International Journal of Innovative
Computing, Information and Control
Volume **16**, Number **2**, April **2020**

ICIC International ©2020 ISSN 1349-4198
pp. 585–595

DETERMINATION OF MODEL STRUCTURE VIA CYCLO-STATIONARITY BASED NEURAL NETWORK

Lianming Sun[1], Xinyu Liu[2] and Jianing Liu[2]

[1]Faculty of Environmental Engineering
[2]Graduate School of Environmental Engineering
The University of Kitakyushu
1-1 Hibikino, Wakamatsu-ku, Kitakyushu 808-0135, Japan
sun@kitakyu-u.ac.jp; { z8dcb001; z8mcb402 }@eng.kitakyu-u.ac.jp

Received August 2019; revised December 2019

Abstract. *Determination of model structure commonly influences the performance of system identification and model applications. It has to be performed by the data-driven methods if the priori structure information is not available, whereas the data are sometimes collected under severe experiment conditions. In this paper, a cyclo-stationarity based neural network is applied to determining the model structure through compounding information indices obtained by the output over-sampling scheme. It is illustrated that several distinct information indices on the cyclo-stationarity are detected from the experimental data. Then, different indices are compounded by a neural network to improve the determination performance. The effectiveness of the proposed approach is demonstrated through an identification experiment on a magnetic levitation system, while the performance of conventional methods degrades largely due to the severe numerical conditions.*
Keywords: Model structure, Output over-sampling, Cyclo-stationarity, System identification

1. Introduction.

Determination of model structure is a key problem in system identification, model-based system analysis and design where the mathematical model is explicitly applied. Two main categories, i.e., knowledge-based and data-driven methods, are used to determine the model structure. However, in many practical engineering systems, the priori information is not sufficient enough to construct an effective model; consequently, the data-driven method is necessary in engineering applications [1]. If the experimental data or the model residuals are regarded as stochastic ones, then some information criteria such as the final prediction-error (FPE), Akaike's information criterion (AIC), the minimum description length (MDL), or Bayesian information criterion (BIC) can be used to determine the model orders by minimizing the information criteria on both the mean square residual and model complexity [2]. An alternative approach is to find the principal eigenvalues in decomposition of data space [3]. After the model structure is determined, the model parameters are then estimated through the estimation algorithms.

Informativeness of experimental data is an essential condition in data-driven methods. However, there are many restrictions to obtain informative data in practical applications. For example, the frequency band of some process signals in physical systems is often limited in a narrow specified range; the experimental data are corrupted by uncertainty or unmeasurable noises [4]; the input and output signals of a system in the closed-loop systems have not sufficient independent excitations due to the feedback loop. If the experimental data are not informative enough, there exist some local minima in the information

DOI: 10.24507/ijicic.16.02.585

criteria, while the conventional methods are largely affected by the noise terms and uncertainties so they cannot extract satisfactory structure information from the experimental data. It is expected to introduce additional information such as some well-designed test signals that help to improve the informativeness; nevertheless, external test signals cannot be arbitrarily added into the practical system for some safety or economic reasons. Therefore, to detect much extra information from the data obtained under the feasible experiment conditions is an important and challenging task in these issues. For example, more frequency information and spatial information are strongly desired by using the techniques such as signal processing, sensor networks or multi-rate sampling schemes.

On the other hand, it has been shown that a distinct signal characteristic called cyclo-stationarity can be detected from some specified signals, such as the modulated signals in communication systems [5], the test signals designed for system identification [6], the output signals sampled at multiple times higher rate than that of the input signals [7, 8, 9], and the measured signals by multi-sensors [10]. In contrast with the stationarity, the relation functions of cyclo-stationary signals are periodic, and the cyclo-stationary relation functions contain unique magnitude, phase and spatial information. The cyclo-stationary characteristics have been successfully applied into blind channel identification and blind equalization in digital communication systems through the phase information or the spatial information [8], instrumentation systems using the spatial information [10, 11], closed-loop identification to mitigate the severe numerical conditions [9, 12, 13], model validation [14], etc. Inspired by these results, we introduce the cyclo-stationarity into model structure determination for closed-loop systems [15], model structure identification using cyclo-stationary relation functions and spectra for unstable process without extra test signal [16], where the cyclo-stationary components are easily distinguished from the stationary signals. Although the information of cyclo-stationarity complements the mean square residual, the existing cyclo-stationarity based methods use a single type of cyclo-stationary index, which may still be affected by the data number, the excitations, especially under severe experiment conditions such as closed-loop identification. In order to guarantee the determination performance, the information criterion for structure determination should compound several different cyclo-stationary indices. Moreover, in some practical applications, since a linear criterion cannot well adapt to the complicated numerical conditions, an appropriate nonlinear compounded criterion is necessary for model structure determination, but there is not an explicit formulation to compound the different indices yet. Consequently, the effective approximation of the compounded criterion is strongly expected.

