Generating synthetic wind speed scenarios using artificial neural networks for probabilistic analysis of hybrid energy systems

Jun Chen*

Department of Electrical and Computer Engineering,
Oakland University,
Rochester, MI 48309, USA
Email: junchen@oakland.edu
*Corresponding author

Junhui Zhao

Department of Electrical, Computer Engineering and Computer Science,
University of New Haven,
West Haven, CT 06516, USA
Email: jzhao@newhaven.edu

Abstract: Hybrid energy systems (HESs) have been proposed to include and co-optimise multiple energy inputs and multiple energy outputs to enable increasing penetration of clean energy such as wind power. To optimise the system design, extensive datasets of renewable resources for the given location are required, whose availability may be limited. To address this limitation, this paper proposes an innovative methodology to generate synthetic wind speed data. Specifically, artificial neural networks are adopted to characterise historical wind speed data and to generate synthetic scenarios. In addition, Fourier transformation is used to capture the characteristics of the low frequency components in historical data, allowing the synthetic scenarios to preserve seasonal trends. The proposed methodology enables the possibility of Monte Carlo simulation of HES for probabilistic analysis using large volumes of heterogeneous scenarios. Case study of probabilistic analysis is then performed on a particular HES configuration, which includes nuclear power plant, wind farm, battery storage, electric vehicle charging station, and desalination plant. Wind power availability and requirements on component ramping rate are then investigated.

Keywords: artificial neural networks; ANNs; Fourier transformation; hybrid energy systems; HESs; synthetic scenarios; wind energy.

Reference to this paper should be made as follows: Chen, J. and Zhao, J. (2022) 'Generating synthetic wind speed scenarios using artificial neural networks for probabilistic analysis of hybrid energy systems', *Int. J. Modelling, Identification and Control*, Vol. 41, No. 3, pp.183–192.

Biographical notes: Jun Chen received his BS in Automation from the Zhejiang Uniersity, Hangzhou, China, in 2009, and PhD in Electrical Engineering from the Iowa State University, Ames, IA, USA, in 2014. He was with Idaho National Laboratory from 2014 to 2016, and with General Motors from 2017 to 2020. He joined Oakland University in 2020, where he is currently an Assistant Professor at the ECE Department. His research interests include advanced control and optimisation, model predictive control, machine learning, with applications in automotive and energy systems. He is currently a senior member of the IEEE and a member of SAE.

Junhui Zhao received his MSc and PhD in Electrical Engineering in 2009 and 2014, respectively from the Chongqing University, Chongqing, China and Wayne State University, Detroit, USA. He is currently an Associate Professor of University of New Haven, USA. His current research interests include modelling and control of distributed generation (DG) sources, operation of smart grid with high penetration of DG and energy storage, and voltage stability of power systems.

This paper is a revised and expanded version of a paper entitled 'Synthetic wind speed scenarios generation using artificial neural networks for probabilistic analysis of hybrid energy systems' presented at 2021 IEEE International Symposium on Industrial Electronics (ISIE), Kyoto, Japan, 20–23 June 2021.

1 Introduction

Hybrid energy systems (HESs) have been proposed in literature (Chen and Garcia, 2016a; Kim et al., 2016; Garcia et al., 2016; El Fadil et al., 2020; Chen et al., 2016, 2021a; Chen and Garcia, 2016b; Zioui and Mahmoudi, 2020; Di Silvestre et al., 2014; Zhu et al., 2015; Graditi et al., 2015) to enable higher level of renewable energy penetration. HES usually consist of multiple energy generations and utilisation units, and have been shown to be flexible to accommodate the high variability introduced from renewable generation, modern loads [such as electric vehicles (EVs)], and electric markets (Kim et al., 2016; Garcia et al., 2016; Chen et al., 2016; Chen and Garcia, 2016b). In addition, HES can also participate in both day-ahead and real-time electricity markets, as well as ancillary service markets (Chen and Garcia, 2016b) to increase their economic viability and improve power systems reliability. However, these prior analyses were performed based on historical measurements data on renewable energy, whose availability is very limited for a given location. For example, the studies carried out in Chen and Garcia (2016b) were based on one year of wind speed data, and therefore the conclusion drawn there could be strongly biased towards that particular dataset. To address the data scarcity issue, this paper proposes an innovative methodology to generate synthetic wind speed scenarios, which need to be statistically conformed to historical measurements.

The topic of generating synthetic wind speed scenarios has been studied in the literature to some extent. For example, Meibom et al. (2011) uses autoregressive moving average (ARMA) model to generate synthetic residues, which are then added to the historical data to provide dataset with different waveforms. Similarly, Chen and Rabiti (2017), Chen et al. (2017), Morales et al. (2010), Papavasiliou et al. (2011) and Ma et al. (2013) use ARMA or AR model, together with sampled white noise, to generate scenarios. One of the assumptions of ARMA is that the underlying time series is normally distributed. To satisfy such assumption, the measurement data needs to be recast into Gaussian distribution before being used to train ARMA model. On the other hand, artificial neural networks (ANN) have also been used for this task in literature. For example, Gonzalez-Romera et al. (2006) decomposes the load data into two components, one with low frequency to represent deterministic seasonal trend and the other with high frequency to represent the intermittency of the renewable resources. Both high and low frequency components are used to train two ANN, one for each part. Amjady and Keynia (2009) combines wavelet transform with ANN approach, while Steckler et al. (2013) adopts Bayesian belief network to improve load prediction. Fourier transform has also been utilised to capture the seasonal trend from the low frequency components (Soares and Medeiros, 2008; Sumer et al., 2009). In addition, Gaussian process has also been found to be effective for scenarios generation. For example, Mori and Kurata (2008) trains a Gaussian process model based on metrological data and wind speed measurements, and the trained model can then be used for prediction. Note that Lee and Baldick (2014a) adopts similar methodology. Even though this work focuses on point prediction rather than scenarios generation, it is worth noting that the reported methodologies can be straightforwardly extended for scenarios generation, since the estimated model essentially characterises a distribution. Furthermore, Lee and Baldick (2013) uses factor analysis for scenario synthesis, which statistically models the variability among observed correlated variables in terms of a potentially lower number of unobserved variables. Finally, Lee and Baldick (2014b) utilises the power spectrum density (PSD), as extracted from measurement data, to synthesise PSD based on future capacity, and generates sample waveform by inverse fast Fourier transformation of synthesised PSD.

Though the aforementioned works provide good synthetic results, they do require domain knowledge about the data being generated, and the methodologies are generally portable from one dataset to another. In this paper, a new methodology to generate synthetic wind speed scenarios is proposed. In particular, ANN is utilised to characterise historical wind speed measurements and then to synthesise scenarios. Furthermore, the seasonal trend in the historical data is isolated by using Fourier transformation, allowing the synthetic scenarios to preserve seasonal trend. After training the model over historical data by finding optimal parameters, the combined model (ANN and Fourier series) can then be used to generate synthetic scenarios by first generating residue terms for each time step using the trained ANN model and then adding the Fourier terms representing low frequency components (seasonal trends). Since the synthetic high frequency components have introduced sufficient variability compared to dataset, the low frequency components, as identified by Fourier transformation, are adopted from historical data, i.e., no training and model evaluation are required and hence saving computation. To validate the quality of the synthetic wind speed scenario, statistical analysis (e.g., mean, variance, and empirical cumulative distribution function) as well as frequency analysis are performed to compare the synthetic scenario to the historical measurements. The generated synthetic wind speed scenarios will in turn be utilised to analyse a particular HES configuration, which includes nuclear power plant, wind farm, battery storage, EV charging station, and desalination plant. Wind power availability and requirements on component ramping rate are then investigated.

