2017 American Control Conference
Sheraton Seattle Hotel
May 24–26, 2017, Seattle, USA

Probabilistic Analysis of Hybrid Energy Systems Using Synthetic Renewable and Load Data

Jun Chen, *Member, IEEE*, Jong Suk Kim and Cristian Rabiti

***Abstract*—This paper focuses on probabilistic analysis of hybrid energy systems (HES), which integrate multiple energy inputs and multiple energy outputs for effective management of variability in renewable energy and grid demand. To characterize the volatility, a statistical model combining Fourier series and autoregressive moving average (ARMA) is used to generate synthetic weather condition (e.g., wind speed) and grid demand data. Specifically, Fourier series is used to model the seasonal trends in historical data, while ARMA is applied to characterize the autocorrelation in residue time series (e.g., measurements with seasonal trends subtracted). The synthetic data is shown to have same statistic characteristics with historical measurements, but possesses different temporal profile. The probabilistic analysis of a particular HES configuration is then performed, which consists of nuclear power plant, wind farm, battery storage, and desalination plant. Requirements on component ramping rate, and the effects of deploying different sizes of batteries in smoothing renewable variability, are all investigated.**

***Index Terms*—Hybrid energy systems, renewable energy integration, synthetic data generation, autoregressive moving average**

I. Introduction

Hybrid energy systems (HES) that consist of multiple energy generations/utilizations have been proposed in literature [1]–[7] to address our energy concerns and to enable higher level of renewable energy utilization. By integrating more than one energy resources and consumptions, HES can act as a highly responsive device and can be operated under flexible operations schedules to accommodate the variability introduced from renewable generation, modern loads (such as electric vehicles), and markets [4]–[7]. For example, [7] shows that by utilizing an operations optimization, HES can participate in both day ahead and real time electricity markets as well as ancillary service market. However, these prior analyses were performed based on historical measurements on weather condition and electric demand, and hence only a limited number of measurements database were used. The objective of this paper is then to study the technical performance of HES under *synthetic* weather and grid load data, which are statistically conformed to actual measurement and allows probabilistic analysis of HES.

The synthetic wind speed data generation has been studied in the literature. For example, [8] uses autoregressive moving average (ARMA) model to generate noises, and adds the sampled noises to the historical data, while [9]–[11] use trained ARMA or AR model, together with sampled white noise, to generate scenarios. To fulfill the assumption of ARMA that time series are normally distributed, the measurement data may need to be transformed into Gaussian distribution before being used to train the model. Reference [10] further normalizes the transformed time series with respect to each hour for each month. Reference [12] uses AR-GARCH (autoregressive generalized autoregressive conditional heteroskedasticity) model for wind speed prediction, which allows the regression of both mean and variance. The prediction outputs, in terms of the mean and variance of the wind speed/power, can be used to sample synthetic scenarios. On the other hand, synthetic electric load data generation has also been investigated in literature. For example, [13] uses AR model to fit available sub-hour load data. The linear and Fourier terms are used to fit the seasonal trend, with remaining irregular load modeled by a AR model. Reference [14] uses ARMA to fit the residues that are resulted by detrending with seasonal latent variables. Reference [15] also decomposes the load data into deterministic seasonal trend part and irregular part. Both parts are used to train neural network (NN), one for each, which are used to generate forecast. Reference [16] combines wavelet transform with NN approach, while [17] uses Bayesian belief network to improve available load forecasting.

In this paper, a combined model with Fourier series and autoregressive moving average is utilized to model the measurement data statistically and to generate synthetic wind speed and electric load data. In particular, Fourier terms are used to capture the seasonal trends in yearly measurements and ARMA is then used to model the autocorrelation in residues (i.e., measurements with seasonal trends subtracted). After training the model over historical data by finding optimal parameters, the combined model can then be used to generate synthetic data, which consists of generating independent white noise for each time step, utilizing ARMA model and the synthesized white noise to compute residues for each time step, and finally adding the Fourier terms representing seasonal trends. In order to validate the proposed model, several key statistics, including mean, variance, and empirical cumulative distribution function, will be computed for both actual measurements database and synthetic data, to verify the match between database and synthetic data. The synthetic data for both renewable and load will in turn be utilized to analyze a particular HES configuration, which includes a nuclear power plant, wind farm, battery storage, and desalination plant. In particular, 3000 synthetic scenarios of wind speed and electric

This research is supported by the Energy Security Initiative and the N-R HES (Nuclear-Renewable Hybrid Energy Systems) Program at INL (Idaho National Laboratory) under the U.S. Department of Energy contract DE-AC07-05ID14517.

