Synthesize Phasor Measurement Unit Data Using Large-Scale Electric Network Models

Ti Xu, Hanyue Li, Adam B. Birchfield, Thomas J. Overbye
Department of Electrical and Computer Engineering
Texas A&M University
College Station, United States

Abstract—Big data analytic applications using phasor measurements help improve the situation awareness of grid operators to better operate and control the system. Phasor measurement unit (PMU) data from actual grids is viewed as highly confidential and is not publicly available to researchers and educators. This paper develops a methodology to synthesize PMU data that can be accessed and shared freely, with a focus on input data preparation. Time series of demand- and generation-side input data are generated using public data from different utilities and statistical analysis methods. Detailed load dynamics modeling is also performed in this paper to extend the synthetic electric network models.

Index Terms—synthetic phasor measurement, data analysis, load modeling

I. INTRODUCTION

POWER system operation data can be obtained via multiple technologies and sources. Supervisory control and data acquisition (SCADA) is an industrial control system built to process, gather and monitor power system operation data [1]. Implementations of phasor measurement units (PMUs) generate large amounts of time-synchronized data measured at a higher speed (e.g., thirty observations per second), compared with traditional SCADA measurements (e.g.,one observation every several seconds) [2]. Big data analytics using those measurements have been applied to various diagnostics and decision support applications in power systems [3], [4] to improve the security, reliability and efficiency of large-scale interconnected power systems.

However, due to security concerns, the measurement data from actual grids is highly confidential, and so researchers have limited access to those data and even actual grid models, which limits big data researches and applications in the power community. Therefore, the research goal of this work is to create synthetic power system operation data using those public large-scale synthetic networks, which can be freely shared and used for various applications.

Researchers in [5]–[9] used public data and statistics obtained from actual system models to build several large-scale synthetic electric networks that mimic the functional and structural features of actual systems without revealing any confidential information. This paper will address the applications of those large-scale network models for synthesizing time-synchronized phasor measurements primarily from two perspectives: time-varying input data and detailed load dynamic models. Time-dependent renewable energy outputs and electricity demands are the key driver of variations observed in phasor measurements. Since geographic information is also available in those synthetic networks, inclusion of those input data will consider both temporal and spatial correlations among different sites. The WECC composite load model - CMPLDW - with different parameter settings will be assigned to load buses, according to their residential/commercial/industrial (RCI) ratios. This paper uses - ACTIVSg2000 - a 2000-bus case on the footprint of the Electric Reliability Council of Texas (ERCOT) as an example to illustrate the process [10], while the proposed method is general enough for other power system test cases and different time periods with various duration.

II. DEMAND-SIDE MODELING AND INPUT DATA CREATION PROCESS

This section first reviews the hourly load modeling procedure, through which each bus will obtain hourly load data for its residential, commercial and industrial classes, respectively. Statistical analysis performed on historical load data in 1-minute resolution will be applied to assign typical load variation patterns to each bus. According to the pre-defined RCI ratio, dynamic model components will be determined statistically for each bus.

A. Overview of Hourly Load Modeling

A bottom-up aggregation method is developed to create hourly load time series for the synthetic grid, where each load bus has a unique load profile [11]. Among three levels of load hierarchy - building load, bus load, and area/system load, lower-level loads are aggregated to build the higher level load.

Building loads are generated under residential, commercial and industrial sectors, any of which have several prototypical

load time series built from building-level energy consumption simulation data [12]. Those building load time series represent load sector typical behaviors and potential variations. In the 2000-bus test case, residential, commercial and industrial sectors have 61, 60 and 12 prototypical building load time series, respectively. As shown in Fig.1, building load time series under the same load sector differ based on load size, weather region, and temporal behaviors.
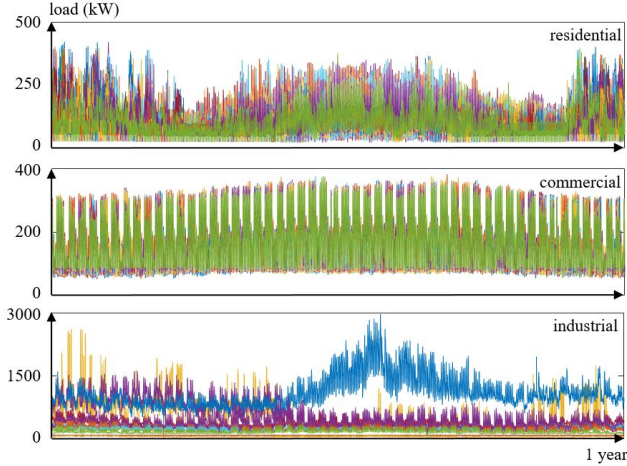