Note that our approach differs from Gonzalez-Romera et al. (2006), which uses ANN to model both high frequency and low frequency components of electricity load data. It is also worth noting that our approach differs from that of Chen et al. (2018) which use generative adversal net (GAN) to generate wind power profile instead of wind speed. Since wind turbine can be seen as a low pass filter, wind power usually has lower variation compared to wind speed profile. Therefore, the problem addressed in Chen et al. (2018) can be considered as a less difficult task. Furthermore, we demonstrate that a

shallow network is sufficient to model the wind speed, hence saving computational time compared to the GAN approach as adopted in Chen et al. (2018). A preliminary version (Chen and Zhao, 2021) of this paper has been presented at 2021 IEEE International Symposium on Industrial Electronics (ISIE), held 20–23 June 2021 in Kyoto, Japan. This paper differs from its conference version by complementing literature review, including more details on algorithm description, and adding more simulation results to demonstrate the effectiveness of the proposed methodology.

The rest of this paper is organised as follows. Section 2 presents the algorithms to train and evaluate ANN for generating synthetic wind speed scenarios, together with numerical results, while Section 3 presents the revised algorithms to extract low frequency components from historical dataset. Probabilistic analysis on HES is discussed in Section 4, and the paper is concluded in Section 5.

2 Synthetic wind speed scenarios based on ANN

We start this section by giving a brief introduction of ANNs, which is later employed to capture the nonlinear autocorrelation function of wind speed. Readers can refer to Bishop et al. (1995) and Mohri et al. (2018) for more details.

2.1 Artificial neural network

ANN is a popular machine learning technique for modelling complex nonlinear relationship between inputs and outputs (Bishop et al., 1995). Figure 1 shows a typical feedforward ANN model with two inputs, one output, and multiple hidden layers (HLs). Each neuron in ANN receives multiple inputs and produces one output. The arrow from neuron i to neuron j indicates that the output of neuron i becomes one of the inputs to neuron j. Denote x as the input to a neuron i, its output y is computed according to the following activation function

$$y = f_{ac}(w_i^T x + b_i)$$

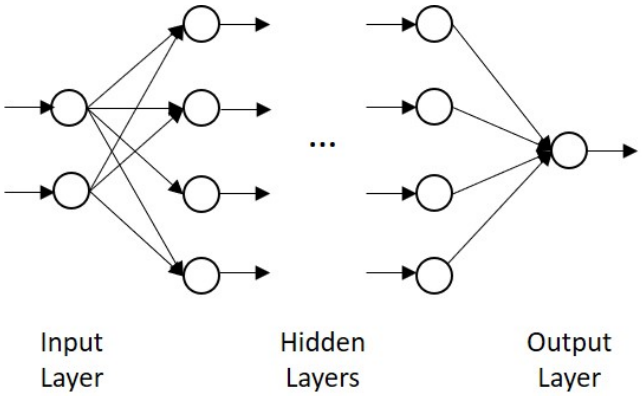
where the weights w_i and bias b_i are model parameters to be identified from data through training process. Denote the number of input neurons as p, and the number of output neurons as l, to train a neural network requires training data consisting of input $X_{p\times N}$ and output $Y_{l\times N}$, where N is the number of samples. Given a uni-variate time series $\mathbf{x} = x_1, x_2, ..., x_L$, in order to capture the autocorrelation, one can construct the following training data with $N = L - p$ and $l = 1$, where p can be chosen as the lag of autocorrelation,

$$X_{p\times(L-p)} = \begin{bmatrix} x_1 & x_2 & \cdots & x_{L-p} \\ x_2 & x_3 & \cdots & x_{L-p+1} \\ \vdots & \vdots & & \vdots \\ x_p & x_{p+1} & \cdots & x_{L-1} \end{bmatrix} \quad (1)$$

$$Y_{1\times(L-p)} = \begin{bmatrix} x_{p+1} & x_{p+2} & \cdots & x_L \end{bmatrix} \quad (2)$$

In this paper, the MATLAB Deep Learning Toolbox is utilised to implement, train, and evaluate ANN model. Specifically, tangent hyperbolic function is used as activation function, i.e., $y = \tanh(w_i^T x + b_i)$, and Levenberg-Marquardt algorithm (Levenberg, 1944; Marquardt, 1963) is utilised to train the ANN model.

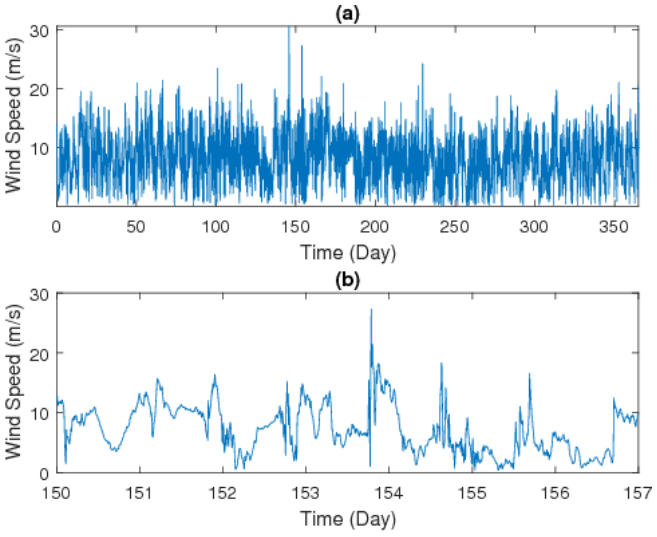
Figure 1 An ANN model with two inputs, one output, and multiple HLs



2.2 Historical dataset and ANN Training

To capture the autocorrelation of wind speed, a one-year historical wind speed dataset[1] is utilised to train ANN, which is a discrete time uni-variate time series with five minutes interval. Figure 2(a) shows the time profile of the wind speed being used, while Figure 2(b) shows the time profile of wind speed for a selected period of seven days (e.g., day 150–156 of the year).

