



# Heuristic Feature Selection for Wind Power Anomaly Events Study

Peiwen Yu<sup>1</sup> and Anping Lin<sup>2\*</sup>

<sup>1</sup>Maritime College, Guangdong Ocean University, Zhanjiang, China, <sup>2</sup>School of Physics and Electronic Electrical Engineering, Xiangnan University, Chenzhou, China

Wind power ramp events are typical harmful anomaly events in wind engineering, which bring new threat to the safety operation of power systems. To in-depth understand ramps and mitigate their harms, suitable ramp characteristics are crucial in many studies, e.g., ramp definition, classification, prediction and so on. However, due to ramps' specificity on event feature, more profound characteristics are needed besides basic ramp morphological characteristics. In this paper, an approach for extracting and selecting ramp characteristics is proposed for ramp study. First, according to ramps' causation on energy change, wavelet transformation is introduced to analyze ramp categories, and used to extract ramp energy characteristics. Then, heuristic feature selection methods are proposed to select ramp characteristics based on specific ramp application contexts. The objective of feature selection is to remove redundant characteristics, and to improve ramp studies' performance. Finally, combining basic ramp characteristics and wavelet characteristics, ramp studies on category classification and prediction of appointed characteristics are implemented on industrial data. The computational results validate the usefulness of wavelet characteristics, the feasibility of the proposed approach, and that performance of ramp study could be improved by using ramp characteristics in this paper.

**Keywords:** wind power ramp events, wavelet transform, feature selection, anomaly detection, feature exaction

## INTRODUCTION

The generation of energy from wind is growing across the world, especially in China where large-scale and highly-concentrated wind projects prevail (Ouyang et al., 2017a). Due to the renewability feature, wind energy offered lots of opportunities, e.g., proving clean energy and reducing environment pollution. On the other hand, due to wind's fluctuation and intermittent, serious anomaly challenges threaten the safety and stability of power grid. For example, wind power ramp events are typical anomaly events bringing one of the greatest threat, which is namely the large and unexpected changes of wind power over a short time period (Wang et al., 2017). In 2008, a down-ramp event was reported in the State of Texas causing serious economic loss to the grid operated by Electric Reliability Council of Texas (ERCOT) (Francis, 2008). Therefore, it is significantly important to study ramp events for mitigating their negative impacts.

Ramp study mainly involves definitions, detection, prediction and classification. Generally, ramp definition and ramp detection are the basis of ramp study. While ramp events are usually detected by combining ramp definitions and specific detection methods. For example, dynamic programming recursion and the swinging door algorithm were proposed to detect ramp events from wind power data in (Florita et al., 2013; Sevlian and Rajagopal, 2013; Ouyang et al., 2017b). In (Xiong et al., 2017), a data mining method using affinity of weather data was also proposed for ramp detection. However, the mainstream definitions up to now mainly focus on three characteristics (Zha et al., 2016) (e.g.,

## OPEN ACCESS

### Edited by:

Tinghui Ouyang,  
National Institute of Advanced  
Industrial Science and Technology  
(AIST), Japan

### Reviewed by:

Zhenhao Tang,  
Northeast Electric Power University,  
China  
Huangke Chen,  
National University of Defense  
Technology, China

### \*Correspondence:

Anping Lin  
anping719@126.com

### Specialty section:

This article was submitted to  
Wind Energy,  
a section of the journal  
Frontiers in Energy Research

**Received:** 11 August 2021

**Accepted:** 14 September 2021

**Published:** 01 October 2021

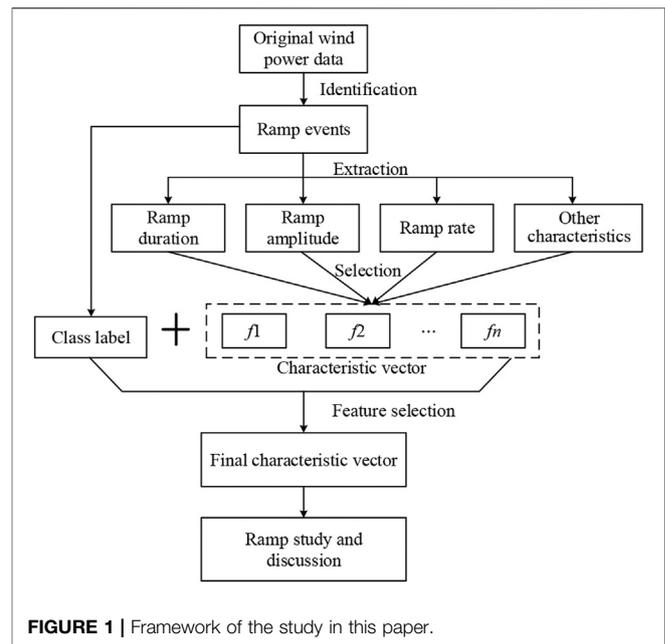
### Citation:

Yu P and Lin A (2021) Heuristic Feature  
Selection for Wind Power Anomaly  
Events Study.  
Front. Energy Res. 9:756733.  
doi: 10.3389/fenrg.2021.756733

ramp amplitude, ramp duration, and ramp rate) which are superficial characteristics from ramp events' basic forms. Copying with complex power system operations in the future, more targeted and effective control strategies need to be made, which are essentially based on study of more profound ramp characteristics. On the other hand, ramp prediction and classification are two major objectives of ramp study. Generally, ramp prediction can be divided into event prediction and ramp's categories prediction which is actually ramp classification. Nowadays, ramp classification has been studied via many data mining algorithms, e.g., *k-means*, support vector machine (SVM), extrema learning machine (ELM), neural networks (NN) and so on (Couto et al., 2013; Florita et al., 2013; Tang et al., 2020; Shen et al., 2021). Ramp classification combined with weather regimes was also studied in (Chen et al., 2018). No matter ramp prediction or classification, proper input features are the premise of constructing high-performance models. However, most of these studies are based on superficial characteristics, no profound physical characteristics are considered to improve ramp studies' performance. Therefore, it is significant useful for studying extraction and selection of ramp characteristics in wind power ramp researches.

In modeling process, determination of input features is primarily based on original data points. Sometimes feature extraction, selection and transformation are involved according to specific criterions. For example, principal component analysis (PCA) is a commonly-used method to reduce dimension of feature space, which was also used in feature analysis for wind forecasting (He et al., 2013a). Other methods based on mathematical transforms were also useful to extract characteristics, e.g. wavelet-transform was utilized in wind power forecasting (Singh and Tewari, 2015). Moreover, Pearson correlation coefficient, Gini index, wavelet transformation and other intelligent tools were also applied to feature processing in engineering (Huang et al., 2018). Nowadays, with more industrial signals are collected from complicated systems and more un-researchable objects are analyzed, these situations lead to the urgent requirement of profound characteristics reflecting structural or physical features in modeling. Study on wind power ramp events is a representative problem among these issues. According to ramps' concept, ramp events involve a period of wind power values and variance, and they don't always have a unified time duration (Tang et al., 2021). Based on the traditional tools, basic characteristics (e.g., ramp amplitude, duration and ramp rate) could be obtained from wind power data. However, ramps' harmful effect on power grid is not just identified by ramp duration and amplitude. The difference of ramp amplitude, ramp rate, energy storage and other factors may also affect the stability of power system at different degree. Therefore, besides the basic ramp characteristics, how to acquire more profound characteristics is an urgent topic in ramp study, such as in ramp classification and prediction.

According to the outlined problems above, the objective of this paper is to propose an approach to extract and select ramp



characteristics for wind power ramp study. Considering ramp events involves the variance of time series and energy change process of wind power, three basic ramp characteristics are extracted based on definitions firstly. These characteristics are able to distinguish ramps and non-ramp events. Then, a method transforming time series into energy forms is proposed to extract extra characteristics. Wavelet transformation has been applied for feature expression in literatures due to its superior description ability at both time and frequency domains. For example, the wavelet transform was utilized to analyze the features of ramp events in (Gallego et al., 2013). On the other hand, wavelet decomposition has the property of multiresolution which is helpful to study the allocation of ramp event's energy. Summarizing these two properties, wavelet transformation is proposed to extract profound characteristics for distinguishing refined ramp categories in depth. Moreover, for reducing dimension, mitigating noise's influence, improving computation efficiency, a heuristic intelligent algorithm is proposed in the feature selection. Finally, based on the refined characteristic vector, ramp studies (classification, prediction) on industrial datasets are discussed, and validate the approach in extracting and selecting ramp characteristics. The framework of the major work in this paper is sketched in the following figure.