Fig. 1. Building-level load time series (residential, commercial and industrial)

All prototype building-level loads are tagged with geographic coordinates. To aggregate from building loads to bus load, each bus is assigned with the geographically closest prototypical building load time series data from each of the residential, commercial and industrial sectors. The residential, commercial and industrial sales values [13] are used to determine the electric utility companies load class ratios. Given the geographic information of utility companies and synthetic grid buses, each bus is assigned to its closest electric utility company, inheriting its RCI ratio (the class ratio of electricity consumption from those sectors).

The assigned prototypical building load time series are combined based on the optimal weights determined from a least square optimization problem in [11], where the peak of the weighted building load sum from three sectors should be close to the bus peak requirement from the base case, and the ratio of residential, commercial and industrial load satisfies the RCI ratio constraints.

The summation of all the bus level load time series aggregate to the area/system level load. System level synthetic load time series exhibit similar characteristics compared to real electric power system of the same size. Fig.2 shows 2000-bus test case system load, and ERCOT system load [5] in 2016.

B. Load Time Series with Higher Time Resolution

Running dynamic simulations requires time series input data with finer time resolution. Historical minutely load data [14] is fully utilized in this paper to obtain characteristic patterns of time-dependent load variations. Due to the time-dependent nature of human activities and weather conditions, electricity demand may establish distinct variation patterns at different
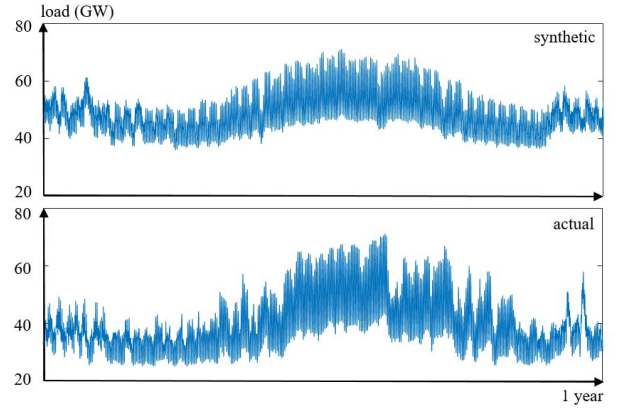

Fig. 2. Comparisons of 2000-bus case and ERCOT system load time series

periods of a day. Thus, for any simulation hour h, this paper collects load data of the specific hour from all days within the same season. Denote $[L_{d,1}, L_{d,2}, ..., L_{d,m}, ..., L_{d,M}]^T$ as the dth load data in minutely time interval. For meaningful comparisons among load data obtained from different days, a scaling procedure is applied to process the collected data:

$$\hat{L}_{d,m} = \frac{L_{d,m}}{\frac{L_{d,M}-L_{d,1}}{M-1} \cdot (m-1) + L_{d,1}}, \text{for } \forall d, m \quad (1)$$

Since clustering techniques have been widely used for feature extraction and pattern analysis, this paper applies the *K-means* clustering algorithm [15] to obtain four characteristic patterns of the scaled load data set, as displayed in Fig.3. For references, Fig.3 also visualizes the percentile plots of the scaled load data set. Fig.3 clearly shows that such a scaling procedure is able to capture the minute-by-minute load variations within each simulation hour, and the clustering results explicitly represent salient characteristics of different load variation patterns.
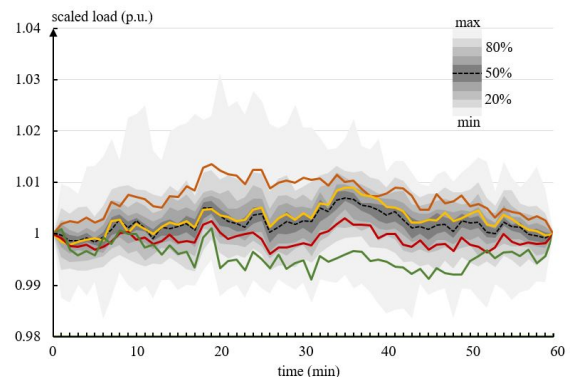
Fig. 3. Percentile visualization and characteristic patterns of the scaled load data set