The authors are with Idaho National Laboratory, 750 University Blvd., Idaho Falls, ID 83415, USA. Email: {jun.chen, jongsuk.kim, cristian.rabiti}@inl.gov

978-1-5090-5992-8/$31.00 ©2017 AACC

demand are generated to simulate the HES configuration, while various probabilistic analysis are performed to understand the requirements on component power consumption and ramping rate in order to accommodate variability introduced by renewable generation and electric demand. Furthermore, the effects of employing different sizes of batteries for variability smoothing will also be investigated.

The rest of this paper is organized as follows. Section II presents the theoretical foundation of the proposed model. The HES configuration under study is presented in Section III, and results on model validation and probabilistic analysis of HES are given in Section IV. Finally, Section V concludes the paper.

II. Synthetic Data Generation

A. ARMA Model and Identification

Autoregressive moving average (ARMA) model with orders p and q, often referred to as ARMA(p,q), is given as [18]:

$$x_t = \sum_{i=1}^{p} \phi_i x_{t-i} + \alpha_t + \sum_{j=1}^{q} \theta_j \alpha_{t-j}, \quad (1)$$

where the process variable x is a vector of dimension n, and parameters ϕ_i for $i = 1, \dots, p$ and θ_j for $j = 1, \dots, q$ are both n by n matrices. The noise term α is usually assumed to be white noise. Given an ARMA(p,q) model, its parameters ϕ_i's and θ_j's can be estimated by the following maximum likelihood estimator (MLE). Denote the covariance of α to be Σ, and

$$\Sigma = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{bmatrix}.$$

Define the parameters to be estimated as $\eta := (\phi_1, \dots, \phi_p.\theta_1, \dots, \theta_q, \sigma_1^2, \dots, \sigma_n^2)$. Given T number of measurements of the process variable x, denoted as $x_1, x_2, \dots, x_T$, the the likelihood function over η can be written as:

$$\begin{aligned} L(\eta) &= f(x_1, \dots, x_T|\eta) \\ &= f(x_1|\eta) f(x_2|\eta, x_1) \cdots f(x_T|\eta, x_1, \dots, x_{T-1}) \\ &= \prod_{t=1}^{T} \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2} \hat{\alpha}_t' \Sigma^{-1} \hat{\alpha}_t\right), \end{aligned}$$

where the estimation of the error term at time t, i.e., $\hat{\alpha}_t$ can then be recursively computed as:

$$\begin{aligned} \hat{\alpha}_1 &:= x_1 \\ \hat{\alpha}_2 &:= x_2 - (\phi_1 x_1 + \theta_1 \hat{\alpha}_1) \\ \hat{\alpha}_3 &:= x_3 - (\phi_1 x_2 + \phi_2 x_1 + \theta_1 \hat{\alpha}_2 + \theta_2 \hat{\alpha}_1) \\ &\vdots \\ \hat{\alpha}_t &:= x_t - \left(\sum_{i=1}^{\min(p,t-1)} \phi_i x_{t-i} + \sum_{j=1}^{min(q,t-1)} \theta_j \hat{\alpha}_{t-j} \right). \end{aligned}$$

Furthermore, the log-likelihood function, i.e., $\log L(\eta)$, can be expressed as:

$$\log L(\eta) = -\frac{nT}{2} \log(2\pi) - \frac{T}{2} \log |\Sigma| - \frac{1}{2} \sum_{t=1}^{T} \hat{\alpha}_t' \Sigma^{-1} \hat{\alpha}_t. \quad (2)$$

The MLE of η is then given by

$$\hat{\eta} := \arg\max \log L(\eta). \quad (3)$$

Furthermore, the order p and q are selected by Bayesian information criterion (BIC), defined as:

$$BIC(p,q) = \log(\hat{\sigma}^2) + \frac{\log(T)(p+q)}{T},$$

where $\hat{\sigma}^2$ is the determinant of $\Sigma^T \Sigma$. It is not hard to see that the first term on the right hand size captures how the model fits to the data, while the second term penalizes larger model to prevent overfitting. For a more detailed treatment of time series and ARMA model please refer to [18].