In **Figure 1**, the identification of historical ramp events is implemented through ramp definitions and detection. Class labels of ramp events are determined by a given classification environment which could be based on weather sceneries, control requirement and so on. The other characteristics are energy characteristics extracted by wavelet transformation in this paper, and  $f_n$  represents the  $n$ th characteristic in the formed characteristic vector in **Figure 1**. According to the above description, we can conclude the novelties of this paper as following three points:

- i) This paper proposes to consider characteristics of ramp events from multiple aspects, including the basic morphological characteristics from time domain and characteristics in frequency domain. Moreover, due to ramp events' specificity, characteristics are also extracted in terms of energy. This is the first time to consider ramp events' energy characteristics.
- ii) Wavelet transformation is proposed to extract ramp characteristics. By utilizing the frequency and multiresolution properties in wavelet decomposition, ramp energy characteristics are expressed by energy at different frequency spaces. Meanwhile, these energy characteristics are selected and refined. Its purpose is to delete the energy of noise which may affect ramp classification.
- iii) Heuristic intelligent algorithms are tried in the selection process, e.g., the sequential forward floating search (SFFS) and sequential backward floating search (SBFS) methods. The heuristic feature selection could refine the characteristic vector in related studies more effectively than unsupervised methods reducing dimension.

Besides the above introduction, rest of this paper is organized as follows. *Feature Extraction* addresses the processes of extracting ramp characteristics, including the basic ramp characteristics based on ramp definitions, and ramp energy characteristics based on wavelet transformation, for ramp studies are also extracted. *Feature Selection* proposes the feature selection approach which is based on dispersion matrix and heuristic method. The detailed processes are also presented in this section. *Ramp Anomaly Analysis and Evaluation* aims at designing ramp studies, e.g., ramp classification and ramp prediction. It also gives out some indicators for evaluating ramp study. In *Experiments and Discussion*, industrial wind power data is utilized in case study. Performance on ramp classification and prediction are compared with models using different feature sets and that using PCA for feature selection. Computational results validate the feasibility of wavelet characteristics and the proposed feature selection method. Finally, *Conclusion* concludes this paper.

## FEATURE EXTRACTION

### Basic Characteristics of Ramp Events

Wind power ramp events bring great harm to system operation as more and more wind power integrated into power grid. Copying with these new events, a series of studies has been carried out, such as ramp definition, ramp prediction and classification. Currently, there are four mainstream definitions widely used in ramp analysis (Zha et al., 2016). These definitions are defined as follows.

**Definition 1.** When the change of wind power in time duration  $\Delta t$  exceeds a given threshold  $Wp_{vab}$  there is a ramp event occurring, as expressed as below.

$$|Wp(t + \Delta t) - Wp(t)| \geq Wp_{val} \tag{1}$$

where:  $Wp(t+\Delta t)$  and  $Wp(t)$  are wind power values at time  $t+\Delta t$  and  $t$ , respectively;  $Wp_{val}$  represents the threshold of ramp amplitude. When the criterion in Eq. 1 meets under a given time period, a ramp event is identified.

**Definition 2.** When the largest difference of wind power in time duration  $\Delta t$  exceeds the given threshold  $Wp_{vab}$  then a ramp event is regarded as occurring.

$$\left| \max_{i \in [t, t+\Delta t]} Wp_i - \min_{i \in [t, t+\Delta t]} Wp_i \right| \geq Wp_{val} \tag{2}$$

where,  $\max_{i \in [t, t+\Delta t]} Wp_i$  and  $\min_{i \in [t, t+\Delta t]} Wp_i$  represent the maximum and minimum value of wind power in a time duration  $\Delta t$ .

**Definition 3.** When the average change of wind power in a given time duration  $\Delta t$  exceeds the given threshold  $Wp_{vab}$  a ramp event is identified.

$$\left( \sum_{i=1}^h |Wp(t + \Delta t + i) - Wp(t + i)| \right) / h \geq Wp_{val} \tag{3}$$

where,  $h$  is the time horizon in ramp identification.

**Definition 4.** When the ratio between the wind power change  $|Wp(t+\Delta t)-Wp(t)|$  and time duration  $\Delta t$  exceeds a given threshold  $R_{vab}$  then a ramp event is regarded as occurring.

$$|Wp(t + \Delta t) - Wp(t)| / \Delta t > R_{val} \tag{4}$$

where, this definition in Eq. 4 pays more attention on ramp rate while that in Eq. 1 emphasizes only ramp amplitude. Summarizing the above four definitions, we can see they mainly focus on three major characteristics (Zha et al., 2016), such as ramp amplitude  $Var$ , ramp duration  $T$ , and ramp rate  $R$ . Based on these three characteristics, the variance of wind power in a given period could be determined as ramps or non-ramps. Therefore, they consist of the initial characteristic vector in ramp study, as  $C_{v0} = [Var, T, R]$ . As we know, sometimes ramp direction is also used in characteristic analysis. However, through above numerical definitions, ramp direction is not necessary, and could be expressed by sign of  $R$  when needed. Considering that ramp analysis in some cases needs the detailed information about power range rather than only fluctuation amplitude, so we replace ramp amplitude  $Var$  with maximum and minimum values. Assuming there are  $N$  studied ramp events in dataset, the initial characteristic matrix  $X_0$  is expressed as below.

$$X_0 = \begin{bmatrix} Wp_{1,max} & Wp_{1,min} & T_1 & R_1 \\ Wp_{2,max} & Wp_{2,min} & T_2 & R_1 \\ \vdots & \vdots & \vdots & \vdots \\ Wp_{N,max} & Wp_{N,min} & T_N & R_N \end{bmatrix} \tag{5}$$

where,  $Wp_{i,min}$  and  $Wp_{i,max}$  represent the minimum and maximum value of wind power in the  $i$ th studied ramp event.

## Wavelet Transformation and Characteristics Extraction

As the research on ramp events get more in depth, profound characteristics are needed to describe ramps' formation and categories. For example, the initial characteristics in Eq. 4 is used to identify the basic formation of ramps and non-ramps. The detailed fluctuation inside ramps needs structural characteristics, and the essence of ramps' occurrence involves energy characteristics. In order to extract more profound information in ramp study, wavelet transformation is applied in this paper. Wavelet transformation is an advanced mathematical technique in signal analysis (Mohanty et al., 2015). It has advantages at decomposing a signal into various time and frequency domains, so it is useful to study the structural characteristics of ramps in different domains. Wavelet transformation also has advantages at detecting abrupt changed values (e.g., in edge detection) and analyzing signal in a specific time window. While ramp events certainly have large power charge in finite time durations, so it is relatively suitable to utilize wavelet transformation in ramp analysis (Escalante Soberanis and Mérida, 2015). On the other hand, due to the multiresolution feature of wavelet decomposition, energy of a given signal could be allocated into different frequency spaces. In this way, we are inspired to extract characteristics for expressing ramps' energy characteristics. This is also a key reason for considering wavelet transformation in extracting ramp characteristics.

The theory of discrete wavelet transformation is described as below. Assuming a discrete signal (e.g., wind power time series of a ramp event) is expressed as  $\{x(t); t = 1, 2, \dots, T\}$ , it could be reconstructed by elements of wavelet transformation, as expressed below.

$$x(t) = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \delta_{j,k} \cdot \psi_{j,k}(t) \tag{6}$$

where:  $x(t)$  is actually the signal of wind power;  $\delta_{j,k}$  is the wavelet coefficient;  $\psi_{j,k}$  is the child wavelet transformed from mother wavelet, denoted as follows.

$$\begin{cases} \psi_{j,k}(t) = a^{-j/2} \psi(a^{-j} \cdot t - kb) \\ \delta_{j,k} = \sum_{-\infty}^{\infty} \psi_{j,k}(t) \cdot x(t) \\ a > 1, b > 0 \end{cases} \tag{7}$$

where:  $j$  and  $k$  represent the scale and shift parameters of child wavelet  $\psi_{j,k}$ ;  $\psi(t)$  is the mother wavelet function;  $a$  and  $b$  are real parameters. From Eqs 6, 7 it implies that the reproduction of original signal could be realized by the weighted sum of wavelet components at different scales.

