Available online at www.sciencedirect.com
ScienceDirect
Energy Reports 6 (2020) 237–249
www.elsevier.com/locate/egyr

Tmrees, EURACA, 13 to 16 April 2020, Athens, Greece

Features extraction of wind ramp events from a virtual wind park

Sambeet Mishra[a,*], Esin Ören[a], Chiara Bordin[b], Fushuan Wen[a], Ivo Palu[a]

[a] *Talllinn University of Technology, Estonia*
[b] *The Arctic University of Norway, Norway*

Received 3 August 2020; accepted 30 August 2020

Abstract

In the European renewable energy portfolio, wind has a sizeable share in the total energy production. The Nordic and Baltic energy systems in particular are benefiting from wind energy to reach the greenhouse gas emissions reduction objectives set by the EU. The wind energy production varies with time, and this intermittent characteristic imposes a challenge for full utilization of renewable energy potential. The power system operator needs to ensure timely power supply of demand. An accurate estimation of power output from a non-dispatchable generation resource such as a wind farm is essential for the operator to ensure the supply–demand balance and adequate sizing of reserve power capacity. Existing methods of feature extraction and prediction such as linear regression often overlook the significant variations or do not utilize in the model building. However, this method misinterprets the trend in data. Understanding the properties of the variations in more details would reduce the uncertainty and significantly improve the feature extraction to aid in decision making. Furthermore, as the volume, shape and type of dataset start to increase and new methods are required to extract meaningful information from the patterns in the big data. The objective of the paper is to present a novel Ramping Behaviour Analysis ($\mathbf{RBA}_{\boldsymbol{\theta}}$) model that identifies and quantifies the variations in a time-varying dataset. The variations are classified into significant and stationary events. The former refers to the significant swings beyond a set threshold range and the latter refers to the swings that are relatively within the threshold limits. The features associated to each event include start time, end time, change in magnitude, persistence of an event, angle at which the event took place and frequency of occurrences of the features. In addition, the rain-flow cycles count is extracted from the original data for each event as a sum of half cycles and full cycles. The model is validated using simulated wind power production data from a virtual wind park spread across Estonia and the results are elaborated. The spatial dynamics of the virtual windfarm are captured through localized spatial autocorrelation of the events with the geospatial locations of the turbines. The results demonstrate that $\mathbf{RBA}_{\boldsymbol{\theta}}$ precisely and accurately identify and quantify the time varying power generation into events with subsequent features. The volume of the data is significantly reduced in the process of summarizing time series data into a series of events. Thereby $\mathbf{RBA}_{\boldsymbol{\theta}}$ can be also used for data compression and reconstruction with minor losses. The system operators can use the proposed algorithm in operational scheduling, maintenance and investment-capacity building decisions.

Peer-review under responsibility of the scientific committee of the Tmrees, EURACA, 2020.

Keywords: Wind power production; Ramp events; Feature extraction; Data science; Time series variations; Rainflow counting; Renewable energy

* Corresponding author.
E-mail address: sambeet.mishra@ttu.ee (S. Mishra).

https://doi.org/10.1016/j.egyr.2020.08.047

Peer-review under responsibility of the scientific committee of the Tmrees, EURACA, 2020.

Nomenclature

t_1	Start time for an event (hours)
t_2	End time for an event (hours)
τ	Set threshold limit
Significant events	
Δt_m	The persistence of a significant event (hours)
w_1	Amplitude of a significant event at time t_1 (kW)
w_2	Amplitude of a significant event at time t_2 (kW)
Δw_m	Change in amplitude for a significant event (up or down ramp) (kW)
θ_m	Slope angle of an event with respect to the time (degree)
σ_m	The mean of the amplitude for a significant event (kW)
λ_m	Frequency of occurrence of features of a significant event
φ_m	Rain-flow cycles count during a significant event (cycles)
Stationary events	
Δt_s	The persistence of a stationary event (hours)
σ_s	Mean of the amplitude for a stationary event (kW)
λ_s	Frequency of occurrence of features of a stationary event
φ_s	Rain-flow cycle counts during a stationary event (cycles)

1. Introduction

Wind energy resources have an increasing share in the total European Union (EU) energy production landscape. The total wind energy share is projected to increase to reach the EU objective of 100% renewable energy system by 2050 [1]. Nordic countries in particular are benefiting from the wind energy to reach the green-house gas emission reduction objective imposed by the EU. Wind energy is affected by multiple natural phenomena and it has therefore a stochastic nature. The weather is affecting wind power and its intermittent nature. The uncertain behaviour of the wind speed results in a high variability in production. This requires additional energy from conventional power stations which will reduce the overall environmental benefits of this renewable resource. Proper forecasting is therefore a key solution to make a better and sustainable use of wind energy, together with interconnected grid, storage technologies and demand side management. The main objective of forecasting is to better handle the uncertainties that renewable energy integration is causing into the power system. Forecasting is therefore a necessary and cost-effective element for the optimal integration of wind power into the energy systems. However, forecasting requires a deep understanding of the main features of the dataset involved, in order to generate accurate predictions. The intermittent nature of wind power can be described as a continuous sequence of changes in the output power. From this point of view, it is possible to define a wind power ramp event as a sudden change in the output power over a set threshold τ. Mathematically, a wind power ramp event can be described as the absolute difference between the power produced w_t in time t and $(t + \Delta t)$ that is above the τ. This concept can be summarized as $\left|w_{(t+\Delta t)} - w_t\right| > \tau$. Within these power changes, the system operation (SO) has to keep the system balanced, i.e. the total generation must meet the demand at each point in time. Wind ramp events can be positive or negative depending on the generation swings. When a ramp event is positive, it might be necessary to shut down the wind turbine in order to avoid accidents or damages to the system. On the other hand, when the ramp event is negative, the SO has to find proper alternatives to meet or mitigate the demand. From an economical point of view, both the energy not used and the energy coming from alternative resources are crucial.

Some of the ramp events are rare events that can damage wind turbines, and they are therefore important to be addressed, identified, and forecast. Indeed, their proper identification, characterization, and forecast is key for a better maintenance planning and a longer lifetime of turbines. In order to properly determine potential investments and optimize real time operations, wind farms usually forecast the wind speed and power production from historical

data over time. The time interval Δt is typically 13h for ramp events. Therefore, events are more significant and accurate in short term forecasting, while they become insignificant when it comes to long term forecasting. For a whole wind park, the τ is usually set to an absolute value. However, it can also be set to a certain percentage of the generation depending on the installed capacity. An issue with this practice is that the peak generation capacity varies through seasons, turbine maintenance or new installations. Even though the τ is dependent on the peculiarities of a specific wind park, the method to classify ramp events is generic.