The key step for creation of load time series with finer time resolution is to assign the obtained minutely load patterns to a pre-defined hourly load curve $[\tilde{L}_h, \tilde{L}_{h+1}]$. Consider a power grid with pre-defined Z zones, each of which includes a group of buses with similar weather conditions and other demand-related factors. From a statistics point of view, the expected

number of zones assigned with the kth minutely load pattern is $Z_k = \text{round}(Z \cdot p_k)$, where p_k is the probability of the kth cluster. In addition, zones assigned to the same load pattern are more likely to be geographically close to each other, while patterns with more significantly different load variations are more possible to be assigned to zones with larger distances. To achieve so, we define a comparison metric $q_{z,z',k,k'} = \frac{D_{z,z'}}{dist_{k,k'}}$[1], and formulate the following problem to complete the assignment process:

$$\max_{c_{z,k}\forall z,k} \quad -\sum_{z<z',k,k'} \frac{\zeta_{z,z',k,k'}}{\zeta_{sum}} \lg\left(\frac{\zeta_{z,z',k,k'}}{\zeta_{sum}}\right) \tag{2a}$$

$$s.t. \quad \sum_z c_{z,k} = Z_k, \ \forall k \tag{2b}$$

$$\sum_k c_{z,k} = 1, \ \forall z \tag{2c}$$

$$c_{z,k} \in \{0,1\}, \ \forall z,k \tag{2d}$$

$$\zeta_{z,z',k,k'} = c_{z,k}c_{z',k'}q_{z,z',k,k'}, \ \forall z<z', k, k' \tag{2e}$$

$$\zeta_{sum} = \sum_{z<z',k,k'} \zeta_{z,z',k,k'} \tag{2f}$$

In constraint (2d), $c_{z,k} = 1$ indicates that the kth load pattern is assigned to zone z. Constraint (2b) defines that each cluster k will be assigned to exactly Z_k zones. Each zone can be assigned with one and only one load pattern as defined in constraint (2c). In (2e), $\zeta_{z,z',k,k'}$ can be either a positive value or zero. In other words, $\zeta_{z,z',k,k'}$ is non zero only when zone z is assigned with the kth load pattern ($c_{z,k} = 1$) and zone z' is assigned with the k'^{th} load pattern ($c_{z',k'} = 1$). Equivalently, ζ_{sum} is the sum of all non-zero $\zeta_{z,z',k,k'}$. Overall, the proposed formulation (2) aims to find an assignment such that ratios of the zone-pair distance to their corresponding cluster-pair distance are as close as possible for all zone pairs z, z'. Solutions to problem (2) can be obtained using any proper optimization techniques, which is outside the scope of this paper and thus won't be addressed here.

Consider a zone z assigned with k^{th} load variation pattern $[\hat{l}_{k,1}, ..., \hat{l}_{k,m}, ..., \hat{l}_{k,M}]^T$. A re-scaling process is proposed to combine the obtained minutely load pattern $[\hat{l}_{k,1}, ..., \hat{l}_{k,m}, ..., \hat{l}_{k,M}]^T$ and the hourly load curve $[\tilde{L}_h, \tilde{L}_{h+1}]$:

$$\tilde{L}_{h,m} = \left(\frac{\tilde{L}_{h+1} - \tilde{L}_h}{M-1} \cdot (m-1) + \tilde{L}_h\right) \cdot \hat{l}_{k,m}, \text{for } \forall h, m \tag{3}$$

By applying the re-scaling procedure to all loads, we obtain geographically and temporally correlated load curves in minutely time interval for all buses. Fig.4 shows zonal load step changes in percentage and the expected system load. For additional modeling purposes, such as forecast errors, statistically random terms can be added to make the load profile more realistic. We note that, although this section focuses on the simulation hour h, the proposed method is also applicable to other hours of different days.

---

[1]$dist_{k,k'}$ is the Euclidean distance between two clusters k, k' and we define $dist_{k,k}$ to be a sufficiently small positive number.
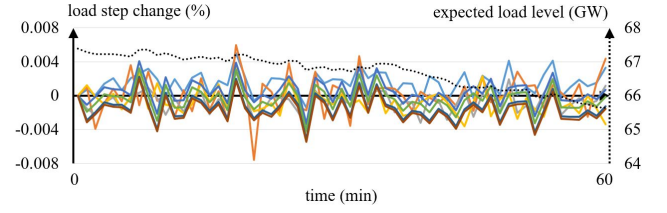

Fig. 4. Load step change in percentage for eight weather zones and the expected system load (dotted line)

C. Modeling Dynamics for Load with Different RCI Ratios

Dynamic load modeling is essential for system transient stability studies [16], [17]. The load time series developed in previous steps can be viewed as the ideal demand trend over time, but a realistic simulation requires a proper representation of load dynamics. Historically the commonly used load dynamic models were constant impedance. However, the report by [18] pointed out that dynamic responses should reflect the actual behavior of aggregate loads. An up-to-date load dynamic model is an important factor to improve the accuracy of grid simulations [19].