To in-depth explain the meaning of each wavelet component in Eq. 6, we could operate wavelet decompose step by step. Assuming the original signal with finite energy is projected on a space  $L$ ,  $x(t)$  with one-level-decomposition could be expressed as below.

$$x(t) = x_1(t) + \sum_k d_k \cdot \psi_k(x) \tag{8}$$

where,  $x_1(t)$  is the estimation of the original signal reflecting variation in time domain;  $\sum_k d(k) \cdot \psi_k(x)$  are detailed signals

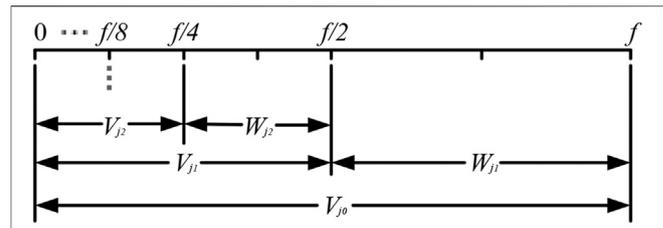


FIGURE 2 | The orthogonal decomposition of the space  $L$ .

expressed by wavelet functions which contain frequency-domain information. By estimating the approximate signal iteratively,  $x_j(t)$  at the  $j$ th decomposition level is expressed as below.

$$x_j(t) = x_{j+1}(t) + \sum_k d_{j,k} \cdot \psi_{j,k}(t) \tag{9}$$

Similarly, we could utilize a series of wavelet functions to describe signal  $x_j(t)$ , as  $\sum c_{j,k} \phi_{j,k}(t)$ . Combining these formulas, transformation of the original signal in Eq. 6 could be rewritten in details as below

$$x(t) = \sum_k c_{j_n,k} \phi_{j_n,k}(t) + \sum_{j=1}^{j_n} \sum_k d_{j,k} \psi_{j,k}(t) \tag{10}$$

where, the auxiliary function  $\phi$  is called father wavelet;  $c_{j_n,k}$  is the coefficients of wavelets  $\phi_{j_n,k}(t)$ ;  $d_{j,k}$  are the coefficients of wavelets  $\psi_{j,k}(t)$ . The formula in Eq. 10 is generally called as the multi-resolution analysis of wavelet transformation (Doucoure et al., 2016).

On the other hand, by corresponding each wavelet component to a frequency space, we could also divide the space  $L$  to series of energy subspaces. As shown in Figure 2,  $V_{j_0}$  represents the original signal space  $L$  with a frequency band  $(0 \sim f)$ . According to Eq. 10, the frequency band is also divided step by step. For example,  $V_{j_0}$  could be divided as orthogonal sum of a low-frequency space  $V_{j_1}$   $(0 \sim f/2)$  and a high-frequency space  $W_{j_1}$   $(f/2 \sim f)$ . The relationship reflecting the division of frequency spaces in wavelet transformation is presented as below.

$$\begin{cases} V_j \oplus W_j = V_{j-1} \\ L = W_{j_1} \oplus W_{j_2} \oplus \dots \oplus W_{j_n} \oplus V_{j_n} \end{cases} \tag{11}$$

where:  $\oplus$  is a denoted operator calculating the orthogonal sum;  $n$  is the number of wavelet decomposition.

Considering signal's energy is generally expressed at frequency subspaces, therefore we could utilize the multi-resolution of wavelet decomposition to analyze the energy distribution of a signal (Ashrafiyan et al., 2017), as described in the following formula.

$$|P_{\{x\}}|^2 = \sum_k |c_{j_n,k}|^2 + \sum_j \sum_k |d_{j,k}|^2 \tag{12}$$

where:  $|P_{\{x\}}|^2$  represent the energy of the given signal  $\{x(t)\}$ ;  $\sum |c_{j_n,k}|^2$  and  $\sum |d_{j,k}|^2$  represent the energy of different subspaces. Since wind process generally involved atmosphere

movements (Mohanty et al., 2015), the occurrence of ramp events could be comprehensibly regarded as the speedy energy release or accumulation in atmosphere systems. Therefore, the energy expressions based on wavelet transform could be utilized to extract ramp energy characteristics which is meaningful in studying ramp's harms on power systems.

Moreover, when the number of decomposition levels is high (e.g.,  $j_n \rightarrow \infty$ ), the value of  $c_{j_n,k}$  is small, which implies the energy of  $V_{j_n}$  becoming small to be ignored. In that case, the energy characteristics of  $W_j$  are mainly selected into the characteristic vector for ramp study in this paper.

$$\begin{cases} P_j = \sum_k |d_{j,k}|^2; \\ C_v = [W_{p_{\max}}, W_{p_{\min}}, T, R, P_j]; \\ j = 1, 2, \dots, j_n; \end{cases} \quad (13)$$

where:  $P_j$  represents the  $j$ th energy characteristic and  $j$  is the decomposition level.  $C_v$  is the characteristic vector combining initial vector  $C_{v0}$  and  $P_j$ . By utilizing this vector including basic ramp characteristics and wavelet energy characteristics as inputs, some advanced ramp study could be implemented besides the identification of ramps and non-ramps.

## FEATURE SELECTION

According to the above wavelet decomposition,  $j_n$  energy characteristics are extracted. Generally, the more the number of decomposition levels, the better the description ability of wavelet characteristics. In Eq. 13, the energy in frequency space  $V_{j_n}$  is excluded since  $V_{j_n}$  is the lowest frequency and its  $c_{j_n,k}$  is very small. However, when more fine-sorted energy characteristics  $P_j$  are generated, it is unavoidable to lead to many superfluous characteristics in a specific study case. Therefore, excluding basic ramp characteristics, feature selection is also necessary in ramp study.

Feature selection could not only select optimal energy characteristics for specific study, but also improve computing performance by reducing data dimension. Generally, feature selection methods are based on specific indicators or criterions to rank all characteristics, then realize selection through ranking scores. Most of these methods do not care about the application context in ranking. The other commonly used methods on dimension reducing is through feature transformation, such as PCA, LDA. This type of methods weakens the physical meaning of selected characteristics, and also ignore actual context. Therefore, in this paper we propose to utilize heuristic selection criterion which combining selection indicators and the application context (e.g., specific ramp study).

## Dispersion Matrix

First, we propose to utilize dispersion matrix (Gu et al., 2017) to create selection indicator. Dispersion matrix is a mathematic tool based on feature distances of different classes, its elements are denoted as below.

$$\begin{cases} \mathbf{B} = \sum_{g=1}^G n_g (\boldsymbol{\mu}_g - \boldsymbol{\mu}_0)^T (\boldsymbol{\mu}_g - \boldsymbol{\mu}_0) \\ \mathbf{W} = \sum_{g=1}^G \sum_{k=1}^{n_g} (\mathbf{x}_{g,k} - \boldsymbol{\mu}_g)^T (\mathbf{x}_{g,k} - \boldsymbol{\mu}_g) \\ (k = 1, 2, \dots, n_g; g = 1, 2, \dots, G; \\ n_1 + n_2 + \dots + n_g = N) \end{cases} \quad (14)$$

where:  $\mathbf{W}$  and  $\mathbf{B}$  are the dispersion matrix representing intra-class and inter-classes, respectively;  $\mathbf{x}_{g,k}$  is characteristic vector of the  $k$ th ramp in the  $g$ th class;  $\boldsymbol{\mu}_g$  and  $\boldsymbol{\mu}_0$  represent the average vector of the  $g$ th class and all classes, respectively;  $n_g$  and  $N$  are the number of ramps the  $g$ th class and all classes, respectively;  $G$  is the number of classes. A Wilks criterion function  $\lambda_p$  could be selected as the reference indicator in feature selection, it is defined as below.

$$\lambda_p = |\mathbf{W}|/|\mathbf{T}| \quad (15)$$

where:  $\mathbf{T}$  is total dispersion matrix, calculated as  $\mathbf{T} = \mathbf{B} + \mathbf{W}$ ;  $p$  is the dimension of feature space. When the value of  $\lambda_p$  is small, implying a small value of  $|\mathbf{W}|$  and a large value of  $|\mathbf{T}|$ , it illustrates that the characteristic is effective to distinguish different classes of samples. Generally, the statistical indicator  $\lambda_p$  is assumed to obey the Wilks distribution. By deciding a testing level  $\alpha$  and its corresponding threshold  $\lambda_{(\alpha)}$ , the hypothesis testing of  $\lambda_p$  could be implemented. For the convenience of calculation, the value of Wilks distribution function could be estimated by the following two common distribution functions.