B. Seasonal Trend and Normality Transform

To model the seasonal trends in historical data, the following Fourier series is used:

$$F_t := \sum_k \{a_k \sin(2\pi f_k t) + b_k \cos(2\pi f_k t)\}. \quad (4)$$

The set of frequency $\{f_k\}$ is user-defined parameters, and the coefficient $\{a_k\}$ and $\{b_k\}$ can be estimated by linear regression.

In general, renewable source and load profile do not satisfy the normality assumption, even after seasonal trends being extracted. To mitigate this problem, the residues are transformed to have Gaussian property [9], [10], [18] before applying the MLE of (3). Define a new stochastic process y, which has a standard normal distribution, as follows:

$$y_t := \Phi^{-1}[f(x_t - F_t)], \quad (5)$$

where f is the empirical cumulative distribution function (CDF) of the residues and Φ is the CDF function of the standard normal distribution. The trained ARMA model can then be used to simulate process y, which is in turn used to generate the scenarios by an inverse transformation, as following:

$$x_t := f^{-1}[\Phi(y_t)] + F_t. \quad (6)$$

III. HES Configuration

The HES under study is shown in Fig. 1, which includes the following components:

- a heat generation plant with 180 MW capacity[1], consisting of a small modular reactor (SMR) and a steam generator,
- a series of steam turbines, feedwater systems, and heaters, paired with an electric generator that converts steam into electricity,

[1]For simplicity, all power calculations will be expressed using the electrical equivalence (in MW) of the particular power stream, assuming fixed thermal-to-electrical conversion efficiency.

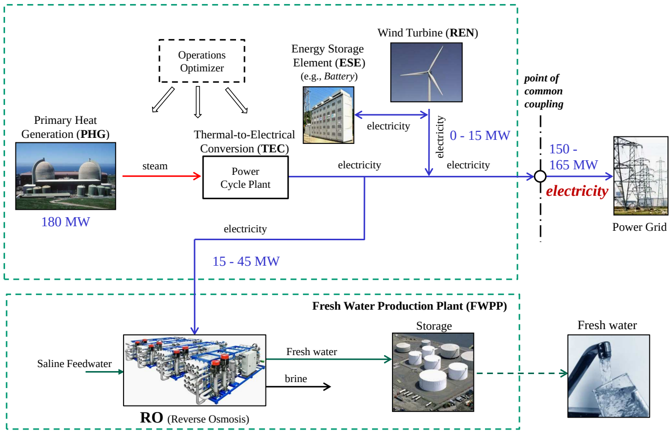

Fig. 1. Topology of the hybrid energy system configuration considered in this paper.

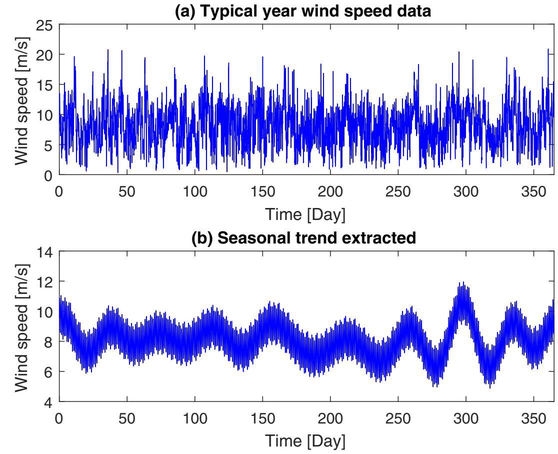

Fig. 2. Wind speed data used to train the model: (a) Whole year data; (b) Seasonal trend extracted.

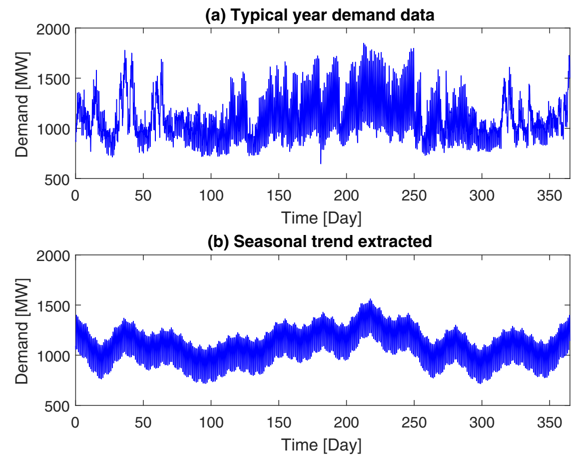

Fig. 3. Demand data used to train the model: (a) Whole year data; (b) Seasonal trend extracted.