The Western Electricity Coordinating Council (WECC) load modeling task force has devoted the past years on development of a composite load model, named as WECC Dynamic Composite Load Model (CMPLDW in PLSF; CMLDBLU1 in PSS/E) [20]. This model is designed to represent behaviors of different end-user components [21]–[23]:

- Motor A: three-phase induction motors operating under constant torque (*i.e.* compressor motors used in large air-conditioners and refrigerators);
- Motor B: three-phase induction motors with high inertia operating under speed-dependent torque (*i.e.* fan motor);
- Motor C: three-phase induction motors with low inertia operating under speed-dependent torque (*i.e.* pump motor);
- Motor D: single-phase induction motors (*i.e.* heat pumps);
- Electronic Load: electronics-driven electricity consumer (*i.e.* personal computers, LED televisions);
- Static Load: a frequency-dependent ZIP model (*i.e.* resistive heating, cocking).

This work adopts the CMPLDW for modeling load dynamics. Table I summarizes the adopted dynamic load composition [24] for different periods of a day. The assignment process refers to Table I and computes appropriate dynamic load component percentages for each load using its RCI ratio. Default parameters published by the North American Electric Reliability Corporation(NERC) Load Modeling Task Force are used for all CMPLDW models.

Through implementing the three steps proposed in this section, a non-conforming load profile with non-uniform dynamic load compositions is obtained. Such a profile captures the geographic and temporal correlations among different sites, and is capable to be used for transient stability studies using large-scale network models.

Period	Class	Dynamic Load Component Percentage (%)					
		Motor				Load	
		A	B	C	D	Electronic	Static
Peak	RES	8	7	2	34	15	34
	COM	12	10	4	25	18	31
	IND	13	22	16	0	27	22
Shoulder	RES	8	7	2	25	19	39
	COM	12	10	4	20	23	31
	IND	13	22	16	0	27	22
Light	RES	10	8	2	0	40	40
	COM	12	10	4	5	38	31
	IND	13	22	16	0	27	22

III. SUPPLY-SIDE MODELING AND INPUT DATA CREATION PROCESS

Time-dependent variations and uncertainty exist in both demand and supply sides. This section aims to build time-varying wind power profiles, by starting with wind speed time series in 5-minute time resolution. Then, analysis on historical secondly wind speed data is performed to obtain statistic characteristics of wind speed variations, which is used to generate synthetic wind speed variation terms in finer time resolution. A procedure is proposed to combine the wind speed measurements in 5-minute time interval and wind speed variation terms in finer time resolution. At last, a common wind power production model is adopted for converting wind speeds to wind farm outputs.

A. 5-min Wind Speed Time Series

Wind Integration National Dataset (WIND) Toolkit is a wind data set developed and maintained by the National Renewable Energy Laboratory (NREL) [25], [26]. In this paper, simulated 5-min wind speed profiles for all synthetic wind farms in the test case are obtained from this toolkit, according to their geographic information - latitudes and longitudes. For further data analysis, we compute correlations among all synthetic wind farms and Fig.5 displays the relationship between correlations and site distances. A two-segment polynomial curve $\rho_{i,j} = \rho(D_{i,j})$ is also obtained to approximate such relationship. The collected 5-min wind speed time series form the basis for inclusion with time-dependent variation terms in finer time resolution, which will be discussed in the next step.
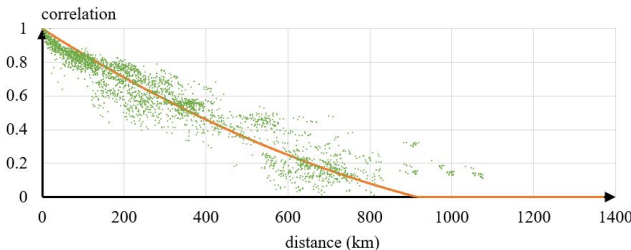