1) Bartlett approximation.

$$-\left(N - \frac{1}{2}(p - g) - 1\right) \cdot \ln \lambda_p \sim \chi^2(p(g - 1)) \quad (16)$$

2) Rao approximation.

$$\frac{N - (p - 1) - g}{g - 1} \cdot \left(\frac{\lambda_{p-1}}{\lambda_p} - 1\right) \sim F(g - 1, N - (p - 1) - g) \quad (17)$$

Through the above two approximation methods, the hypothesis testing of Wilks distribution could be realized by formulas in Eqs 16, 17.

## Heuristic Selection

Combining the selected Wilks indicator  $\lambda_p$  and ramps' categories information, a supervised heuristic method could realize high-performance feature selection. In this paper, we propose to utilize SFFS (sequential floating forward search) and SFBS (sequential floating backward search) algorithms (Gan et al., 2014) which retain the strengths and improve the weakness of SFS (sequential forward selection) and SBS (sequential backward selection).

Considering the specificity of ramp studies, the basic characteristic vector  $C_{v0}$  contains basic ramp characteristics identifying ramps and non-ramps, so it is necessary included in the characteristic subset. Actually, the task of feature selection is to select optimal newly-extracted characteristics in ramp study.

Therefore, in this paper the initial subset is  $\mathbf{X}_0$ , the characteristics that need to be processed are wavelet energy

characteristics  $P_i (i = 1, 2, \dots, j_n)$ . Assuming after the  $m$ th selection step, the feature subset is denoted as  $\mathbf{X}_m$ . If we consider to add a new characteristic  $x_r$  at the  $(m+1)$ th step by SFFS algorithm, the updated dispersion matrix could be calculated as below.

$$\begin{cases} \mathbf{W}^* = \begin{pmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{pmatrix} \\ \mathbf{T}^* = \begin{pmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{pmatrix} \end{cases} \quad (18)$$

where,  $\mathbf{W}_{11}$  and  $\mathbf{T}_{11}$  are the intra-class and the total dispersion matrix of  $\mathbf{X}_m$ , respectively; The rest of sub-matrix are newly introduced matrix related with  $x_r$ , calculated as the following formulas.

$$\begin{cases} \mathbf{W}_{12} = (w_{1r}, w_{2r}, \dots, w_{pr})^T; \mathbf{W}_{21} = \mathbf{W}_{12}^T; \mathbf{W}_{22} = w_{rr}; \\ \mathbf{T}_{12} = (t_{1r}, t_{2r}, \dots, t_{pr})^T; \mathbf{T}_{21} = \mathbf{T}_{12}^T; \mathbf{T}_{22} = t_{rr}; \end{cases} \quad (19)$$

Based on calculation of these sub-matrixes, the updated indicator  $\lambda_{m+1}$  could be calculated as below.

$$\begin{cases} \lambda_{m+1} = \frac{|\mathbf{W}^*|}{|\mathbf{T}^*|} = \frac{|\mathbf{W}_{11}| |\mathbf{W}_{22} - \mathbf{W}_{21} \mathbf{W}_{11}^{-1} \mathbf{W}_{12}|}{|\mathbf{T}_{11}| |\mathbf{T}_{22} - \mathbf{T}_{21} \mathbf{T}_{11}^{-1} \mathbf{T}_{12}|} = \lambda_m \cdot A_r \\ A_r = \frac{|\mathbf{W}_{22} - \mathbf{W}_{21} \mathbf{W}_{11}^{-1} \mathbf{W}_{12}|}{|\mathbf{T}_{22} - \mathbf{T}_{21} \mathbf{T}_{11}^{-1} \mathbf{T}_{12}|} \end{cases} \quad (20)$$

By substituting Eqs 15–18, the expression  $\lambda_m/\lambda_{m+1} - 1$  could be replaced by  $(1 - A_r)/A_r$ . Then, the estimated value of testing  $x_r$  is denoted as  $F_{1r}$ , expressed as below

$$F_{1r} = \frac{1 - A_r}{A_r} \cdot \frac{N - p_m - g}{g - 1} \sim F(g - 1, N - p_m - g) \quad (21)$$

where:  $p_m$  is the number of characteristics in set  $\mathbf{X}_m$ . If  $F_{1r} > F_\alpha(g-1, N-p_m-g)$  at given testing level  $\alpha$ , then the hypothesis is correct and  $x_r$  is added into characteristic vector.

Similarly, if we utilize the SFBS algorithm to delete a characteristic  $x_r$  from the set  $\mathbf{X}_m$ , the final testing value is defined as  $F_{2r}$ , expressed as below

$$F_{2r} = \frac{1 - A_r}{A_r} \cdot \frac{N - (p_m - 1) - g}{g - 1} \sim F(g - 1, N - (p_m - 1) - g) \quad (22)$$

If the formula  $F_{2r} \leq F_\alpha(g-1, N-(p_m-1)-g)$  is satisfied at a given level  $\alpha$ , then the variable  $x_r$  is regarded as invalid and removed from  $\mathbf{X}_m$ .

## RAMP ANOMALY ANALYSIS AND EVALUATION

By taking the selected ramp characteristics as inputs and different types of data as output, we could construct models for different ramp study, e.g., ramp classification and ramp prediction. In this paper, we only do some simple experiments on these two studies for evaluating the selected ramp characteristics.

## Ramp Classification

In ramp classification study, class labels of historical ramp events are taken as the output. The classification model can be constructed by data mining algorithms. To accurately classify ramp events, four data mining algorithms are applied to train ramp classification model, including support vector machine (SVM), neural networks (NN), random forests algorithm (RF) and boosted trees (BT) (Chen et al., 2017; Ouyang et al., 2017c; He et al., 2017; Ouyang, 2021). The optimal classification model could be determined by the comparison of their performance.

To evaluate the classification performance, the confusion matrix introduced from information retrieval (IR) field is widely applied (He et al., 2013b). The detailed expression is presented in the following table.

In Table 1, four types of events are defined, such as true positive event (TP), false negative event (FN), false positive event (FP) and true negative event (TN). Based on these events, several indicators could be defined to evaluate classification performance. Four representative indicators are defined as below.

$$\begin{cases} Rec = num_{TP} / (num_{TP} + num_{FN}) \\ Pre = num_{TP} / (num_{TP} + num_{FP}) \\ Acc = (num_{TP} + num_{TN}) / num_{All} \\ Err = 1 - Acc \end{cases} \quad (23)$$

where,  $num_X$  represents the number of the specific event X;  $Pre$  represents precision indicator implying the percentage of TP in classified true events;  $Rec$  represents recall indicator implying the percentage of TP in observed true events;  $Acc$  is the classification accuracy, and  $Err$  is the classification error. By utilizing these four indicators, we could complete selection of the optimal classification model and the evaluation of classification performance.

