Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Denoising of Industrial Oscillation Data Using EEMD with CCA ⋆

Xun Lang * **Yan Liu** ** **Yufeng Zhang** * **Lei Xie** ***
Alexander Horch **** **Hongye Su** ***

* *Department of Electronic Engineering, Information School, Yunnan University, Kunming 650091, China (e-mail: langxun@ynu.edu.cn (Xun Lang), zhangyf@ynu.edu.cn (Yufeng Zhang)).*
** *Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, China (e-mail: liuyan@siebt.ac.cn).*
*** *State Key Laboratory of Industrial Control Technology, Zhejiang University, 310027 Hangzhou, China (e-mail: leix@iipc.zju.edu.cn (Lei Xie), hysu@iipc.zju.edu.cn (Hongye Su)).*
**** *HIMA Paul Hildebrandt GmbH, Albert-Bassermann-Str. 28, 68782 Bruehl bei Mannheim, Germany (e-mail: a.horch@hima.com).*

Abstract: Industrial oscillation recordings are often contaminated with random noise, process disturbances and underlying nonstationarity, which obscure the useful information of the signal and complicate subsequent oscillation detection and diagnosis. This paper proposes a novel denoising technique to improve the quality of oscillation data, by jointly employing ensemble empirical mode decomposition (EEMD) with canonical correlation analysis (CCA). The proposed method first utilizes EEMD to decompose the single-loop data into a set of intrinsic mode functions (IMFs). Then CCA is applied to isolate the oscillation-dependent components from the decomposed IMFs. We evaluated the performance of the method through both numerical and industrial examples. The results demonstrate that this work is a promising tool for oscillation data preprocessing in the single control-loops.

Keywords: Industrial Oscillation, Denoising, EEMD, CCA.

1. INTRODUCTION

With the increasing demands on product quality and system performance, modern industrial processes are designed more complicated in both structure and automation (Chen et al. (2019b)). As a result, the number of oscillatory loops in process plant keeps steady high even with the growth of technologies and researches in the area of control performance monitoring (Bauer et al. (2016)). Although the statement of Jelali and Huang (2009) affirmed that oscillation is a solved problem, a recent review (Dambros et al. (2019)) reported that oscillation is still one of the most frequent problems in process control, and the last ten years have witnessed a rapid development of research activities in oscillation monitoring (Lang et al. (2018)).

One of the main reasons for above research status is that most of the established oscillation monitoring methods require the analyzed signal to fulfill certain conditions that are not always met in practical application (Zhou et al. (2017); Dambros et al. (2019)). Because of the feedback control systems, random noise, disturbances and the nonstationary trend are often presented in both input and output variables of the process (Dambros et al. (2017); Lang et al. (2018)). The presence of these artifacts can often lead to a decrease in the accuracy of standard oscillation detection methods and subsequently penalize their industrial acceptance (Zhou et al. (2017)). It is important to remove common artifacts from the signal before oscillation detection by most methods, however, the research topic on developing applicable preprocessing techniques is rarely discussed (Dambros et al. (2019)).

Zhou et al. (2017) published the only work exclusively focused on this topic. The proposed technique aims to detect and remove transient changes from the oscillatory time series, which cannot cater for a general denoising task of the industrial oscillation data. There are indeed a number of works that made certain improvements on robustness of their methods to deal with contaminated data (Thornhill et al. (2003); Naghoosi and Huang (2014); Xie et al. (2016b); Aftab et al. (2018); Chen et al. (2019a)), however, such improvements which mostly cover specific problems, are not portable to other methods.

To resolve above issues, a novel noise removal method, namely EEMD-CCA, is developed in this study. The proposed method operates by first decomposing the industrial recordings via EEMD into several IMFs. Then CCA is adopted to linearly unmix all IMFs into their corresponding underlying sources. Thirdly, sources with low autocorrelation are determined as noise artifacts and rejected. Finally, the remained sources are reconstructed

⋆ The authors would like to thank sponsor and financial support from National Natural Science Foundation of China under Grant 81771928 and National Key Research and Development Program of China under Grant 2018YFB1701102.

by inverse CCA to produce the noise-free oscillation data. The effectiveness of the proposed approach is verified by simulations as well as typical industrial case.

The rest of this paper is organized as follows: Section 2 introduces the methodology of the work. In Section 3, detailed performance analysis and comparison are conducted on two representative simulations. An industrial case is studied in Section 4, which is followed by conclusions in Section 5.

2. METHODOLOGY

2.1 Notion and Problem Description

Our choice of using the EEMD-CCA method for signal denoising is not trivial or arbitrary. It is motivated by the prerequisite that an oscillation in the signal is usually highly autocorrelated (Thornhill et al. (2003); Naghoosi and Huang (2014)), while other noisy artifacts have relatively low autocorrelations due to their irregular amplitudes and broader frequency spectrum. This is not conflict with our intuitive knowledge, due to the fact that oscillation is heuristically defined as periodic variation that is not completely hidden in noise (Horch (2007)), or signal that proposes well-defined amplitude and frequency (Choudhury et al. (2008)).