- a series of wind turbines as renewable power generation source with total capacity of 15 MW,
- electrical storage (i.e., a system scale battery set) used for power smoothing of the electricity generated by wind turbines,
- a Brackish water reverse osmosis (RO) desalination plant that converts saline water to fresh water by consuming electricity between 15 MW and 45 MW,
- electric grid connected to HES at a point of common coupling and consuming electricity between 150 MW and 165 MW.

In the following, we briefly describe the modeling of wind power generation unit. For the detailed modeling of the rest components, please refer to [5]–[7]. The wind farms consist of 10 wind turbines, each rated at 1.5 MW and located on a 2 square kilometer site for a maximum of about 15 MW generation at full production. Each wind turbine is then modeled as a mapping function from wind speed to power output, as follows.

$$E_{REN} := \begin{cases} 0 & \text{if } U \leq 3 \text{ m/s or } U \geq 25 \text{ m/s} \\ 0.5\eta\rho U^3 \frac{\pi d^2}{4} & \text{if } 3 \text{ m/s} < U \leq 14 \text{ m/s} \\ 1.5 & \text{if } 14 \text{ m/s} < U < 25 \text{ m/s} \end{cases} \tag{7}$$

where η is the conversion efficiency of the wind turbine, ρ is the density of the air at the site, U is the wind speed, and d is the diameter of the turbine blades. In this study the values used for each parameter in equation (7) are: $\eta = 35\%$, $\rho = 1.17682\ g/m^3$, $d = 58.13\ m$.

IV. Results and Discussion

A. Model Characterization and Validation

This section presents the results on model characterization. The wind speed database[2] used to train the model for wind speed data generation is shown in Fig. 2(a), while the seasonal trends, modeled as Fourier series of equation (4), is given in Figure 2(b). Likewise, the demand database[3] used to train the model for load data generation is shown in Fig. 3(a), while the seasonal trends is given in 3(b).

Fig. 4 plots the synthetic wind speed scenario and the actual database for selected 7 days, which exhibit similar dynamics and volatility. Furthermore, Table I compares several key statistics (mean, standard deviation, etc) between the synthetic and actual database, while empirical cumulative distribution functions (CDF) of synthetic wind speed and database are compared in Fig. 5, suggesting good match between synthetic wind speed scenarios and database. Likewise, Fig. 6 plots the synthetic electric demand scenario and the actual database for selected 7 days, which exhibit similar dynamics and volatility. Furthermore, Table II compares several key statistics (mean, standard deviation, etc) between the synthetic and actual database, while empirical CDF of synthetic electric demand

[2]Downloaded from the Eastern Wind dataset maintained by NREL (National Renewable Energy Laboratory) at http://www.nrel.gov/electricity/transmission/eastern_wind_dataset.html on November 21, 2014.

[3]Downloaded from Electric Reliability Council of Texas at http://www.ercot.com/gridinfo/load/load_hist/ on December 11, 2014.

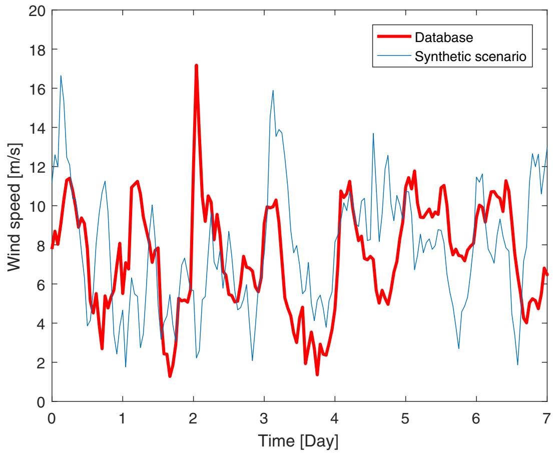

Fig. 4. Synthetic wind speed scenario and the actual database for selected 7 days.