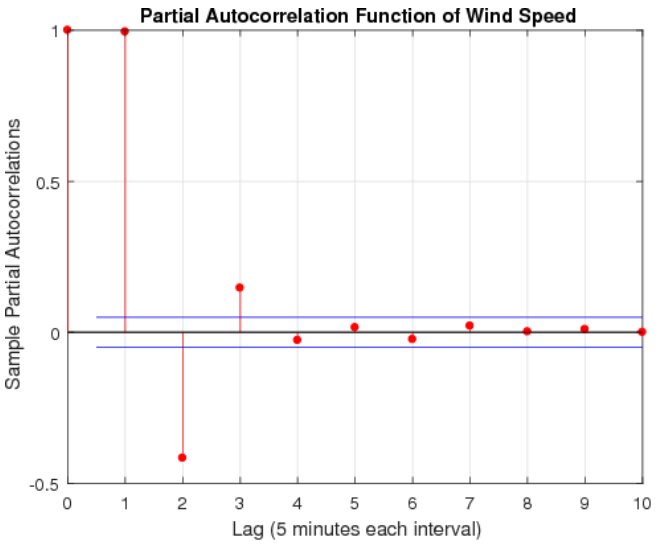
Figure 2 (a) Wind speed data for one year (b) Wind speed data for a selected period of seven days (see online version for colours)



To train ANN, this wind speed dataset, denoted as $\mathbf{x}_w = x_1, x_2, ..., x_L$, where L = 105,120, can be transformed into training input $X_{p\times(L-p)}$ and output Y_{L-p} according

to equations (1) and (2). The number of input neurons, p, which captures the lag of the autocorrelation of wind speed, can be selected based on the partial autocorrelation function of wind speed. In particular, Figure 3 plots the empirical partial autocorrelation function of $\mathbf{x}_w$. Each lag indicates a five minutes interval, corresponding to the highest resolution available in the historical dataset. The blue lines correspond to ± 0.05, indicating 5% partial autocorrelation. As indicated in Figure 3, a lag of 3 is sufficient to capture the autocorrelation of the data.

Figure 3 Sample partial autocorrelation function of wind speed data (see online version for colours)



Notes: Each lag represents 5-minute interval. The blue lines correspond to the value of ± 0.05, indicating a lag of 3 is sufficient to capture the autocorrelation of the data.

Remark 1: It is a well known fact that wind speed is not a stationary process, which makes it impractical to use time series techniques like auto-regressive moving average (ARMA) for its modelling, as it would require the underlying time series to possess stationarity property. Note that in Chen and Rabiti (2017); Chen et al. (2017), an ARMA model is used to fit into wind speed data and used for synthetic data generation. Despite of the reasonable results, the approach studied in Chen and Rabiti (2017); Chen et al. (2017) clearly violates the theoretical foundation of ARMA.

After construction of training dataset $X_{p\times(L-p)}$ and Y_{L-p}, an ANN can be trained by using MATLAB Deep Learning Toolbox. In particular, the feedforward ANN used in this paper consists of one input layer, two HLs, and one output layer, with $3p$ neurons on each HL. Algorithm 1 summarises the procedure for data preprocessing and model training for synthetic wind speed scenarios generation.

Remark 2: Note that the primary contribution of this work is to propose a framework for wind speed scenarios generation. Hence a simple feedforward ANN is used in Algorithm 1 for the sake of simplicity. The proposed framework and Algorithm 1 are flexible to accommodate more complex ANN architecture, such as recurrent neural networks, which remains as a future work. See conclusions section.

Algorithm 1 Algorithm to train ANN for generating synthetic wind speed scenarios

procedure SYNTHETIC_WIND_TRAIN($\mathbf{x}_w$)
  Compute empirical partial autocorrelation `pacf` of $\mathbf{x}_w$;
  $p \leftarrow \arg\max_k |\texttt{pacf}(k)| > 0.05$;
  Construct $X_{p\times(L-p)}$ and Y_{L-p} according to equations (1)–(2);
  Instantiate `net` $\leftarrow$ an ANN with p inputs and 1 output;
  Train `net` on $X_{p\times(L-p)}$ and Y_{L-p};
  return `net`
end procedure

2.3 Synthetic scenarios generation using trained ANN

Once the ANN is trained, it can be utilised to generate synthetic wind speed scenarios by recursively evaluating it over the past p data point in the synthetic time series. Denote the synthetic wind speed data as $\hat{\mathbf{x}}_w$, and assume the first p elements (i.e., $\hat{x}_1, ..., \hat{x}_p$) are given (whose generation will be detailed shortly). Then for $k = p+1, ..., L$,

$$\hat{x}_k = \texttt{net}(\left[\hat{x}_{k-p} \cdots \hat{x}_{k-1}\right]^T) + \sigma_k \tag{3}$$

where σ_k denotes a random noise to avoid degeneration.

Remark 3: Note that the purpose of adding σ_k in equation (3) is to sufficiently excite the recursive generation process. Otherwise the recursion may degenerate and the synthesised time series may converge to a constant value. Note that σ_k can be sampled either through a pre-selected distribution, or through the error vector between the trained ANN (denoted as `net`) and Y_{L-p}. Denote the error vector $\mathbf{e}$ with length $L-p$, then for $k = 1, ..., L-p$,

$$e_k = \texttt{net}(\left[x_k \cdots x_{k+p-1}\right]^T) - x_{k+p}. \tag{4}$$

Algorithm 2 summarises the procedure that generates N synthetic wind speed scenarios based on trained ANN `net`.

Algorithm 2 Algorithm to generate synthetic wind speed scenarios

procedure SYNTHETIC_WIND_GENERATE(`net`, N)
  $\mathbf{e} \leftarrow \texttt{net}(X_{p\times(L-p)}) - Y_{L-p}$;
  for $n = 1, ..., N$ **do**
    Obtain $\mathbf{e}_n$ by sampling from elements of $\mathbf{e}$ for L times;
    for $k = 1, ..., p$ **do**
      $\hat{\mathbf{x}}_n(k) \leftarrow \mathbf{x}_w(k) + \mathbf{e}_n(k)$
    end for
    for $k = p+1, ..., L$ **do**
      $d \leftarrow \left[\hat{\mathbf{x}}_n(k-p) \cdots \hat{\mathbf{x}}_n(k-1)\right]^T$
      $\hat{\mathbf{x}}_n(k) \leftarrow \texttt{net}(d) + \mathbf{e}_n(k)$
    end for
  end for
  $\hat{\mathbf{x}} \leftarrow \left[\hat{\mathbf{x}}_1 \cdots \hat{\mathbf{x}}_N\right]$
  return $\hat{\mathbf{x}}$
end procedure

2.4 Numerical results and discussion

This section presents numerical results of the procedures presented in Algorithms 1 and 2. Recall that the model depth p is selected to be 3, and the ANN consists of two HLs, each of which has nine neurons.

Figure 4 shows the model error **e** of the trained ANN, which will be sampled to generate the noise signal in equation (3). Figure 5 compares seven days of the synthetic wind speed and that of the historical dataset. As can be seen, the synthetic scenario presents completely different time profile compared to that of the dataset, which can be utilised to perform probabilistic analysis of wind power generation.

