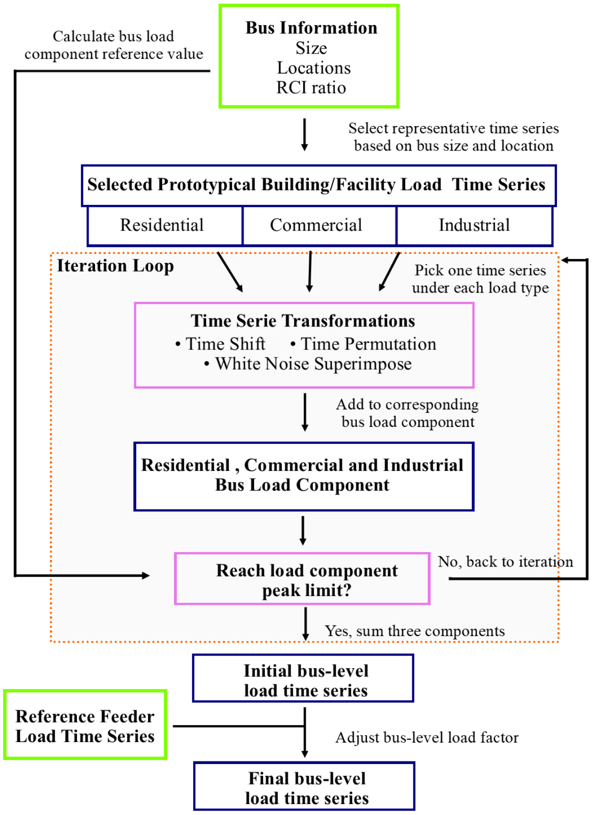
Calculate bus load component reference value
Bus Information
Size
Locations
RCI ratio
Select representative time series based on bus size and location
Selected Prototypical Building/Facility Load Time Series
Residential
Commercial
Industrial
Iteration Loop
Pick one time series under each load type
Time Serie Transformations
• Time Shift • Time Permutation
• White Noise Superimpose
Add to corresponding bus load component
Residential , Commercial and Industrial Bus Load Component
Reach load component peak limit?
No, back to iteration
Yes, sum three components
Initial bus-level load time series
Reference Feeder Load Time Series
Adjust bus-level load factor
Final bus-level load time series

Fig. 5. Flow chart of heuristic load aggregation approach

Under each load type, within one iteration, only one building or facility load time series is picked based on a predetermined probabilistic distribution. In the work of [29], a combined transmission and distribution synthetic data set is created using commercially obtained parcel data as one of the inputs, where different building types within the geographic footprint of the test case are extrapolated based on the parcel usage categories. This publicly available data set [30] is utilized to summarize the typical percentage of building types within the service territory of a transmission substation or bus. The region of a transmission substation is defined as the polygon boundary of all the distribution feeders that are serviced by this substation. This selected building- or facility-level load time series is then processed through three types of transformations: time shift, time permutation, and noise insertion. Those transformations diversify the load profiles of end users, so that the smoothing effect for load aggregation can be produced.

The original prototypical load time series can be shifted both forward and backward up to 12 hours following predefined probability mass functions, where the time to be shifted is a discrete integer variable. For residential load class, the distribution of time shift is summarized from publicly available household metering data [31]. The Pecan Street electricity consumption data has 15-minute resolution, and is collected from residential homes mostly located in the state of Texas, California and New York. To be consistent with the prototypical load time series, the metered load time series is downsampled to hourly granularity. Detrended cross-correlation analysis is then conducted for the non-stationary residential load time series [32], where the time lag yields to the peak cross-correlation is considered to be the hour shifted in between two time series. The distribution of shifting hours is then summarized among all the time series pairs as probability mass functions. Due to the lack of reference data in commercial and industrial load classes, heuristically, we assume the probabilities of shifting the time series of those two load classes to be 30% and 50% lower than the residential load class.

To imitate the random surges or drops of load for individual customers, certain hours (100, 100 and 50 hour pairs for residential, commercial and industrial respectively) are randomly chosen within the year to be permutated. As the prototypical load time series used as the input to this process is simulated data, it reflects an expected level of electricity consumption every hour, but does not account for the stochastic behavior of electricity users. For example, figure 6 shows a comparison between the prototypical and actual residential load time series from the same geographic region. The upper plot in figure 6 is the simulated data used to create bus-level load time series in Austin, TX. The lower plot is the one residential electricity consumption measurement from the same city [31]. It is observed that while the two time series are on the same load level, and share similar daily trend, the actual load time series exhibits more jitters than the simulated data.