1.1. Literature review

Ramp events forecasting has received wide attention in literature both for short term purposes as in [2,3] and for long term predictions and corrections [4,5]. A survey on techniques for wind power uncertainty quantification in smart grids is proposed in [6]. In [7] a 6-h and 24-h binary (ramp/non-ramp) prediction based on reservoir computing methodology is proposed, to avoid damages in the turbines. Simulations of the system are performed, and the results show that the proposed algorithm can predict about 60% of ramp events in both 6-h and 24-h prediction cases. In [8] the authors proposed a model to forecast ramp events as well. They modelled observed wind speeds into forecast models and converted this into power forecasts with the help of the power curve of the wind turbines. It suggests that the same method could be implemented for solar power plants. The work in [9] proposed a probabilistic forecasting method, utilizing a Neural Network (NN) to generate possible future scenarios, employing an objective function based on cumulative distribution functions and auto-correlation functions to train the NN, primarily teaching it their distribution. Again another work [10] proposed a model to synthesized wind speed scenarios based on statistical parameters of wind and Markov chains. A robust approach for estimating the probability of wind power ramp events is proposed in [11] where authors describe an uncertainty quantification model to estimate the probability of ramp events with distributional robustness guarantee, based on Gaussian mixture model. Similarly, a continuous gaussian mixture model is proposed in [12] where a probabilistic forecasting method based on scenario generation is investigated. In contrast, Kaut [13] proposes a new heuristic to generate scenarios that use copulas instead of common correlation functions. Copula-based probabilistic forecast are proposed also in [14] and numerical simulation on publicly available wind power data show the high level of reliability of the method. A data driven optimization approach for the estimation of wind power ramp events is available in [15].

However, the traditional studies available in literature, tend to focus mainly on statistical values (i.e. mean, median, average variance and the like), which can identify the overall time series properties, but are not enough to thoroughly describe the individual ramping events. Indeed, some ramping events are rare events, and they should therefore be studied as single events, by identifying the specific relevant properties of each of them. For instance, traditional forecasting methods like regression, or ARIMA models, are not accurate when it comes to rare event detections like wind ramps. When the number of events is low, relative to the number of predictors, standard regression could produce overfitted risk models that make inaccurate predictions [16]. Authors in [17] show a diagram where the normalized wind speed variation is represented together with a regression curve on top. The proposed figure is a clear example of the inaccuracy of traditional regression methods when it comes to wind ramp events. Indeed, the regression curve proposed in the diagram, completely ignores the positive ramp event and the negative ramp event that are clearly visible toward the end of the dataset trend. Moreover, despite recent advances, machine learning models in general, as well as deep learning models in particular, are still limited when it comes to life-long and one-shot learning, especially in remembering rare events, like wind ramps [18]. There is therefore high interest among the scientific community, in further developing new frameworks for wind power predictions, and specifically for wind ramp identification and predictions.

In order to improve the accuracy of wind ramp predictions, it is not enough to rely on time-series, but it is important to properly identify some key features of the available historical dataset. Indeed, each wind power ramp event can be described by specific properties (such as the peak, length of time required to reach a certain level of peak, length of time spent at a certain peak, where a peak represents the sudden change in the wind power value from an increasing trend towards a decreasing trend or vice-versa). Such features can provide additional relevant information to better quantify wind power variations and extract trends from time-series dataset.

The identification of key features behind the wind ramp events falls under the research area of “wind ramp events identification, detection and characterization”. This represents a preliminary step to take before the actual forecasting tasks, in order to better understand the key properties of a dataset, and better identify which features are more relevant to be predicted and why.

As opposed to the wind power ramp predictions area, the wind ramp event identification, detection and characterization has not been addressed much in literature. The studies in [19] and [20] introduced the terminology for identification of ramp events, ramping behaviour analysis (RBA), which comprises the perspective used in this study. They also filtered and extracted events and clustered them into groups. More studies have been performed on identifying ramp events in [21] and [22]. A detection and characterization of extreme wind speed ramps is proposed in [23]. Another study focused on large wind power ramps characterization can be found in [24] where generation data from a wind cluster in conjunction with meteorological observations are analysed to determine the magnitude and frequency of ramping events. Wind power ramp events detection is also addressed in [25] where authors propose a hybrid classifier to perform a features selection and improve the machine learning models training. Wind characteristic analysis based on Weibull distribution are proposed in [26].

1.2. Key contribution

The available scientific literature shows that the identification of key features behind the wind ramp events has not yet been widely addressed by researchers. Therefore, the primary objective of this work is to propose and test a methodology for wind ramp events identification, detection and characterization. The secondary objectives are: (a) develop an improved rain-flow counting cycle to find the sum of half cycles and full cycles in events as a unique feature; (b) extract frequency of occurrences of the events as insight into the dataset and as unique features; (c) test the proposed methodology on near-real dataset from a wind park with hourly resolution, to show the performance.

The key contribution of this paper is to propose a new algorithm to extract explanatory features from wind data (information of wind variations over time). Such algorithm can be used for predictions in place of traditional wind power time-series. The main novelty of this paper is the development of a novel algorithm that focuses on ramp events identification and characterization, instead of just wind power time-series like traditional works in literature. By properly identifying the most relevant features of wind ramp events, the algorithm is capable to develop improved training dataset that can be utilized within forecasting tools or machine learning algorithms. The novel approach of the proposed study can be summarized as follows: compared to the traditional literature that aims at understanding "how a time-series dataset will be", the proposed study takes an innovative approach by shifting focus on "what kind of events' features to expect".

2. Ramping Behaviour Analysis (RBA_θ) model

The RBA_θ is an algorithm that searches for trends of variations in a discrete time series based on set threshold. The algorithm determines features of the variation that defines the trend. Moreover, the algorithm classifies the variations into stationary and significant events. The distinction comes from whether the variation is bigger or smaller than a predefined τ, where the first is identified as significant and the latter as stationary event. The significant events features are $t_1, t_2, \Delta t_m, w_1, w_2, \Delta w_m, \theta_m, \sigma_m, \lambda_m, \varphi_m$, where t_1 is the beginning and t_2 is the ending point of the event; Δt_m the persistence of the event; w_1 is the amplitude of the production at t_1; w_2 is the amplitude of the production at t_2; Δw_m is the amplitude of the significant event; θ_m is the slope; σ_m is the mean of the amplitude; λ_m is a set of the frequencies per features $\Delta t_m, \Delta w_m, \theta_m,$ *and* σ_m of the event over the given dataset; φ_m is the total cycle counted by Rainflow-counting algorithm during the significant event. The subscript "m" and "s" refer to the feature belonging to a significant or a stationary event respectively. Significant events stay on slope of θ_m for Δt_m, with Δw_m of change in amplitude, where Δw per consecutive t is bigger than chosen τ. The stationary events features are $t_1, t_2, \Delta t_s, \sigma_s, \lambda_s, \varphi_s$, where; t_1 is the beginning point of the event; t_2 is the ending point of the event; Δt_s the length of the event; σ_s is the mean of the amplitude and; λ_s are a set of the frequencies per features Δt_s and σ_s of the event over the given dataset; φ_s is the total cycle counted by Rainflow-counting algorithm during the stationary event. Stationary events stay in τ for Δt_m around mean amplitude of σ_s. Note that the stationary events are the events which are less than the τ and therefore the change in power is not captured. On the same note, the angle is not extracted.

2.1. Conceptualization of RBA_θ model

Fig. 1 depicts the concept of RBA_θ. Fig. 1(a) presents the features of the events where the blue area shows one stationary event and few up and down significant events. The first significant event took place with Δw_m magnitude change, θ_m slope angle for Δt_m time. Fig. 1(b) shows the rainflow cycles (φ_m, φ_s) extracted per subset of the data where there was an event, e.g. within the first significant down-ramp event, there are three half cycles and two full cycles in the original data. Note that in case if rainflow cycles extracted from the whole dataset, the cycle counts will differ. To derive the localized spatial autocorrelation among the turbine locations and the features a spatial Markov-chain method is implemented for the features $\Delta t_m, \Delta w_m, \theta_m, \sigma_m$ for significant events and $\Delta t_s, \sigma_s$ for stationary events. The interdependency between production from intermittent sources in two regions where the relationships is complementary would be more beneficial for the overall production.

