

Article

# A Harmonic Impedance Identification Method of Traction Network Based on Data Evolution Mechanism

Ruixuan Yang, Fulin Zhou \* and Kai Zhong

School of Electrical Engineering, Southwest Jiaotong University, Chengdu 610031, China

\* Correspondence: fulin-zhou@swjtu.edu.cn; Tel: +86-186-0804-8810

Received: 31 January 2020; Accepted: 25 March 2020; Published: 13 April 2020

**Abstract:** In railway electrification systems, the harmonic impedance of the traction network is of great value for avoiding harmonic resonance and electrical matching of impedance parameters between trains and traction networks. Therefore, harmonic impedance identification is beneficial to suppress harmonics and improve the power quality of the traction network. As a result of the coupling characteristics of the traction power supply system, the identification results of harmonic impedance may be inaccurate and controversial. In this context, an identification method based on a data evolution mechanism is proposed. At first, a harmonic impedance model is established and the equivalent circuit of the traction network is established. According to the harmonic impedance model, the proposed method eliminates the outliers of the measured data from trains by the Grubbs criterion and calculates the harmonic impedance by partial least squares regression. Then, the data evolution mechanism based on the sample coefficient of determination is introduced to estimate the reliability of the identification results and to divide results into several reliability levels. Furthermore, in the data evolution mechanism through adding new harmonic data, the low-reliability results can be replaced by the new results with high reliability and, finally, the high-reliability results can cover all frequencies. Moreover, the identification results based on the simulation data show the higher reliability results are more accurate than the lower reliability results. The measured data verify that the the data evolution mechanism can improve accuracy and reliability, and their results prove the feasibility and validation of the proposed method.

**Keywords:** harmonic impedance; traction network; harmonic impedance identification; linear regression model; data evolution mechanism

---

## 1. Introduction

With a rapid development of railway electrification systems (RESs), especially high-speed railways, harmonic distortion problems have attracted increasing attention. At present, electrical locomotives and electric multiple units (EMUs) (collectively called trains) based on pulse-width-modulation (PWM) controlled converters are widely applied in practice [1]. These trains could inject wider and higher high-order harmonic currents into traction power supply systems. The frequencies of harmonic currents can cover the resonance frequencies of traction networks. This will lead to a lot of abnormal problems, such as harmonic resonance [2,3] and harmonic instability [4,5]. Under these conditions, the large components of high-order harmonics could not only easily cause temporary overvoltage, but also even in extreme cases cause some serious incidents, such as the burst of on-board arresters [6]. Thus, the harmonic problem, a huge

impact on the normal operation of trains, is a hidden danger to the security of RESs. Based on the above, proper harmonic suppression [7,8] and good matching characteristics of the harmonic impedance [9] are the key to improving harmonic problems and power quality [10,11], while the harmonic impedance of the traction network is an important parameter. Therefore, a method that can accurately identify the harmonic impedance of the traction network is needed.

In recent years, various methods have been used to calculate harmonic impedance in a traction network. Reference [2,12] built the impedance models of traction networks and calculated harmonic impedance according to the structure and parameters of a traction network. These models can obtain the resonance frequencies and research some parameters which could influence the harmonic impedance. However, in practice, it is very difficult to obtain all system parameters accurately. Thus, the model methods are usually used in simulation research, but they have some limitations in measured harmonic data. Reference [13] proposed a method to estimate harmonic impedance by injecting harmonic currents with specific spectrum into the power grid. However, this method needs a harmonic source device and it is possible to have an impact on the normal operation of the power system. Moreover, some identification methods of harmonic impedance are widely used in the utility power grid (UPG). Fluctuation methods [14,15] can identify the harmonic impedance through the ratio of the increments of harmonic voltage to current at a point of common coupling (PCC) with high measurement accuracy. Reference [16–19] proposed linear regression methods. In essence, they build an equivalent circuit model for harmonic analysis and establish the regression equation by deriving the harmonic voltage and current correlation at PCC. Then, using large amounts of measured harmonic data, they calculate the harmonic impedance and harmonic emission level by regression estimation. Moreover, [20] proposed a method to estimate harmonic parameters and harmonic responsibility with the harmonic amplitude and phase difference of harmonic voltage and current. These methods have already had quite mature applications in the UPG, so in these research works it is possible to use the linear regression method in RESs.

However, compared with the traditional UPG, RESs are a special power grid and hold some unique characteristics. These mean the linear regression method has some limitations in the application process so that the identification results of harmonic impedance may be inaccurate. For further analysis, the main limitations of linear regression methods is shown as follows:

1. The traditional UPG mainly has low-order harmonics, while in the RESs the harmonic problems usually focus on the high frequencies, such as 20th–60th (1000–3000 Hz). However, in practice, it is more difficult to accurately measure high-order harmonic information than low-order, especially the phases of high-order harmonics. In the application of linear regression methods, the measurement errors could be converted into fluctuations of calculated data, which will influence the identification results of harmonic impedance.
2. In the RESs, it is necessary to consider the train-network coupling [21]. The train-network coupling is a dynamic electrical interaction between fast moving trains and the static traction network. Therefore, in this situation, the harmonic impedance of a traction network could change with the fast movement of trains. Meanwhile, some system parameters, which are essential for calculating impedance, could also change. As a result of the dynamic coupling between trains and the traction network, it is not conducive to the accurate identification of harmonic impedance.

To solve these problems, a method combining linear regression with data elimination and data evolution mechanism is proposed in this paper. The main contributions of this paper are as follows:

1. The harmonic impedance identification model based on an electrical circuit and linear regression method is derived in this paper, which is the electrical theoretical basis for calculating impedance.
2. The data elimination based on the Grubbs criterion is introduced to eliminate the outliers of the measured data in order to reduce the influence of error data on linear regression.
3. In this paper, the presented data evolution mechanism serves two purposes. Firstly, based on the sample coefficient of determination, the data evolution mechanism is used to evaluate the

reliability of regression results. Then the regression results can be divided into different reliability levels. Results with high reliability level are more accurate and more valuable than those with low reliability. Secondly, taking further advantage of the reliability level, data evolution mechanism could supplement the results of a high-reliability level by adding new measured data, and can even replace the results of a low-reliability level at some frequencies.

Section 2 introduces the train-network coupling system and derives the equivalent circuit and regression equations. In Section 3, the mathematical theory and application of data elimination and data evolution mechanism are illustrated, and the identification process of harmonic impedance is given. Section 4 carries on the simulation of harmonic impedance identification to verify the effectiveness and accuracy of the proposed method, and defines calculation parameters. In Section 5, measured data is used to demonstrate the application case, and the results of harmonic impedance identification is discussed. Finally, Section 6 is a summary of the full paper.

## 2. Harmonic Impedance Identification Model

The traction power supply system, including power grid, traction substation, traction network, rails and trains, is a complex network structure. For example, a train-network coupling structure, based on a typical two-phase network, is shown in Figure 1.