To introduce the stochastic behaviour back into the simulated data, and avoid bus load time series being overly conforming due to the use of similar prototypical building or facility time series, a small noise is also imposed. Since the prototypical building and facility load time series already included the seasonal, weekly and daily variations, a Gaussian noise is added to the base loading to not introduce seasonality [33]. This transformed time series is then added to the corresponding type of load component, and the iteration would stop once after the load component maximum value calculated in previous step has been reached.

To mimic the effect of increasing load factor as load aggregates to a higher level [28], public feeder load time series

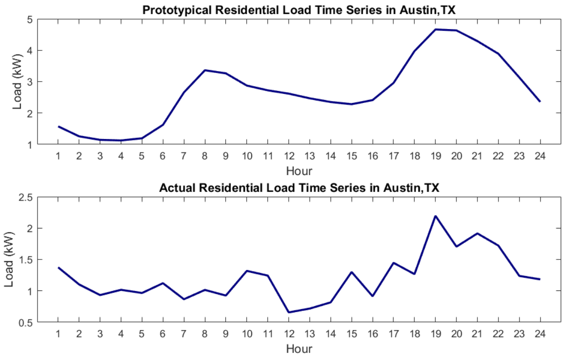

Fig. 6. Simulated and Actual Residential Load Time Series

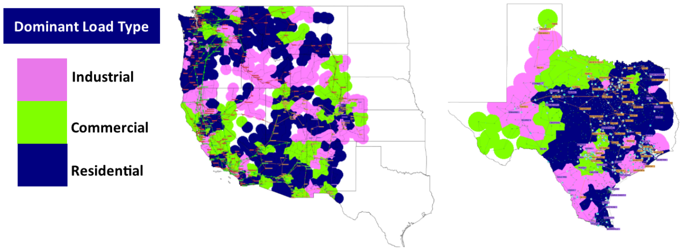

Fig. 7. Dominant load type contour for ACTIVSg synthetic systems

managed by National Renewable Energy Laboratory (NREL) [34] is used to adjust the bus-level load factors to a realistic range. This distribution feeder load time series is populated from the taxonomy distribution feeders of different geographic regions.

The geographic coordinates of each load bus in the synthetic system are used to randomly select a subset of taxonomy feeders from the same geographic region, so that the summation of feeder load time series is on the same scale as the bus load. The load factor of the aggregated feeder load time series is calculated to be the reference value. A constant component is added to the created bus-level load time series to adjust its load factor to a realistic value according to equation (1).

$$\frac{Constant + Average\ Load}{Constant + Max\ Load} = Reference\ Load\ Factor \tag{1}$$

III. Example Results

The load time series created for ACTIVSg2000 and ACTIVSg10K synthetic power systems are presented in this paper. There are 1125 and 4170 load buses in those two cases, with 71 and 132 GW of system peak load respectively. While statistical validation of the load time series are discussed in details in Section IV of this paper, figure 8, 9 and 10 provide an overview of load profiles on both individual bus level and aggregated system level.

The plots in figure 7 show the dominant bus load type using the method discussed in section II.B, where the load type with highest composition percentage is considered to be the dominant type of its load bus. In ACTIVSg2000 system, 67.4% of the buses are dominated by residential load, 18.8% are primary composed of commercial load, and 13.8% for industrial load. For the ACTIVSg10K system, 45.4%, 33.7% and 20.9% of buses have residential, commercial and industrial load as primary composition respectively.

On the bus-level, each load time series is unique based off the location and load composition ratio of the load bus. Average bus-level load time series of different dominant load types are shown in figure 8. Residential-dominated bus load time series exhibit noticeable seasonal differences, where the electricity consumption in summer and winter seasons tend to have higher average values as well as higher variations. Commercial-dominated bus load time series have distinct daily patterns, while the electricity consumption base line stays relatively constant throughout the year. Industrial-dominated bus loads usually have the lowest variation and highest load factor. The average size of industrially-dominated buses are larger than the other two types.

The system-level synthetic load time series and the actual system load from their footprint regions are shown in figure 9 and figure 10. Although duplicating system-level load time series is not the desired outcome, synthetic load time series on the system level should exhibit similar general load shapes and trends compared to the actual system.