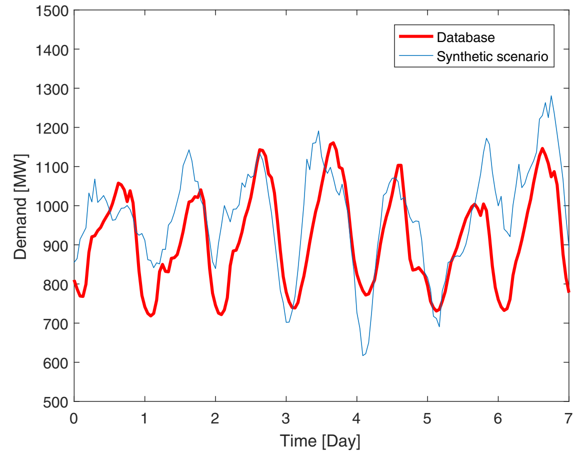

Fig. 6. Synthetic electric demand and the actual database for selected 7 days.

TABLE I
COMPARISON BETWEEN SYNTHETIC WIND SPEED AND DATABASE

Variable	Database	Synthetic
Mean (wind speed)	8.078	8.088
Standard deviation (wind speed)	3.392	3.372
Mean (step to step difference)	0	0
Standard deviation (step to step difference)	0.659	0.642

TABLE II
COMPARISON BETWEEN SYNTHETIC ELECTRIC DEMAND AND DATABASE

Variable	Database	Synthetic
Mean (demand)	1102.3	1103.4
Standard deviation (demand)	222.2	223.8
Mean (step to step difference)	0	0
Standard deviation (step to step difference)	48.4	54.2

and database are compared in Fig. 7, suggesting good match between them.

B. Probabilistic Analysis of HES

To demonstrate the benefit of the synthetic scenarios, Fig. 8 plots the actual wind speed together with 50 synthetic scenarios, while Fig. 9 plots the actual electric demand with 50 synthetic scenarios, both for 48 hours. As can be seen, for both wind speed and electric demand, each synthetic scenario possesses very different temporal profile from the database. Considering that the synthetic data possesses same statistical characteristics with database while having different temporal profile, they can be used for Monte Carlo simulation and probabilistic analysis of energy integration systems, while avoiding bias introduced by using the same database. In the rest of this sections 3000 synthetic wind speed and electric demand scenarios will be generated to simulate the HES configuration introduced in Section III. In all of the simulations, the HES is operated in such a way that the nuclear power plant provides constant maximum output of 180 MW.

Fig. 10 plots the turbine output, grid demand, and RO set point for 7 days. As can be seen, in order to follow the variability in grid demand while absorbing the volatility in wind turbine output, the RO plant needs to be operated

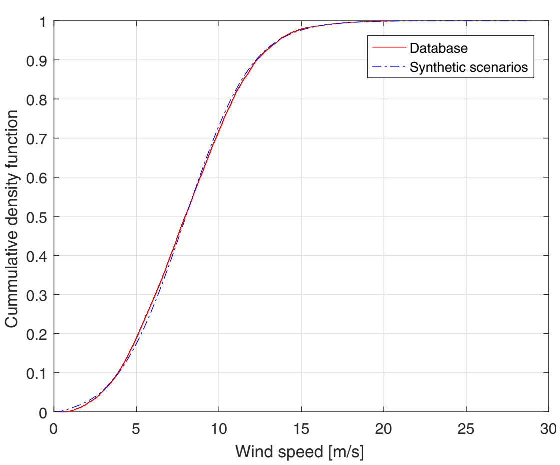

Fig. 5. Empirical CDF of synthetic wind speed and actual database.

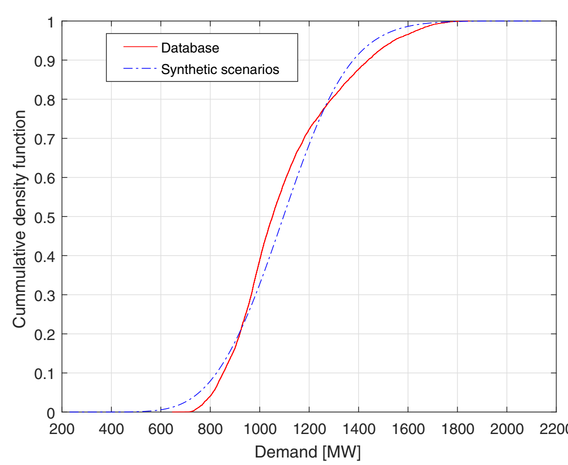

Fig. 7. Empirical CDF of synthetic electric demand and actual database.