Based on above observation, the CCA technique can be adopted to isolate noise and disturbance from the ongoing signal, since it aims to find the sources maximally autocorrelated and meanwhile mutually uncorrelated (Chen et al. (2015)). However, an important property required to run CCA is that the number of the signal channels should greater than or equal to the number of unknown underlying sources, making it unavailable for single-channel (single control loop) denoising (Sweeney et al. (2012)).

To address this problem, additional signal decomposition step is needed in order to extend the input channel (Liu et al. (2019)). Then a two-step denoising strategy can be developed which involves: (i) A single-channel signal is first decomposed into multi-channel components using popular methods such as wavelet analysis or empirical mode decomposition (EMD) (Huang et al. (1998)); (ii) Those decomposed multichannel signals are further processed by CCA to generate meaningful sources. Without *a priori* knowledge of the signal of interest, it is hard to select mother wavelets and decomposition level for wavelet analysis (Xie et al. (2016a)). In contrast, EMD is completely data-driven that has been widely used in nonlinear and nonstationary processes.

In this work, instead of EMD, we propose to use the EEMD method (Wu and Huang (2009)) for oscillation data decomposition. The main reason is that EEMD as a noise-assisted EMD version solves the mode mixing problem over the univariate time-frequency tools, making it viable in handling industrial data which is full of noise and signal intermittency.

2.2 Empirical mode decomposition

The standard EMD can adaptively decompose the input into a finite set of oscillatory components known as IMFs. More specifically, for a real valued signal $x(t)$, the application of EMD yields M sets of IMFs, denoted as $\{d_i(t)\}_{i=1}^{M}$, and a monotonic residue $r(t)$, so that

$$x(t) = \sum_{i=1}^{M} d_i(t) + r(t), \tag{1}$$

where $r(t)$ is a monotonic function. Detailed procedures of EMD are summarized in Algo. 1 (Huang et al. (1998)).

Algorithm 1 Algorithm of EMD.

Input: $x^1(t) = x^2(t) = x(t)$, $i = 1$

1: Find the locations of all the extrema of $x^1(t)$;
2: Interpolate all the maxima (minima) to obtain the upper (lower) envelop, $e_{\max}(t)$ ($e_{\min}(t)$);
3: Find the local mean, $m(t) = [e_{\min}(t) + e_{\max}(t)]/2$;
4: Subtract the mean from the signal to obtain an oscillatory mode, $s(t) = x^1(t) - m(t)$;
5: If $s(t)$ obeys the stoppage criteria, $d_i(t) = s(t)$ becomes an IMF, go to step 6. Otherwise set $x^1(t) = s(t)$ and repeat the process from step 1;
6: Subtract the so derived IMF from $x^2(t)$, so that $x^2(t) := x^2(t) - d_i(t)$. If $x^2(t)$ becomes a monotonic function, stop the sifting process with $r(t) = x^2(t)$. Otherwise, $x^1(t) = x^2(t)$, $i = i + 1$ and go to step 1;
7: **return** $\{d_i(t)\}_{i=1}^{M}$ and $r(t)$.

2.3 Ensemble Empirical Mode Decomposition

The frequent appearance of noise and intermittency in real-world data usually causes mode mixing in EMD, where mode mixing is defined as either a single IMF consisting of widely disparate scales, or an amplitude frequency modulated oscillation residing in different IMFs (Wu and Huang (2009)). To alleviate this problem, ensemble empirical mode decomposition (EEMD) has been proposed by performing the EMD over an ensemble of the signal plus white Gaussian noise (wGn), and then treats the ensemble means of the corresponding IMFs as the final EEMD results, mode-by-mode. The aim of adding noise is to homogenize the scale in time-frequency space, which enables the natural filter bank of EMD to filter the intrinsic local oscillations adaptively to proper scales and thus restrain mode mixing.

Consider a real valued signal $x(t)$ with predefined noise amplitude ε and ensemble size N (they usually take values of $\varepsilon = 0.2$ and $N = 100$ as recommended by Wu and Huang (2009)), the procedures of EEMD is outlined in Algo. 2 (Wu and Huang (2009)).

2.4 Canonical Correlation Analysis

CCA is a blind source separation technique which solves the separation by forcing the sources to be maximally autocorrelated and mutually uncorrelated (Chen et al. (2015)). Let $\mathbf{X}_1(t) \in \mathcal{R}^{M\times T}$ be the observed data matrix with M channels and T observations, and $\mathbf{X}_2(t) = \mathbf{X}_1(t-1)$ be a temporally delayed version of the original data, CCA aims to find two sets of basis vectors $\mathbf{w}_1$ and $\mathbf{w}_2$, one for $\mathbf{X}_1(t)$ and another for $\mathbf{X}_2(t)$, such that the correlations ρ between the projections of the variables onto

Algorithm 2 Algorithm of EEMD.

Input: $x(t)$, ε, N.