Fig. 5. Correlations versus distances among wind farms

B. Wind Speed Time Series with Finer Time Resolution

Creation of wind speed time series in finer time resolution faces a key challenge - the inherent uncertainty, variability and intermittency features of wind energy. This subsection applies a set of secondly wind speed measurements from [27] for statistical analysis. Equations (4) are formulated to process one secondly data series $[W_1, ..., W_s, ..., W_S]^T$ over a period of a predefined duration.

$$\hat{W}_s = W_s - (\frac{W_S - W_1}{S - 1} \cdot (s - 1) + W_1), \ \forall s \quad (4a)$$

$$\hat{\hat{W}}_s = \hat{W}_s - \hat{W}_{s-1}, \ \forall s \geq 2 \quad (4b)$$

Here $\hat{\hat{W}}_1$ is defined to be zero. Clearly, the set $\{\hat{\hat{W}}_s, \forall s\}$ over one period has a zero mean and a non-zero standard deviation. This paper uses normal distribution $\mathcal{N}(0, \sigma)$ to approximate the distribution of the set $\{\hat{\hat{W}}_s, \forall s\}$. Thus, we repeat this process for secondly wind speed data set from each of all same-duration periods on a rolling basis.
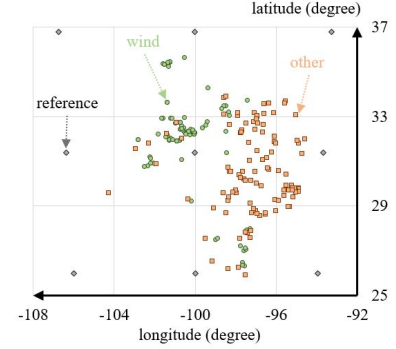

Fig. 6. Geographic representation of generators and selected reference points

The locations of wind and non-wind generators are shown in Fig.6. In this paper, we consider a reference network with N points uniformly distributed over the footprint. According to Fig.5, two wind farm sites with a distance at least 600 km is viewed as trivially correlated. Thus, as shown in Fig.6, those reference points are selected such that two neighboring sites in each East-West row and each North-South column have a distance of 600 km. For each reference point n, we statistically select a $\sigma \sim \psi(\cdot)$ and generate a variation set $[\hat{\hat{w}}_{n,1}, ..., \hat{\hat{w}}_{n,s}, ..., \hat{\hat{w}}_{n,S'}]^T$ over a 5-minute period from $\mathcal{N}(0, \sigma)$, where $\hat{\hat{w}}_{n,1}$ is defined to be zero. Correspondingly, we can compute the wind speed variation terms $[\hat{w}_{n,1}, ..., \hat{w}_{n,s}, ..., \hat{w}_{n,S'}]^T$ by taking $\hat{w}_{n,s} = \hat{w}_{n,s-1} + \hat{\hat{w}}_{n,s}$ and $\hat{w}_{n,0} = 0$, which can be viewed as an inversion of (4b). As such, we can further build the wind speed variation terms $[\hat{v}_{e,1}, ..., \hat{v}_{e,s}, ..., \hat{v}_{e,S'}]^T$ for each synthetic wind farm e over a 5-minute period using those reference points:

$$\hat{v}_{e,s} = \sum_n \frac{\pi_{e,n} \hat{w}_{n,s}}{\Omega_e}, \text{where } \Omega_e = \sum_n \pi_{e,n}, \forall e \quad (5)$$

$\pi_{e,n}$ is the weight factor for wind farm e and reference point n. In this paper, the correlation value computed based on Fig.5 is used as the weight factor: $\pi_{e,n} = \rho_{e,n} = \rho(D_{e,n})$.

Suppose that $V_{e,\tau}$ and $V_{e,\tau+1}$ are wind speeds from Section III.A - two end points of a 5-minute time period. The wind speed vector $[v_{e,1}, ..., v_{e,s}, ..., v_{e,S'}]^T$ for each synthetic wind farm e over a 5-minute period can be computed by:

$$v_{e,s} = (\frac{V_{e,\tau+1} - V_{e,\tau}}{S' - 1} \cdot (s-1) + V_{e,\tau}) + \hat{v}_{e,s}, \forall s \in (1, S') \quad (6)$$

where $v_{e,1}$ and $v_{e,S'}$ are set to be equal to $V_{e,\tau}$ and $V_{e,\tau+1}$, respectively. Overall, equation (6) combines the wind speed data $[V_{e,\tau}, V_{e,\tau+1}]^T$ in the 5-minute time interval (from Section III.A) with the 5-minute-duration wind speed variation vectors $[\hat{v}_{e,1}, ..., \hat{v}_{e,s}, ..., \hat{v}_{e,S'}]^T$ in a finer time interval (from Section III.B), to formulate the wind speed data $[v_{e,1}, ..., v_{e,s}, ..., v_{e,S'}]^T$ in a finer time interval for a 5-minute period. Such a procedure is then repeated to generate multi-site wind speed profile over a longer period (e.g., one hour or a day).