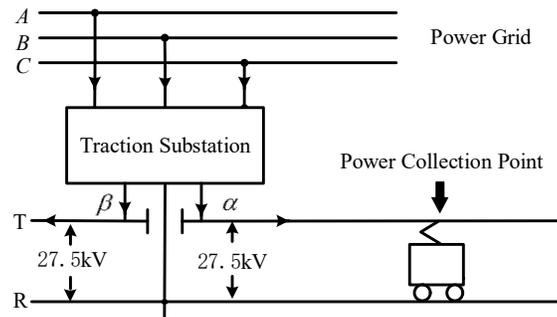


Figure 1. Train-network coupling system.

Figure 2 is the equivalent circuit of train-network coupling system in the harmonic state. The traction substation (SS) and the section post (SP) are the edges of the traction power supply system. The train at the power collection point (PCP) is simply consisted of a current source  $I_T^h$  and a harmonic impedance  $Z_c^h$ .  $Z_{T1}^h$  and  $Y_{T1}^h$  are the T-type equivalent circuit parameters of the traction network between PCP and SS, and  $Z_{T2}^h$  and  $Y_{T2}^h$  are the parameters of the traction network between PCP and SP. Moreover,  $Z_{SS}^h$  is the harmonic impedance of the traction substation, including the leakage impedance of the traction transformer, the harmonic impedance of the external power grid, etc.  $U_s^h$  is the system harmonic voltage, which in other words is the equivalent background harmonic voltage source.

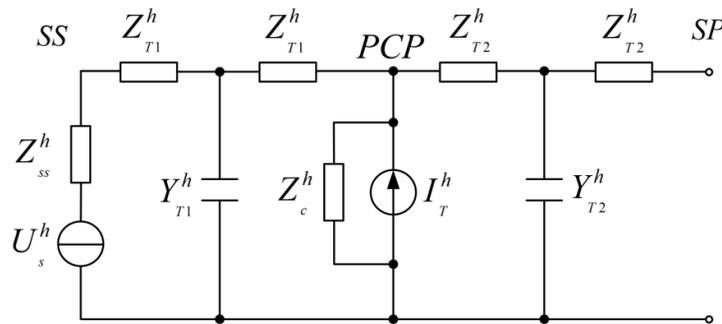


Figure 2. The equivalent circuit of train-network coupling system.

The Thevenin equivalent circuit is established in Figure 3 by the equivalent parameters except the parameters of the train.  $Z_S^h$  is the impedance of the traction network, which needs to be calculated and be identified.  $U_P^h$  and  $I_P^h$  are the voltage and current at the PCP, which could be measured by the potential transformer (PT) and the current transformer (CT) of the train.

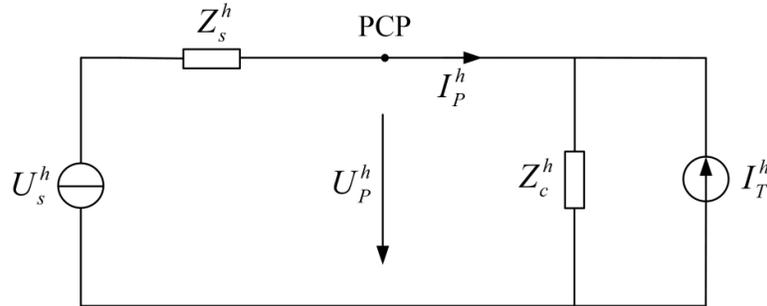


Figure 3. Thevenin equivalent circuit.

Therefore, the relationship equation can be obtained in (1):

$$\dot{U}_S^h = \dot{I}_P^h Z_S^h + \dot{U}_P^h \quad (1)$$

Expansion (1) with the real and imaginary parts, there are two regression equations:

$$U_{Sx}^h = U_{Px}^h + Z_{Sx}^h I_{Px}^h - Z_{Sy}^h I_{Py}^h \quad (2)$$

$$U_{Sy}^h = U_{Py}^h + Z_{Sy}^h I_{Px}^h + Z_{Sx}^h I_{Py}^h \quad (3)$$

where  $U_{Px}^h$ ,  $U_{Py}^h$ ,  $Z_{Sx}^h$  and  $Z_{Sy}^h$  are the regression coefficients that can be estimated by a regression method. Linear regression is a common method in many fields [22]. In this paper, we use partial least squares regression (PLSR) [23] to identify  $Z_S^h$ .

In this section, through circuit derivation from the train-network coupling system, the Thevenin equivalent circuit and regression equations are obtained. It is easy to calculate harmonic impedance via the PLSR method. This is a simple and common methodology, which is capable of calculating harmonic impedance in traditional power grids. However, as described in Section 1, the two limitations are so unavoidable that this methodology straightly applied to RESs could cause great calculation error. So, taking the coupling characteristics of the train-network system as guidance, the method needs specific improvement. The improvement is designed under the framework of the above harmonic impedance identification model.

### 3. Data Elimination and Data Evolution Mechanism

Based on the above, the improvements of the data elimination and data evolution mechanisms are designed in this section. The Grubbs criterion is introduced to eliminate outliers for increasing the accuracy of regression results, and the data evolution mechanism based on reliable estimation is introduced to investigate whether the identification results are reliable and uncontroversial.

#### 3.1. Elimination of Outliers Based on Grubbs Criterion

In order to reduce the influence of the outliers on the calculation accuracy of PLSR, for each sample we use the Grubbs criterion, which can process data consistency, to recognize and eliminate the abnormal error data.

Assuming that a group of samples  $Y = \{y_i | y_i \in R, y_1 < y_2 < \dots < y_n\} (i = 1, 2, \dots, n)$  is normally distributed, and calculating the statistics value in (4) and (5):

$$G_1 = \frac{(\bar{y} - y_1)}{s} \quad (4)$$

$$G_n = \frac{(y_n - \bar{y})}{s} \quad (5)$$

where  $\bar{y}$  and  $s$  are the sample mean and standard deviation, respectively. Then, identifying the statistic critical value  $G(\alpha, n)$  by looking up the Grubbs critical table.

If  $[G_1 \geq G_n] \wedge [G_1 > G(\alpha, n)]$ ,  $y_1$  is an outlier and should be eliminated. Correspondingly, if  $[G_n \geq G_1] \wedge [G_n > G(\alpha, n)]$ ,  $y_n$  should be eliminated. Then, we proceed by recalculating the sample mean and standard deviation with the remaining samples, and identifying the new statistics value  $G'_1$  and  $G'_n$ , until there are no outliers anymore.

In this paper, we use the Grubbs criterion to recognize outliers of the current  $I_p^h$  shown in Figure 3. Because in practice the amplitudes of measured harmonic currents are more susceptible to change than the voltages due to the dynamic change of train operation, and the outliers of currents could have a greater impact on the PLSR calculation results. Moreover, the amplitudes of harmonic currents are approximated as a normal distribution [24], while the Grubbs criterion is a classical statistical treatment of outliers in the normally distributed samples.

### 3.2. The Reliability Estimation for PLSR Calculation Results

Firstly, the regression Equations (2) and (3) are the general expression of the multivariate linear regression model in (6):

$$y = \lambda_0 + \lambda_1 \cdot x_1 + \lambda_2 \cdot x_2 \quad (6)$$

where  $\lambda_0$  is the undetermined constant,  $\lambda_1$  and  $\lambda_2$  are the undetermined coefficients. These three variables can be estimated by PLSR.  $y$  and  $x$  are the dependent variable and independent variable, which correspond to the voltage and current, respectively.