Figure 9 and figure 10 show that the synthetic loads share similar size with the load of the actual system in the corresponding service territory. ACTIVSg2000 synthetic system has 71.1 GW of peak load, and 48.7 GW of average load, and the actual load of ERCOT system has 71.2 GW of maximum load and 41.0 GW of average load. For ACTIVSg10K system, there is 132.5 GW of peak load and 88.1 GW of average load. The corresponding actual system, the United States Western Interconnection, has 136.2 GW of peak load and 83.8 GW of average load.

Daily and weekly patterns can be seen from the ACTIVSg2000 and ACTIVSg10K synthetic load time series. It is also observed that the synthetic system has similar seasonal trends compared to the actual system. ACTIVSg2000 and ERCOT systems both experience peak load in summer, and also have some high load days weeks in winter. ACTIVSg10K and WI system also peak in summer, while their profile in the winter season is flatter.

IV. Validation of Synthetic Time Series

Since synthetic time series are fictitious, the validation of the created data against the actual data is critical to determine the quality and realism of the time series. A statistical based validation approach is implemented in this paper, as synthetic time series aim to realistically represent behaviors of load over time, instead of being an exact duplicate or forecast of the actual system time series.

A comprehensive set of validation metrics enables researchers to use synthetic time series with ease, but at the same time to be aware of the underlying assumptions.

Validation metrics that are generic and independent from geographic locations are summarized using statistical characteristics found in public load data of 37 European countries

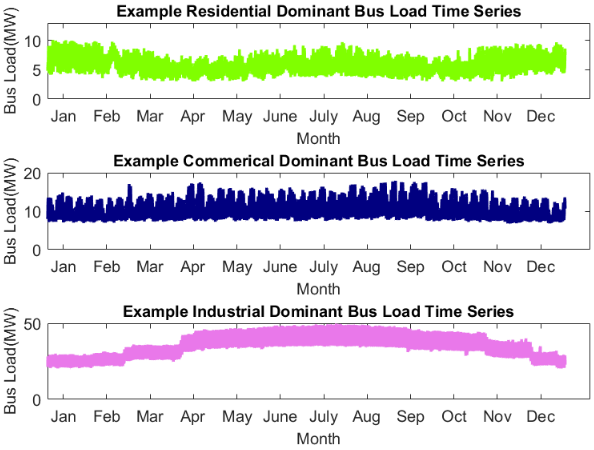

Fig. 8. Bus-level synthetic load time series of different dominant load type

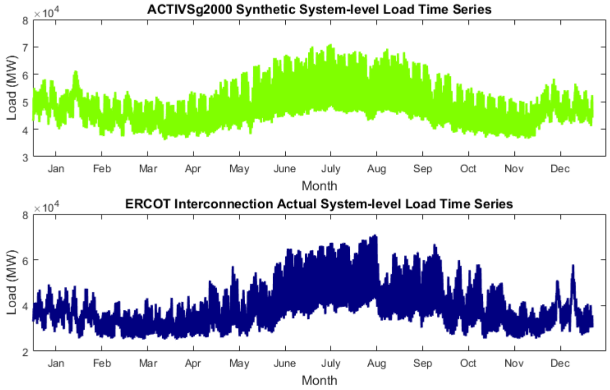

Fig. 9. ACTIVSg2000 synthetic system load V.S. ERCOT Interconnection system Load

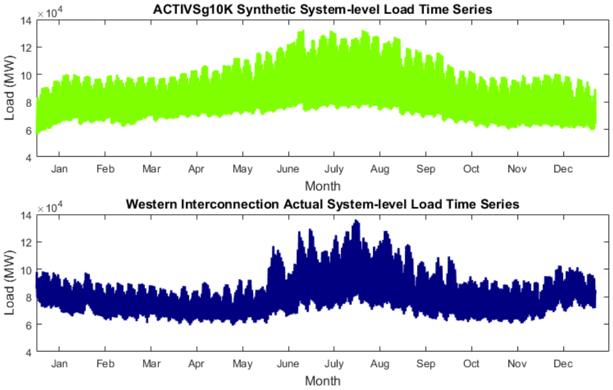

Fig. 10. ACTIVSg10K synthetic system load V.S. Western Interconnection system load

[9] and 66 United States Balancing Authorities [35], so that synthetic load time series without a geographic footprint or have no availability of actual load time series can also be validated.