1: Generate the ensemble $y_n(t) = x(t) + \varepsilon w_n(t)$ for $n = 1, \ldots, N$, where $w_n(t) \sim \mathcal{N}(0, 1)$;
2: Decompose every member of $y_n(t)$ into M_n IMFs using EMD (Algo. 1), to yield the set $\{d_i^n(t)\}_{i=1}^{M_n}$;
3: Assemble same-index IMFs across the ensemble using the mean operator to obtain the final IMFs within EEMD; for instance, the mth IMF is computed as $d_i(t) = \text{mean}\left\{d_i^1(t), d_i^2(t), \ldots, d_i^N(t)\right\}$;
4: **return** $\{d_i(t)\}_{i=1}^{M}$ and $r(t)$.

these basis vectors are mutually maximized. This leads to the following objective function maximizing the correlation between the linear combination of the components in $\mathbf{X}_1(t)$ and $\mathbf{X}_2(t)$:

$$\begin{aligned}\arg\max_{\mathbf{w}_1,\mathbf{w}_2} \rho(s_1(t), s_2(t)) &= \frac{E\left[s_1^T s_2\right]}{\sqrt{E\left[s_1^T s_1\right]E\left[s_2^T s_2\right]}} \\ &= \frac{\mathbf{w}_1^T \mathbf{C}_{12} \mathbf{w}_2}{\sqrt{\mathbf{w}_1^T \mathbf{C}_{11} \mathbf{w}_1}\sqrt{\mathbf{w}_2^T \mathbf{C}_{22} \mathbf{w}_2}}\end{aligned}, \qquad (2)$$

where $s_1(t) = \mathbf{w}_1^T \mathbf{X}_1(t)$ and $s_2(t) = \mathbf{w}_2^T \mathbf{X}_2(t)$ denote the first pair of canonical variates (sources), $\mathbf{C}_{11}$ and $\mathbf{C}_{22}$ are the auto-covariance matrices of $\mathbf{X}_1(t)$ and $\mathbf{X}_2(t)$, respectively, and $\mathbf{C}_{12}$ is the cross-covariance matrix. The solutions to this maximization problem can be derived by setting the derivatives of (2) with respect to $\mathbf{w}_1$ and $\mathbf{w}_2$ to zero, yielding following eigen decomposition problem (Sweeney et al. (2012)):

$$\begin{cases} \mathbf{C}_{11}^{-1}\mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21}\mathbf{w}_1 = \rho^2 \mathbf{w}_1 \\ \mathbf{C}_{22}^{-1}\mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{C}_{12}\mathbf{w}_2 = \rho^2 \mathbf{w}_2 \end{cases}, \qquad (3)$$

where ρ^2 denotes the eigenvalue and $\mathbf{w}_1$ and $\mathbf{w}_2$ are the corresponding eigenvectors. By following the same idea, the second pair of canonical variates can be obtained by solving the same eigen decomposition problem with additional constraint that they are uncorrelated with the first pair of variates. The subsequent pairs of canonical variates can again be derived according to (3), restricted to the condition that the variates are uncorrelated with all previously found variates. In doing so, typically M pairs of canonical variates (sources) can be derived from the combination of $\mathbf{X}_1(t)$ and $\mathbf{X}_2(t)$, which are denoted as:

$$\begin{cases} \mathbf{S}_1(t) = \left[s_1^1(t); s_1^2(t); \ldots; s_1^M(t)\right] \\ \mathbf{S}_2(t) = \left[s_2^1(t); s_2^2(t); \ldots; s_2^M(t)\right] \end{cases}, \qquad (4)$$

with their corresponding unmixing matrices:

$$\begin{cases} \mathbf{W}_1 = \left[\left(\mathbf{w}_1^1\right)^T; \left(\mathbf{w}_1^2\right)^T; \ldots; \left(\mathbf{w}_1^M\right)^T\right] \\ \mathbf{W}_2 = \left[\left(\mathbf{w}_2^1\right)^T; \left(\mathbf{w}_2^2\right)^T; \ldots; \left(\mathbf{w}_2^M\right)^T\right] \end{cases}. \qquad (5)$$

According to their definitions, we have $\mathbf{S}_1(t) = \mathbf{W}_1\mathbf{X}_1(t)$ and $\mathbf{S}_2(t) = \mathbf{W}_2\mathbf{X}_2(t)$. Note that the corresponding rows between $\mathbf{S}_1(t)$ and $\mathbf{S}_2(t)$ are highly correlated, while the rows within each individual matrix are uncorrelated with each other. In addition, the rows in $\mathbf{S}_1(t)$ are sorted in terms of their autocorrelation ρ^i, which is computed as

$$\rho^i = \frac{1}{(N-1)(\sigma^i)^2} \sum_{t=2}^{T} \left(s_1^i(t) - \mu^i\right)\left(s_1^i(t-1) - \mu^i\right), \quad (6)$$

where σ^i and μ^i denote the standard deviation and mean of $s_1^i(t)$, respectively. Due to the relatively low autocorrelation, CCA is able to isolate the noisy artifacts into the first several source components. Consequently, removal of the noise can be accomplished in a manner that the rows of $\mathbf{S}_1(t)$, which represent the artifacts in the industrial recordings, are set to zero before performing the reconstruction (Liu et al. (2019)).

2.5 Proposed EEMD-CCA Method

The proposed EEMD-CCA based denoising scheme is a combination of EEMD and CCA, which is actually a two-step modeling method.

Firstly, the industrial oscillation data $x(t)$ is decomposed into a multichannel signal $\mathbf{X}(t)$ using the EEMD algorithm. In order to reduce the impact of nonstationary trend on oscillation monitoring, the residual signal $r(t)$ is removed. As a result, $\mathbf{X}(t)$ only contains the M sets of IMFs, i.e., $\mathbf{X}(t) = [d_1(t); d_2(t); \ldots; d_M(t)]$.