In order to evaluate the reliability of the result, the sample coefficient of determination (SCD)  $\gamma^2$  is introduced in (7). It is the ratio of the regression sum of squares (SSR) to sum of squares for total (SST):

$$\gamma^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

where  $y_i$  is the measured data, and  $\hat{y}_i$  and  $\bar{y}$  respectively denote the estimate value and average value of the measured data  $y_i$ .

In (7), SST reflects the uncertainty of the dependent variable  $y$  and SSR reflects the uncertainty of the estimate value depended on the independent variable  $x$ . In other words, SCD  $\gamma^2$  determines the fluctuation in the dependent variable caused by the variation of the independent variable. Obviously, SCD ranges from 0 to 1. The closer SCD gets to 1, the higher reliability the result of regression estimation holds (i.e., the more information of independent variable  $x$  the multivariate linear regression model utilizes).

### 3.3. The Data Evolution Mechanism with Reliability Estimation

As for measured voltage and current, we use the fast Fourier transform (FFT) algorithm with 10 cycles of waveform data to obtain the amplitude and phase information of harmonic voltage and current varying with time. Then we can calculate the real part and imaginary part of harmonic voltage and current with the amplitude and phase information. It can form a matrix  $H_{1000 \times L}$  in

which the 1000 rows denote the frequencies from 5 Hz to 5000 Hz and the length  $L$  of columns denotes the sample number of measured harmonic data. Based on the matrix  $H_{1000 \times L}$ , we set a calculation window of 100 columns as one calculation group, and slide only one column every time to the end, i.e., the first calculation group is columns 1 through 100, the second group is columns 2 through 101, and so on. This is assuming that the number of the calculation groups is  $m$  and taking PLSR with calculation groups by (2) and (3). Thus, we can get  $m$  results of regression estimation and each result holds a SCD. We can evaluate the reliability of regression results and divide them into four reliability levels shown in Table 1.

**Table 1.** The reliability levels of sample coefficient of determination (SCD).

Reliability Levels	Ranges of SCD	Priority of Data Processing
High reliability	$\gamma^2 \in [\alpha_1, 1]$	Highest
Medium reliability	$\gamma^2 \in [\alpha_2, \alpha_1)$	Medium
Low reliability	$\gamma^2 \in [\alpha_3, \alpha_2)$	Lowest
No reliability	$\gamma^2 \in [0, \alpha_3)$	Data eliminating

In Table 1,  $\alpha_1$ ,  $\alpha_2$  and  $\alpha_3$  are the critical values of reliability levels. Obviously, we have  $0 < \alpha_3 < \alpha_2 < \alpha_1 < 1$ .

Based on the above, the calculation results (the harmonic impedance) with high reliability, medium reliability and low reliability are aggregated in sets  $W_1$ ,  $W_2$  and  $W_3$ . Therefore, because the reliability levels show the accuracy of PLSR, we first consider using high-reliability data  $W_1$ , then medium reliability data  $W_2$  and finally low reliability data  $W_3$ . In other words, the high reliability data samples could be used to calculate the harmonic impedance in priority. The harmonic impedance can be calculated as the mean value of vectors in (8).

$$\bar{Z}_s^h = \frac{1}{p_i} \left( \sum_{k=1}^{p_i} Z_{sx}^h(k) + j \cdot \sum_{k=1}^{p_i} Z_{sy}^h(k) \right) \quad (8)$$

where  $p_i$  ( $i=1,2,3$ ) is the sample number of  $W_i$ ,  $Z_s^h \in W_i$  and  $h=5,10,15,20 \dots 5000\text{Hz}$ .

If the reliability estimation results show that there are no or less high reliability data sets  $W_1$  due to the lack of data or the rapid change of system parameters, we could consider using medium reliability data set  $W_2$  to calculate the harmonic impedance. Furthermore, if the sample numbers of  $W_1$  and  $W_2$  are both equal to zero or close to 0, the medium reliability data set  $W_3$  could be used to calculate the harmonic impedance, although there is an inaccuracy in harmonic impedance. In this paper, we consider 5% of  $m$  results as the critical value of whether the data quantity is sufficient. It means that if  $p_1 > 0.05m$ , we use  $W_1$  to finish the calculation. Through this process, the harmonic impedance at 5–5000 Hz could be obtained preliminarily.

However, owing to the data elimination and data evolution, maybe there are two problems which could exist in practical industry scenarios:

1. The high-reliability result may not cover all the range of frequencies (e.g., the critical resonance frequency may be emitted);
2. The total number of the results for some certain harmonic impedance may be very small so that it is not convincing.

Fortunately, as for a certain power supply section (PSS) that is desired to obtain the harmonic impedance, the two problems can be solved by adding new measured data from the same vehicle moving through the same PSS at other times to improve and supplement the last calculation result. With the increasing number of measured data: 1) the reliability of the results is improved; 2) the

high-reliability result covers a much wider range of the spectrum; 3) the absolute number of the high-reliability data will increase.

This solution is reasonable and feasible, because for one electrified railway line during one day or one week, there are plenty of scheduled trains running through the PSS and for a certain train, it will run several times. With the increasing of the new measured data, the calculation result of harmonic impedance will be an increasingly accurate approach to the best optimal solution.

### 3.4. The Identification Process of Harmonic Impedance

The identification steps of harmonic impedance is shown as follows:

- **Step 1**—Calculate the harmonic voltage  $U_p^h$  and current  $I_p^h$  by FFT algorithm with measured waveform data.
- **Step 2**—Eliminate the outliers of  $I_p^h$  by Grubbs Criterion.
- **Step 3**—Calculate the real parts  $Z_{Sx}^h$  and the imaginary parts  $Z_{Sy}^h$  of the harmonic impedance by PLSR with each 100 samples and calculate the corresponding SCD  $\gamma^2$ .
- **Step 4**—Repeat **Step 3** until you get  $m$  results of PLSR and SCD.
- **Step 5**—Select the proper critical values  $\alpha_1$ ,  $\alpha_2$  and  $\alpha_3$  of reliability levels and classify the  $m$  results into four reliability levels in Table 1.
- **Step 6**—Judge whether the number of the high-reliability data set  $W_1$  is higher than 5% of  $m$ , then we use data set  $W_1$  to calculate the harmonic impedance by (8). If not, we use medium reliability data set  $W_2$ . If the high-reliability data set and medium-reliability data set both are insufficient, we only use data set  $W_3$ .
- **Step 7**—Judge whether the reliability data is sufficient and whether the reliability results cover all the range of frequency. If not, we could add data new measured data.

In order to express the identification process more clearly, the flowchart of harmonic impedance identification is shown in Figure 4. In general, the process consists of four parts: data acquisition, data elimination, regression and the data evolution mechanism. The main purpose of data acquisition is to obtain harmonic voltage and current data which can be used for regression, and indeed data acquisition is a practical engineering aspect. Data elimination can reduce the error of regression calculation caused by measurement error and can ensure the high accuracy of regression. In addition, to calculate harmonic impedance in this paper regression is to provide the reliability for data evolution mechanism. The data evolution mechanism can be further subdivided into two parts. Firstly, data grouping is based on the reliability estimation. Secondly, the previous results of data grouping can be supplemented and replaced by new data.