C. Wind Power Production Model

This paper adopts standard terminology [28] for wind power production model: cut-in (v_{cutin}), rated (v_{rated}), and furling wind speed (v_{furl}) of a wind turbine. The following continuous function [29] is used to model the power output $p(\cdot)$ (in p.u.) of the wind turbine, as a function of wind speed v:

$$p(v) = \begin{cases} 1 - (1 + \exp(\frac{v - v_{mid}}{\alpha_1}))^{-1}, \ 0 \le v \le v_{lim} \\ 1 - \alpha_2(v - v_{lim})^{\alpha_3}, \ v_{lim} < v \le v_{furl} \\ 0, \ \text{otherwise} \end{cases} \quad (7a)$$

$$\alpha_1 = (v_{rated} - v_{mid})(\ln(\frac{\beta}{1 - \beta}))^{-1} \quad (7b)$$

$$\alpha_2 = (v_{furl} - v_{lim})^{-\alpha_3} \quad (7c)$$

v_{mid} is the wind speed where the wind power output is expected at half the rated value. Shape parameter α_1 is tuned by changing β (typically close to 1) such that the power output at the rated speed can almost achieve the rated power. Shape parameters α_2 and α_3 are selected to model a drop-off power output function after the wind speed exceeds the limiting value v_{lim} but still within the v_{furl}. Parameter values from [29] are modified for uses in this paper with $\beta = 0.95$.

As a summary, Fig.7 summarizes the simulated wind speeds at multiple wind farms in 15-s time resolution for one hour and the corresponding total wind power output. Those curves show that the proposed method is capable to generate correlated wind speed and output profiles that capture the uncertainty and variability nature of wind resources.

IV. SYNTHETIC PMU DATA CREATION EXAMPLE

A simulation of 10-minute with half cycle time step is performed on the ACTIVSg2000 system to create the example synthetic PMU data. PowerWorld Dynamic Studio, an interactive environment for transient stability level simulation, is used [30].

For consistent input data granularity, the 1-min interval load data is superimposed with Gaussian noise of 1% standard deviation every 15 seconds. The load and wind variation data
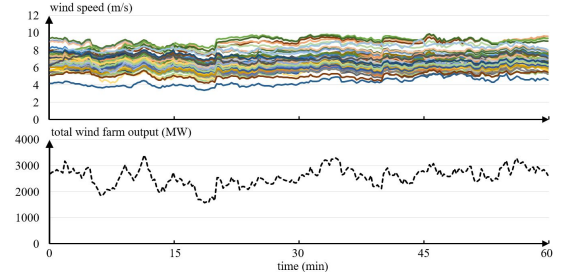

Fig. 7. Multi-site wind speeds and system total wind power

is then formatted to be a time series binary (tsb) file as the simulation input. System control actions, such as automatic generation control, and economic dispatch, are also taken into consideration for the realistic creation of synthetic PMU data. Generators in the synthetic system are re-dispatched every 15 minutes to mimic real time operations from the control rooms.

Figure 8 shows the created PMU data of voltage magnitude for all the buses in the ACTIVSg2000 synthetic system. It is observed that the synthetic PMU voltage data reflects the variations in load, renewable generation and generator controls.
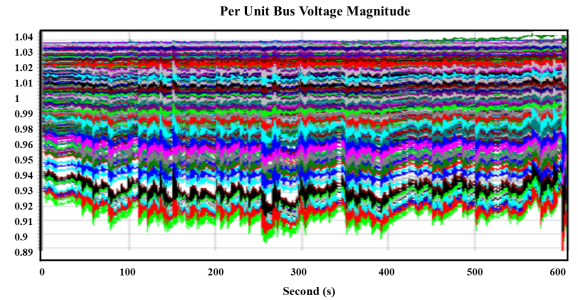

Fig. 8. Per unit bus voltage magnitude

V. CONCLUSION AND FUTURE WORKS

This paper proposed a methodology of creating PMU data from large-scale synthetic electric systems, with a focus on the input data preparation. The creation of time-varying input data of both demand and generation, as well as the detailed dynamic load models are discussed. An example of generating synthetic PMU data using created time-varying inputs is also provided. The produced synthetic PMU data can be accessed and distributed freely without confidentiality concerns.