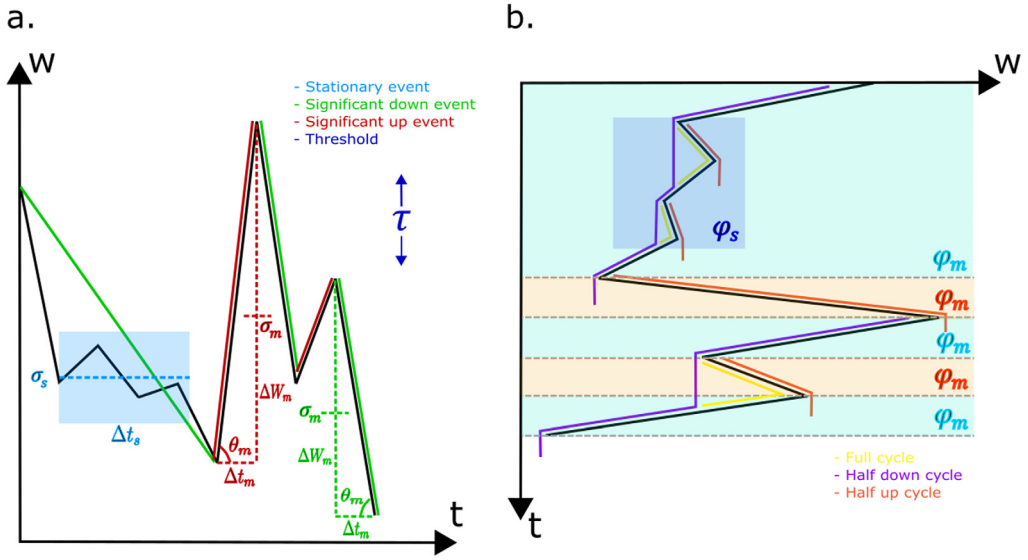

Fig. 1. (a) Wind ramp event extraction (b) rainflow cycles counting.

A modified version of the Rainflow-counting algorithm is presented in [27–30]. There are two definitions which are highly used in literature namely top-level-up cycle and rainflow cycle [29]. This paper uses the later definition. The Rainflow-counting algorithm was initially introduced for the estimation of stress/strain in a material, essentially the effect of vibration. The algorithm identifies cycles in form of either full or half. A half cycle can be upward trending or downward. The algorithm uses the amplitude as the metrics to find the cycles. The input to the algorithm is a simple series of peaks and valleys, i.e., local maxima and minima, that form hysteresis loops. Closed loops are full cycles, and unclosed loops are half cycles. The algorithm uses a change in slope as an indicator that the time series is going through a peak or valley. Only the magnitude of the peak or valley is then entered into the Rainflow-counting algorithm. The drawback of this method for a time series application is that it identifies one pattern in the whole time series. This process undermines the small variations and there are high chances for one long cycle. However, if the time series is broken into pieces then more granular results can be achieved. For this reason, the algorithm is applied to the ranges, subsets of the original data per event, subsets between t_1 and t_2.

A first-order Markov chain is the realization of the stochastic process in the discrete data x where every discrete point is attained to a discrete state value S = 1, …, m. Here, the state of the process depends only on the previous state and a conditional probability. The theory of spatial dynamics states that the power production from an individual turbine can be quite different even though they are physically located in one farm and quite close to each other. The difference might arise from wake effect, terrain condition and other environmental effects. To capture this relation, a spatial dynamics approach has been considered.

Specifically, spatial Markov chain model [31–33]. A Markov chain has a series of states that are mutually exclusive of each other thus applicable to discrete dataset. A transition matrix has transition probabilities from

one state to another where each event placed in to a state. For this study a local indicator of spatial association (LISA) Markov [33] which is based on Moran's I and show the transition probabilities while placing each areas behaviour into quantiles based on whether they behave similarly to their neighbours or not, is chosen.

2.2. RBA_θ model description

The Algorithm 1 presents the pseudo code of the $\boldsymbol{RBA_\theta}$.

Algorithm 1: $\boldsymbol{RBA_\theta}$ (τ, shapefile, w(t) ∈ R+)
Result: Significant events, Stationary events
procedure Significant Events (τ, $w(t)$):
  $\Delta w(t)$ = difference of consecutive items in $w(t)$
  for each Δ in Δw do:
    if the sign of Δ == sign of next Δ then
      events ← concatenate events where Δs point
    end if
  end for
  for each event in events do:
    if event > τ then
      w_m ← event
    end if
  end for
  Δw_m = difference of consecutive items in w_m
  for each Δ in Δw_m do:
    if the sign of Δ == sign of next Δ then
      significant events ← concatenate events where Δs are
    end if
  end for
  for each event in *significant events* do:
    event ← $t_1, t_2, \Delta t_m, w_1, w_2, \Delta w_m, \theta_m, \sigma_m$
    event ← $\lambda(\Delta t_m), \lambda(\Delta w_m), \lambda(\theta_m), \lambda(\sigma_m), \lambda(\theta_m, \sigma_m)$
    event ← φ_m
  end for
  return *significant events*
procedure Stationary Events (τ, $w(t)$):
  start = 0
  while start ← length of $w(t)$ do
    forward = 0
    for each point forward in $w(t)$ from start do
      if point is between in τ from $w(t)_{start}$ then
        forward += 1
      else then
        break *for*
      end if
      if forward > 0 then
        stationary events ← t_1 = start,
                t_2 =start + forward,
                Δt_s = forward,
                σ_s = mean of the points in event
      end if
    end for
    start += forward + 1
  for each event in stationary events do
      event ← $\lambda(\Delta t_s), \lambda(\sigma_s), \lambda(\Delta t_s, \sigma_s)$
      event ← φ_s
  end for
  return stationary events

procedure λ (events, τ):
  k = 1 // τ
  if events are stationary then:
    bins of Δt_s← linearly spaced k ranges between [1, max(Δt_s)]
    bins of σ_s ← linearly spaced k ranges between [-1, 1]
    $\lambda_{\Delta t_s}$← frequencies of Δt_s in their bin
    λ_{σ_s}← frequencies of σ_s in their bin
  else if events are significant then:
    Δt_m ← linearly spaced k ranges between [1, max(Δt_m)]
    Δw_m ← linearly spaced k ranges between [-1, 1]
    θ_m ← linearly spaced k ranges between [-90, 90]
    σ_m ← linearly spaced k ranges between [-1, 1]
    $\lambda_{\Delta t_m}$← frequencies of Δt_m in their bin
    $\lambda_{\Delta w_m}$← frequencies of Δw_m in their bin
    λ_{θ_m}← frequencies of θ_m in their bin
    λ_{σ_m}← frequencies of $\sigma_m$in their bin
  end if
  return λ

procedure φ ($w(t)$, events):
  for event in events do:
    section ← section of $w(t)$ between t_1 and t_2 of event
    φ ← **RainflowCounting**(section)
  end for
  return φ

procedure Lisa Markov(Major events, Stationary events, shapefile):
  $weights$ ← **libpysal.DistanceBand**(shapefile, distance threshold)
  for feature in [$\Delta t_m, \Delta w_m, \theta_m, \sigma_m, \Delta t_s, \sigma_s$]:
    $transition\ matrix_{feature}$ ← **pysal.LisaMarkov**($attribute\ matrix$, $weights$)
    $transition\ matrices$ ← $transition\ matrix_{attribute}$
  end for
return $transition\ matrices$

procedure Reconstruction(Major events):
  for event in Major events:
    $reconstructed$ ← interpolated $w(t)$ between t_1 and t_2
    if t_2 of the current event != t_1 of next event:
    $reconstructed$ ← interpolated $w(t)$ between t_2 and next t_1