This paper investigates a neural network based nonlinear function approximation to compound different indices of cyclo-stationary information. It is shown that the cyclo-stationarity can be detected from the experimental data under output over-sampling scheme, and the information indices on the maximal magnitude, the mean magnitude and the distribution of cyclo-stationary components are compounded through a neural network to yield a nonlinear criterion for structure determination. The numerical example of an experiment to determine the model structure of levitation system illustrates the effectiveness of the proposed approach.

The rest of the paper is organized as follows. In the next section, the main problem of the model structure determination is overviewed. In Section 3, the cyclo-stationary characteristics of the experimental data are analyzed, and the estimation algorithm to detect the cyclo-stationarity from experimental data is presented. Then the model structure determination approach based on the neural network is illustrated in Section 4, and the numerical example of a magnetic levitation experiment is shown in Section 5. Finally, the conclusion and the future research work are given in Section 6.

2. **Problem Statement.** In system analysis and design, the model-based approaches generally use mathematical models to describe the dominant system properties quantitatively. Clearly the model structure largely influences the model's effectiveness and efficiency; however, sometimes it has to be determined by experiments in the practical applications where less priori information is available. The problem of model structure determination is stated first in this section.

2.1. **Model description.** Consider a system that works in a specified operation range, where the system model is approximated by a linear discrete-time model whose input is a piece-wise signal with a holding period T. If the system is stable, then it is often described by the following discrete-time transfer function expression

$$y(m) = G^{\mathrm{o}}\left(z^{-1}\right) u(m) + H^{\mathrm{o}}\left(z^{-1}\right) w(m), \tag{1}$$

where $u(m)$, $y(m)$ and $w(m)$ are the input, output and the noise term at the instant mT, respectively, and the input signal $u(m)$ is assumed to be treated as a pseudo stationary stochastic signal at the interval T. $G^{\mathrm{o}}\left(z^{-1}\right)$ and $H^{\mathrm{o}}\left(z^{-1}\right)$ are the true theoretical rational transfer functions with numerator orders n_B, n_C and denominator orders n_F, n_D, respectively, and z^{-1} is a backward shift operator. For the uniqueness of the model, $H^{\mathrm{o}}\left(z^{-1}\right)$ is specified as monic. For an unstable process, the model is often expressed by

$$A^{\mathrm{o}}_{\mathrm{us}}\left(z^{-1}\right) y(m) = G^{\mathrm{o}}_{\mathrm{s}}\left(z^{-1}\right) u(m) + H^{\mathrm{o}}\left(z^{-1}\right) w(m) \tag{2}$$

for the numerical stability, where $A^{\mathrm{o}}_{\mathrm{us}}\left(z^{-1}\right)$ is an n_Ath order monic polynomial that contains all the unstable poles of the process model, and $G^{\mathrm{o}}_{\mathrm{s}}\left(z^{-1}\right)$ is the remainder of the process transfer function after removing $A^{\mathrm{o}}_{\mathrm{us}}\left(z^{-1}\right)$. In many model-based practical applications, not only the model parameters, but the model orders are required to be identified from the experimental data. Therefore, the data of $\{u(m), y(m)\}$ must be informative enough to construct the system model, and then the estimated model orders, which are summarized in a vector $\hat{\boldsymbol{n}}^*$, are conventionally determined by

$$\hat{\boldsymbol{n}}^* = \arg \min_{\hat{\boldsymbol{n}}} V(\hat{\boldsymbol{n}}), \tag{3}$$