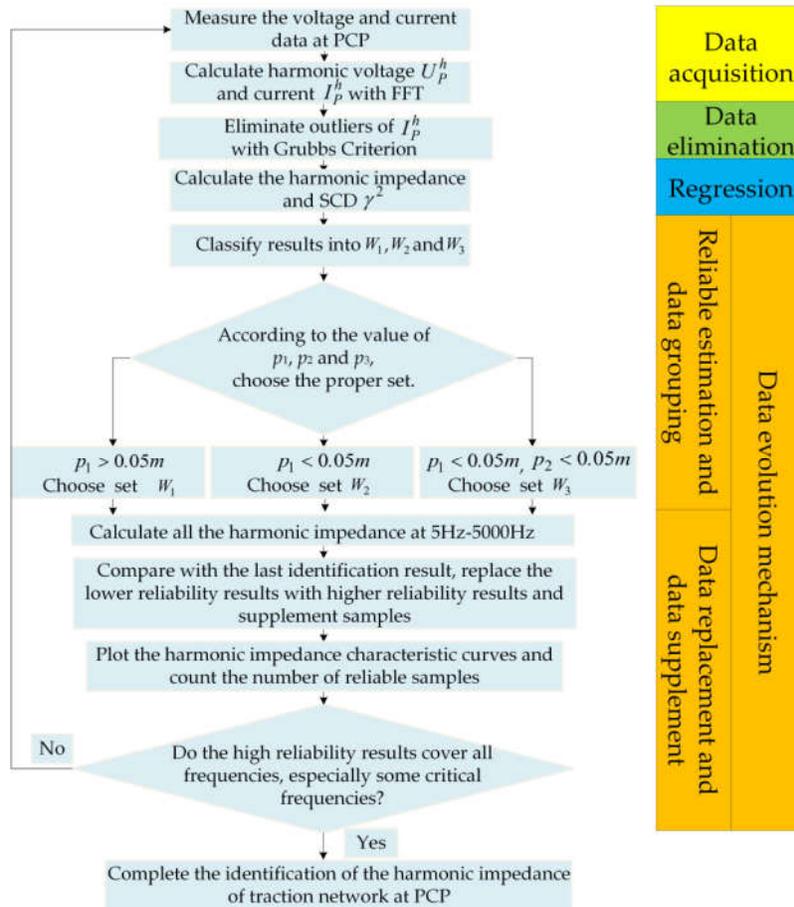


Figure 4. The flowchart of harmonic impedance identification.

#### 4. Simulation Verification

In Figure 5, we build a simulation model of a train-network coupling system in Matlab/Simulink program, based on the direct power supply system.

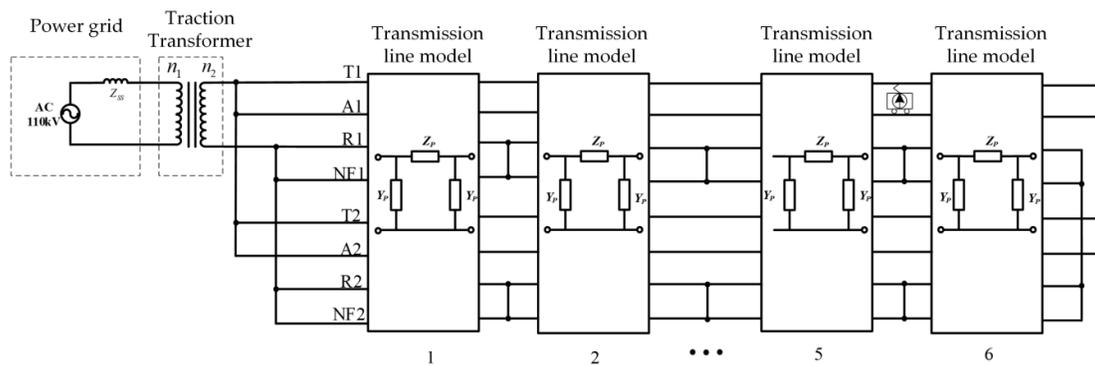
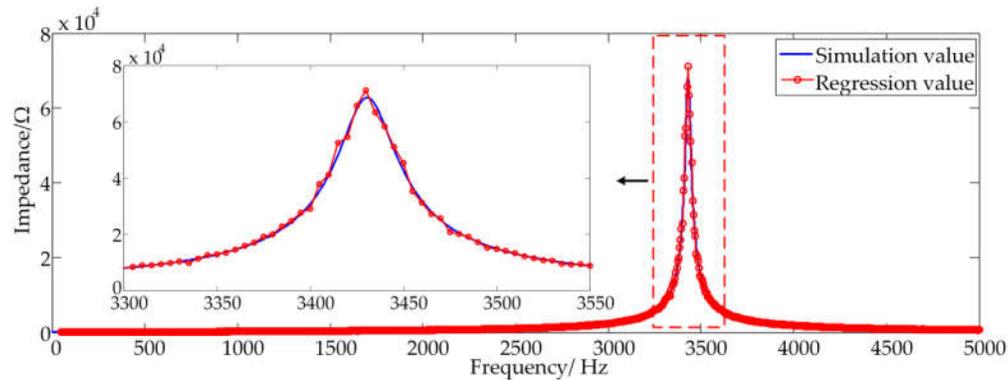


Figure 5. The simulation model of a train-network coupling system.

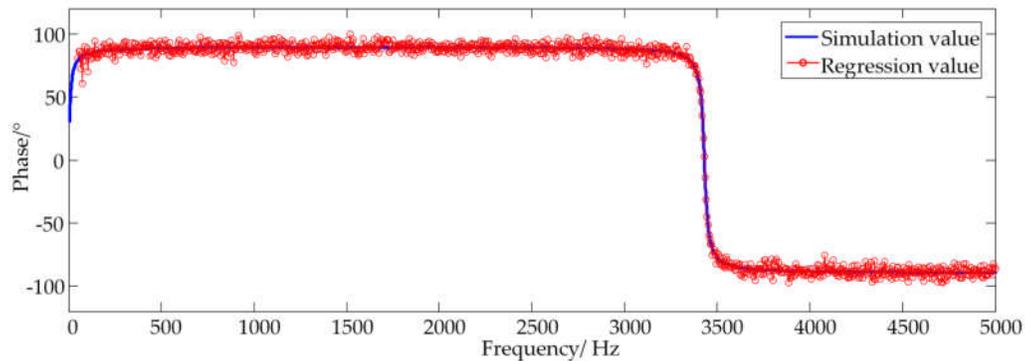
The power grid is a 110 kV voltage source and a short-circuit impedance in series connection. Traction transformer is a single-phase transformer and its voltage is 27.5 kV. In the simulation, we take the harmonic current source as a train, located 5 km away from the end of traction network. In order to simulate the ideal situation of obtaining sufficient harmonic data covering all the range of

frequency (5–5000 Hz), the equivalent harmonic current sources (train) injects 2nd–100th order harmonic currents, 5 A amplitudes and 0 phases, into the traction network. Moreover, the length of the traction network is 30 km and the distance between each connection point of the rail and the return line is 5 km. Referring to the eight-port representation model, we use eight conductors to build the transmission line model of the traction network. T1 and T2 are contact lines; R1 and R2 are the rail line; A1 and A2 are the reinforced lines; NF1 and NF2 are the return lines.