The algorithm has 5 procedures — Significant Events, Stationary Events, Frequency of occurrences (λ), Rainflow cycles count (φ), LISA Markov, and Reconstruction. The proposed algorithm is a hybrid architecture meaning procedure 1 and 2 are parallel. Procedure 3, 4, 5 and 6 can be parallel to each other but are sequental to 1 and 2 as they take the output from 1 and/or 2 as input. The procedure Significant Events and Stationary Events extract the features for individual events. The λ procedure identifies the frequency of occurrence of features such as $\boldsymbol{\Delta w_m}, \boldsymbol{\Delta t_m}, \boldsymbol{\theta_m}, \boldsymbol{\sigma_m}$ and $\boldsymbol{\Delta t_s}, \boldsymbol{\sigma_s}$ in specified bins. Each bin is calculated by dividing the total range of the feature into windows. The LISA Markov procedure evaluates the transition matrices per major feature from all turbines taking into account a given shapefile that expresses their locations.

3. RBA_θ application to the wind energy production

This section outlines the results obtained from applying the model to synthetic wind energy production data from virtual wind farm. The section begins with introducing the data, followed by τ tests. Consequently, the first 5 significant and stationary events with features are presented for one wind turbine.

3.1. The virtual wind farm with simulated wind energy dataset

The data are obtained from the virtual wind park with hourly time resolution for 3 years, 2017 to 2019. Simulated wind turbine power production data are obtained with 80 metres hub height from [34,35]. Fig. 2(a) presents the power output from the virtual wind park for 8760 h. The data has certainly many power swings and it can be seen more clearly when a smaller time-window is chosen. However, the data is demonstrated with a top view to show the pattern in seasons. For example, the winter season has dense power production in comparison to lighter one in summer.

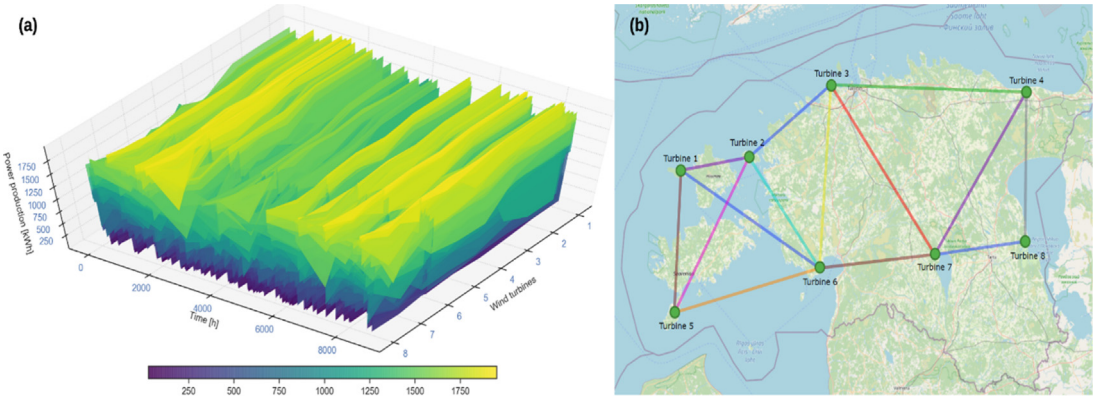

Fig. 2. (a) Hourly wind power production from 8 turbines for a year (b) distance based neighbouring relationships of the virtual wind farm.

A shapefile containing the geospatial locations of the turbines in the coastal regions of Estonia is chosen using the QGIS as an input. Fig. 2(b) shows the location of the virtual wind farm and the neighboring relationships of the turbines to one another. The calculation of the spatial weights that express the neighbouring relationships of the turbines are done using the mentioned shapefile and are based on a threshold distance of around 180 km. If two locations are closer than the threshold distance, then the turbines are considered neighbouring. These weights are used to calculate spatial relations between turbines per event feature using LISA Markov [33] method.

3.2. Results and discussion

The RBA algorithm has been tested using the dataset outlined in the sub-section 3.1. The number of events is increasing with decreasing τ. Note that the results are for the first wind turbine. A threshold test is conducted considering 10 τ values from 0.5 to 0.9 to identify how the number of events extracted change along the increasing threshold. Then the change for maximum of σ_m with threshold values is investigated. Fig. 3 presents the changes in significant and stationary events with reference to ascending threshold for 8 wind turbines. With increase in threshold the number of identified events decrease, so as the means. In order to investigate the full spectrum of events 0.1 threshold is chosen for the study. Different time varying data have different sensitivity, thereby different threshold values should be tested to find the optimum threshold. Rainflow-cycle counting algorithm reduces a spectrum of variations into an equivalent set of simple reversals. These reversals are classified as half-cycle and full-cycles depending on whether the variation is a half or a whole hysteresis loop. The rainflow cycles count, φ for every event as the sum of the cycles that are exclusively taking place over the duration of the particular event is extracted. Fig. 4 presents the rainflow cycles by applying the cycle-counting between t_1 and t_2 to extract φ values per event, on the right side for the significant events and on the left side for the stationary events, for the first turbine. Fig. 5 presents

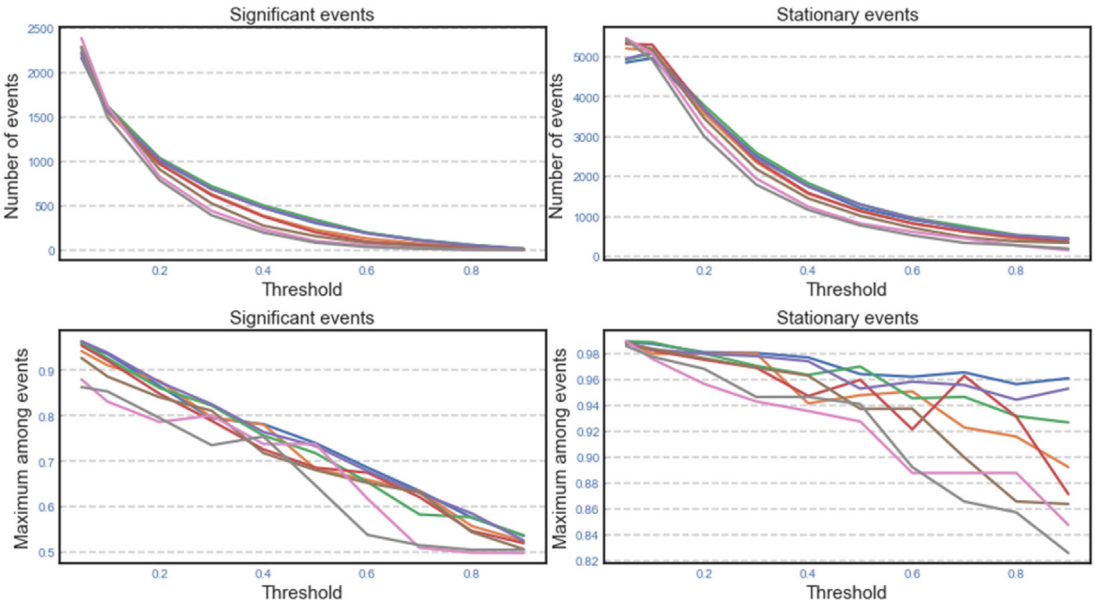

Fig. 3. Threshold tests for change in events.