where $\hat{\boldsymbol{n}} = [\hat{n}_A, \hat{n}_B, \hat{n}_C, \hat{n}_D, \hat{n}_F]$, $\hat{n} = \hat{n}_A + \hat{n}_B + \hat{n}_C + \hat{n}_D + \hat{n}_F$, $V(\hat{\boldsymbol{n}})$ is a specified information criterion on the mean square residual and the penalty of model complexity. For example, the final prediction error criterion (FPE) defines the following criterion

$$V_{\mathrm{FPE}}(\hat{\boldsymbol{n}}) = \underbrace{\frac{1 + \hat{n}/K}{1 - \hat{n}/K}}_{V_2(\hat{n})} \underbrace{\left(\frac{1}{K}\sum_{k=1}^{K} \varepsilon^2(k, \hat{\boldsymbol{n}})\right)}_{V_1(\hat{\boldsymbol{n}})}, \tag{4}$$

where K is the data number, $V_1(\hat{\boldsymbol{n}})$ associates with the model residual $\varepsilon(k, \hat{\boldsymbol{n}})$ so it indicates the model accuracy, while $V_2(\hat{n})$ indicates the model complexity since it associates with the model orders directly. Furthermore, Akaike's information criterion (AIC) and minimum description length criterion (MDL) use the modified $V_2(\hat{n})$ to evaluate the model complexity in (5), (6), respectively [2].

$$V_{\mathrm{AIC}}(\hat{\boldsymbol{n}}) = \log\left(\left(1 + \frac{2\hat{n}}{K}\right) \sum_{k=1}^{K} \varepsilon^2(k, \hat{\boldsymbol{n}})\right), \tag{5}$$

$$V_{\mathrm{MDL}}(\hat{\boldsymbol{n}}) = \left(1 + \frac{2\hat{n}}{K} \log K\right) \sum_{k=1}^{K} \varepsilon^2\left(k, \hat{\boldsymbol{n}}\right). \tag{6}$$

If the experimental data $\{u(m), y(m)\}$ are informative, and they contain sufficient exciting components, $\hat{\boldsymbol{n}}^*$ obtained by (3) is close to the theoretical ones, and the model is possible to effectively describe the dominant characteristics of the system. Furthermore, a large number of system models have to be constructed by the data-driven methods, whereas the experimental data are collected under severe experiment conditions in many practical applications due to the economic or safety reasons. For example, the input signals in many industrial processes are restricted in such a specified frequency band that the experimental data cannot provide sufficiently exciting components for identification or model structure determination; the process is regulated by a feedback controller in closed-loop where a part of the information of the input signals is overlapped with that of the output so the independent information may be insufficient to determine the model structure; the experimental data are contaminated by strong noise terms in the instrumentation applications. Under the severe conditions, the conventional criteria of (4)-(6) have several local minima, which lead to poor performance of model structure determination, and ultimately a poor system model [3]. How to extract, compound and utilize the extra different types of information indices from the available experimental data are the key challenging points in the data-driven approaches that have application potential.

In order to improve the performance under the severe experiment conditions, an output over-sampling scheme is applied to collecting the experimental data, and then more information on cyclo-stationarity can be detected from the over-sampled data even under the severe conditions [14]. In output over-sampling scheme, the input signal remains the same as in (1) or (2) whose sampling rate is $1/T$, while the output is sampled P times faster than the input signal. Let the output sampling interval T/P be denoted as Δ, correspondingly, the input and output signals at $k\Delta$ are denoted as $u_\Delta(k)$ and $y_\Delta(k)$, respectively, and the system model in (1) can also be described with respect to interval Δ

$$y_\Delta(k) = G^{\mathrm{o}}_\Delta\left(q^{-1}\right) u_\Delta(k) + H^{\mathrm{o}}_\Delta\left(q^{-1}\right) w_\Delta(k), \tag{7}$$

for the stable model, where q^{-1} indicates the backward shift operator corresponding to the interval Δ, $H^{\mathrm{o}}_\Delta\left(q^{-1}\right)$ is a monic rational transfer function, while for an unstable process, the model is written as

$$A^{\mathrm{o}}_{\mathrm{us},\Delta}\left(q^{-1}\right) y_\Delta(k) = G^{\mathrm{o}}_{\mathrm{s},\Delta}\left(q^{-1}\right) u_\Delta(k) + H^{\mathrm{o}}_\Delta\left(q^{-1}\right) w_\Delta(k). \tag{8}$$