It is important to note that the validation of synthetic load time series is only conducted on the aggregated system level due to the lack of available real data on bus-level load time series. For an unbiased validation, aggregated reference data used during the construction process and the real data used for the validation process are independent and kept separate.

A. Load factors

Load factor is defined as the ratio of average and peak value of a load time series. It is one effective metric to quantitatively validate the overall shape of the synthetic load profile. For profiles with relatively constant load level, such as regions with a high industrial composition, load factors are usually higher; while heavily residential areas tend to have lower load factors due to light occupation during the day [36].

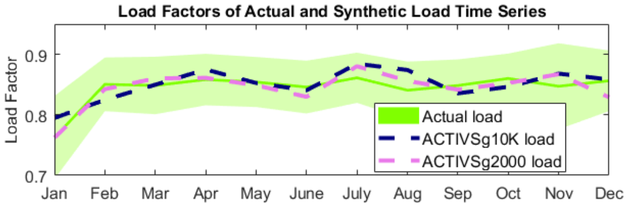

Fig. 11. Monthly load factors of actual and synthetic load time series

The range of load factors of each month is summarized from public load data mentioned above and shown in figure 11 as the green shaded region. It is observed that as a general trend, the value of load factors are slightly higher in summer months, due to the increase of base electricity consumption from spacing cooling. There is also a consistent difference between the lowest and highest load factors of actual load every month, where systems with smaller size and less industrial load often have lower load factors.

The load factors of ACTIVSg10K and ACTIVSg2000 load time series lie inside the range observed from actual load time series, and also follow the same monthly trend.

B. Load distribution curves

Load distribution curves show the percentage of time that load is at different levels relative to its mean value. The load time series is normalized based off its mean value, where load levels exceeding yearly average would have per unit values larger than one. The vertical axis of a load distribution curve is the percentage of time points.

The green shaded band in Figure 12 shows the range of load distribution curves found in real load time series, where load levels are scattered in between 0.4 and 1.8 per unit, with a denser distribution in the range from 0.8 to 1.2. The distribution of ACTIVSg10K and ACTIVSg2000 load time series follow the same general trend, load at most of the time points are within 0.8 to 1.2 times its yearly average.

C. Autocorrelations

Autocorrelation exhibits the relationship between time points of load time series that are certain time lags apart. It provides a validation perspective in time sequence order, instead of observing time series values as if they are independent

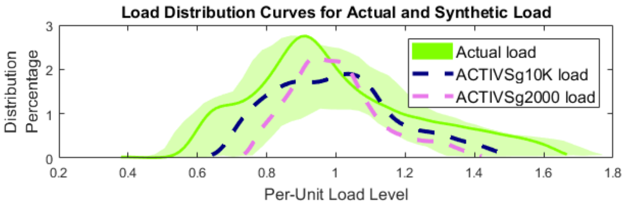

Fig. 12. Load distribution curve validation

recordings. Since the load time series data is not stationary, where the average value is not constant,and the variance grows with the level of the time series, the log and differencing transformation are used to stabilize and remove the mean trend from the original time series.

Figure 13 shows the autocorrelations of actual and ACTIVSg synthetic load time series for time lags up to 48 steps. According to the real load time series data, the autocorrelation plot appears to be periodic with a 24-hour cycle, with its magnitude slightly decreasing every cycle. All the load time series autocorrelation exhibit a similar trend, within each cycle, the autocorrelation drops from 1 to below 0 and then increases from negative correlation back to almost unity correlation by the end of the cycle. It is also interesting to note that during the middle of each cycle, around 12 hour time lag, the autocorrelation of the differenced, logged load time series has a local maximum. The plots of synthetic load's autocorrelation lie within the upper and lower bound established by the autocorrelation of actual load time series.

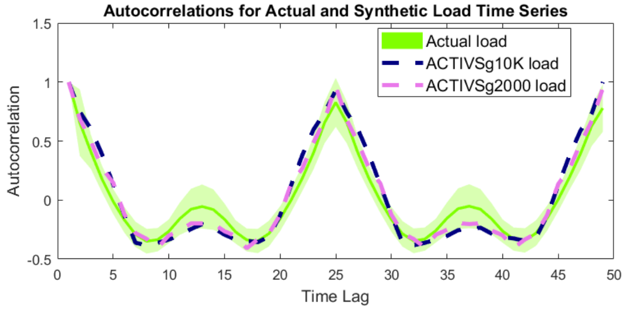

Fig. 13. Load autocorrelation validation

D. Power spectral density

Power spectral density measures the the distribution of power content versus frequency of a time series. It is a technique that enables us to discover underlying periodic behaviors. The spectral density can be estimated using periodogram, which establishes the squared correlation between the targeted time series and sinusoidal waves at different frequencies spanned by the time series. Similar to the autocorrelation analysis, the log and differencing transformation are used to stabilize and remove the mean trend from the original time series.