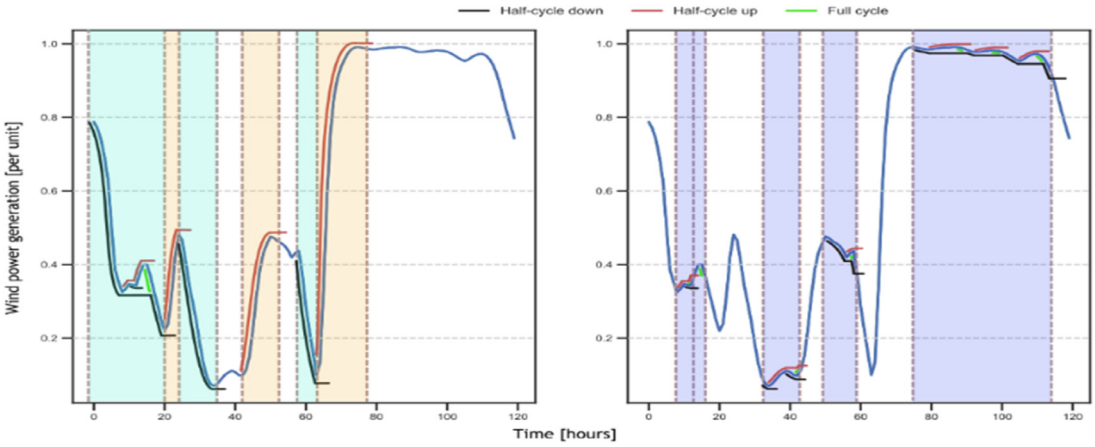

Fig. 4. Rain-flow cycles on the wind power plot.

the turbine-wise extracted events and corresponding features. The events are further classified into significant and stationary events. The significant events are comprised of up and down ramp events. Essentially the upward or rising events are up-ramp and downward or falling events are down-ramp events. The relatively less variations are classified as stationary events and often persistent. These events are highlighted on top of the hourly wind power production dataset. Notice that the figures are for a 120-hours period to provide a clear picture. The significant events are associated with a sudden and significant changes in energy.

While stationary events are associated with relatively minor changes in energy. This is an improvement to the existing form of event extraction presented in [36,37] by introducing the context of significant and stationary along with the features. Wind ramp prediction techniques presented in [4,7,38,39] can be further improved by introducing the features extracted in this paper. Similarly the wind farm controls as in [16] to either avoid a catastrophe or pose to extract the maximum potential energy from the wind flow can be realized.

Beyond that the features can be embedded into wind turbine model for investment planning as in [40,41]. The features corresponding to significant and stationary events are demonstrated in Tables 1 and 2 respectively.

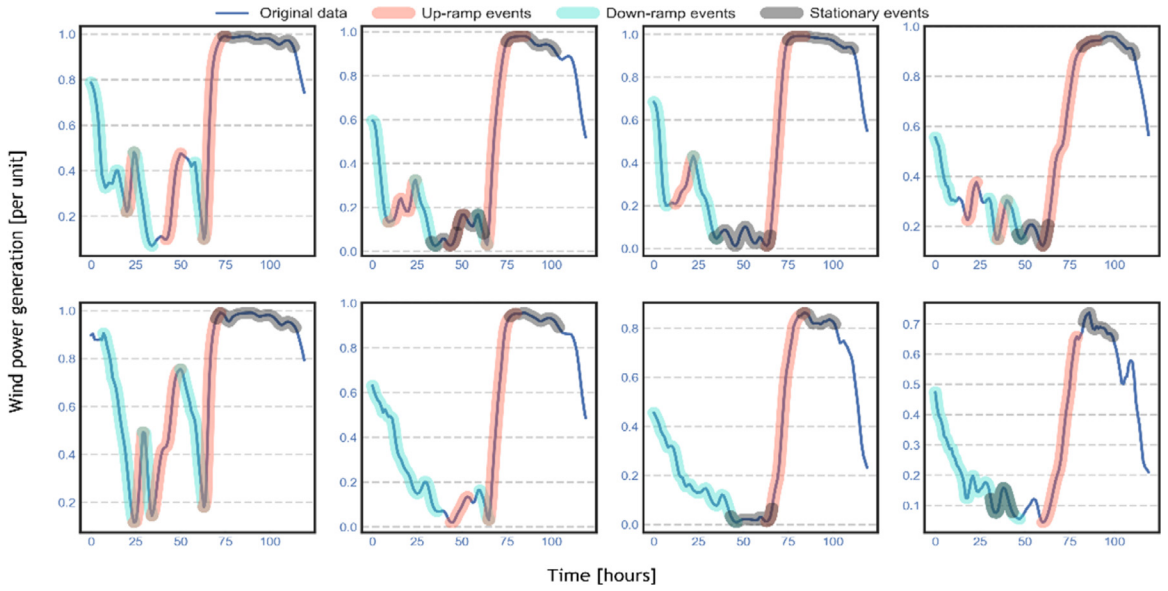

Fig. 5. Significant and Stationary events extracted from turbine-wise wind power production data.

Table 1. RBA results of significant events.

Event	t_1	t_2	Δt_m	w_1	w_2	Δw_m	σ_m	θ_m	$\lambda\,(\Delta t_m)$	$\lambda\,(\Delta w_m)$	$\lambda\,(\theta_m)$	$\lambda\,(\sigma_m)$	$\lambda\,(\theta_m, \sigma_m)$	φ_m
1	0	20	20	0.8	0.2	−0.6	0.5	−70.6	242	144	462	653	45	2.5
2	20	24	4	0.2	0.5	0.3	0.4	81.3	605	335	488	653	68	0.5
3	24	34	10	0.5	0.1	−0.4	0.3	−76.3	492	205	462	330	30	0.5
4	42	50	8	0.1	0.5	0.4	0.3	78.1	605	198	488	330	35	0.5
5	58	63	5	0.4	0.1	−0.3	0.3	−81.6	605	205	462	330	34	0.5

Table 2. RBA results of stationary events.

Event	t_1	t_2	Δt_s	σ_s	$\lambda\,(\Delta t_s)$	$\lambda\,(\sigma_s)$	$\lambda\,(\Delta t_s, \sigma_s)$	φ_s
1	8	12	4	0.34	4374	1105	368	2
2	13	17	4	0.37	4374	1105	350	1
3	33	43	10	0.09	457	374	55	2.5
4	49	58	9	0.45	457	1105	120	2
5	74	113	39	0.98	9	1020	3	5

The first 5 extracted events are listed in Table 1 for the first turbine. For instance, the first significant event is a down ramp event which lasted for 20 h with power variation of 0.6 with an angle of −70.6 and 0.69 mean value. Similarly, the first stationary event lasted for 4 h with mean of 0.34. Stationary events that have persisted longer than 3 h are taken into account. The frequency of occurrences (λ) is a measure to count how many times a given feature repeats within a corresponding time window. The time window is chosen for this study is 10, meaning that the total range is sliced into 10 parts. Note that the slope angle can be positive and negative making it an up or down ramp event. The frequency of occurrences for features of significant events $\lambda\,(\Delta t_m)$, $\lambda(\Delta w_m)$, $\lambda\,(\theta_m)$, $\lambda\,(\sigma_m)$, $\lambda\,(\theta_m, \sigma_m)$ are calculated and presented in Table 1. Similarly, $\lambda\,(\Delta t_s)$, $\lambda\,(\sigma_s)$, $\lambda\,(\Delta t_s, \sigma_s)$ for stationary events are presented in Table 2. Fig. 6 presents the relation among significant and stationary events mapped through parallel coordinates for each turbine. Note that the dataset for the significant events are widely and densely distributed across Δw_m while sharply divided for θ_m and dense across σ_m. The major significant event is extracted from fifth turbine on November 2018 with amplitude of 0.99 and persisted for 43 h. The major stationary event occurred on eighth turbine on October 2017 for a period of 75 h. The significant event with maximum persistence, 111 h, occurred on eighth turbine on October 2019.