For future work, the post-modification and adaptation of PMU signals, as well as statistical validation of the created data are critical to improve the realism of the synthetic PMU data.

NOMENCLATURE

Section II

d	index of day from 1 to D
m	index of minute from 1 to M
$L_{d,m}$	the measured load level at minute m of the specific hour for day d (MW)

$\hat{L}_{d,m}$	the scaled load level at minute m of the specific hour for day d (p.u.)
$\hat{l}_{k,m}$	the scaled load level at minute m in the kth load variation pattern (p.u.)
k	index of cluster from 1 to K
p_k	probability of the kth cluster
z	index of zone from 1 to Z
$q_{z,z',k,k'}$	comparison metric defined as $\frac{D_{z,z'}}{dist_{k,k'}}$
$c_{z,k}$	status indicator equal to 1 if and only if the kth load pattern is assigned to zone z
$\zeta_{z,z',k,k'}$	defined as $c_{z,k}c_{z',k'}q_{z,z',k,k'}$
ζ_{sum}	defined as $\sum_{z<z',k,k'}\zeta_{z,z',k,k'}$
$\tilde{L}_h, \tilde{L}_{h+1}$	two end points of the hourly load curve (MW)
$\tilde{L}_{h,m}$	the re-scaled load level at minute m of the specific hour h (MW)

Section III

$\rho_{i,j}$	correlation between two sites i, j
s	index of second from 1 to S
W_s	the measured wind speed at second s (m/s)
$\hat{W}_s, \hat{\tilde{W}}_s$	variation terms defined in equation (4)
σ	the standard deviation of $\hat{\tilde{W}}_s$
$\psi(\cdot)$	approximated probability distribution function for σ
n	index of reference point from 1 to N
$\hat{\tilde{w}}_{n,s}$	sampled variation term
$\hat{w}_{n,s}$	defined as $\hat{w}_{n,s} = \hat{w}_{n,s-1} + \hat{\tilde{w}}_{n,s}$ and $\hat{w}_{n,0} = 0$
e	index of wind farm from 1 to E
$\hat{v}_{e,s}$	computed wind speed variation term at farm e
$v_{e,s}$	wind speed at farm e (m/s)
$\pi_{e,n}$	weight factor for wind farm e and reference point n with $\Omega_e = \sum_n \pi_{e,n}$
$V_{e,\tau}, V_{e,\tau+1}$	wind speeds - two end points of a 5-minute time period.
v_{cutin}	cut-in wind speed (m/s)
v_{mid}	wind speed where the wind power output is expected at half the rated value (m/s)
v_{rated}	wind speed where the wind power output is expected at the rated value (m/s)
v_{lim}	limiting wind speed (m/s)
v_{furl}	furling wind speed (m/s)
α_1, β	shape parameters for wind power production model when $0 \leq v \leq v_{lim}$
α_2, α_3	shape parameters for drop-off power output function when $v_{lim} < v \leq v_{furl}$

References

[1] M. S. Thomas and J. D. McDonald, *Power System SCADA and Smart Grids*. CRC Press, 2015.

[2] "North AmericanSynchroPhasor Initiative." [Online]. Available: https://www.naspi.org/

[3] M. Kezunovic, L. Xie, and S. Grijalva, "The role of big data in improving power system operation and protection," in *2013 IREP Symposium Bulk Power System Dynamics and Control - IX Optimization, Security and Control of the Emerging Power Grid*, Aug 2013, pp. 1–9.

[4] "Frequency Monitoring Network." [Online]. Available: http://fnetpublic.utk.edu/

[5] 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, July 2017.

[6] A. B. Birchfield, K. M. Gegner, T. Xu, K. S. Shetye, and T. J. Overbye, "Statistical considerations in the creation of realistic synthetic power grids for geomagnetic disturbance studies," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 1502–1510, March 2017.

[7] T. Xu, A. B. Birchfield, K. M. Gegner, K. S. Shetye, and T. J. Overbye, "Application of large-scale synthetic power system models for energy economic studies," *Hawaii International Conference on System Sciences (HICSS-50)*, pp. 3123–3129, Jan 2017.