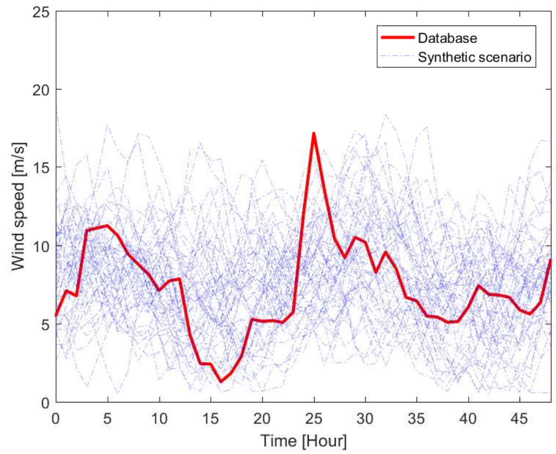

Fig. 8. 50 synthetic wind speed scenarios and database for 48 hours.

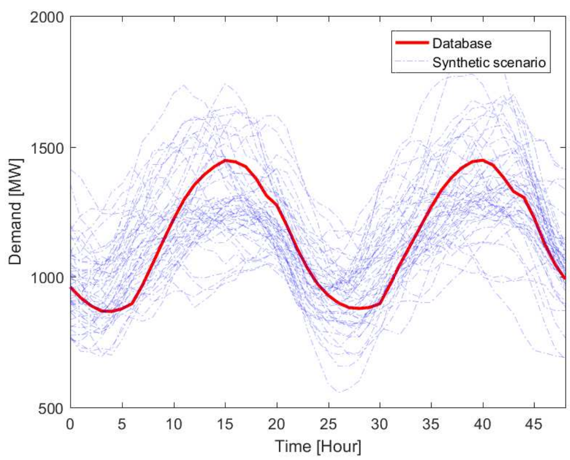

Fig. 9. 50 synthetic electric demand scenarios and database for 48 hours.

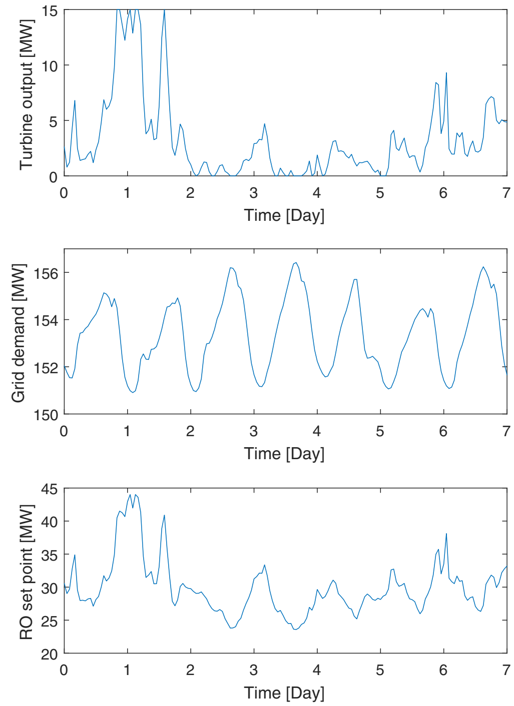

Fig. 10. Turbine output, grid demand, and RO set point for 7 days.

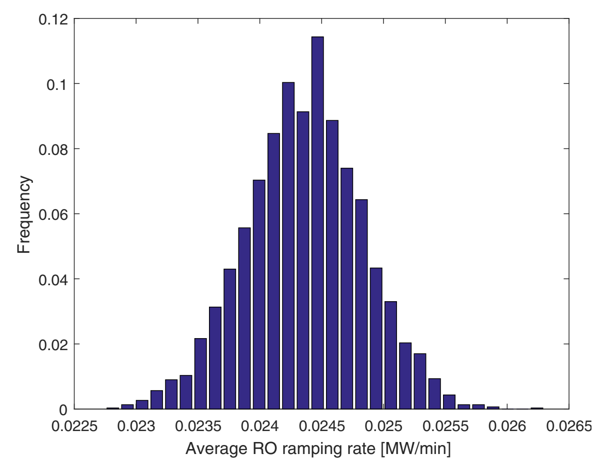

Fig. 11. Histogram of yearly average RO ramping rate for 3000 synthetic scenarios.