## Ramp Prediction

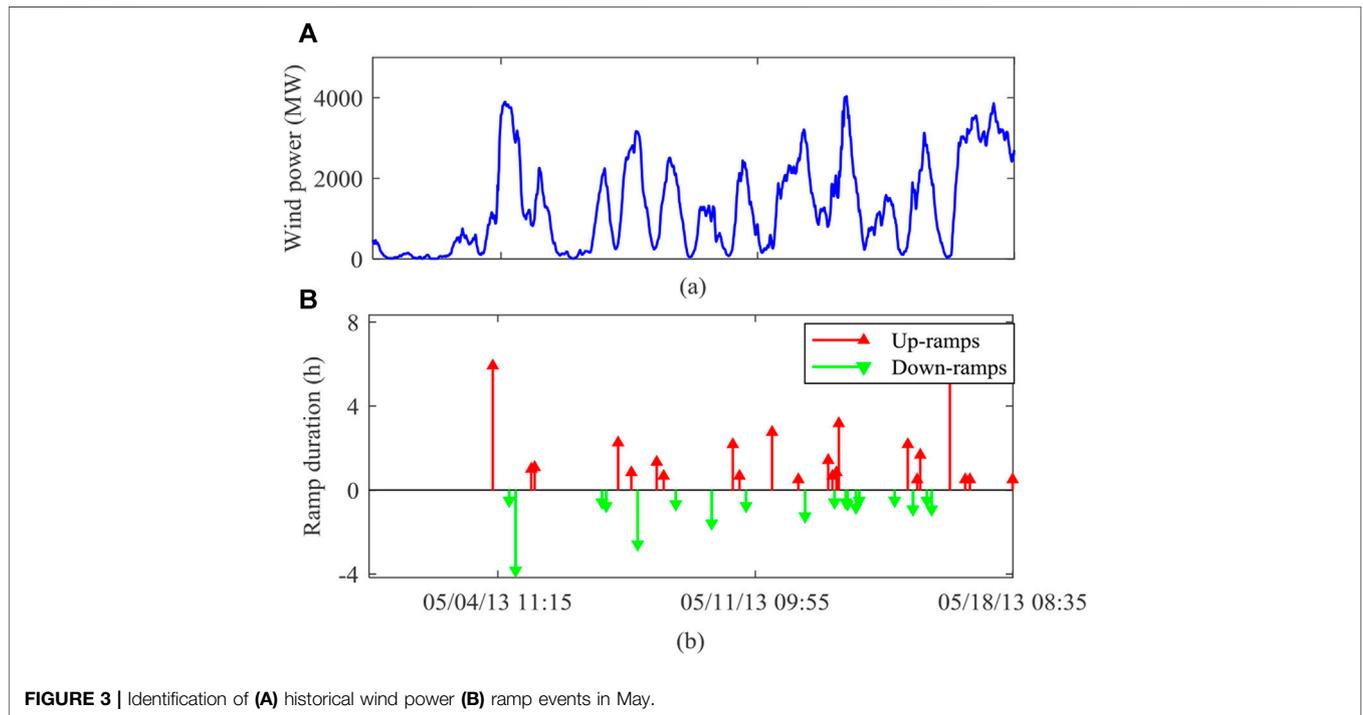
Ramp prediction is usually divided into two types: event prediction and regression prediction. Event prediction includes ramp detection and ramp classification study. Regression prediction mainly focus on utilizing traditional regression models to predict ramp characteristics, e.g., the ramp rate prediction in (Zheng and Kusiak, 2009). In this paper, we consider predicting two characteristics: ramp amplitude and ramp rate, which implies values of these two characteristics are taken as the output in modeling. Since the above study on ramp characteristics' extraction and selection are based on historical ramp events which randomly occur in wind power time series. Therefore, we propose to utilize these characteristics extracted from wind power in a given time window as inputs, then predict one appointed characteristic in the predicted time window, e.g., to predict ramp amplitude or ramp rate in the future 1-h horizon. Since this type of prediction is still based on regression models, the performance indicator could be decided by the commonly used root-mean-square error (RMSE), which is defined as below.

$$RMSE = \sqrt{\sum_{k=1}^K (\hat{y}_k - y_k)^2 / K} \quad (24)$$

where,  $y_k$  and  $\hat{y}_k$  are the  $k$ th values predicted and observed of an appointed ramp characteristic;  $K$  is the number of tested samples.

**TABLE 1** | Confusion matrix.

		Observation		Total
		True	False	
Classification	True	TP	FP	True classification
	False	FN	TN	False classification
Total		True observation	False observation	All



## EXPERIMENTS AND DISCUSSION

In this paper, the industrial wind power data from Bonneville Power Administration (BPA) website ([bpa.gov/transmission](http://bpa.gov/transmission), 2013) is taken as the studied case. The data set spanning from 01/01/13 00:00 to 12/31/13 23:55 totally has 105,120 data points with a sampling interval of 5 min. First, we need to detect historical ramp events from wind power time series for the following ramp characteristics study. According to the definitions in Eqs 1–4, the fourth definition can reflect three basic ramp characteristics more conveniently, so that it is selected to identify historical ramps in this paper. In (Tang et al., 2021), the value of  $R_{val}$  was chosen as 50% of the installed capacity within 4 h, so the value of  $R_{val}$  is computed by considering the capacity of 4,500 MW in the studied case, as below.

$$R_{val} = \frac{50\% \cdot P_{total}}{4hours} = \frac{50\% \cdot 4500MW}{4h} = 562.5 MW/h \quad (25)$$

Considering ramps always have a duration larger than 0.5 h, so we assume the minimum threshold as  $\Delta t = 0.5 h$  there are totally 526 ramp events are detected from data of former 6 months. One

part of ramp identification results is shown in Figure 3. The subpicture 1) and 2) depict historical wind power and ramp events, respectively. In Figure 3B, up-ramps and down-ramps are expressed by lines above and below the X-axis, respectively, their durations are reflected by values in Y-axis. It is seen that most of ramps have duration around hours, some even reach 6 h.

Historical wind power and ramp events are depicted in Figure 3A and Figure 3B, respectively. In Figure 3B, up-ramps and down-ramps are expressed by lines above and below the X-axis, respectively, their durations are reflected by values in Y-axis.

### Selection of Ramp Characteristics

Through analysis on ramp definitions, basic ramp characteristics (amplitude, duration, and ramp rate) could be extracted, as (5). These basic characteristics could be used to identify ramps and non-ramps, as two typical signals in Figure 4. To identify more detailed division of ramp categories, wavelet transformation is proposed to extract profound characteristics.

In order to illustrate the feasibility of wavelet characteristics, we firstly utilize wavelet coefficients as an index to qualitatively analyze ramps and non-ramps.

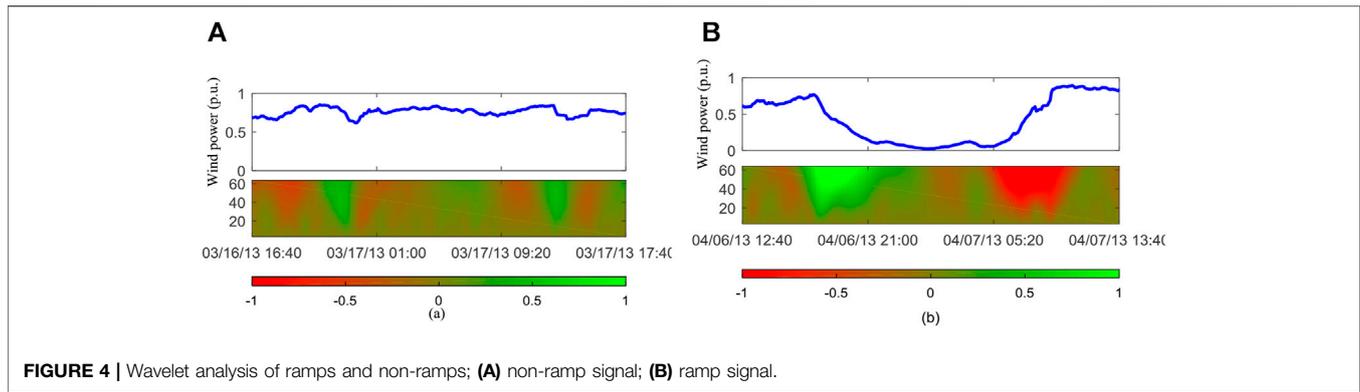


FIGURE 4 | Wavelet analysis of ramps and non-ramps; (A) non-ramp signal; (B) ramp signal.

In Figure 4, the two typical ramp and non-ramp events are identified by characteristics in Eq. 5. The function applied here is Haar wavelet which was validated useful in ramp analysis (Gallego et al., 2013). It is seen from Figure 4 that wavelet coefficients which reflect signal’s energy could obviously distinguish ramps and non-ramps. For example, ramps have larger coefficients than non-ramps. It verifies the validity of wavelet transformation in ramp characteristics analysis, therefore we could further extract more profound wavelet characteristics for ramp recognition.

According to the heuristic methods in selecting characteristics, the class labels are required. Therefore, we propose to construct ramp classification according to a specific context. Considering wind process are formed by different meteorological phenomena, so the categories of ramp events are related to division wind processes to some extent. In (Wang et al., 2013), five types of wind process are discussed, namely small wind, small fluctuation wind, large fluctuation wind, double peak wind and multi-peaks wind. However, there are a few ramp events attributed to the small wind. Double peaks wind can be regarded as a special type of multi-peaks wind. Based on these assumptions, we can group historical ramp events into the following category library  $L$ , denoted as below.

$$L = \{A, B, C\} \begin{cases} A = \text{small wind, small fluctuation wind;} \\ B = \text{large fluctuation wind;} \\ C = \text{double peaks wind, multi - peaks wind;} \end{cases} \quad (26)$$

Combing with the description of wind process in (Wang et al., 2013) and the constructed library  $L$ , historical ramp events of training set are classified into three classes, as presented below.

Table 2 shows the statistical results of ramp events belonging to three categories in  $L$ , where number 1, 2 and 3 are defined as the class labels of A, B, C. It is seen that most ramp events are associated with the large fluctuation wind, a few of ramp events associated with the small wind and multi-peak wind. These results agree with the concept of ramp events involving a large change of wind power. Therefore, the constructed application context is reasonable for studying ramp characteristics. Based on the constructed library of ramp categories, then the extraction and selection of ramp characteristics could be implemented.

TABLE 2 | Number of ramps in three categories.