Figure 4 Histogram of ANN model errors that are within 5 and 95 percentile (see online version for colours)

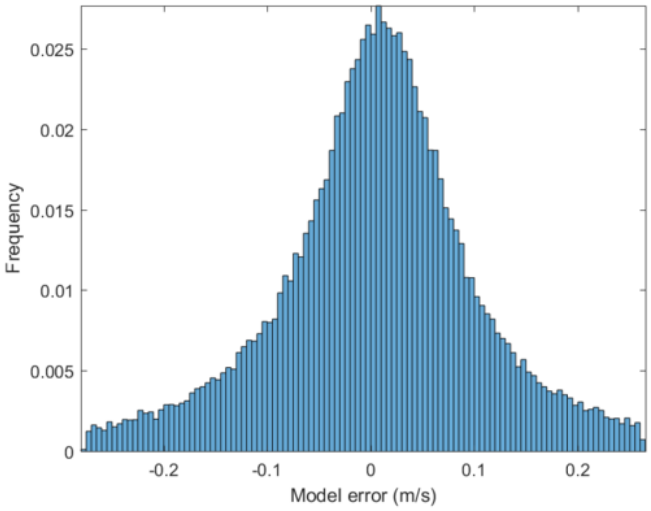
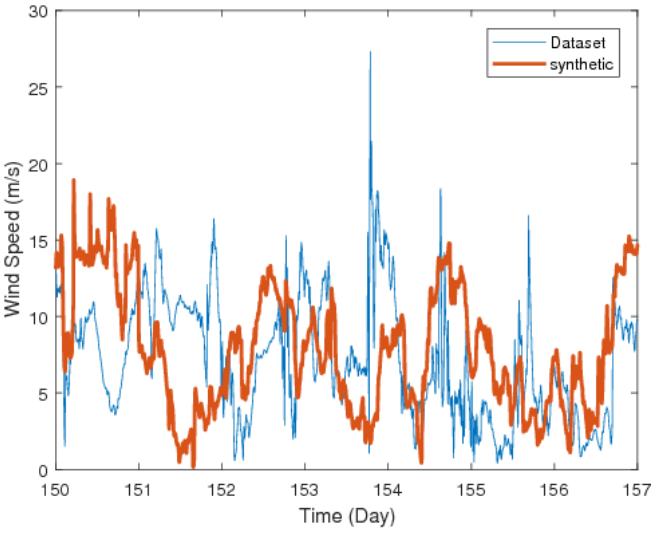


Figure 5 Synthetic wind speed versus historical dataset for a selected period of seven days (see online version for colours)



Despite the completely different time profiles, the synthetic wind speed and historical dataset actually possess similar statistical characteristics. Several key statistics of the synthetic scenarios and dataset are summarised in Table 1, showing good alignment between the synthetic scenarios and dataset. Furthermore, Figure 6 plots the empirical cumulative density function (CDF) between synthetic wind speed and historical dataset, where satisfactory statistical conformance can also be found, despite of the completely different time profile as shown in Figure 5. Furthermore, Table 2 compares the power density of the synthetic scenarios and historical dataset, again showing good alignment between the two.

Figure 6 Empirical CDF for synthetic wind speed versus dataset (see online version for colours)

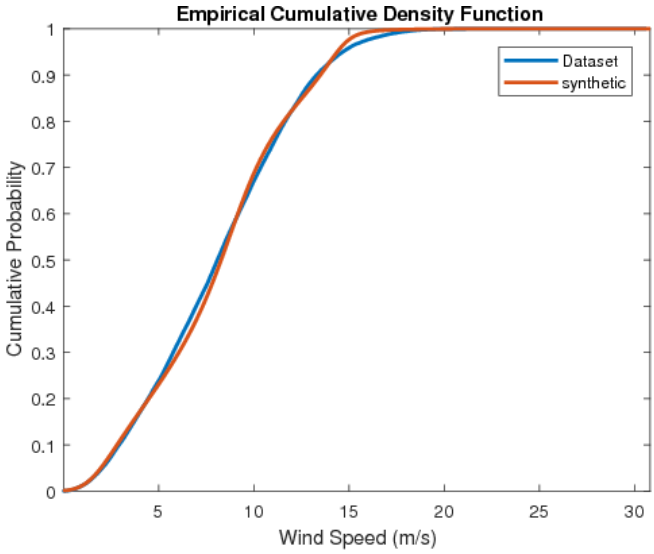


Table 1 Comparison of key statistics of synthetic scenarios and historical dataset

Stats	*Dataset*	*Synthetic*
Max	30.60	30.82
95 percentile	14.67	14.38
Mean	8.16	8.16
Median	8.09	8.29
Standard deviation	3.91	3.78
5 percentile	2.04	1.97
Min	0.02	0.01

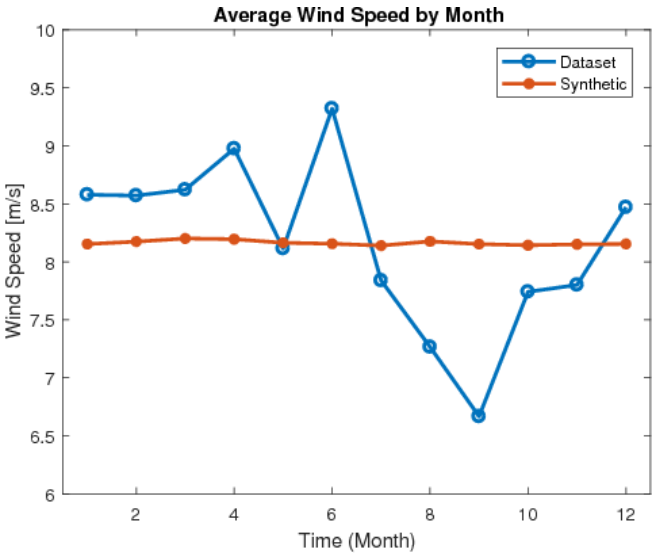
Table 2 Comparison of power density of synthetic scenarios and historical dataset

Period	*Dataset (dB/Hz)*	*Synthetic (dB/Hz)*
≥3 months	75.53	72.47
≥1 month	62.18	57.12
≥1 week	59.20	60.24
≥1 day	55.27	55.54
≥1 hour	33.06	33.45
≥0.5 hour	21.70	22.30
<0.5 hour	10.50	9.92

Despite the good statistical conformance discussed above, the proposed procedure does present one drawback. As can be noticed from Algorithms 1 and 2, the timing

information (e.g., time of the year) is not taken into account when generating the synthetic scenarios. The lack of such information does not impact the overall statistical distribution; however, it does impact the monthly (or even weekly) trend that the historical dataset presents. In particular, Figure 7 compares the monthly average wind speed of synthetic scenarios and historical dataset. It is clear that the synthetic wind speed does not preserve the seasonal trend that exists in the dataset. Note that this may not be an issue for analysis that does not depend on seasonal trend, but for analysis that does, such limitation can be problematic. To overcome this, in the following section, we proposed a revised procedure that removes the low frequency components from the training and generation process.

Figure 7 Monthly trend of the historical dataset and synthetic scenario (see online version for colours)



3 Synthetic wind speed scenarios generation preserving seasonal trend

In this section, we propose a revised procedure that removes the low frequency components from the training and generation process. In other words, the ANN is only used to generate synthetic high frequency components, while the low frequency components are taken from the historical dataset. We start this section by giving a brief introduction of discrete Fourier transform (DFT) and fast Fourier transform (FFT). More details can be found in Oppenhiem et al. (1983).