dynamically (since the SMR is operated in constant full production mode). Fig. 11 shows the histogram of yearly average ramping rate of RO for 3000 synthetic scenarios, while Fig. 12 shows the histogram of hourly fresh water production for a whole year using database. In particular, it requires RO to ramp averagely 24 kW/min in order to absorb the volatility. As discussed in [5]–[7], a battery storage can be used to smooth the variability of the wind farm production before sending the renewable power to HES. To analyze the effects of such battery storage, multiple Monte Carlo simulations are performed, each with different battery storage capacity, namely, no battery, 5 MWh, 10 MWh, and 15 MWh. Fig. 13 plots the box plots of maximum yearly rates on RO ramping up and down, suggesting less RO ramping is needed if larger battery is employed to smooth renewable generation. Fig. 13 also suggests that, for this particular HES configuration, the battery has larger effects on relaxing the RO ramping down requirement than that of ramping up. Finally, Fig. 14 plots the box plots of RO hourly fresh water production under synthetic scenario for different battery sizes. Since the hourly production depends on the total energy consumption within that hour instead of its volatility, it is only marginally affected by the battery size.

V. Conclusion

This paper proposed a computational model, based on Fourier series and autoregressive moving average, to generate synthetic wind speed and electric demand data, which are shown to possess the same statistical characteristics with historical measurements. Probabilistic analysis of a particular

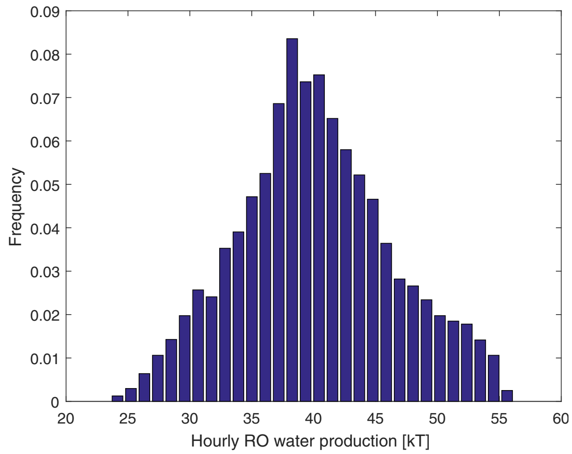

Fig. 12. Histogram of RO hourly fresh water production under actual database.

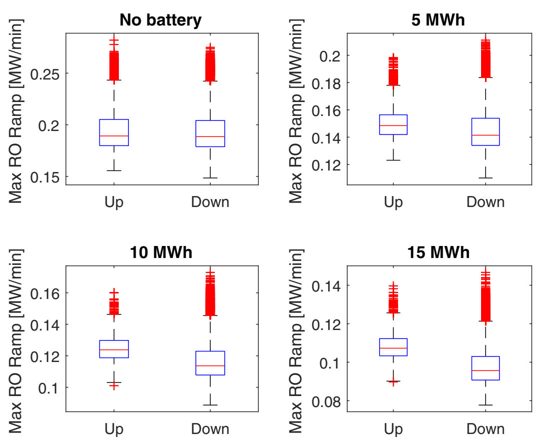

Fig. 13. Box plots of yearly maximum RO ramping up and down rates for 3000 synthetic scenarios with different battery size.

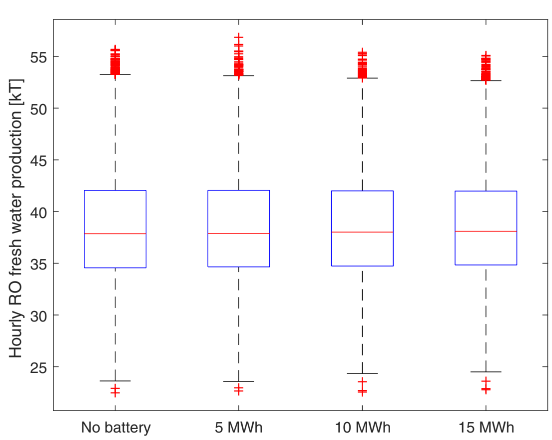

Fig. 14. Box plot of RO hourly fresh water production under synthetic scenario for different battery sizes.

hybrid energy systems configuration was performed based on synthetic data, to understand the component ramping requirement and the effects of using larger battery storage. Such analysis over synthetic data, which are statistically conformed to database while possessing different temporal profile, avoids the bias introduced by using the same database. Future work include synthetic data generation for other renewable resources such as solar and hydro power.