Figure 14 shows the power spectral density of the actual system-level and synthetic system-level load time series. The horizontal axis of this figure is the frequency, which presents periodic behaviors with longest period of every year, and shortest period of every 10 hours. The reference range of power spectral density of at each frequency is summarized from the public system load data, shown in the green shaded region in Figure 14. It is observed that the power spectral density of ACTIVSg2000 and ACTIVSg10K synthetic load generally lie in the defined upper and lower bounds. The power spectral density of the synthetic loads also exhibit distinctive sharp peaks around the same frequencies as those of the actual load time series, where the highest spike occurs at the daily 24-hour period with frequency equals to 1/(24 hour) = 0.0417/hour. Power spectral density spikes are also observed around periods such as half-day (12 hours), half-week (84 hours),one-week (168 hours), three-months (2160 - 2190 hours) and six-months (4320 - 4380 hours) .

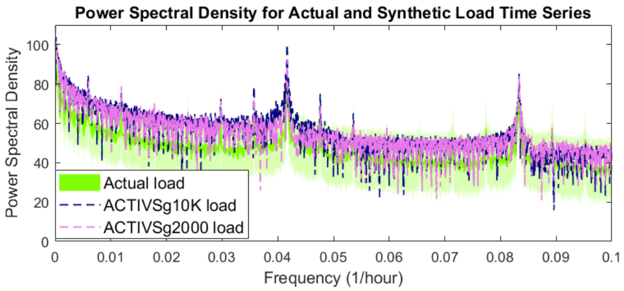

Fig. 14. Load power spectral density validation

V. Synthetic Time Series Applications Example

The creation of synthetic time series enables the potential of scenario development, time sequence simulations, and wide-area visualizations of large-scale power systems. Those time series data reflect the typical behavior of electric load in a hourly manner for the whole year, which is fundamental for power system steady-state analysis desired at different times and of different durations. The buses each load time series locates also cover a wide range of North America regions and have longitudes and latitudes. This allows non-uniform alterations of the time series to construct realistic power system scenarios. The alterations can be made considering the coupling of power system and location- based factors such as weather and major events. Along with transmission system models, generator cost functions, and other system data, synthetic load time series can facilitate power flow, optimal power flow analysis, as well as unit commitment of various scenarios.

As an example, this paper utilizes the bus-level load time series in the ACTIVSg10K system as the benchmark and develops a high behind-the-meter (BTM) solar scenario in the Western United States region. The BTM solar energies are solar generating units on the consumer's side of the retail meter that serve all or part of the customer's retail load with electric energy [37]. They are often treated as ”negative loads” to be subtracted from the total load at the demand side. The increasing capacity of BTM solar installation changes the shape of daily net load on the system-level, when solar generation peaks at midday, the net load is low and when solar generation trails off at the end of the day, the total demand ramps quickly upward [38]. This new load shape is often referred as a ”duck curve”. It has raised concerns on a conventional power system’s ability to accommodate the ramp rate and range needed to effectively supply the load and fully utilize the renewable energy [39].

The bus-level load time series in the synthetic power system is utilized to construct rare and extreme scenarios that can be useful to study this impact over time. Based off the composition of bus load type and the location-based solar potentials, the benchmark load time series at each bus is altered so that a system-level ”duck curve” is created for the ACTIVSg10K system. This ”duck curve” scenario can be used as the input of power flow analysis and unit commitment to provide analytical insights on transmission line loadings, generator dispatches schedules, system costs, and other system conditions.

This example uses a 24-hour time period in late spring from the benchmark hourly time series to develop a daily duck curve since such scenario usually occurs during the spring and summer seasons [38]. The BTM solar generation capacity is set to be 30,000 MW in the ACTIVSg10K system, and is distributed among load buses weighting their load sizes and the documented average solar resource outputs. Since most BTM solar installations are in the non-industrial sectors, only the size of residential and commercial load on each bus are considered to calculate the weights indicating bus solar potential, where buses with higher combined load are assigned with a higher peak BTM solar generation. On the other hand, as the solar potential is also dependent on solar radiance that varies with geographic locations, a solar resource data set from National Renewable Energy Laboratory (NREL) is also utilized to determine the weight to distribute the system BTM solar capacity [40]. This data set provides the monthly average solar output $(kWh/m^2/day)$ at each zip code location.