3.1 Discrete Fourier transform

Given a time series $\mathbf{x} = x_0, x_1, ..., x_{L-1}$, the DFT transforms it into another sequence $\bar{\mathbf{x}} = \bar{x}_0, \bar{x}_1, ..., \bar{x}_{L-1}$ defined by

$$\bar{x}_k = \sum_{n=0}^{L-1} x_n e^{-\frac{i2\pi}{L}kn}$$

$$= \sum_{n=0}^{L-1} x_n \left[\cos\left(\frac{2\pi}{L}kn\right) - i\sin\left(\frac{2\pi}{L}kn\right)\right].$$

Denote this transform by $\mathcal{F}$, we have $\bar{\mathbf{x}} = \mathcal{F}(\mathbf{x})$. The inverse DFT, denoted as $\mathcal{F}^{-1}$, transforms $\bar{\mathbf{x}}$ back to $\mathbf{x}$, i.e., $\mathbf{x} = \mathcal{F}^{-1}(\bar{\mathbf{x}})$, and can be computed according to

$$x_k = \frac{1}{N}\sum_{n=0}^{L-1} \bar{x}_n e^{\frac{i2\pi}{L}kn}$$

$$= \frac{1}{N}\sum_{n=0}^{L-1} \bar{x}_n \left[\cos\left(\frac{2\pi}{L}kn\right) + i\sin\left(\frac{2\pi}{L}kn\right)\right].$$

It is easy to see that $\bar{x}_n$ denotes the amplitude of the component at period of $\frac{L}{n}T_s$ (or equivalently at frequency of $\frac{n}{LT_s}$), where T_s is the sampling interval of the time series $\mathbf{x}$.

Evaluating $\mathcal{F}$ according to its definition requires computational complexity of $O(N^2)$, while an FFT algorithm can compute it with complexity of $O(N \log N)$. In this paper, the MATLAB implementation `fft` and `ifft` are utilised to compute $\mathcal{F}$ and $\mathcal{F}^{-1}$, respectively, which are based on Frigo and Johnson (1998).

3.2 Revised algorithm preserving seasonal trend

According to Subsection 3.1, given a wind speed time series $\mathbf{x}_w$ with length L, its Fouries transformation can be computed by $\bar{\mathbf{x}}_w = \mathcal{F}(\mathbf{x}_w)$ such that the nth element $\bar{x}_n$ of $\bar{\mathbf{x}}_w$ indicates the amplitude of the component at frequency $\frac{n}{LT_s}$. Given q such that $\frac{q}{LT_s}$ (approximately) equals to a cutoff frequency f_c, and

$$\bar{\mathbf{x}}_L = \left[\bar{x}_1, ...\bar{x}_q, 0, ..., 0\right] \quad (5)$$

$$\bar{\mathbf{x}}_H = \left[\bar{0}, ...\bar{0}, \bar{x}_{q+1}, ..., \bar{x}_L\right], \quad (6)$$

then the high frequency components $\mathbf{x}_H$ and low frequency components $\mathbf{x}_L$ of $\mathbf{x}_w$ can be computed by

$$\mathbf{x}_L = \mathcal{F}^{-1}(\bar{\mathbf{x}}_L) \quad (7)$$

$$\mathbf{x}_H = \mathcal{F}^{-1}(\bar{\mathbf{x}}_H). \quad (8)$$

Figure 8 shows the low frequency and high frequency components of the historical wind speed dataset, where f_c = 1.649 μHz, corresponding to a period of one week.

Once $\mathbf{x}_H$ and $\mathbf{x}_L$ are obtained, an ANN `net_h` can be trained over $\mathbf{x}_H$ utilised Algorithm 1, which can then be used to generate synthetic high frequency components using Algorithm 2. The synthetic high frequent components are then shifted by $\mathbf{x}_L$ to obtain the synthetic wind speed. This procedure is summarised in Algorithms 3 and 4.

3.3 Numerical results and discussion

Figures 9 and 10 summarise the results using Algorithms 3 and 4 with f_c = 1.649 μHz. In particular, Figure 9 plots

the synthetic wind speed versus historical dataset for a selected period of seven days, where only the high frequency components are synthetic. Figure 10 compares the monthly average wind speed of synthetic scenarios and historical dataset, which shows that the synthetic wind speed can preserve the monthly trend exhibited in the dataset. Therefore, the proposed algorithms in this section can generate a set of wind speed profiles that possess rich variation in terms of time series, while at the same time preserves the seasonal trends exhibited in the dataset.

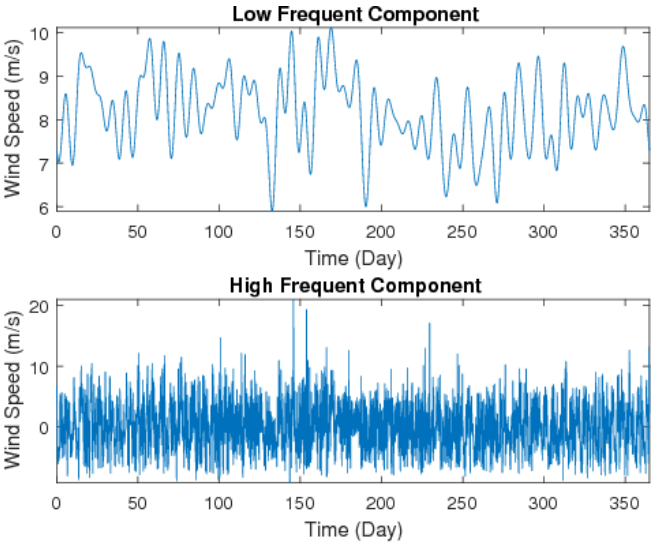
Algorithm 3 Algorithm to train ANN using only high frequency components

procedure SYNTHETIC_WIND_HIGH_TRAIN($\mathbf{x}_w$, f_c)
  $\bar{\mathbf{x}}_w \leftarrow \mathcal{F}(\mathbf{x}_w)$; $\bar{\mathbf{x}}_H \leftarrow \bar{\mathbf{x}}_w$; $\bar{\mathbf{x}}_L \leftarrow \bar{\mathbf{x}}_w$;
  $q \leftarrow \lceil f_c L T_s \rceil$;
  for $k = 1, ..., q$ **do**
    $\bar{\mathbf{x}}_H(k) \leftarrow 0$
  end for
  for $k = q+1, ..., L$ **do**
    $\bar{\mathbf{x}}_L(k) \leftarrow 0$
  end for
  $\mathbf{x}_L \leftarrow \mathcal{F}^{-1}(\bar{\mathbf{x}}_L)$; $\mathbf{x}_H \leftarrow \mathcal{F}^{-1}(\bar{\mathbf{x}}_H)$
  net_h $\leftarrow$ SYNTHETIC_WIND_TRAIN($\mathbf{x}_H$)
  return net_h, $\mathbf{x}_L$
end procedure