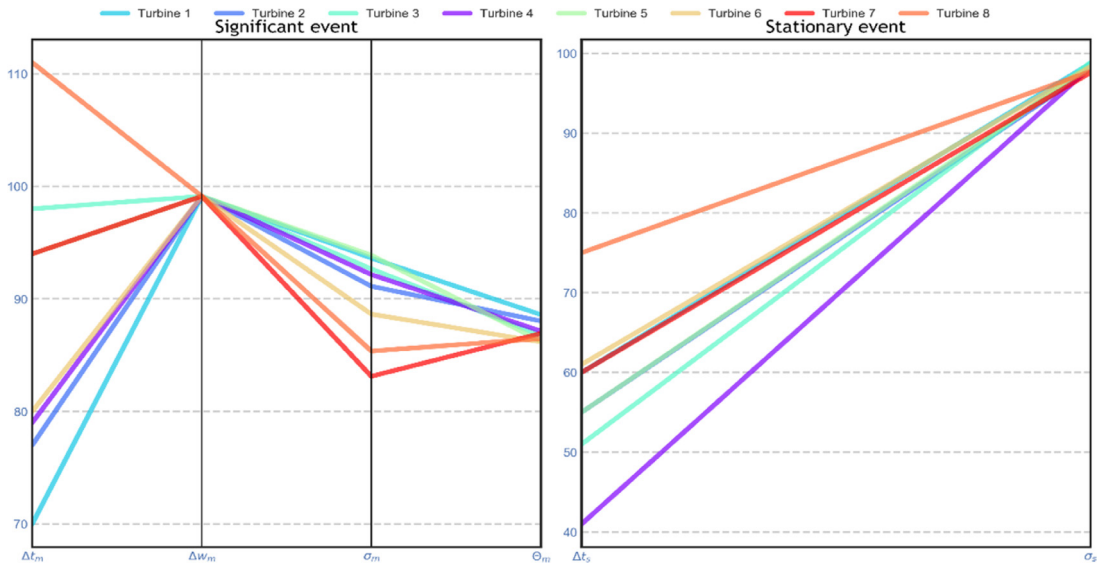

Fig. 6. Turbine wise significant and stationary events and their association.

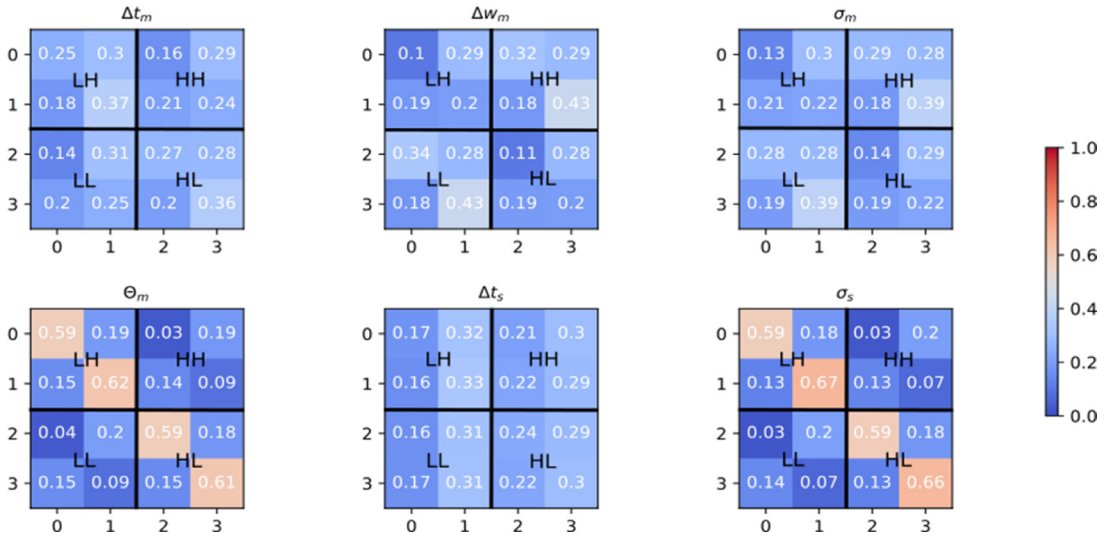

Fig. 7. Spatial dynamics autocorrelation of features.

The LISA-Markov [32] constructs the transition matrix that contains the probabilistic inter relationship among turbines utilizing the weights from the shapefile of the locations of the turbines and the series of features. The transition probabilities for the features $\Delta t_m, \Delta w_m, \theta_m, \sigma_m, \Delta t_s, \sigma_s$ are evaluated. These probabilities represent the spatial autocorrelation on Moran's I. Moran's I has four quadrants, first referring to observations of high values surrounded by high values, third referring to low values surrounded by low values, second referring to observations of low values surrounded by high values, and fourth referring to observations of high values surrounded by low values. The results as presented in Fig. 7 identifies that Δw_m has more mobility for observations in first (HH) and third (LL) quadrants. While θ_m, Δt_s and Δt_m are evenly distributed across the quadrants. The σ_s and σ_m are distinctly present in second (LH) and fourth (HL) quadrants. This implies the amplitude of events throughout the virtual wind farm are positively correlated but the mean of these events is dissimilar. The spatial dynamics of wind turbines shows that they have similar variations on different levels. Furthermore, the turbines in closer proximity has more impact on each other while all turbines influence each other. A null hypothesis test is conducted on the

transition matrices of the features and the results clarify that there is statistical significance in the dataset. Since the wind is a phenomenon that is influenced by various global factors, it is important to consider the spatial dynamics in relation to the features extracted. There is further room for investigation considering wind parks in different locations in place of wind turbines.

The extracted features are used to reconstruct the original data with losses. For reconstruction the features starting time, ending time, magnitude at t1, and magnitude at t2 ($t_1, t_2, \Delta w_m^{t1}, \Delta w_m^{t2}$) are given as input that defines the outer boundary for an significant event. In the process of reconstruction, the $t_1, t_2, \Delta w_m^{t1}, \Delta w_m^{t2}$ are joined together by using linear interpolation. The RMSE loss between the original and reconstructed dataset is 0.008. The shape of original data for 3 years hourly wind power production data is 26 480*1, while the shape of compressed dataset is 1558*4. The size of original data and the data needed for reconstruction when saved in csv format are 46 kB and 178 kB respectively. The original data size is reduced by 74% in the process of compression. Fig. 8 presents the original data and the reconstructed data visualization with difference as a bar plot. As visually noticeable, the reconstruction is with losses. Furthermore, different curve fitting methods can be used to improve the efficiency.