REFERENCES

[1] H. E. Garcia, A. Mohanty, W. Lin, and R. Cherry, "Dynamic analysis of hybrid energy systems under flexible operation and variable renewable generation - part i: Dynamic performance analysis," *Energy*, vol. 52, pp. 1–16, 2013.

[2] ——, "Dynamic analysis of hybrid energy systems under flexible operation and variable renewable generation - part ii: Dynamic cost analysis," *Energy*, vol. 52, pp. 17–26, 2013.

[3] J. Chen and H. E. Garcia, "Operations optimization of hybrid energy systems under variable markets," in *Proc. 2016 American Control Conference*. Boston, MA: IEEE, July, 2016, pp. 3212–3218.

[4] J. S. Kim, J. Chen, and H. E. Garcia, "Modeling, control, and dynamic performance analysis of a reverse osmosis desalination plant integrated within hybrid energy systems," *Energy*, vol. 112, pp. 52–66, 2016.

[5] H. E. Garcia, J. Chen, J. S. Kim, R. B. Vilim, W. R. Binder, S. M. B. Sitton, R. D. Boardman, M. G. McKellar, and C. J. Paredis, "Dynamic performance analysis of two regional nuclear hybrid energy systems," *Energy*, vol. 107, pp. 234–258, 2016.

[6] J. Chen, H. E. Garcia, J. S. Kim, and S. M. Bragg-Sitton, "Operations optimization of nuclear hybrid energy systems," *Nuclear Technology*, vol. 195, no. 2, pp. 143–156, 2016.

[7] J. Chen and H. E. Garcia, "Economic optimization of operations for hybrid energy systems under variable markets," *Applied Energy*, vol. 177, pp. 11–24, 2016.

[8] P. Meibom, R. Barth, B. Hasche, H. Brand, C. Weber, and M. O'Malley, "Stochastic optimization model to study the operational impacts of high wind penetrations in ireland," *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 1367–1379, 2011.

[9] J. M. Morales, R. Minguez, and A. J. Conejo, "A methodology to generate statistically dependent wind speed scenarios," *Applied Energy*, vol. 87, no. 3, pp. 843–855, 2010.

[10] A. Papavasiliou, S. S. Oren, and R. P. O'Neill, "Reserve requirements for wind power integration: A scenario-based stochastic programming framework," *IEEE Trans. Power Syst.*, vol. 26, no. 4, pp. 2197–2206, 2011.

[11] X.-Y. Ma, Y.-Z. Sun, and H.-L. Fang, "Scenario generation of wind power based on statistical uncertainty and variability," *IEEE Trans. Sustainable Energy*, vol. 4, no. 4, pp. 894–904, 2013.

[12] J. W. Taylor, P. E. McSharry, and R. Buizza, "Wind power density forecasting using ensemble predictions and time series models," *IEEE Trans. Energy Conversion*, vol. 24, no. 3, pp. 775–782, 2009.

[13] L. J. Soares and M. C. Medeiros, "Modeling and forecasting short-term electricity load: A comparison of methods with an application to brazilian data," *International Journal of Forecasting*, vol. 24, no. 4, pp. 630–644, 2008.

[14] K. K. Sumer, O. Goktas, and A. Hepsag, "The application of seasonal latent variable in forecasting electricity demand as an alternative method," *Energy policy*, vol. 37, no. 4, pp. 1317–1322, 2009.

[15] E. Gonzalez-Romera, M. A. Jaramillo-Moran, and D. Carmona-Fernandez, "Monthly electric energy demand forecasting based on trend extraction," *IEEE Trans. Power Syst.*, vol. 21, no. 4, pp. 1946–1953, 2006.

[16] N. Amjady and F. Keynia, "Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm," *Energy*, vol. 34, no. 1, pp. 46–57, 2009.

[17] N. Steckler, A. Florita, J. Zhang, and B. Hodge, "Analysis and synthesis of load forecasting data for renewable integration studies," in *Proc. 12th International Workshop on Large-Scale Integration of Wind Power into Power Systems*, London, England, Oct. 22-24, 2013.

[18] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.