Class label	1	2	3
Number	101	378	47

TABLE 3 | Process of feature selection.

	$P_1$	$P_2$	$P_3$	$P_4$	$P_5$	Selection
Step 1	0.3064	0.3032	0.2969	0.2996	0.3146	$P_3$
Step 2	0.2963	0.2944	\	0.2917	0.2960	$P_4$
Step 3	0.2911	0.2901	\	\	0.2900	None

Assuming each signal is decomposed into five wavelet layers, the energy of each wavelet layer is extracted as ramp characteristics by Eq. 13, expressed as  $P = [P_1, P_2, P_3, P_4, P_5]$ . Here, the number of decomposition layers is set as 5 since ramp events have only three categories in this paper. In other application context which requires more refined ramp classification, the level of decomposition could be higher. As the description in Feature Selection, the purpose of feature selection is to delete redundant characteristics, reduce dimension and computation cost. Combining ramp basic characteristics and extracted wavelet energy characteristics, the characteristic vector is expressed as  $C_v = [Wp_{max}, Wp_{min}, T, R, P_1, P_2, P_3, P_4, P_5]$ . For all historical ramp events, the feature set constructed by  $C_v$  is denoted as  $X_{p0}$ . Then, according to the selection algorithm in Feature Selection, the process is shown in the following table.

At each step of Table 3, the value of  $\lambda_p$  is calculated for the rest characteristics  $P_i$  first. Then the minimum one is applied for  $F_r$  testing based on Eqs 15, 16. If the result is satisfied, adding the corresponding characteristic into characteristic vector  $C_v$ . It is seen from Table 3 that  $P_3$  and  $P_4$  is selected, so the final characteristic is re-written as  $C_{vp} = [Wp_{max}, Wp_{min}, T, R, P_3, P_4]$ , and the final feature set for all ramp events as  $X_p$ .

### Ramp Study and Discussion

By taking the feature set  $X_p$  as inputs and the class labels from Table 2 as output, five data mining algorithms (SVM, NN, RF,

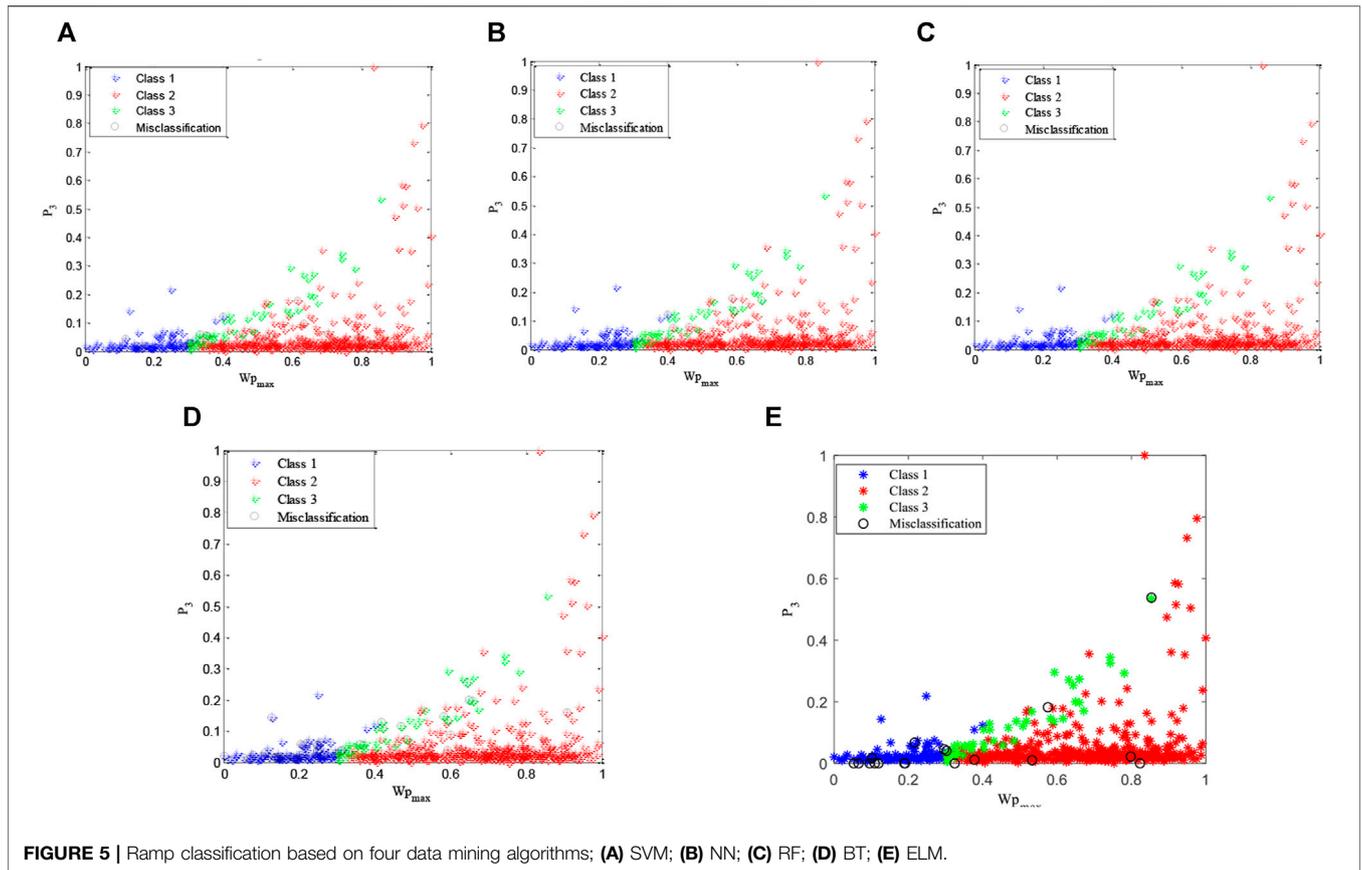


FIGURE 5 | Ramp classification based on four data mining algorithms; (A) SVM; (B) NN; (C) RF; (D) BT; (E) ELM.

TABLE 4 | Performance of ramp classification.

		Pre	Rec	Acc	Err
Class 1	SVM	0.9000	0.9802	0.9753	0.0247
	NN	1.0000	0.9714	0.9946	0.0054
	RF	1.0000	0.9683	0.9944	0.0056
	BT	1.0000	0.6126	0.9183	0.0817
	ELM	0.9381	0.9010	0.9696	0.0304
Class 2	SVM	0.9688	0.9868	0.9677	0.0323
	NN	0.9886	0.9886	0.9838	0.0162
	RF	0.9961	0.9846	0.9859	0.0141
	BT	0.8651	0.9729	0.8745	0.1255
	ELM	0.9739	0.9868	0.9715	0.0285
Class 3	SVM	0.8065	0.5319	0.9468	0.0532
	NN	0.8684	0.9167	0.9784	0.0216
	RF	0.8333	0.9677	0.9802	0.0198
	BT	0.7442	0.6957	0.9525	0.0475
	ELM	0.9565	0.9362	0.9905	0.0095

BT, ELM) are applied in ramp classification modeling. The classification results of three categories of ramps are shown in following figure.

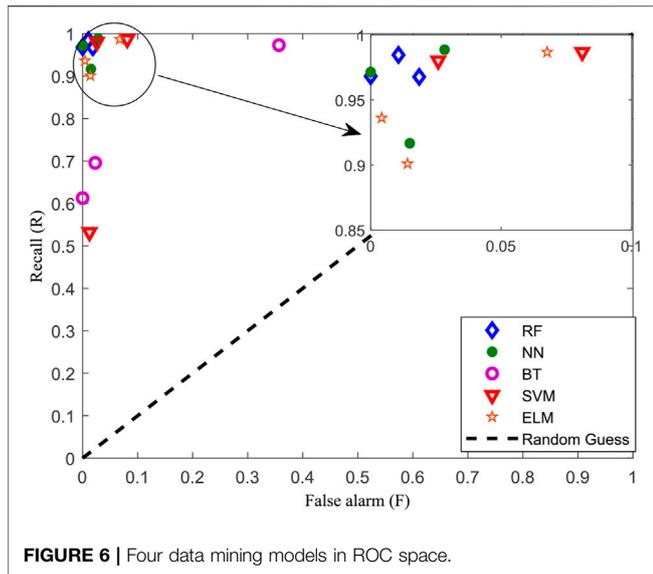
For the convenience of presentation, Figure 5 utilizes two characteristics ( $Wp_{max} \times P_3$ ) to show ramp classification results, as X-axis representing values of  $Wp_{max}$  and Y-axis representing values of  $P_3$ . Red, blue and green points represent ramps of Class1, Class2 and Class3, respectively. Black points represent

ramp events classified into incorrect class. Based on these classification results, the performance indicators defined in Eq. 23 could be calculated, as presented below.