Algorithm 4 Revised algorithm to generate synthetic wind speed scenarios

procedure SYNTHETIC_WIND_HIGH_GENERATE(net_h, $\mathbf{x}_L$, N)
  for $n = 1, ..., N$ **do**
    $\hat{\mathbf{x}}_1 \leftarrow$ SYNTHETIC_WIND_GENERATE(net_h, 1)
    $\hat{\mathbf{x}}_1 \leftarrow \mathbf{x}_L + \hat{\mathbf{x}}_1$
  end for
  $\hat{\mathbf{x}} \leftarrow \left[\hat{\mathbf{x}}_1 \cdots \hat{\mathbf{x}}_N\right]$
  return $\hat{\mathbf{x}}$
end procedure

Figure 8 Low frequency and high frequency components of historical wind speed dataset (see online version for colours)



Finally, Table 3 analyses the impact of f_c in Algorithms 3 and 4, by sweeping f_c through different values. Specifically, the root mean square error (RMSE) between synthetic wind speed and historical dataset is used to quantified the impact of f_c. As can be seen, as f_c increase, higher frequency components are removed from the training and evaluation process, leaving less components being synthetic. Consequently, the synthetic wind speed presents a more similar time profile compared to dataset, resulting a smaller RMSE, as demonstrated in Table 3.

Figure 9 Synthetic wind speed versus historical dataset for a selected period of seven days, with only high frequency components being synthetic (see online version for colours)

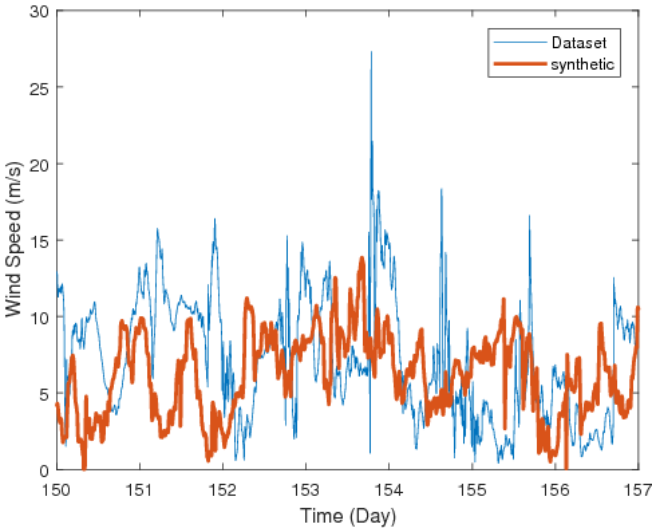
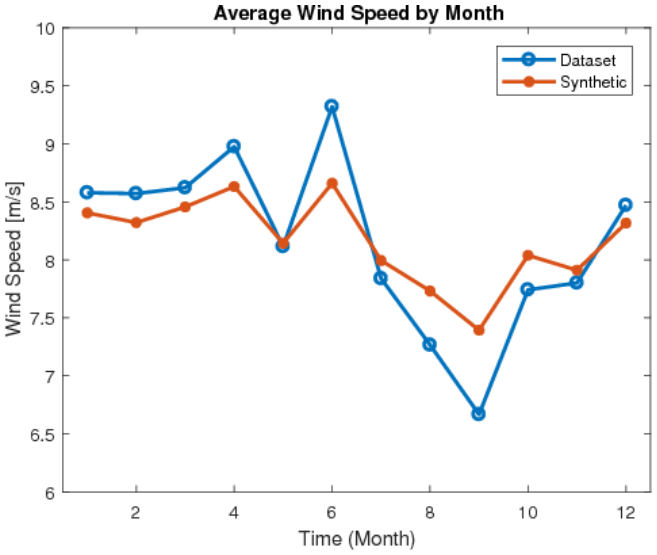


Figure 10 Monthly trend of the historical dataset and synthetic scenarios, with only high frequency components being synthetic (see online version for colours)



Remark 4: Note that RMSE is used as a metric to measure the variation of the synthetic data. Note also that here the larger RMSE the better, as the objective is to generate a time profile that is *different* from historical dataset. However, as illustrated in Figures 7 and 10, seasonal trend may loss when the RMSE is too large, which can be

important for some studies such as economic analysis where the electrical price presents strong seasonal trend.

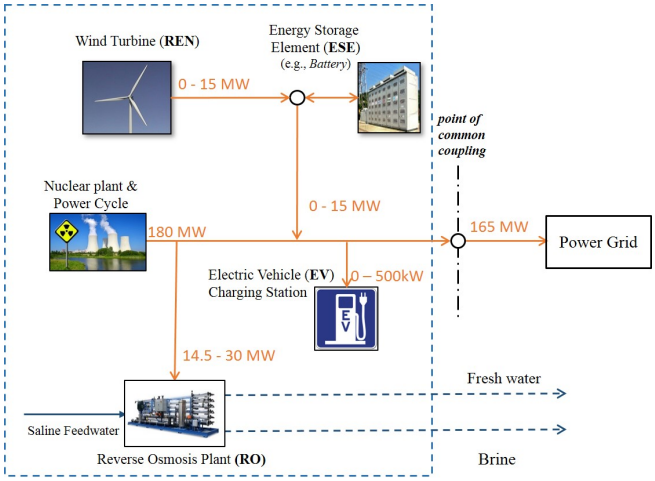
Remark 5: The selection of frequency f_c depends on the desired variation among generated datasets. In general, if higher variation is desirable, lower frequency can be used for Fourier series.

Remark 6: Note that instead of using Fourier series to preserve seasonality, it is use conditional machine learning that integrates the seasonal trend as part of ANN training process. We reserve this as future study.

Table 3 Root mean square error between synthetic wind speed and historical dataset

f_c (μHz)	Period	RMSE (m/s)
0.0634	3 months	5.3503
0.3805	1 month	5.2525
1.649	1 week	4.8264
11.57	1 day	3.5168
277.8	1 hour	2.5016

Figure 11 Topology of the HES configuration under study in this paper (see online version for colours)



4 Probabilistic analysis of HES

This section applies the proposed method to probabilistic analysis of a specific HES configuration, whose topology is shown within the dotted line in Figure 11. Specifically, the HES under study includes the following components:

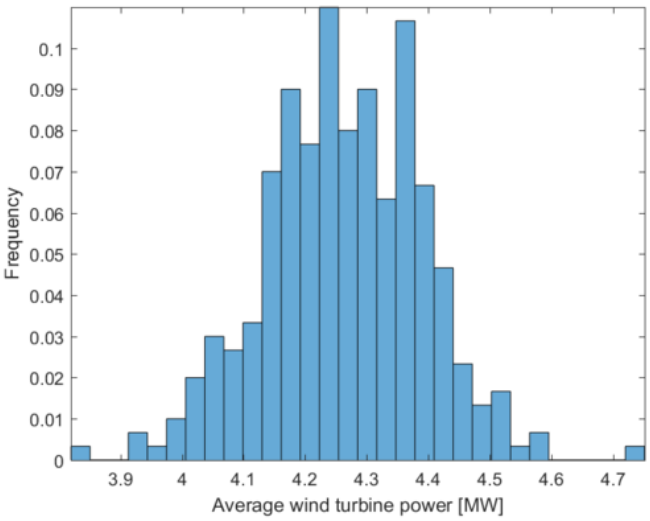
- A baseline electricity generation unit with 180 MW capacity. Specifically, this generation unit consists of a small modular reactor (SMR), a steam generator, a power cycle converting steam into electricity.
- A series of wind turbines as renewable power generation unit with total capacity of 15 MW.
- An energy storage element (ESE) used for power smoothing of the electricity generated by wind turbines.
- A reverse osmosis (RO) plant used to convert saline water into potable water and consumed electricity between 14.5 MW and 30 MW.
- An EV charging station consuming electricity between 0 and 500 kW.
- Finally, an electric grid connected to HES at a point of common coupling to consume 165 MW electricity from HES.