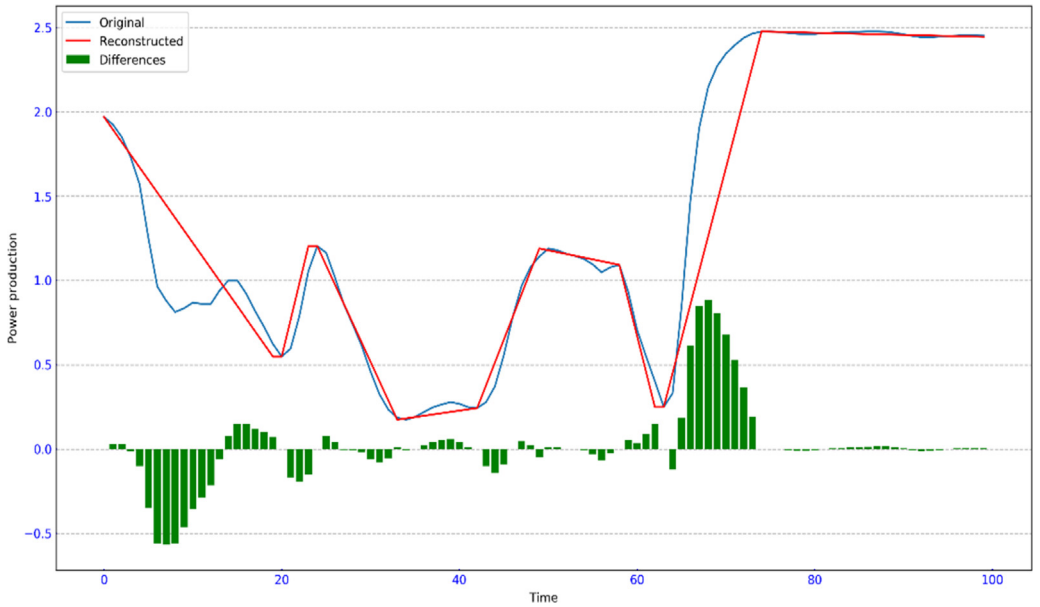

Fig. 8. Data reconstruction from features.

4. Conclusion and future work

This paper proposes a novel algorithm, Ramping Behaviour Analysis (RBA_θ) for detection and quantification of changes in time varying dataset. Thereby summarizing the time-series data to a series of events with associated features. RBA_θ classifies variations into significant or stationary events given a threshold. The threshold acts as a range of magnitude, beyond which significant event takes place and stationary events within. The features include the start time, end time, magnitude change, average magnitude change, persistence (time duration), slope angle at which the event took place and frequency of occurrences. Thereafter, a modified rainflow cycle counting is applied to extract cycles per event as sum of half and full cycle counts. The spatial effects on the temporal events are calculated using localized spatial autocorrelation with location of the turbines and extracted features as inputs. Simulated wind power production data for a virtual wind park located in high wind speed regions in Estonia are used as input for a case study to validate the proposed model. To demonstrate how threshold changes the result, a threshold test considering 10 thresholds are conducted. To extract the full spectrum of events a low, 0.1, threshold value is chosen for this study. RBA_θ accurately and precisely identified and quantified the events residing in the dataset. The extracted events can be used by the system operator for making decisions such as- scheduling the maintenance of the wind turbines and capacity building of the wind park. Moreover, the extracted events and associated features can be used in both operational and investment decision making process. The events extracted by RBA_θ are used to

reconstruct the original data with minor losses. The volume of the data is significantly reduced thereby the proposed model can also be used to compress the time varying data volume. For instance, if an event took place for 20 h then the RBA_θ summarizes it in its amplitude and persistence. The proposed model is applicable to any time varying data set with different objectives such as data compression, significant events, stationary events, etc.

Future work will be dedicated to optimizing the computational time of the proposed algorithm and perform benchmarking in comparison to other existing relevant algorithms. Further investigation is required to identify optimal threshold considering the inter-relation among the features such as the slope angle with varying threshold. In a future direction of research, the extracted events will be used for prediction, meaning predication of ramp events rather than whole time series to benchmark the model in parallel with traditional prediction techniques. In this study localized spatial autocorrelation is explored to enumerate the inter-relation among space and time on the wind power data. Further investigation is required to better understand the relationship and causality of the events.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the Estonian Research Council grant PUTJD915.

References

[1] Zappa W, Junginger M, van den Broek M. Is a 100% renewable european power system feasible by 2050? Appl. Energy 2019;(1027):233–4. http://dx.doi.org/10.1016/j.apenergy.2018.08.109.

[2] Zhang J, Cui M, Hodge BM, Florita A, Freedman J. Ramp forecasting performance from improved short-term wind power forecasting over multiple spatial and temporal scales. Energy 2017;122:528–41. http://dx.doi.org/10.1016/j.energy.2017.01.104.

[3] Ferreira M, Santos A, Lucio P. Short-term forecast of wind speed through mathematical models. Energy Rep 2019;5:1172–84. http://dx.doi.org/10.1016/j.egyr.2019.05.007.

[4] Ouyang T, Huang H, He Y. Ramp events forecasting based on long-term wind power prediction and correction. IET Renew Power Gener 2019;13:2793–801. http://dx.doi.org/10.1049/iet-rpg.2019.0093.

[5] Mishra S, Bordin C, Taharaguchi K, Palu I. Comparison of deep learning models for multivariate prediction of time series wind power generation and temperature. Energy Rep 2020;6:273–86. http://dx.doi.org/10.1016/j.egyr.2019.11.009.

[6] Quan H, Khosravi A, Yang D, Srinivasan D. A survey of computational intelligence techniques for wind power uncertainty quantification in smart grids. IEEE Trans Neural Netw Learn Syst 2019;1–18. http://dx.doi.org/10.1109/tnnls.2019.2956195.

[7] Dorado-Moreno M, Cornejo-Bueno L, Gutiérrez PA, Prieto L, Hervás-Martínez C, Salcedo-Sanz S. Robust estimation of wind power ramp events with reservoir computing. Renew Energy 2017;111:428–37. http://dx.doi.org/10.1016/j.renene.2017.04.016.

[8] Bianco L, Djalalova IV, Wilczak JM, Cline J, Calvert S, Konopleva-Akish E, Finley C, Freedman J. A wind energy ramp tool and metric for measuring the skill of numerical weather prediction models. Weather Forecast 2016;31:1137–56. http://dx.doi.org/10.1175/WAF-D-15-0144.1.

[9] Cui M, Zhang J, Florita AR, Hodge BM, Ke D, Sun Y. An optimized swinging door algorithm for identifying wind ramping events. IEEE Trans Sustain Energy 2016;7:150–62. http://dx.doi.org/10.1109/TSTE.2015.2477244.

[10] Karatepe S, Corscadden KW. Wind speed estimation: Incorporating seasonal data using Markov chain models. ISRN Renew Energy 2013;2013:657437. http://dx.doi.org/10.1155/2013/657437.

[11] Cao Y, Wei W, Mei S, Zhang X, Huang S. Estimating the probability of wind power ramp events: A distributionally robust approach. In: IEEE power and energy society general meeting. IEEE Computer Society; 2019, http://dx.doi.org/10.1109/PESGM40551.2019.8973995.