The solar potential that determines the BTM solar capacity at each bus is calculated as the weighted summation of normalized load size and normalized solar resource output in equation (2) - (3). To construct the 24-hour ”duck curve” scenario in the ACTIVSg10k synthetic system, a BTM solar output time series is created and subtracted from the original load time series of each bus. The bus-level BTM solar capacity is considered as the peak value of each BTM solar output time series that would occur at a random time step in between 1 pm and 3 pm. The starting and ending time step of solar output are randomly chosen from 6 to 8 am, and 6 to 8 pm respectively. Before the starting and after the ending time point, the BTM solar output is zero.

$$\begin{aligned} BTM\ solar\ potential(i) = {} & X_1 \frac{load\ size(i)}{max(load\ size)} \\ & + X_2 \frac{solar\ resource(i)}{max(solar\ resource)} \end{aligned} \tag{2}$$

$$\begin{aligned} BTM\ solar\ capacity(i) = {} & system\ BTM\ solar\ capacity \times \\ & \frac{BTM\ solar\ potential(i)}{\sum_{i=1}^{Nload} BTM\ solar\ potential(i)} \end{aligned} \tag{3}$$

where:

X_1, X_2 weights on bus load size and solar resource

The load size and solar resources data, as well as the BTM solar capacity determined for buses in ACTIVSg10K system are shown in the contour plots Figure 15. It can be observed that most of the buses with higher BTM solar capacity are locations with high solar resources. Besides, major metropolitan areas with dense residential and commercial demands also have higher BTM solar capacities. The benchmark and duck curve load time series on the system-level are shown in Figure 16.

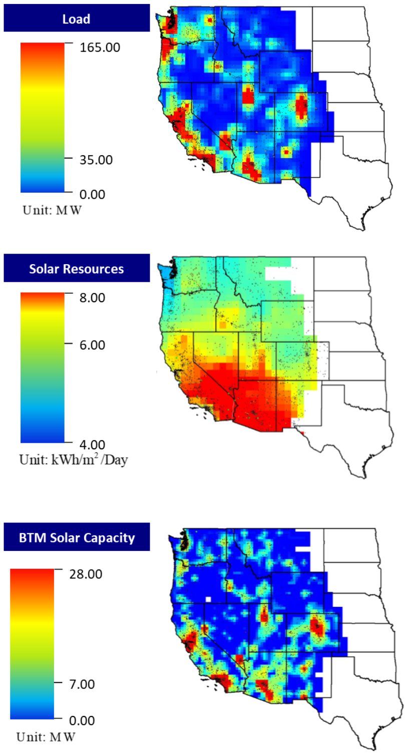

Fig. 15. Contours of bus load, average solar resource, and BTM solar capacity in ACTIVSg10K synthetic system

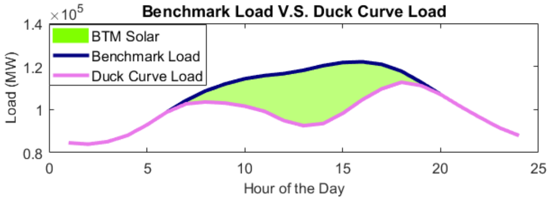

Fig. 16. Benchmark load and duck curve load for ACTIVSg10K synthetic system

VI. Conclusions

This paper proposed a methodology to synthesize and validate bus-level load time series in the existing synthetic power systems. The creation of time series uses an iterative bottom-up approach. Based on the geographic location and load type composition of each bus, prototypical building and facility level time series are integrated to construct a bus-level time series with unique profiles. Each time series has hourly resolution, and spans for a year. To validate and improve the realism and quality of synthetic load time series, actual load time series obtained from electric systems of different sizes are analyzed statically so that representative and comprehensive set of validation metrics can be developed.

Since the data set utilized in the synthesizing process is publicly available, the created time series can be accessed and distributed freely without any confidentiality concerns. The wide geographic coverage, time resolution and duration of bus-level time series enable its versatile applications in system scenario development and studies. As an example, this paper demonstrated the construction of a "duck curve" scenario in ACTIVSg10K system using the bus-level load time series as the benchmark.