[8] T. Xu, A. B. Birchfield, K. S. Shetye, and T. J. Overbye, "Creation of synthetic electric grid models for transient stability studies," in *2017 IREP Symposium Bulk Power System Dynamics and Control*, Sept 2017.

[9] T. Xu, A. B. Birchfield, and T. J. Overbye, "Modeling, tuning and validating system dynamics in synthetic electric grids," *IEEE Transactions on Power Systems*, pp. 1–1, 2018.

[10] "Electric Grid Test Case Repository - Synthetic Electric Grid Cases." [Online]. Available: https://electricgrids.engr.tamu.edu/

[11] 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)*, Feb 2018, pp. 1–6.

[12] "Annual Electric Power Industry Report (EIA-861 data file)." [Online]. Available: https://www.eia.gov/electricity/data/eia861/

[13] Electric Reliability Council of Texas, "Hourly Load Data Archives." [Online]. Available: http://www.ercot.com/gridinfo/load

[14] "Historical Load Data." [Online]. Available: http://www.ceps.cz/ENG/Data/Vsechna-data/Pages/Zatizeni.aspx

[15] R. Xu and D. Wunsch, *Clustering*. Hoboken, NJ: John Wiley and Sons, Inc., 2009, pp. 67–72.

[16] IEEE Task Force on Load Representation for Dynamic Performance, "Load representation for dynamic performance analysis (of power systems)," *IEEE Transactions on Power Systems*, vol. 8, no. 2, pp. 472–482, May 1993.

[17] ——, "Standard load models for power flow and dynamic performance simulation," *IEEE Transactions on Power Systems*, vol. 10, no. 3, pp. 1302–1313, Aug 1995.

[18] North American Electric Reliability Corporation, "Reliability guideline - developing load model composition data." [Online]. Available: https://www.nerc.com/pa/RAPA/rg/ReliabilityGuidelines/Reliability_Guideline_-_Load_Model_Composition_-_2017-02-28.pdf

[19] R. Bravo, R. Yinger, D. Chassin, H. Huang, N. Lu, I. Hiskens, and G. Venkataramanan, "Final project report - load modeling transmission research." [Online]. Available: https://web.eecs.umich.edu/~hiskens/publications/LM_Final_Report.pdf

[20] N. Lu and A. Qiao, "Composite load model evaluation." [Online]. Available: https://www.pnnl.gov/main/publications/external/technical_reports/PNNL-16916.pdf

[21] Y. Liu, Y. Zhang, Q. Huang, S. Kundu, Y. Tang, D. James, P. Etingov, B. Mitra, and D. Chassin, "Impact of building-level motor protection on power system transient behaviors." [Online]. Available: https://arxiv.org/abs/1806.07050

[22] North American Electric Reliability Corporation, "Technical reference document - dynamic load modeling." [Online]. Available: https://www.nerc.com/comm/PC/Pages/default.aspx

[23] K. Zhang, H. Zhu, and S. Guo, "Dependency analysis and improved parameter estimation for dynamic composite load modeling," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3287–3297, July 2017.

[24] Midcontinent Independent System Operator, "MISO MOD-032 model data requirements & reporting procedures." [Online]. Available: https://www.misoenergy.org/planning/planning-models/mod-032-1/

[25] C. Draxl, A. Clifton, B.-M. Hodge, and J. McCaa, "The wind integration national dataset (wind) toolkit," *Applied Energy*, vol. 151, pp. 355 – 366, 2015.

[26] National Renewable Energy Laboratory, "Wind integration national dataset toolkit." [Online]. Available: https://www.nrel.gov/grid/wind-toolkit.html

[27] UCAR/NCAR - Earth Observing Laboratory, "PCAPS ISFS 1 second data." [Online]. Available: https://www.eol.ucar.edu/field_projects/pcaps/one-second-dataset

[28] G. Masters, *Renewable and Efficient Electric Power System*. Hoboken, NJ: John Wiley & Sons, Inc., 2013.

[29] S. V. Dhople and A. D. Domnguez-Garca, "A framework to determine the probability density function for the output power of wind farms," in *2012 North American Power Symposium (NAPS)*, Sept 2012, pp. 1–6.

[30] T. J. Overbye, Z. Mao, K. S. Shetye, and J. D. Weber, "An interactive, extensible environment for power system simulation on the pmu time frame with a cyber security application," in *2017 IEEE Texas Power and Energy Conference (TPEC)*, 2017.