[12] Cui M, Feng C, Wang Z, Zhang J, Wang Q, Florita A, Krishnan V, Hodge B-M. Probabilistic wind power ramp forecasting based on a scenario generation method. Institute of Electrical and Electronics Engineers (IEEE); 2018, p. 1. http://dx.doi.org/10.1109/pesgm.2017.8274394.

[13] Kaut M. A copula-based heuristic for scenario generation. Comput Manag Sci 2014;11:503–16. http://dx.doi.org/10.1007/s10287-013-0184-4.

[14] Cui M, Krishnan V, Hodge BM, Zhang J. A copula-based conditional probabilistic forecast model for wind power ramps. IEEE Trans Smart Grid 2019;10:3870–82. http://dx.doi.org/10.1109/TSG.2018.2841932.

[15] Cao Y, Wei W, Wang C, Mei S, Huang S, Zhang X. Probabilistic estimation of wind power ramp events: A data-driven optimization approach. IEEE Access 2019;7:23261–9. http://dx.doi.org/10.1109/ACCESS.2019.2899404.

[16] Pavlou M, Ambler G, Seaman SR, Guttmann O, Elliott P, King M, Omar RZ. How to develop a more accurate risk prediction model when there are few events. BMJ 2015;351:7–11. http://dx.doi.org/10.1136/bmj.h3868.

[17] Shimada T, Kawamura H, Shimada M, Watabe I, Iwasaki SI. Evaluation of JERS-1 SAR images from a coastal wind retrieval point of view. IEEE Trans Geosci Remote Sens 2004;42:491–500. http://dx.doi.org/10.1109/TGRS.2003.821268.

[18] Kaiser L, Roy A, Nachum O, Bengio S. Learning to remember rare events. In: 5th int. conf. learn. represent. ICLR 2017 - conf. track proc. 2019, p. 1–10.
[19] Mishra S, Leinakse M, Palu I. Wind power variation identification using ramping behavior analysis. In: Energy procedia. Elsevier Ltd; 2017, p. 565–71. http://dx.doi.org/10.1016/j.egypro.2017.11.075.
[20] Mishra S, Leinakse M, Palu I, Kilter J. Ramping behaviour analysis of wind farms. In: Proc. - 2018 IEEE int. conf. environ. electr. eng. 2018 IEEE ind. commer. power syst. Eur. EEEIC/I CPS Eur. 2018. 2018, p. 1–5. http://dx.doi.org/10.1109/EEEIC.2018.8493720.
[21] Bossavy A, Girard R, Kariniotakis G. Forecasting ramps of wind power production with numerical weather prediction ensembles. Wind Energy 2013;16(1):51–63.
[22] Bossavy A, Girard R, Kariniotakis G. A novel methodology for comparison of different wind power ramp characterization approaches. In: Eur. wind energy conf. exhib. EWEC 2013, Vol. 2. 2013, p. 709–14.
[23] Hannesdóttir Á, Kelly M. Detection and characterization of extreme wind speed ramps. Wind Energy Sci 2019;4:385–96. http://dx.doi.org/10.5194/wes-4-385-2019.
[24] Drew DR, Barlow JF, Coker PJ. Identifying and characterising large ramps in power output of offshore wind farms. Renew Energy 2018;127:195–203. http://dx.doi.org/10.1016/j.renene.2018.04.064.
[25] Cornejo-Bueno L, Aybar-Ruiz A, Camacho-Gómez C, Prieto L, Barea-Ropero A, Salcedo-Sanz S. A hybrid neuro-evolutionary algorithm for wind power ramp events detection. In: Rojas I, Joya G, Catala A, editors. Advances in computational intelligence. Cham: Springer International Publishing; 2017, p. 745–56.
[26] Mahmood FH, Resen AK, Khamees AB. Wind characteristic analysis based on Weibull distribution of al-salman site. Iraq Energy Rep 2020;6:79–87. http://dx.doi.org/10.1016/j.egyr.2019.10.021.
[27] Downing SD, Socie DF. Simple rainflow counting algorithms. Int J Fatigue 1982;4:31–40. http://dx.doi.org/10.1016/0142-1123(82)90018-4.
[28] Harral BB. The application of a statistical fatigue life prediction method to agricultural equipment. Int J Fatigue 1987;9:115–8. http://dx.doi.org/10.1016/0142-1123(87)90053-3.
[29] Rychlik I. A new definition of the rainflow cycle counting method. Int J Fatigue 1987;9:119–21. http://dx.doi.org/10.1016/0142-1123(87)90054-5.
[30] Schluter LL, Sutherland HJ. SANDIA REPORT User's guide for LIFE2's rainflow counting algorithm SFYCOCI(S-81), 1991.
[31] Carle SF, Fogg GE. Modeling spatial variability with one and multidimensional continuous-lag Markov chains. Math Geol 1997;29:891–918. http://dx.doi.org/10.1023/A:1022303706942.
[32] Rey SJ. Spatial empirics for economic growth and convergence. Geogr Anal 2010;33:195–214. http://dx.doi.org/10.1111/j.1538-4632.2001.tb00444.x.
[33] Rey SJ, Anselin L. PySAL: A python library of spatial analytical methods. Rev Reg Stud 2007;37:5–27. http://dx.doi.org/10.1007/978-3-642-03647-7_11.
[34] Staffell I, Pfenninger S. Using bias-corrected reanalysis to simulate current and future wind power output. Energy 2016;114:1224–39.
[35] Staffell I, Pfenninger S. Using bias-corrected reanalysis to simulate current and future wind power output. Energy 2016;114:1224–39. http://dx.doi.org/10.1016/j.energy.2016.08.068.
[36] Kamath C. Understanding wind ramp events through analysis of historical data. In: 2010 IEEE PES transmission and distribution conference and exposition: smart solutions for a changing world. 2010, http://dx.doi.org/10.1109/TDC.2010.5484508.
[37] Mishra S, Leinakse M, Palu I, Kilter J. Ramping behaviour analysis of wind farms. In: Proceedings - 2018 IEEE international conference on environment and electrical engineering and 2018 IEEE industrial and commercial power systems Europe, EEEIC/I and CPS Europe 2018. Institute of Electrical and Electronics Engineers Inc; 2018, http://dx.doi.org/10.1109/EEEIC.2018.8493720.
[38] Dorado-Moreno M, Navarin N, Gutiérrez PA, Prieto L, Sperduti A, Salcedo-Sanz S, Hervás-Martínez C. Multi-task learning for the prediction of wind power ramp events with deep neural networks. Neural Netw 2020;123:401–11. http://dx.doi.org/10.1016/j.neunet.2019.12.017.
[39] Zhang Y, Ai Q, Xiao F, Hao R, Lu T. Typical wind power scenario generation for multiple wind farms using conditional improved wasserstein generative adversarial network. Int J Electr Power Energy Syst 2020;114:105388. http://dx.doi.org/10.1016/j.ijepes.2019.105388.
[40] Mishra S, Bordin C, Tomasgard A, Palu I. A multi-agent system approach for optimal microgrid expansion planning under uncertainty. Int J Electr Power Energy Syst 2019;109:696–709. http://dx.doi.org/10.1016/j.ijepes.2019.01.044.
[41] Qiao B, Liu J. Multi-objective dynamic economic emission dispatch based on electric vehicles and wind power integrated system using differential evolution algorithm. Renew Energy 2020;154:316–36. http://dx.doi.org/10.1016/j.renene.2020.03.012.