Secondly, CCA is applied to the multichannel signal $\mathbf{X}(t)$ to obtain the unmixing matrix $\mathbf{W}$ and source matrix $\mathbf{S}(t)$, in which the rows of the latter are sorted orderly from low to high according to their autocorrelation.

Thirdly, the first several sources of $\mathbf{S}(t)$ that corresponding to the noise components are set to zero, yielding the cleaned source matrix $\tilde{\mathbf{S}}(t)$. By leveraging the observation that noise component has lower autocorrelation as compared to the oscillatory component, a threshold of $\rho_0 = 0.9$ is designated to separate the noise from the oscillation of interest. In practice, according to our experience, the threshold between 0.75 and 0.9 will lead to good results for most cases.

Finally, the cleaned multichannel signal $\tilde{\mathbf{X}}(t)$ can be reconstructed by multiplying the mixing matrix $\mathbf{A} = \mathbf{W}^{-1}$ with the cleaned source matrix $\tilde{\mathbf{S}}(t)$. The denoised oscillation data $\tilde{x}(t)$ are eventually computed by summing the updated IMFs (rows) in $\tilde{\mathbf{X}}(t)$. To elaborate the details of the proposed EEMD-CCA method, we summarize the entire procedures of the algorithm in Algo. 3.

We highlight that by doing the cleaning and reconstruction using only the EEMD decomposed IMFs is not enough to achieve a desired cancellation of the noise. Since these IMFs usually contain both the oscillatory components and the noise artifacts, the direct rejection of any IMF will result in the potential loss of oscillations or residue of noise activities. To verify this claim, an oscillation plus wGn signal is constructed as given by

$$x_1(t) = \sin(2\pi f t) + \delta w(t), \qquad (7)$$

in which δ denotes the amplitude of the added noise and $w(t) \sim \mathcal{N}(0, 1)$. If we set $\delta = 0.3$, $f = 6$, $f_s = 1000$ (sampling rate), and $T = 1000$ (data length), a typical decomposition of $x_1(t)$ via EEMD is shown in Fig. 1. It is observed that the fundamental wave is split into $d_5(t)$ and $d_6(t)$, and each of them is a mixture of the oscillation and noise. Since the morphological characteristics of the oscillation in these two modes are slightly distorted, it is hard to accurately reconstruct the cleaned signal by directly using the decomposed IMFs.

Algorithm 3 Algorithm of EEMD-CCA.

Input: $x(t)$ and ρ_0.

1: Decompose a set of averaged IMFs as the final output of EEMD, i.e., $\mathbf{X}(t) = [d_1(t); d_2(t); \ldots; d_M(t)]$;
2: Utilize $\mathbf{X}(t)$ as the input of CCA to derive the source matrix $\mathbf{S}(t) = [s_1(t); s_2(t); \ldots; s_M(t)]$ and the unmixing matrix $\mathbf{W}$;
3: Calculate the autocorrelation for each row of $\mathbf{S}(t)$, yielding $\rho = [\rho_1; \rho_2; \ldots; \rho_M]$.
4: Initialize $\tilde{\mathbf{S}}(t) = \mathbf{S}(t)$, then:
for $i = 1 : M$ **do**
  if $\rho_i < \rho_0$ **do**
    $\tilde{s}_i(t) = 0$;
  end if
end for
5: Reconstruct the cleaned multichannel signal $\tilde{\mathbf{X}}(t) = \left[\tilde{d}_1(t); \tilde{d}_2(t); \ldots; \tilde{d}_M(t)\right]$ by performing $\mathbf{W}^{-1}\tilde{\mathbf{S}}(t)$;
6: Compute the denoised data using $\tilde{x}(t) = \sum_{i=1}^{M} \tilde{d}_i(t)$;
7: **return** $\tilde{x}(t)$.

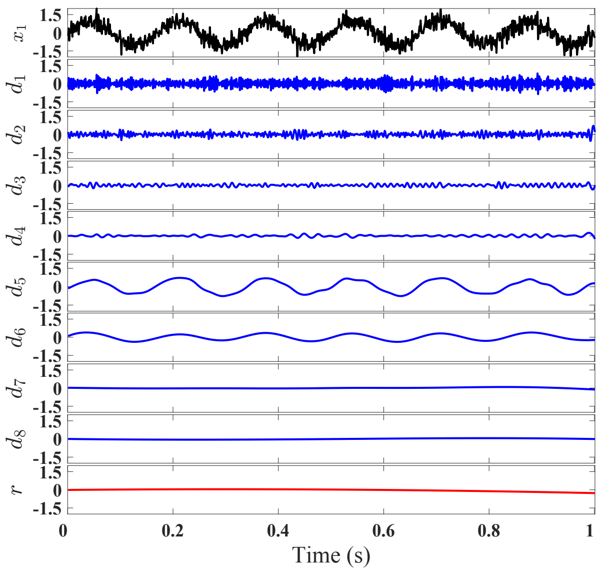

Fig. 1. Typical decomposition of $x_1(t)$ using EEMD. Note that all IMFs ($d_1(t)$-$d_8(t)$) will be treated as the input of the CCA method for calculating the canonical variates, although the last two IMFs are almost zero.