Table 4 presents values of four performance indicators in classification of three ramp categories. According to the definitions of four matrixes, a classification system performs well with large value of Pre, Rec, Acc, and small value of Err. In Table 4, it is seen that NN, RF and ELM algorithms outperform on three classes, respectively. However, it is difficult to choose the best to classify all three ramp categories. To determine the final optimal model in ramp classification, the receiver operating characteristic curve (ROC) is introduced to compare performance further. ROC space is constructed by Recall (Rec) in the X-axis and false alarm (F) in the Y-axis. The definition of F is also based on Table 1, as expressed below.

$$F = \text{num}_{FP} / (\text{num}_{FP} + \text{num}_{TN}) \quad (27)$$

where, F calculates the percentage of FP in observed false events. According to these two indicators' definitions, it is easily comprehended that a classifier having a large Rec and a small F performs better, which implies the upper-left corner of the ROC space means the better performance. For a discrete classification system, a classifier is usually represented by a point in ROC space. Therefore, points representing all classifiers in Table 4 are shown in the following figure.



**TABLE 5 |** Classification performance of different feature sets.

		Pre	Rec	Acc	Err
Class1	$X_0$	0.6111	1.0000	0.9157	0.0843
	$X_{p0}$	0.6804	1.0000	0.9357	0.0643
	$X_p$	0.6875	1.0000	0.9378	0.0622
	$X_{pca}$	0.6735	1.0000	0.9398	0.0602
Class2	$X_0$	0.9621	0.8945	0.9016	0.0984
	$X_{p0}$	0.9914	0.8934	0.9157	0.0843
	$X_p$	1.0000	0.8958	0.9137	0.0863
	$X_{pca}$	0.9940	0.9089	0.8916	0.1084
Class3	$X_0$	0.2813	0.1875	0.8574	0.1426
	$X_{p0}$	0.3095	0.2708	0.8715	0.1285
	$X_p$	0.3231	0.4375	0.8755	0.1245
	$X_{pca}$	0.3088	0.4375	0.8514	0.1486

In **Figure 6**, three points having the same type present classifiers of three classes by a same algorithm. The points representing SVM and BT perform worse than the other three algorithms again. By comparing the points of NN, RF and ELM, it is seen that points of RF obviously are concentrated and closer to the upper-left corner, implying their classifiers have a better performance. Therefore, RF algorithm is finally chosen for modeling ramp classification in this paper.

Then, taking these trained RF models as ramp classifiers, 100 ramp events are selected from July to December as testing samples. To discuss the performance with and without wavelet characteristics in ramp classification, three different input sets are considered in the case study, such as initial feature set  $X_0$  consisting of vector  $C_{v0}$ , feature set  $X_{p0}$  consisting of vector  $C_v = [C_{v0}, P]$  which contains basic ramp characteristics and all wavelet characteristics, refined feature set  $X_p$  consisting of  $C_{vp} = [C_{v0}, P_3, P_4]$  which reduces dimension by feature selection. Classification performance of testing data is presented in the following table.

**Table 5** shows values of four indicators at classification of three ramp categories. Since three ramp categories have

**TABLE 6 |** Performance of ramp prediction.

		1 h	2 h	3 h	4 h	5 h
Var	$X_0$	21.94	40.80	79.49	120.42	156.71
	$X_{p0}$	19.02	38.05	81.63	115.94	151.46
	$X_p$	18.68	35.31	86.21	113.93	149.46
	$X_{pca}$	20.27	35.63	79.12	122.02	160.87
R	$X_0$	1.8635	1.9418	2.4102	2.5717	2.6601
	$X_{p0}$	1.7527	1.6638	2.0607	2.5032	2.4825
	$X_p$	1.4786	1.4928	1.9097	2.3972	2.3897
	$X_{pca}$	1.8355	1.8644	2.2158	2.5242	2.5728

imbalance distribution as **Table 2II**, so the classification performance of Class 2 is the best. By comparing classification performance of using three feature sets, it is seen that using  $X_{p0}$  has better performance than using  $X_0$ , having an average improvement of 10.17% (8.14% on Pre, 14.77% on Rec, 1.80% on Acc, 15.98% on Err). These results imply wavelet characteristics are useful in ramp classification. Also, using  $X_p$  improves a little again (with an average improvement of 6.05%) than using  $X_{p0}$ , which implies that the proposed feature selection approach is feasible and effectual in ramp classification. While, for comparing with other feature selection methods, the commonly used PCA is applied. After the analysis of PCA on  $P = [P_1, P_2, P_3, P_4, P_5]$ , two principal components are selected to keep a same dimension with  $X_p$ , these two components contribute 84.00% explanation in classification, and consist of the feature set  $X_{pca}$ . **Table V** also presents the performance of  $X_{pca}$ . By comparing  $X_p$  and  $X_{pca}$ , it is seen that using  $X_p$  has an average outperformance of 3.69% than using  $X_{pca}$ . Summarizing all these results, the proposed approach on selecting wavelet characteristics is validated to be feasible.

On the other hand, by utilizing these selected ramp characteristics, we could do some try on the study of ramp prediction. Since ramp classification has validated the effectiveness on distinguishing different categories, so two basic ramp characteristics are taken as target output in prediction, such as ramp amplitude (*Var*) and ramp rate (*R*). As the design of ramp prediction in *Ramp Anomaly Analysis and Evaluation*, the historical feature extraction window and the predicted time window are set as the same for convenience. In this paper, ramp prediction is designed to predict two variables (*Var* and *R*) within future horizon of 1, 2, ..., 5 h, the prediction performance is presented in the following table.

**Table 6** shows the performance of ramp prediction by RMSE of two ramp characteristics. In these two variables' prediction, a typical NN with three layers is used in modeling. For comparison study, four feature sets discussed in **Table 5** are also utilized as inputs of prediction models. It is seen from results of **Table 6** that models using feature sets containing wavelet characteristics (e.g.,  $X_{p0}$ ,  $X_p$ ,  $X_{pca}$ ) outperforms than that only containing basic ramp characteristics (e.g.,  $X_0$ ). Through the proposed feature selection in this paper, the model using  $X_p$  has an improvement of 5.97% than using  $X_0$ , 1.29% than using  $X_{p0}$ , and 2.70% than using  $X_{pca}$  on prediction of *Var*. Similarly, using  $X_p$  has an improvement of 16.30% than using  $X_0$ , 8.24% than using  $X_{p0}$ , and 13.07% than using  $X_{pca}$  on prediction of *R*. Through the discussion on results

of Table VI, it is concluded that the proposed approach on extracting and selecting ramp characteristics is also useful for constructing inputs of ramp prediction, and acquires good prediction performance.

## CONCLUSION

The study in this paper focus on extracting and selecting profound ramp characteristics for in-depth ramp researches. First, based on wavelet transformation's properties on time-frequency domains and multiresolution, wavelet decomposition is validated useful in analyzing ramps and non-ramps, also different categories of ramps. Then, ramp characteristics are extracted based on the energy decomposition at different wavelet layers. Combining with given ramp categories from wind process, heuristic feature selection methods (e.g., SFFS, SFBS) are applied to select valid characteristics, to remove redundant characteristics and reduce feature dimension. Based on basic ramp characteristics and selected wavelet characteristics, ramp studies on classification and prediction acquire better performance than that without wavelet characteristics and that using PCA in feature selection. Therefore, the conclusion could be summarized in this paper that wavelet transformation is useful to extract profound ramp characteristics, and that selecting ramp characteristics by the proposed approach is feasible to improve performance of ramp studies.

However, besides the above conclusions, there is also a number of conceptual alternatives worth discussing and pursuing: 1) ramp categories in this paper are determined by wind process. Therefore, the selected wavelet characteristics are

not completely applicable to other ramp contexts. The approach involving feature extraction and selection in this paper could be still referential. 2) Ramp events generally involve complicated weather movement, it is reasonable that considering meteorological variables in ramp studies could improve the performance. While, for the limitation of data sources in this paper, we only consider ramp characteristics from wind power data. More work on exogenous variables will be studied in future. 3) Based on the selected ramp characteristics and results of some ramp studies, power system's operation associated with ramp events could be studied further. Besides these points, more studies are needed to in-depth understand ramp events.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

PY write the paper and implement the experiments. AL supervises the whole process.