The kinetic energy in wind is captured and converted into electricity by wind turbines. This process can be modelled as a static mapping function from wind speed to wind power, as follows.

$$P_{REN} := \begin{cases} 0 & \text{if } V \leq 3 \text{ m/s or } V \geq 25 \text{ m/s} \\ 0.5\ \eta\rho V^3 \frac{\pi d^2}{4} & \text{if } 3 \text{ m/s} < V \leq 14 \text{ m/s} \\ 1.5 & \text{if } 14 \text{ m/s} < V < 25 \text{ m/s} \end{cases} \quad (9)$$

where η is the conversion efficiency of the wind turbine, ρ is the density of the air at the turbine, V is the wind speed, and d is the diameter of the turbine blades. In this study the values used for each parameter in equation (9) are: η = 35%, ρ = 1.17682 g/m^3, d = 58.13 m. Note that with these parameters, a wind turbine has rated maximum capacity of 1.5 MW.

Figure 12 Histogram of average wind turbine power generation using synthetic wind speed (see online version for colours)



The operations of the HES is as follows. First, a constant of 165 MW of electricity is supplied to the power grid, with the volatility from EV charging station and wind farm being absorbed by RO plant and ESE. To flexibly operate RO plant to accommodate the volatility, an operation optimisation framework as report in Chen and Garcia (2016a, 2016b) is used. To test this operation scheme, Algorithms 3 and 4 with f_c = 1.649 μHz are used to generate 3,000 synthetic wind speed profiles and the HES operations are simulated under these generated scenarios.

Figure 13 Histogram of average wind turbine down time (see online version for colours)

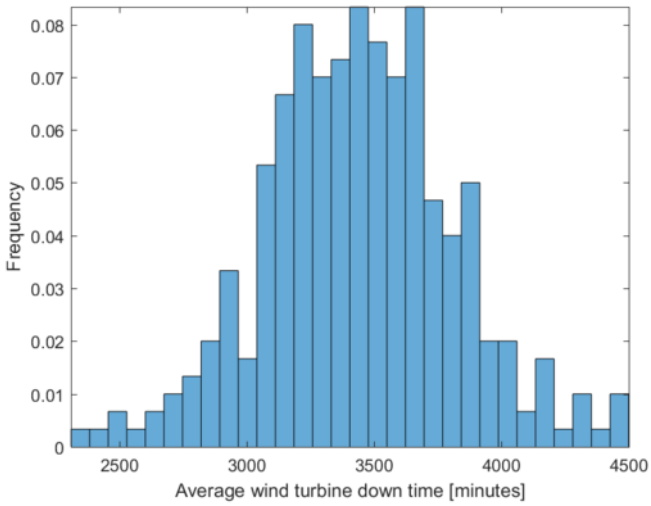


Figure 14 Histogram of average ramp-up/-down rates for RO chemical plant (see online version for colours)

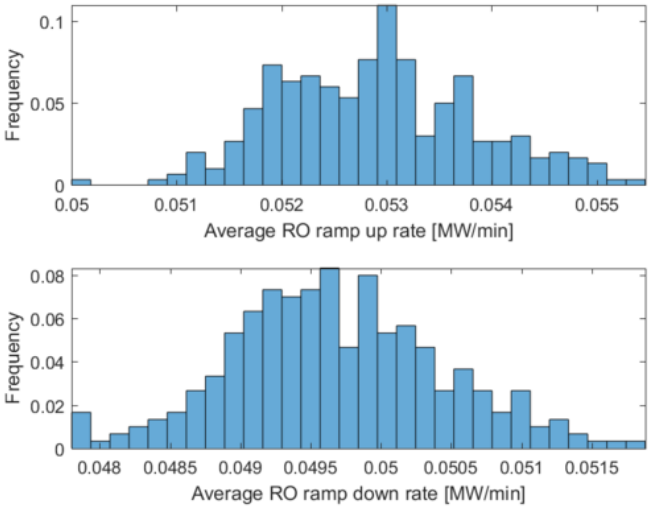


Figure 12 plots the histogram of the wind farm power generation, with a mean of 4.264 MW and standard deviation of 0.129 MW. Figure 13 depicts the average down time of wind turbine (recall that according to equation (9), wind turbine is shut off and produces zero power when the wind speed is below 3 m/s or above 25 m/s). Note that, results shown in Figure 13 can be used to determine the amount of maintenance required and its variation, which is critical in calculating the operational cost when conducting economic analysis of HES (Chen and Garcia, 2016b).

Finally, Figure 14 plots the histogram of the maximum RO ramp-up and ramp-down rate required to absorb the volatility. Note that the effect of ESE has been considered in these plots. These results can be helpful when considering the resiliency of the HES to accommodate both the average and worst case scenarios.

5 Conclusions

This paper focuses on synthetic scenarios generation for wind speed. More specifically, ANNs are utilised to capture the characteristics in the historical wind speed measurements and to generate synthetic data. In addition, Fourier transformation is used to model the low frequency components in historical datasets to preserve seasonal trends. The usefulness of the synthetic wind speed scenarios is demonstrated by performing probabilistic analysis of a HES that includes nuclear power plant, wind farm, battery storage, EV charging station, and desalination chemical plant. Wind power availability and requirements on component ramping rate are then investigated. Future work includes

1 parallel computing to allow fast scenarios generation and HES simulation

2 investigating the impact on the EV adoption (Chen et al., 2021b, 2021c)

3 using the synthetic dataset for state estimation (Chen, 2020)

4 finally, designing the optimal ANN architecture is another future direction.

References

Amjady, N. and Keynia, F. (2009) 'Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm', *Energy*, Vol. 34, No. 1, pp.46–57.

Bishop, C.M. et al. (1995) *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford, UK.

Chen, J. (2020) 'Extended kalman filter steady gain scheduling using k-means clustering', *International Journal of Modelling, Identification and Control*, Vol. 34, No. 2, pp.158–162.

Chen, J. and Garcia, H.E. (2016a) 'Operations optimization of hybrid energy systems under variable markets', in *Proc. 2016 American Control Conference*, Boston, MA, July, pp.3212–3218.