We next investigate the autocorrelations of all EEMD-based IMFs, where the coefficients against the IMF indices are plotted in the black dashed-line shown in Fig. 2. Obviously, the largest autocorrelation coefficient among them is 0.22. Such a low value further confirms our statement that EEMD is unable to directly separate the oscillation of interest from the original data. In comparison, the autocorrelations of the CCA transferred sources are calculated as depicted by the red solid-line. Owing to the ability of CCA of forcing the sources to be maximally autocorrelated and mutually uncorrelated, the underlying oscillation structure can be captured, ensuring a good denoising performance.

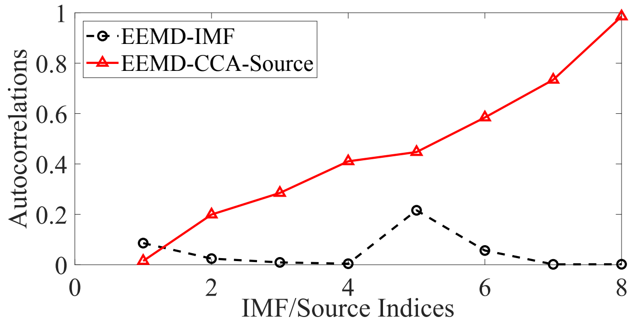

Fig. 2. Autocorrelations with different approaches.

2.6 Contribution of This Paper

The main contribution of this work is that we propose a novel denoising scheme for oscillation data preprocessing from industrial control systems. In addition to improving the accuracy and reliability of existing methods for oscillation monitoring, the proposed method also can help reduce uncertainties in the measurement process, thus relaxing instrument front-end design constraints and enhancing the reliability of instrument systems.

Although the combination of EEMD and CCA has already shown effectiveness for motion and muscle artifacts removal in single-channel EEG recordings (Sweeney et al. (2012)), this work is the first attempt to introduce EEMD-CCA into the field of oscillation monitoring for industrial data denoising. Moreover, by comparing with the work proposed by Sweeney et al. (2012), our method highlights following differences:

(i) To reconstruct the clearer EEG from the high-frequency dominated signal, Sweeney et al. (2012) used lower noise amplitude and ensemble size for EEMD implementation, i.e., $\varepsilon = 0.1$ and $N = 5$. For oscillation data denoising, however, the method needs much higher parameters ($\varepsilon = 0.2$, $N = 100$) since the signal is mainly dominated by low-frequency oscillations (Wu and Huang (2009)).

(ii) The residual signal $r(t)$ is removed from the EEMD decomposition, enabling the proposed method to detrend the industrial data, thus reduce the impact of nonstationary trend on oscillation monitoring.

(iii) The pure EEG is mainly located in frequency bands ranging from 1 to 30 Hz. In contrast, the oscillation is considered as a signal with well-defined and fixed frequency (or narrow-band signal with a small frequency variation). This observation indicates that the difference between the oscillation and the noise component on autocorrelation is more significant than the difference between EEG and artifact, which suggests that the effectiveness and advantages of EEMD-CCA may be more prominent for denoising the industrial oscillation data.

3. PERFORMANCE ANALYSIS AND COMPARISON

In this section, we set out to analyze the denoising performance of the proposed method on two representative signals: (i) a mono-component oscillation contaminated by both noise and trend, (ii) a nonlinearity-induced oscillation which is synthesized with noise, fundamental wave and two harmonics. To demonstrate the effectiveness of this work, the EEMD-based reconstruction is included as a competing technique. Since the autocorrelations of EEMD-IMFs are relatively low, a threshold of 0.15 is empirically used for the EEMD scheme to separate the noise artifacts from oscillation components.

3.1 Oscillation with Noise and Trend

The numerical signal used in this section is similar to (7), except that one quadratic trend is added, as given by

$$x_2(t) = \sin(2\pi f t) + 0.5t + 0.5t^2 + \delta w(t). \tag{8}$$

By initializing $\delta = 0.5$, $f = 5$, $f_s = 1000$ and $T = 1000$, typical realizations of both the EEMD and EEMD-CCA based denoising methods are shown in Fig. 3. From visual inspection, both methods show abilities of denoising and detrending, while the EEMD-CCA based scheme seems to yield a better reconstruction. However, more general and quantitative evidences should be provided in order to show the true usefulness of the proposed method.

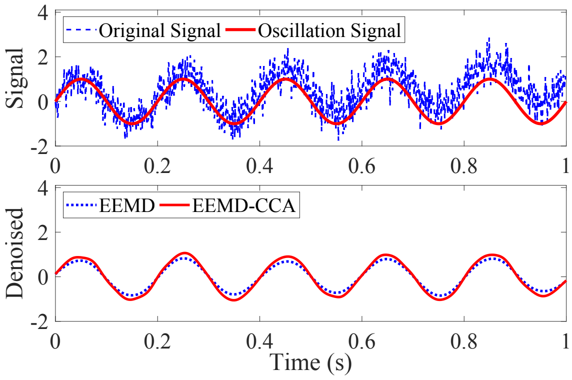

Fig. 3. The original signal and the denoised ones. Top: dashed-line for the original signal $x_2(t)$, solid-line for the oscillation $\sin(2\pi f t)$; Bottom: dotted-line for the EEMD denoised signal, solid-line for the EEMD-CCA denoised signal.