Based on the above simulation model, we use the parameter method to calculate the harmonic impedance of the traction network as the ‘simulation value’. Then we use PLSR to calculate the harmonic impedance as the ‘regression value’. The result is shown in Figure 6.



(a) Amplitudes of the harmonic impedance



(b) Phase of the harmonic impedance

**Figure 6.** The calculation results of the harmonic impedance.

Obviously, the comparison in Figure 6 between ‘simulation value’ and ‘regression value’ indicates that the PLSR results of the harmonic impedance are accurate. Although the calculation results of phases fluctuate slightly, the overall trend is relatively accurate. Therefore, the PLSR model can correctly calculate the harmonic impedance of the traction network with the measured harmonic data at PCP.

Furthermore, to verify the validity of data evolution mechanism, we calculate the harmonic impedance and SCD under fluctuations. In this simulation, we set noise signal to the amplitude of the harmonic current source (train) to simulate the practical disturbance. In this paper, we define that the critical values  $\alpha_1$ ,  $\alpha_2$  and  $\alpha_3$  of reliability levels are 0.9, 0.7 and 0.3, respectively. Taking 19th-order harmonic as an instance, through adding different amounts of noise, we respectively calculate the harmonic impedance and plot its trend in with  $\gamma^2 \approx 1$ ,  $\gamma^2 \approx 0.9$  and  $\gamma^2 \approx 0.7$ . The results are shown in Figure 7. and Table 2.

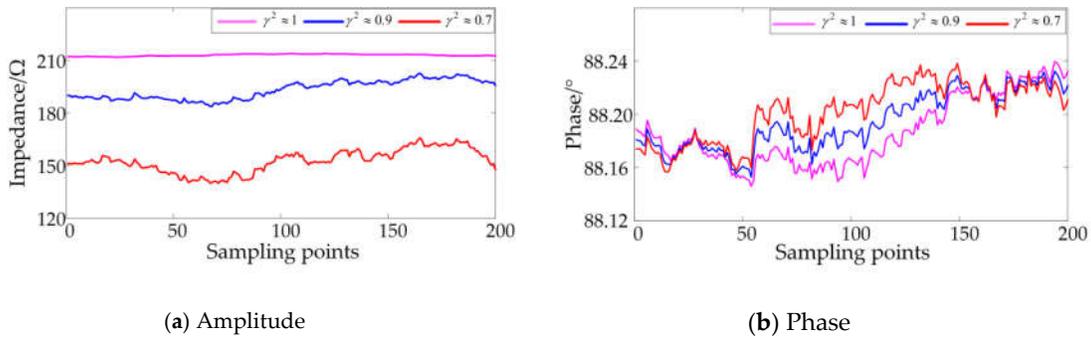


Figure 7. The 19th-order harmonic impedance under noise.

Table 2. Analysis of calculation results.

Harmonic Impedance	Simulation Value	Regression Value					
		$\gamma^2 \approx 1$		$\gamma^2 \approx 0.9$		$\gamma^2 \approx 0.7$	
		Mean Value	Error/%	Mean Value	Error/%	Mean Value	Error/%
Amplitude/Ω	212.39	212.91	0.24	193.08	9.09	154.47	28.21
Phase/°	89.17	88.19	1.10	88.19	1.10	88.19	1.10

The results of simulation verification show that:

1. Without noise, the ‘regression value’ is close to the ‘simulation value’. The error of the harmonic impedance amplitude is very low and the error of the phase is only about 1%. This indicates that the PLSR method can be used to identify the harmonic impedance accurately without any disturbance.
2. The calculation error increases rapidly after the disturbance noise is injected into the traction network. It means that in practice the error of the calculation results could be very high owing to the characteristics of the train-network coupling system. Thus, the proposed data evolution mechanism could show the reliability of harmonic impedance identification.
3. There is a certain correlation between the accuracy of the calculation results and the reliability of PLSR, so it is proper to select SCD  $\gamma^2$  as the index to evaluate the reliability.

### 5. Application Case

In this research, the voltage and current waveform data was from PT and CT of a HXD1 locomotive, shown in Figure 8. The sampling frequency of measurement devices was 20,000 Hz, and the voltage and current waveform is shown in Figure 9.

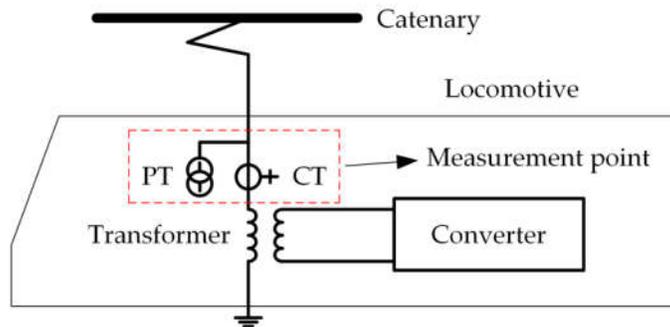


Figure 8. Schematic diagram of measurement.

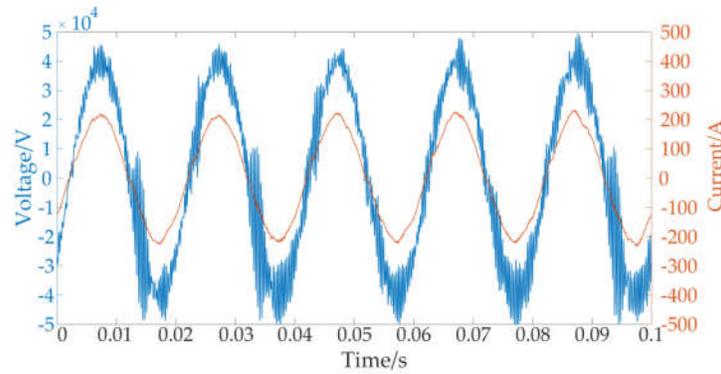
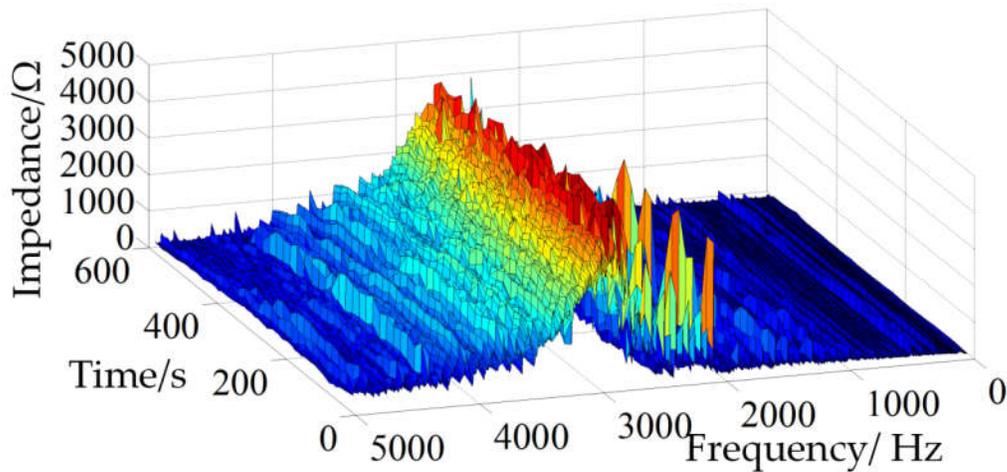


Figure 9. Voltage and current waveform.