Acknowledgment

This work was supported in part by the Department of Energy Advanced Research Projects Agency-Energy (ARPA-E) under the GRID DATA program, and in part by the National Science Foundation under Award NSF Grant 1916142.

References

[1] Z. Wang, A. Scaglione, and R. J. Thomas, "Generating statistically correct random topologies for testing smart grid communication and control networks," *IEEE Trans. Smart Grid*, vol. 1, no. 2, pp. 28–39, 2010.

[2] ——, "The node degree distribution in power grid and its topology robustness under random and selective node removals," in *Proc. 2010 IEEE Int. Conf. Commun. Workshops*, 2010, pp. 1–5.

[3] K. M. Gegner, A. B. Birchfield, T. Xu, K. S. Shetye, and T. J. Overbye, "A methodology for the creation of geographically realistic synthetic power flow models," in *2016 IEEE Power and Energy Conference at Illinois (PECI)*, 2016.

[4] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye, "Grid structural characteristics as validation criteria for synthetic networks," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3258–3265, 2014.

[5] D. Deka, S. Vishwanath, and R. Baldick, "Analytical models for power networks: The case of the western us and ercot grids," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2794–2802, 2016.

[6] S. Soltan and G. Zussman, "Generation of synthetic spatially embedded power grid networks," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, 2016.

[7] S. Soltan, A. Loh, and G. Zussman, "A learning-based method for generating synthetic power grids," *IEEE Systems Journal*, pp. 1–10, 2018.

[8] N. Hutcheon and J. W. Bialek, "Updated and validated power flow model of the main continental european transmission network," in *Proc. IEEE Power Tech Conference*, 2013, pp. 1–5.

[9] "Open Power System Data." [Online]. Available: https://open-power-system-data.org/

[10] "ISO New England Express." [Online]. Available: https://www.iso-ne.com/markets-operations/iso-express/

[11] "Independent Electricity System Operator Power Data." [Online]. Available: http://www.ieso.ca/power-data/

[12] J. V. Paatero and P. D. Lund, "A model for generating household electricity load profiles," *International Journal of Energy Research*, 2006.

[13] C. Bucher and G. Andersson, "Generation of domestic load profiles - an adaptive top-down approach," in *Proc. of PMAPS 2012*, 2012.

[14] T. Zufferey, D. Toffanin, D. Toprak, A. Ulbig, and G. Hug, "Generating stochastic residential load profiles from smart meter data for an optimal power matching at an aggregate level," in *2018 Power Systems Computation Conference (PSCC)*, 2018.

[15] Y. Chen, Y. Wang, D. Kirschen, and B. Zhang, "Model-free renewable scenario generation using generative adversarial networks," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 3265–3275, 2018.

[16] P. Chen, T. Pedersen, B. Bak-Jensen, and Z. Chen, "Arima-based time series model of stochastic wind power generation," *IEEE transactions on power systems*, vol. 25, no. 2, pp. 667–676, 2009.

[17] R. Aguiar and M. Collares-Pereira, "Tag: a time-dependent, autoregressive, gaussian model for generating synthetic hourly radiation," *Solar energy*, vol. 49, no. 3, pp. 167–174, 1992.

[18] J. Obeysekera and J. Salas, "Modeling of aggregated hydrologic time series," *Journal of Hydrology*, vol. 86, no. 3-4, pp. 197–219, 1986.

[19] J. D. Salas and J. Obeysekera, "Conceptual basis of seasonal streamflow time series models," *Journal of Hydraulic Engineering*, vol. 118, no. 8, pp. 1186–1194, 1992.

[20] H. Li, A. L. Bornsheuer, T. Xu, A. B. Birchfield, and T. J. Overbye, "Load modeling in synthetic electric grids," in *2018 IEEE Texas Power and Energy Conference (TPEC)*, 2018.