To cater for this requirement, we study the performance of EEMD and EEMD-CCA quantitatively, using multiple realizations of $x_2(t)$ with different frequency f and wGn sequences. The frequency value of $\sin(2\pi f t)$ is designed to vary from 3 to 10 with an interval of 1. Then for each of the frequency setting, 1000 realizations of EEMD and EEMD-CCA based denoising are carried out to yield their respective denoised oscillation data. The performance is measured using the Pearson correlation coefficient and root mean square error ($RMSE$) in the task of reconstructing $\sin(2\pi f t)$ from the original signal. Given paired data $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ consisting of n pairs, the Pearson correlation metric r_{xy} is defined as

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}, \tag{9}$$

where $\bar{x}$ and $\bar{y}$ denote the sample means. Additionally, $RMSE$ is defined as

$$RMSE = \frac{1}{\sqrt{n}}\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}. \tag{10}$$

Accordingly, the one with the higher Pearson correlation and lower $RMSE$ is the better method for industrial data denoising. The averaged correlation and $RMSE$ indices computed under different frequencies are plotted in Fig. 4 and Fig. 5, respectively.

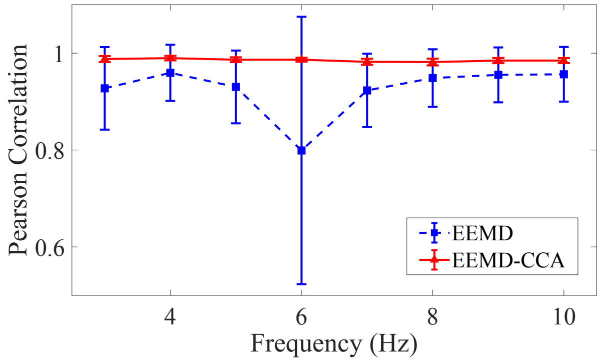

Fig. 4. The performance comparison between EEMD and EEMD-CCA in terms of Pearson correlation coefficient for $x_2(t)$ denoising.

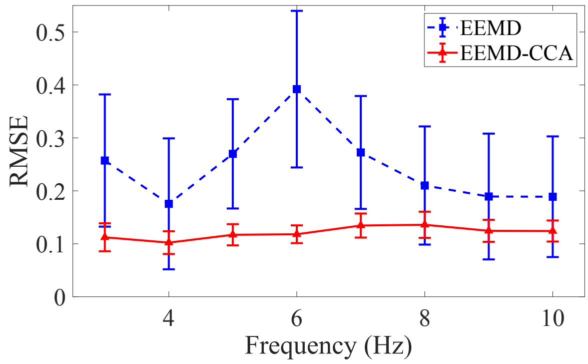

Fig. 5. The performance comparison between EEMD and EEMD-CCA in terms of $RMSE$ for $x_2(t)$ denoising.

As evidenced, the proposed EEMD-CCA method exhibits superior performance for all of the tested frequencies. More specifically, it observes higher mean but lower standard deviation on the Pearson correlation statistics, while showing lower mean and standard deviation on $RMSE$ statistics. This experiment indicates that EEMD-CCA is an effective enhancement tool for industrial oscillation data. We highlight that the extent of the nonstationarity trend has little effect on denoising performance of the method, since it has been extracted and removed as the residual during the EEMD process.

3.2 Nonlinearity-induced Oscillation

Here, the oscillation model is borrowed from Chen et al. (2019a), which is formulated as

$$x_3(t) = \frac{4}{\pi}\left[\sin(2\pi f t) + \sin(6\pi f t) + \sin(10\pi f t)\right] + \delta w(t). \tag{11}$$

Through this simulation, the proposed method can be more comprehensively examined in terms of multiple oscillations detection and nonlinearity-related harmonics extraction. The denoising results of a typical realization of

$x_3(t)$ for both EEMD and EEMD-CCA based methods are plotted in Fig. 6, with the configuration of $\delta = 0.25$, $f = 6$, $f_s = 1000$ and $T = 1000$. In this case, the proposed method shows considerable superiority, due to the observation that the EEMD-based denoising only preserved the fundamental wave, while the EEMD-CCA one retained both the fundamental and harmonic components.

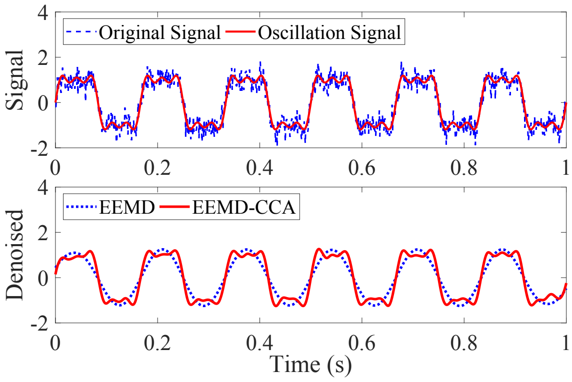

Fig. 6. The original signal and the denoised ones for $x_3(t)$. Top: dashed-line for the original signal, solid-line for the combination of the oscillations; Bottom: dotted-line for the EEMD denoised signal, solid-line for the EEMD-CCA denoised signal.