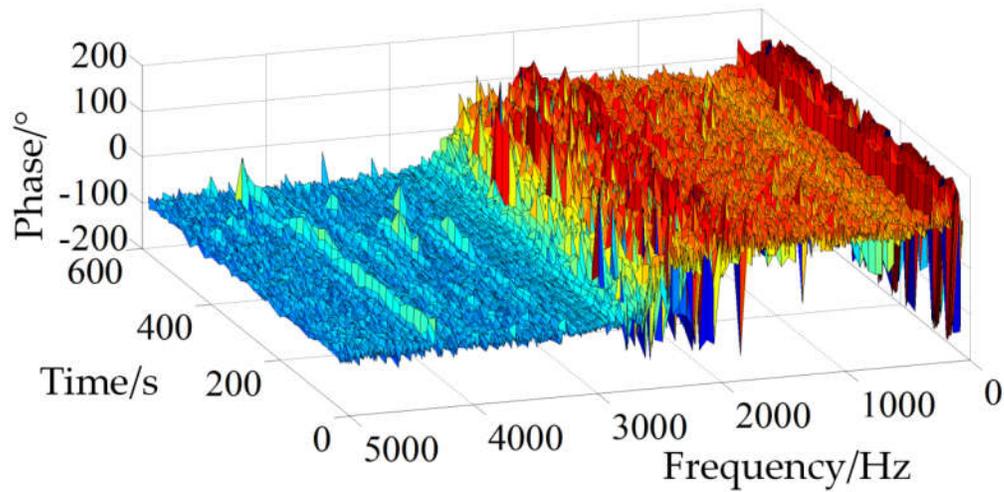
Obviously, voltage was extremely distorted owing to a large amount of high-order harmonic. By FFT, the total harmonic distortion (THD) of voltage has reached 22.09% and the characteristic frequencies approximately ranged from 2500 Hz to 3000 Hz. Therefore, in this case, it is necessary for in-depth analysis to identify the harmonic impedance of this traction network.

Firstly, without a data evolution mechanism, the PLSR method is used to calculate the harmonic impedance at 5–5000 Hz in Figure 8.

In Figure 10, the amplitudes of harmonic impedance peak at about 2750 Hz (55th-order), 2550 Hz (51st-order) and 2050 Hz (41st-order) and, respectively, reach nearly 4723.84  $\Omega$ , 2382.86  $\Omega$  and 4023.89  $\Omega$  (95% probability value). Combined with the phase spectrum, we find that the phase trend is close to the zero crossing point at 2550 Hz and the harmonic impedance changes from the inductive impedance to capacitive. Therefore, we preliminarily conclude that the harmonic resonance frequencies in this train-network coupling system range from 2550 Hz to 2750 Hz.



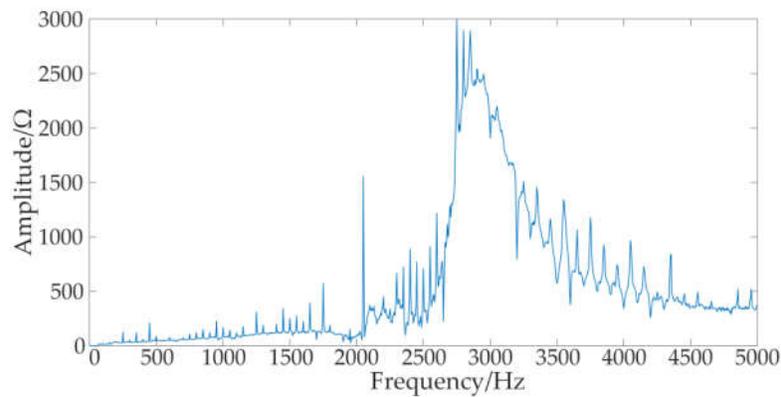
(a) Amplitude of the harmonic impedance with measured data.



(b) Phase of harmonic impedance with measured data.

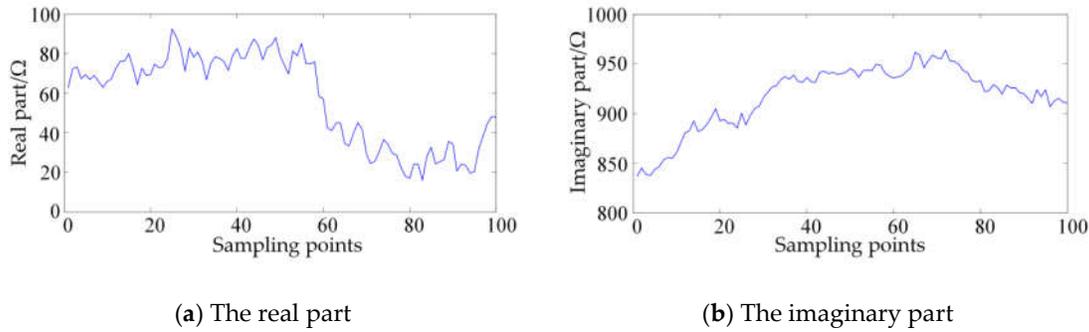
**Figure 10.** Harmonic impedance spectrum with measured data.

Moreover, the mean values of the harmonic impedance is calculated by (8) and the amplitude of them is shown in Figure 11. Same with the above analysis, the characteristic frequencies range from 2750 Hz to 3000 Hz. And the maximum amplitude of the mean values is about 3000  $\Omega$  at 2750 Hz, while the maximum amplitude in Figure 10 is about 4800  $\Omega$ . In addition, the maximum amplitude of harmonic impedance at other characteristic frequencies mostly decreases, because of large changes in phases. This demonstrates that the harmonic impedance at the characteristic frequencies easily changes, which means it is difficult to identify it accurately.



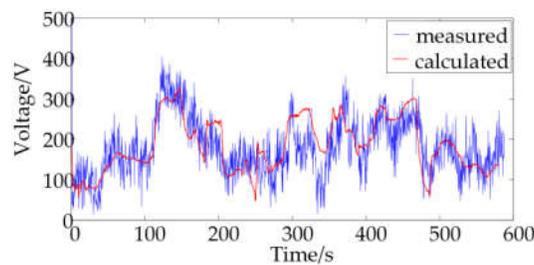
**Figure 11.** Amplitude of the mean value of harmonic impedance.