[21] "Electric Grid Test Case Repository." [Online]. Available: https://electricgrids.engr.tamu.edu/

[22] "Annual Electric Power Industry Report." [Online]. Available: https://www.eia.gov/electricity/data/eia861/

[23] "Commercial and Residential Hourly Load Profiles." [Online]. Available: https://openei.org/datasets/dataset/commercial-and-residential-hourly-load-profiles-for-all-tmy3-locations-in-the-united-states/

[24] R. Hendron and C. Engebrecht, "Building america house simulation protocols." [Online]. Available: https://www.nrel.gov/docs/fy11osti/49246.pdf

[25] M. Deru, K. Field, D. Studer, K. Benne, B. Griffith, and P. Torcellini, "Department of energy commercial reference building models of the national building stock," Tech. Rep. [Online]. Available: https://www.nrel.gov/docs/fy11osti/46861.pdf

[26] M. Starke and N. Alkadi, "Assessment of industrial load for demand response across u.s. regions of the western interconnect," Tech. Rep. [Online]. Available: https://info.ornl.gov/sites/publications/files/pub45942.pdf

[27] "Industrial Assessment Centers." [Online]. Available: https://iac.university/

[28] D. Toffanin, "Generation of customer load profiles based on smart-metering time series, building-level data and aggregated measurements," Ph.D. dissertation, Zurich, 2016.

[29] H. Li, J. Wert, A. Birchfield, T. Overbye, C. Mateo, F. Postigo, P. Duenas, T. Gmez, T. Elgindy, and B. Palmintier, "Building highly detailed synthetic electric grid data sets for combined transmission and distribution systems," *submitted to IEEE Open Access Journal of Power and Energy. Manuscript available upon request*, 2020.

[30] "Syn-Austin-TDGrid." [Online]. Available: https://electricgrids.engr.tamu.edu/combined-td-synthetic-dataset/

[31] "Pecan Street DataPort." [Online]. Available: https://dataport.pecanstreet.org/

[32] D. Horvatic, H. E. Stanley, and B. Podobnik, "Detrended cross-correlation analysis for non-stationary time series with periodic trends," *EPL (Europhysics Letters)*, vol. 94, no. 1, p. 18007, 2011.

[33] T. Odun-Ayo and M. L. Crow, "Structure-preserved power system transient stability using stochastic energy functions," *IEEE Transactions on Power Systems*, vol. 27, no. 3, pp. 1450–1458, 2012.

[34] "Taxonomy Feeders Load Time Series." [Online]. Available: https://openei.org/datasets/dataset/randomized-hourly-load-data-for-use-with-taxonomy-distribution-feeders

[35] "Electric System Operation Data." [Online]. Available: https://www.eia.gov/realtime_grid/

[36] D. Hostick, D. Belzer, S. Hadley, T. Markel, C. Marnay, and M. Kintner-Meyer, "End-use electricity demand. Vol. 3 of renewable electricity futures study." Tech. Rep., 2012. [Online]. Available: https://www.nrel.gov/docs/fy12osti/52409-3.pdf

[37] "NERC distributed energy resources report," North American Electric Reliability Corporation, Tech. Rep., 2017. [Online]. Available: https://www.nerc.com/comm/Other/essntlrlbltysrvcstskfrcDL/Distributed_Energy_Resources_Report.pdf

[38] "CAISO Duck Curve Fast Facts." [Online]. Available: https://www.caiso.com/Documents/FlexibleResourcesHelpRenewables_FastFacts.pdf

[39] P. Denholm, M. OConnell, G. Brinkman, and J. Jorgenson, "Overgeneration from solar energy in california: a field guide to duck chart," Tech. Rep., 2015. [Online]. Available: https://www.nrel.gov/docs/fy16osti/65023.pdf

[40] "NREL Solar Data." [Online]. Available: https://www.nrel.gov/gis/data-solar.html

Hanyue Li (S'14) received the B.Sc. degree in electrical engineering from Illinois Institute of Technology, Chicago, IL, USA, in 2016, and the M.Sc. degree in electrical and computer engineering in Carnegie Mellon University, Pittsburgh, PA, USA, in 2017. She is currently working toward the Ph.D. degree in electrical engineering at Texas A&M University, College Station, TX, USA.

Ju Hee Yeo (S'17) received the B.Sc. degree in electrical engineering from Sangmyung University, Seoul, South Korea, in 2017. She is currently working toward the Ph.D. degree in electrical engineering at Texas A&M University, College Station, TX, USA.

Ashly L. Bornsheuer (S'17) received the B.Sc. degree in electrical engineering from Texas A&M University, College Station, TX, USA in 2018.

Thomas J. Overbye (S'87-M'92-SM'96-F'05) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the University of Wisconsin-Madison, Madison, WI, USA. He is currently a TEES distinguished research Professor in the Department of Electrical and Computer Engineering at Texas A&M University, College Station, TX, USA.