Chen, J. and Garcia, H.E. (2016b) 'Economic optimization of operations for hybrid energy systems under variable markets', *Applied Energy*, Vol. 177, pp.11–24.

Chen, J. and Rabiti, C. (2017) 'Synthetic wind speed scenarios generation for probabilistic analysis of hybrid energy systems', *Energy*, Vol. 120, pp.507–517.

Chen, J. and Zhao, J. (2021) 'Synthetic wind speed scenarios generation using artificial neural networks for probabilistic analysis of hybrid energy systems', in *30th IEEE International Symposium on Industrial Electronics*, Kyoto, Japan, 20–23 June.

Chen, J., Garcia, H.E., Kim, J.S. and Bragg-Sitton, S.M. (2016) 'Operations optimization of nuclear hybrid energy systems', *Nuclear Technology*, Vol. 195, No. 2, pp.143–156.

Chen, J., Kim, J.S. and Rabiti, C. (2017) 'Probabilistic analysis of hybrid energy systems using synthetic renewable and load data', in *Prof. 2017 American Control Conference*, Seattle, WA, May, pp.4723–4728.

Chen, Y., Wang, Y., Kirschen, D. and Zhang, B. (2018) 'Model-free renewable scenario generation using generative adversarial networks', *IEEE Transactions on Power Systems*, Vol. 33, No. 3, pp.3265–3275.

Chen, J., Li, Z. and Yin, X. (2021a) 'Optimization of energy storage size and operation for renewable-EV hybrid energy systems', in *2021 IEEE Green Technologies Conference*, Denver, CO, 7–9 April.

Chen, J., Liang, M. and Ma, X. (2021) 'Probabilistic analysis of electric vehicle energy consumption using MPC speed control and nonlinear battery model', in *2021 IEEE Green Technologies Conference*, Denver, CO, 7–9 April.

Chen, J., Behal, A. and Li, C. (2021c) 'Active cell balancing by model predictive control for real time range extension', in *2021 IEEE Conference on Decision and Control*, Austin, TX, USA, 13–15 December.

Di Silvestre, M.L., Graditi, G. and Sanseverino, E.R. (2014) 'A generalized framework for optimal sizing of distributed energy resources in micro-grids using an indicator-based swarm approach', *IEEE Transactions on Industrial Informatics*, Vol. 10, No. 1, pp.152–162.

El Fadil, H., El Idrissi, Z., Intidam, A., Rachid, A., Koundi, M. and Bouanou, T. (2020) 'Nonlinear control and energy management of the hybrid fuel cell and battery power system', *International Journal of Modelling, Identification and Control*, Vol. 36, No. 2, pp.89–103.

Frigo, M. and Johnson, S.G. (1998) 'FFTW: an adaptive software architecture for the FFT', in *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181)*, IEEE, Vol. 3, pp.1381–1384.

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

Gonzalez-Romera, E., Jaramillo-Moran, M.A. and Carmona-Fernandez, D. (2006) 'Monthly electric energy demand forecasting based on trend extraction', *IEEE Trans. Power Syst.*, Vol. 21, No. 4, pp.1946–1953.

Graditi, G., Di Silvestre, M.L., Gallea, R. and Sanseverino, E.R. (2015) 'Heuristic-based shiftable loads optimal management in smart micro-grids', *IEEE Transactions on Industrial Informatics*, Vol. 11, No. 1, pp.271–280.

Kim, J.S., Chen, J. and Garcia, H.E. (2016) 'Modeling, control, and dynamic performance analysis of a reverse osmosis desalination plant integrated within hybrid energy systems', *Energy*, Vol. 112, pp.52–66.

Lee, D. and Baldick, R. (2013) 'Synthesis of sample paths of wind power through factor analysis & cluster analysis', in *Proc. 2013 North American Power Symposium (NAPS)*, Manhattan, KS, 22–24 September, pp.1–6.

Lee, D. and Baldick, R. (2014) 'Short-term wind power ensemble prediction based on gaussian processes and neural networks', *IEEE Trans. Smart Grid*, Vol. 5, No. 1, pp.501–510.

Lee, D. and Baldick, R. (2014b) 'Future wind power scenario synthesis through power spectral density analysis', *IEEE Trans. Smart Grid*, Vol. 5, No. 1, pp.490–500.

Levenberg, K. (1944) 'A method for the solution of certain non-linear problems in least squares', *Quarterly of Applied Mathematics*, Vol. 2, No. 2, pp.164–168.

Ma, X-Y., Sun, Y-Z. and Fang, H-L. (2013) 'Scenario generation of wind power based on statistical uncertainty and variability', *IEEE Trans. Sustainable Energy*, Vol. 4, No. 4, pp.894–904.

Marquardt, D.W. (1963) 'An algorithm for least-squares estimation of nonlinear parameters', *Journal of the Society for Industrial and Applied Mathematics*, Vol. 11, No. 2, pp.431–441.

Meibom, P., Barth, R., Hasche, B., Brand, H., Weber, C. and O'Malley, M. (2011) '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.

Mohri, M., Rostamizadeh, A. and Talwalkar, A. (2018) *Foundations of Machine Learning*, MIT Press, Cambridge, MA.

Morales, J.M., Minguez, R. and Conejo, A.J. (2010) 'A methodology to generate statistically dependent wind speed scenarios', *Applied Energy*, Vol. 87, No. 3, pp.843–855.

Mori, H. and Kurata, E. (2008) 'Application of Gaussian process to wind speed forecasting for wind power generation', in *Proc. 2008 IEEE Int. Conf. Sustainable Energy Techn.*, IEEE, pp.956–959.

Oppenhiem, A.V., Willsky, A.S. and Nawab, S.H. (1983) *Signals and Systems*, Prentice-Hall, Hoboken, NJ.

Papavasiliou, A., Oren, S.S. and O'Neill, R.P. (2011) 'Reserve requirements for wind power integration: a scenario-based stochastic programming framework', *IEEE Trans. Power Syst.*, Vol. 26, No. 4, pp.2197–2206.

Soares, L.J. and Medeiros, M.C. (2008) '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.

Steckler, N., Florita, A., Zhang, J. and Hodge, B.M. (2013) '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, 22–24 October.

Sumer, K.K., Goktas, O. and Hepsag, A. (2009) 'The application of seasonal latent variable in forecasting electricity demand as an alternative method', *Energy Policy*, Vol. 37, No. 4, pp.1317–1322.

Zhu, B., Tazvinga, H. and Xia, X. (2015) 'Switched model predictive control for energy dispatching of a photovoltaic-diesel-battery hybrid power system', *IEEE Trans. Control Syst. Tech.*, May, Vol. 23, No. 3, pp.1229–1236.

Zioui, N. and Mahmoudi, A. (2020) 'Modal analysis and modelling approach for piezoelectric transducers based energy harvesting applications', *International Journal of Modelling, Identification and Control*, Vol. 36, No. 4, pp.304–314.

Notes

1 Downloaded from the Eastern Wind dataset maintained by National Renewable Energy Laboratory (NREL) at http://www.nrel.gov/electricity/transmission/eastern_wind_dataset.html on 21 November 2019.