Secondly, to prove that the harmonic impedance of high reliability is accurate, we choose 100 results of 35th-order harmonic impedance with the highest SCD. They are shown in Figure 12 and their mean value is  $57.58 + j918.06(\Omega)$ .



**Figure 12.** The 35th-order harmonic impedance of 100 results.

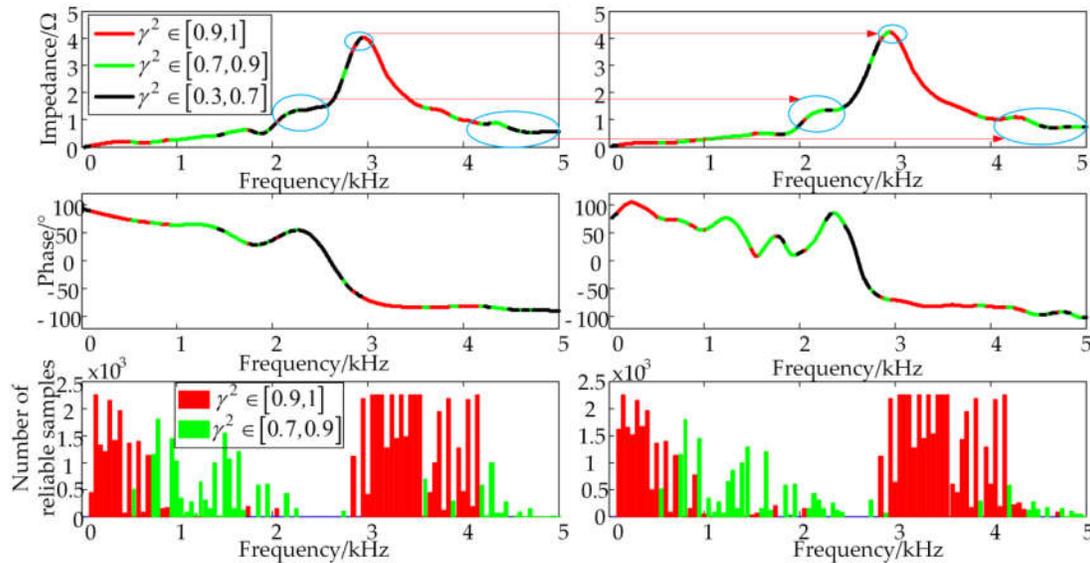
Then, the 35th-order harmonic voltage calculated by (2) and (3) at PCP with the impedance and current is shown in Figure 13, and the mean value of its amplitude is 183.83 V. In addition, the measured data show the voltage is about 179.84 V and the error is 2.30%. Thus, this shows the accuracy is high enough, and the validity of the proposed method is also verified. Furthermore, according to Figures 12 and 13, the trends of harmonic impedance and harmonic voltage vary obviously with time. This demonstrates that due to the train-network coupling the harmonic problem is dynamic and the harmonic impedance can change with the fast movement of trains.



**Figure 13.** The 35th-order harmonic voltage.

Thirdly, according to the flowchart in Figure 4, the data elimination of outliers and the data evolution mechanism are implemented in the harmonic impedance identification. The results of harmonic impedance identification based on the data evolution mechanism are shown in Figure 11.

The amplitudes and phases of harmonic impedance are shown in Figure 14a. It is obvious that the amplitude trend and harmonic resonance frequencies are very similar to the results of Figure 14. Moreover, the reliable samples means they are from the high-reliability data set and medium reliability data set ( $\gamma^2 \geq 0.7$ ). According to the number of reliable samples, there are not enough reliability samples to prove whether the harmonic impedance identification is accurate at some important frequencies close to the harmonic resonance frequencies. Because at these frequencies the train-network coupling system is in an unstable state and this locomotive is moving fast, it could lead to dynamic change of the impedance and system parameters. This is why the results of the harmonic impedance identification at these frequencies are not reliable.



(a) The identification results with a part of data (b) The identification results with all data

**Figure 14.** Harmonic impedance characteristic curves and reliable samples.

Therefore, after adding new measured data of the same locomotive in the same power supply section, we use all measured data to obtain the results in Figure 11. It is easy to find the improvement.

1. The results with high reliability and medium reliability cover a wider spectrum, such as the circled harmonic impedance in the top of Figure 11.
2. A part of the results raises the reliability levels, such as 30th-order and 31st-order harmonic impedance.
3. The amount of reliable data absolutely increases, such as 18th-order and 29th-order.

Moreover, compared with the maximum amplitude of 3000  $\Omega$  in Figure 10, the maximum amplitude of harmonic impedance can reach 4000  $\Omega$  and the reliability of the results is high enough. This means that a part of results of the calculated harmonic impedance are not accurate. If the mean value of all calculation results is directly considered as the identification result, highly reliable and accurate data could be mixed with the wrong data. Therefore, the identification result of introducing the data evolution mechanism is better than the result without any data processing.

Although the limitations of the practical measurement make it impossible to fully demonstrate the entire process of the data evolution mechanism, the validation of it has been proved by comparison between a part of data and all data. With the increase of measured data, the harmonic impedance characteristic curves will be improved.

## 6. Conclusions

In this paper, a method is proposed to utilize the measured data of trains at the PCP to identify the harmonic impedance of the traction network. According to the characteristics of the train-network coupling system in the RESs, the data elimination (Grubbs criterion) and data evolution mechanism (based on the sample coefficient of determination) are introduced into the linear regression method (PLSR), which is based on the Thevenin equivalent circuit of the traction network. Compared to the traditional identification based on linear regression without any data processing, the proposed method can not only improve identification accuracy but also estimate the reliability of identification results. It demonstrates that the identification results are reliable and uncontroversial. Moreover, as a result of the introduction of the data evolution mechanism, with the

addition of new measured data, the identification results can be further supplemented and improved. The harmonic impedance covering all important frequencies, such as characteristic frequencies, can eventually be obtained with high reliability. Therefore, a reliable identification of harmonic impedance can provide a data basis for harmonic suppression (such as harmonic filter design and adjustment of train current spectrum to avoid resonance frequencies), which is beneficial to improving the security and stability of the RESs.

Although the proposed method holds the above advantages, it can only be used for offline analysis at present. The identification method requires harmonic data of all frequencies so that the computation time would be too long. Thus, in addition to improving the robustness of the method, it is necessary to focus on the computation time and data amount. For example, the harmonic impedance only at characteristic frequencies is identified. These problems need to be studied in future research.

**Author Contributions:** Conceptualization, F.Z.; Methodology, F.Z.; Validation, R.Y. and K.Z.; Formal analysis, R.Y.; Investigation, K.Z.; Resources, F.Z.; Writing—original draft preparation, R.Y.; writing—review and editing, R.Y. and F.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the National Key Research and Development Program of China under Grant 2016YFB1200401 and Grant 2017YFB1201103.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest.