Similar to Section 3.1, we also investigated statistical evidences of the performance of both methods in denoising signal $x_3(t)$. Fig. 7 and Fig. 8 are corresponding averaged results (from 1000 independent experiments) for the Pearson correlation coefficient and $RMSE$, respectively, with the objective of separate all nonlinearity-related oscillations from the original data. As evidenced, the proposed method consistently outperforms the EEMD-based denoising for all frequency settings in terms of the higher correlations and lower $RMSE$ values.

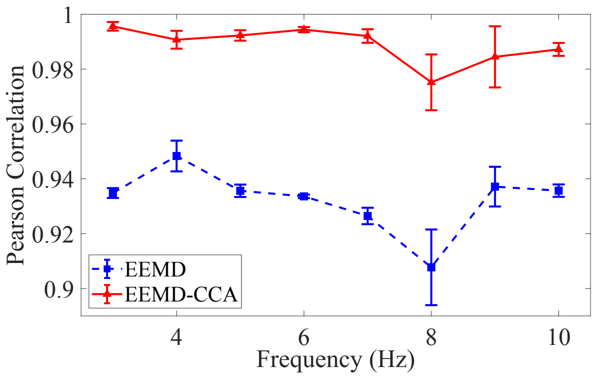

Fig. 7. The performance comparison between EEMD and EEMD-CCA in terms of Pearson correlation coefficient for $x_3(t)$ denoising.

This study shows that the proposed method is empirically a good choice for denoising signals synthesized by harmonics, which may facilitate the subsequent detection and diagnosis of nonlinearity-induced oscillations.

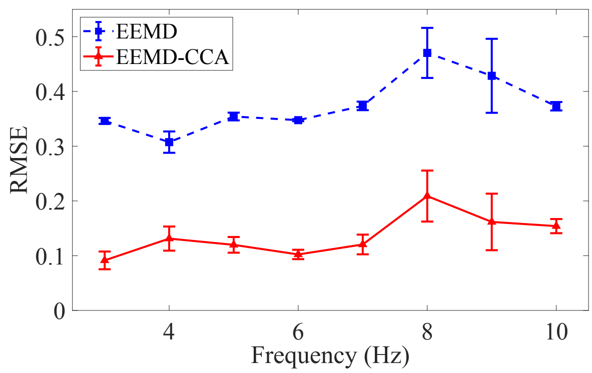

Fig. 8. The performance comparison between EEMD and EEMD-CCA in terms of $RMSE$ for $x_3(t)$ denoising.

4. INDUSTRIAL CASE STUDY

A typical industrial case is presented in this section to demonstrate the effectiveness of the proposed method for industrial oscillatory data denoising. The data set under study is borrowed from Jelali and Huang (2009), which is recorded from a level control loop in a chemical plant (chemicals.loop10.pv). The uncompressed nonlinear data are sampled from the output of the control system every 1 second, and it was known *a priori* that the control valve in this loop contained stiction (Jelali and Huang (2009)). Fig. 9 shows the original time series and its cleaned versions by both the EEMD and EEMD-CCA based schemes.

From manual inspection, it is observed that the process data does not contain obvious harmonics. This is not unusual because this quantified stiction is termed as apparent stiction, whose siction effect has been smoothed due to the influence of loop dynamics and the regulate action from the controller (Jelali and Huang (2009)).

With respect to the denoised data, we observe that the cleaned oscillation obtained from the proposed method exhibits the best fitting to the oscillation behavior underlying the original signal. In contrast, the EEMD-based method shows compromised capability of noise reduction, since the oscillatory component has split into several scales and then mixed with the corresponding noise components.

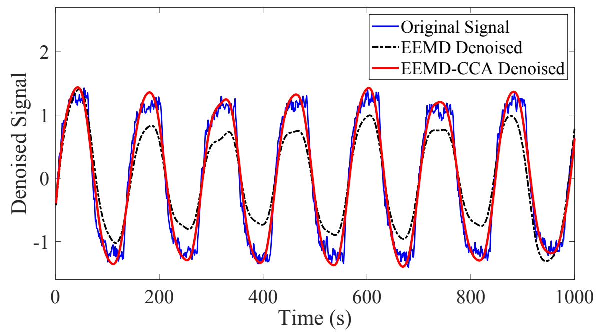

Fig. 9. The original industrial data and its cleaned versions by different methods.

To further verify the usefulness of this work, we use power spectral density (PSD) which describes the distribution of power into frequency components to evaluate the denoising performance. Due to the lack of ground truth of the cleaned oscillation, PSD is an appropriate tool for qualitative performance comparison. The PSD results are shown in Fig. 10, from which we can see that the EEMD-based method has rejected both the oscillation and noise components during the reconstruction. On the contrary, our proposed approach prohibits the loss of useful information, while largely suppresses the noise components.