## FUNDING

This work is supported by Scientific Research Fund of Hunan Provincial Education Department (20A460).

## REFERENCES

- Ashrafiyan, A., Mirsalim, M., and Masoum, M. A. (2017). An Adaptive Recursive Wavelet Based Algorithm for Real-Time Measurement of Power System Variables during Off-Nominal Frequency Conditions. *IEEE Trans. Ind. Inform.* 14, 818–828. doi:10.1109/TII.2017.2727222
- bpa.gov/transmission (2013). [http://transmission.bpa.gov/Business/Operations/Wind/WindGenTotalLoadYTD\\_2013.xls](http://transmission.bpa.gov/Business/Operations/Wind/WindGenTotalLoadYTD_2013.xls).
- Chen, Z., Zhang, L., Cao, Z., and Guo, J. (2018). Distilling the Knowledge from Handcrafted Features for Human Activity Recognition. *IEEE Trans. Ind. Inform.* 14, 4334. doi:10.1109/tii.2018.2789925
- Chen, Z., Zhu, Q., Soh, Y. C., and Zhang, L. (2017). Robust Human Activity Recognition Using Smartphone Sensors via CT-PCA and Online SVM. *IEEE Trans. Ind. Inf.* 13 (6), 3070–3080. doi:10.1109/tii.2017.2712746
- Couto, A., Costa, P., Rodrigues, L., Lopes, V. V., and Estanqueiro, A. (2013). "Impact of Weather Regimes on the Wind Power Ramp Forecast," in 12th International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plants, London, United Kingdom, October 22–24, 2013.
- Doucoure, B., Agbossou, K., and Cardenas, A. (2016). Time Series Prediction Using Artificial Wavelet Neural Network and Multi-Resolution Analysis: Application to Wind Speed Data. *Renew. Energy* 92, 202–211. doi:10.1016/j.renene.2016.02.003
- Escalante Soberanis, M. A., and Mérida, W. (2015). Regarding the influence of the Van der Hoven spectrum on wind energy applications in the meteorological mesoscale and microscale. *Renew. Energy* 81, 286–292. doi:10.1016/j.renene.2015.03.048
- Florita, A., Hodge, B. M., and Orwig, K. (2013). "Identifying Wind and Solar Ramping Events," in IEEE 5th Green Technologies Conference, Denver, CO, 4–5 April 2013 (IEEE). doi:10.1109/GreenTech.2013.30
- Francis, N. (2008). *Predicting Sudden Changes in Wind Power Generation*, 58. Honolulu: North American Wind Power.
- Gallego, C., Costa, A., Cuerva, Á., Landberg, L., Greaves, B., and Collins, J. (2013). A Wavelet-Based Approach for Large Wind Power Ramp Characterisation. *Wind Energy* 16 (2), 257–278. doi:10.1002/we.550
- Gan, J. Q., Awwad Shiekh Hasan, B., and Tsui, C. S. L. (2014). A Filter-Dominating Hybrid Sequential Forward Floating Search Method for Feature Subset Selection in High-Dimensional Space. *Int. J. Mach. Learn. Cyber.* 5 (3), 413–423. doi:10.1007/s13042-012-0139-z
- Gu, B., Sun, X., and Sheng, V. S. (2017). Structural Minimax Probability Machine. *IEEE Trans. Neural Netw. Learn. Syst.* 28 (7), 1646–1656. doi:10.1109/tnnls.2016.2544779
- He, M., Vittal, V., and Zhang, J. (2013). Online Dynamic Security Assessment with Missing Pmu Measurements: A Data Mining Approach. *IEEE Trans. Power Syst.* 28 (2), 1969–1977. doi:10.1109/tpwrs.2013.2246822
- He, S., Li, K., and Zhang, M. (2013). A Real-Time Power Quality Disturbances Classification Using Hybrid Method Based on S-Transform and Dynamics. *IEEE Trans. Instrum. Meas.* 62 (9), 2465–2475. doi:10.1109/tim.2013.2258761
- He, Y., Kusiak, A., Ouyang, T., and Teng, W. (2017). Data-driven Modeling of Truck Engine Exhaust Valve Failures: a Case Study. *J. Mech. Sci. Technol.* 31 (6), 2747–2757. doi:10.1007/s12206-017-0518-1
- Huang, H., Liu, F., Zha, X., Xiong, X., Ouyang, T., Liu, W., et al. (2018). Robust Bad Data Detection Method for Microgrid Using Improved ELM and DBSCAN Algorithm. *J. Energy. Eng.* 144 (3), 04018026. doi:10.1061/(asce)ey.1943-7897.0000544

- Mohanty, S. R., Kishor, N., Ray, P. K., and Catalao, J. P. S. (2015). Comparative Study of Advanced Signal Processing Techniques for Islanding Detection in a Hybrid Distributed Generation System. *IEEE Trans. Sustain. Energ.* 6 (1), 122–131. doi:10.1109/tste.2014.2362797
- Ouyang, T. (2021). Feature Learning for Stacked ELM via Low-Rank Matrix Factorization. *Neurocomputing* 448, 82–93. doi:10.1016/j.neucom.2021.03.110
- Ouyang, T., Kusiak, A., and He, Y. (2017). Modeling Wind-Turbine Power Curve: A Data Partitioning and Mining Approach. *Renew. Energ.* 102, 1–8. doi:10.1016/j.renene.2016.10.032
- Ouyang, T., Zha, X., and Qin, L. (2017). A Combined Multivariate Model for Wind Power Prediction. *Energ. Convers. Manag.* 144, 361–373. doi:10.1016/j.enconman.2017.04.077
- Ouyang, T., Zha, X., Qin, L., and Kusiak, A. (2017). Optimisation of Time Window Size for Wind Power Ramps Prediction. *IET Renew. Power Generation* 11 (8), 1270–1277. doi:10.1049/iet-rpg.2016.0341
- Sevlian, R., and Rajagopal, R. (2013). Detection and Statistics of Wind Power Ramps. *IEEE Trans. Power Syst.* 28, 3610–3620. doi:10.1109/tpwrs.2013.2266378
- Shen, X., Ouyang, T., Yang, N., and Zhuang, J. (2021). Sample-Based Neural Approximation Approach for Probabilistic Constrained Programs. *IEEE Trans. Neural Netw. Learn. Syst.*, 1–8. doi:10.1109/TNNLS.2021.3102323
- Singh, K., and Tewari, A. (2015). Classification of Power Quality Disturbances Using Wavelet Transform and Neural Network. *Int. J. Eng. Res. Tech.* 4 (05), 1423. doi:10.17577/ijertv4is051331
- Tang, Z., Zhao, G., and Ouyang, T. (2021). Two-phase Deep Learning Model for Short-Term Wind Direction Forecasting. *Renew. Energ.* 173, 1005–1016. doi:10.1016/j.renene.2021.04.041
- Tang, Z., Zhao, G., Wang, G., and Ouyang, T. (2020). Hybrid Ensemble Framework for Short-Term Wind Speed Forecasting. *IEEE Access* 8, 45271–45291. doi:10.1109/access.2020.2978169
- Wang, L., Zhang, Z., Long, H., Xu, J., and Liu, R. (2017). Wind Turbine Gearbox Failure Identification with Deep Neural Networks. *IEEE Trans. Ind. Inf.* 13 (3), 1360–1368. doi:10.1109/tii.2016.2607179
- Wang, Z., Wang, W., Liu, C., and Feng, S. (2013). Uncertainty Estimation of Wind Power Prediction Result Based on Wind Process Method. *Power Syst. Tech.* 37 (1), 242–247.
- Xiong, Y., Zha, X., Qin, L., Ouyang, T., and Xia, T. (2017). Research on Wind Power Ramp Events Prediction Based on Strongly Convective Weather Classification. *IET Renew. Power Generation* 11 (8), 1278–1285. doi:10.1049/iet-rpg.2016.0516
- Zha, X., Ouyang, T., Qin, L., Xiong, Y., and Huang, H. (2016). Selection of Time Window for Wind Power Ramp Prediction Based on Risk Model. *Energ. Convers. Manag.* 126, 748–758. doi:10.1016/j.enconman.2016.08.064
- Zheng, H., and Kusiak, A. (2009). Prediction of Wind Farm Power Ramp Rates: A Data-Mining Approach. *J. solar Energ. Eng.* 131 (3), 031011. doi:10.1115/1.3142727

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Yu and Lin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.