## References

- Holtz, S.J.; Keli, H.-J. The propagation of harmonic currents generated by inverter-fed locomotives in the distributed overhead supply system. *IEEE Trans. Power Electron.* **1989**, *4*, 168–174, doi:10.1109/63.24900.
- Lee, H.; Lee, C.; Jang, G.; Kwon, S. Harmonic analysis of the Korean high-speed railway using the eight-port representation model. *IEEE Trans. Power Deliv.* **2006**, *21*, 979–986, doi:10.1109/TPWRD.2006.870985.
- He, Z.; Hu, H.; Zhang, Y.; Gao, S. Harmonic Resonance Assessment to Traction Power-Supply System Considering Train Model in China High-Speed Railway. *IEEE Trans. Power Deliv.* **2014**, *29*, 1735–1743, doi:10.1109/TPWRD.2013.2284233.
- Mollerstedt, E.; Bernhardsson, B. Out of control because of harmonics—analysis of the harmonic response of an inverter locomotive. *IEEE Control Syst. Mag.* **2000**, *20*, 70–81, doi:10.1109/37.856180.
- Wang, X.; Blaabjerg, F. Harmonic Stability in Power Electronic-Based Power Systems: Concept, Modeling, and Analysis. *IEEE Trans. Smart Grid* **2019**, *10*, 2858–2870, doi:10.1109/TSG.2018.2812712.
- Zhou, F.; Xiong, J.; Zhong, K.; Zhu, P.; Huang, Q. Research on the phenomenon of the locomotive converter output current spectrum move based on the coupling of the train net system. In Proceedings of the CSEE, Guelph Canada, 18–21 July 2018; Volume 38, pp. 1818–1825, doi:10.13334/j.0258-8013.pcsee.170612. (In Chinese)
- Liu, S.; Lin, F.; Fang, X.; Yang, Z.; Zhang, Z. Train Impedance Reshaping Method for Suppressing Harmonic Resonance Caused by Various Harmonic Sources in Trains-Network Systems With Auxiliary Converter of Electrical Locomotive. *IEEE Access* **2019**, *7*, 179552–179563, doi:10.1109/ACCESS.2019.2958880.
- Rivas, D.; Moran, L.; Dixon, J.W.; Espinoza, J.R. Improving passive filter compensation performance with active techniques. *IEEE Trans. Ind. Electron.* **2003**, *50*, 161–170, doi:10.1109/TIE.2002.807658.
- Song, K.; Mingli, W.; Yang, S.; Liu, Q.; Agelidis, V.G.; Konstantinou, G. High-Order Harmonic Resonances in Traction Power Supplies: A Review Based on Railway Operational Data, Measurements, and Experience. *IEEE Trans. Power Electron.* **2020**, *35*, 2501–2518, doi:10.1109/TPEL.2019.2928636.
- Liu, Y.; Xu, J.; Shuai, Z.; Li, Y.; Peng, Y.; Liang, C.; Cui, G.; Hu, S.; Zhang, M.; Xie, B. A Novel Harmonic Suppression Traction Transformer with Integrated Filtering Inductors for Railway Systems. *Energies* **2020**, *13*, 473, doi:10.3390/en13020473.
- Xiong, J.; Li, Y.; Cao, Y.; Panasetky, D.; Sidorov, D. Modeling and operating characteristic analysis of MMC-SST based shipboard power system. In Proceedings of the 2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), Xi'an, China, 25–28 October 2016; pp. 28–32, doi:10.1109/APPEEC.2016.7779464.

12. Morrison, R.E.; Corcoran, J.C.W. Specification of an overvoltage damping filter for the National Railways of Zimbabwe. *IEE Proc. B Electr. Power Appl.* **1989**, *136*, 249–256, doi:10.1049/ip-b.1989.0033.
13. Tsukamoto, M.; Kouda, I.; Nasuda, Y.; Minowa, Y.; Nishimura, S. Advanced Method to Identify Harmonics Characteristic between Utility Grid and Harmonic Current Sources. In Proceedings of the 8th International Conference on Harmonics and Quality of Power. Proceedings (Cat. No.98EX227), Athens, Greece, 14–16 October 1998; Volume 1, pp. 419–425, doi:10.1109/ICHQP.1998.759946.
14. Song, K.; Yuan, X.; Chen, B.; Zhao, S. A method for assessing customer harmonic emission level based on improved fluctuation method. In Proceedings of the 2012 China International Conference on Electricity Distribution, Shanghai, China, 10–14 September 2012; pp. 1–5, doi:10.1109/CICED.2012.6508436.
15. Yang, H.; Pirotte, P.; Robert, A. *Harmonic Emission Levels of Industrial Loads Statistical Assessment*; CIGRE Proceedings; International Council on Large Electric Systems: Paris, France, 1996; pp. 36–306.
16. Yao, S.; Yang, J.; Wang, Y.; Liu, L.; Ma, L. The Harmonic Assessment Method Based on the Generalized Ridge Regression. In Proceedings of the 2012 Asia-Pacific Power and Energy Engineering Conference, Shanghai, China, 27–29 March 2012; pp. 1–4, doi:10.1109/APPEEC.2012.6307682.
17. Xu, Y.; Huang, S.; Liu, Y. Partial Least-Squares Regression Based Harmonic Emission Level Assessing at the Point of Common Coupling. In Proceedings of the 2006 International Conference on Power System Technology, Chongqing, China, 22–26 October 2006; pp. 1–5, doi:10.1109/ICPST.2006.321428.
18. Zang, T.; He, Z.; Fu, L.; Wang, Y.; Qian, Q. Adaptive method for harmonic contribution assessment based on hierarchical K-means clustering and Bayesian partial least squares regression. *IET Gener. Transm. Distrib.* **2016**, *10*, 3220–3227, doi:10.1049/iet-gtd.2015.1426.
19. Li, P.; Tao, S.; Yao, L.; Qian, Y.; Sun, J. An assessment method of power system harmonic impedance based on mixed total least squares. In Proceedings of the 2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), Xi'an, China, 25–28 October 2016; pp. 1694–1698, doi:10.1109/APPEEC.2016.7779839.
20. Zhou, F.; Liu, F.; Yang, R.; Liu, H. Method for Estimating Harmonic Parameters Based on Measurement Data without Phase Angle. *Energies* **2020**, *13*, 879, doi:10.3390/en13040879.
21. Liu, Z.; Hu, X.; Liao, Y. Vehicle-Grid System Stability Analysis Based on Norm Criterion and Suppression of Low-Frequency Oscillation with MMC-STATCOM. *IEEE Trans. Transp. Electr.* **2018**, *4*, 757–766, doi:10.1109/TTE.2018.2837623.
22. Giarnetti, S.; Leccese, F.; Caciotta, M. Non recursive multi-harmonic least squares fitting for grid frequency estimation. *Meas. J. Int. Meas. Confed.* **2015**, *66*, 229–237, doi:10.1016/j.measurement.2015.02.021.
23. Geladi, P.; Kowalski, B.R. Partial least-squares regression: A tutorial. *Anal. Chim. Acta* **1986**, *185*, 1–17, doi:10.1016/0003-2670(86)80028-9.
24. Yang, S.; Wu, M. Study on Harmonic Distribution Characteristics and Probability Model of High Speed EMU Based on Measured Data. *J. China Railw. Soc.* **2010**, *32*, 33–38, doi:10.3969/j.issn.1001-8360.2010.03.006. (In Chinese)