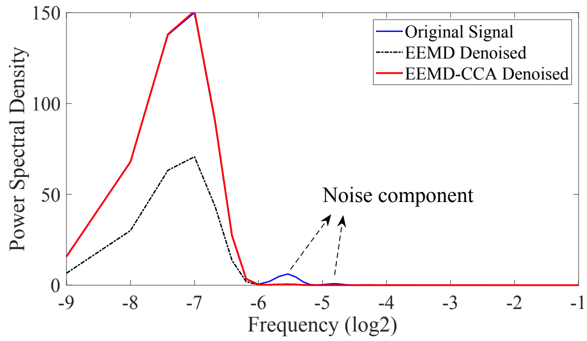

Fig. 10. PSD of the original data and the cleaned ones by different methods. In this figure, the blue curve (input signal) is almost covered by the red curve (EEMD-CCA denoised signal) in frequency interval $f < 2^{-6}$.

5. CONCLUSION

A novel method that incorporated EEMD with CCA has been proposed in this paper to reject noise artifacts of the industrial oscillation data. The approach was used to denoise the process measurements prior to applying the oscillation detection and diagnosis methods. We have verified the practicality of this work through extensive simulations as well as industrial case. Since EEMD-CCA can be applied in nonlinear and nonstationary processes, our contribution may change the conventional ways of data cleaning, which may further benefit the data-driven monitoring of industrial control systems.

Of course, there still exists some issues we have not discussed. For instance, more statistical studies should be carried out in order to determine the proper ensemble size and the amplitude of the added noise for the proposed method. Moreover, this work still needs to be verified in a wide variety of real-world applications.

REFERENCES

Aftab, M.F., Hovd, M., and Sivalingam, S. (2018). Improved oscillation detection via noise-assisted data analysis. *Control Engineering Practice*, 81, 162–171.

Bauer, M., Horch, A., Xie, L., Jelali, M., and Thornhill, N. (2016). The current state of control loop performance monitoring–a survey of application in industry. *Journal of Process Control*, 38, 1–10.

Chen, Q., Lang, X., Xie, L., and Su, H. (2019a). Detecting nonlinear oscillations in process control loop based on an improved vmd. *IEEE Access*, 7, 91446–91462.

Chen, T., Lou, J., Yang, Y., Ren, Z., and Xu, C. (2019b). Vibration suppression of a high-speed macro-micro integrated system using computational optimal control. *IEEE Transactions on Industrial Electronics*.

Chen, X., Liu, A., Chiang, J., Wang, Z.J., McKeown, M.J., and Ward, R.K. (2015). Removing muscle artifacts from eeg data: Multichannel or single-channel techniques? *IEEE Sensors Journal*, 16(7), 1986–1997.

Choudhury, A.A.S., Shah, S.L., and Thornhill, N.F. (2008). *Diagnosis of process nonlinearities and valve stiction: data driven approaches.* Springer Science & Business Media.

Dambros, J.W., Farenzena, M., and Trierweiler, J.O. (2017). Signal preprocessing for stiction detection methods. *Industrial & Engineering Chemistry Research*, 57(1), 302–315.

Dambros, J.W., Trierweiler, J.O., and Farenzena, M. (2019). Oscillation detection in process industries–part i: Review of the detection methods. *Journal of Process Control*, 78, 108–123.

Horch, A. (2007). Benchmarking control loops with oscillations and stiction. In *Process control performance assessment*, 227–257. Springer, London.

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C., and Liu, H.H. (1998). The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. In *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences*, volume 454, 903–995. The Royal Society.

Jelali, M. and Huang, B. (2009). *Detection and diagnosis of stiction in control loops: state of the art and advanced methods.* Springer Science & Business Media.

Lang, X., Zhang, Z., Xie, L., Horch, A., and Su, H. (2018). Time-frequency analysis of plant-wide oscillations using multivariate intrinsic time-scale decomposition. *Industrial & Engineering Chemistry Research*, 57(3), 954–966.

Liu, Y., Zhou, Y., Lang, X., Liu, Y., Zheng, Q., Zhang, Y., Jiang, X., Zhang, L., Tang, J., and Dai, Y. (2019). An efficient and robust muscle artifact removal method for few-channel eeg. *IEEE Access*, 7, 176036–176050.

Naghoosi, E. and Huang, B. (2014). Automatic detection and frequency estimation of oscillatory variables in the presence of multiple oscillations. *Industrial & Engineering Chemistry Research*, 53(22), 9427–9438.

Sweeney, K.T., McLoone, S.F., and Ward, T.E. (2012). The use of ensemble empirical mode decomposition with canonical correlation analysis as a novel artifact removal technique. *IEEE transactions on biomedical engineering*, 60(1), 97–105.

Thornhill, N.F., Huang, B., and Zhang, H. (2003). Detection of multiple oscillations in control loops. *Journal of Process Control*, 13(1), 91–100.

Wu, Z. and Huang, N.E. (2009). Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in adaptive data analysis*, 1(01), 1–41.

Xie, L., Lang, X., Chen, J., Horch, A., and Su, H. (2016a). Time-varying oscillation detector based on improved lmd and robust lempel–ziv complexity. *Control Engineering Practice*, 51, 48–57.

Xie, L., Lang, X., Horch, A., and Yang, Y. (2016b). Online oscillation detection in the presence of signal intermittency. *Control Engineering Practice*, 55, 91–100.

Zhou, B., Chioua, M., and Schlake, J.C. (2017). Practical methods for detecting and removing transient changes in univariate oscillatory time series. *IFAC-PapersOnLine*, 50(1), 7987–7992.