It has been demonstrated in [9] that the model of (1) and (2) can be uniquely determined by (7) and (8), and then, it implies that the model structure can be determined with respect to (7) or (8), rather than (1) or (2) directly if the over-sampled data $\{u_\Delta(k), y_\Delta(k)\}$ are used. Moreover, it is revealed that $\{u_\Delta(k), y_\Delta(k)\}$ have more information than $\{u(m), y(m)\}$ under the same experiment conditions except the output sampling rate [14]. Next, we will theoretically show that by detecting and compounding the cyclo-stationarity of over-sampled data, which improves the performance to deal with the ill-conditioned problems caused by severe experiment conditions, the proposed approach determines the model structure and further constructs system model to meet the practical requirement for application potential.

3. **Cyclo-Stationarity Property Analysis and Estimation.** Assume that the noise terms are the samples of a stationary continuous time stochastic process, hence both $w_\Delta(k)$ and $w(m)$ are stationary signals. Notice that input signal is a pseudo stationary stochastic signal at the interval T, and $u_\Delta(mP) = u_\Delta(mP+1) = \cdots = u_\Delta((m+1)P-1) = u(m)$ in the output over-sampling scheme.

3.1. **Definition of cyclo-stationary relation functions.** Let the correlation function $\mathcal{R}_{x_1,x_2}(k,\tau)$ of two signals $x_1(k)$ and $x_2(k)$ be defined by

$$\mathcal{R}_{x_1,x_2}(k,\tau) = E\left\{x_1(k+\tau)x_2(k)\right\}, \tag{9}$$

where $E\{\cdot\}$ indicates the expectation, and τ indicates the time shift. Then for the stationary signals, the value of correlation functions only depends on the time shift τ rather than the instant k. For example, if $w_\Delta(k)$ is stationary, for a given τ, the following equation

$$\cdots = \mathcal{R}_{w_\Delta,w_\Delta}(k-1,\tau) = \mathcal{R}_{w_\Delta,w_\Delta}(k,\tau) = \mathcal{R}_{w_\Delta,w_\Delta}(k+1,\tau) = \cdots \tag{10}$$

holds with respect to k, i.e., $\mathcal{R}_{x_1,x_2}(k,\tau)$ is a constant for an arbitrary k if both $x_1(k)$ and $x_2(k)$ are stationary. The same property also holds for the input $u(m)$ sampled at interval T when $u(m)$ is stationary. On the other hand, if the input data are recorded by sampling the signal at interval Δ, the value of correlation function $\mathcal{R}_{u_\Delta,u_\Delta}(k,\tau)$ depends not only on the time shift τ, but on the instant k. For example, $\mathcal{R}_{u_\Delta,u_\Delta}(mP,P-1) \neq \mathcal{R}_{u_\Delta,u_\Delta}(mP+1,P-1)$. Moreover, since $u_\Delta(mP) = \cdots = u_\Delta\left((m+1)P-1\right) = u(m)$, $\mathcal{R}_{u_\Delta,u_\Delta}(k,\tau)$ is a periodic function with respect to k, i.e.,

$$\mathcal{R}_{u_\Delta,u_\Delta}(k,\tau) = \mathcal{R}_{u_\Delta,u_\Delta}(k+P,\tau) \tag{11}$$

holds for arbitrary k and τ. For the periodic correlation function $\mathcal{R}_{x_1,x_2}(k,\tau)$ with period P, the following Fourier transform with respect to k at an angle α

$$\mathcal{C}_{x_1,x_2}(\alpha,\tau) = \lim_{K\to\infty}\frac{1}{K}\sum_{k=0}^{K-1}\mathcal{R}_{x_1,x_2}(k,\tau)e^{-i\alpha k},\ 0 \le \alpha < 2\pi \tag{12}$$

is called as cyclo-stationary relation function, which satisfies

$$\mathcal{C}_{x_1,x_2}(\alpha,\tau) = \begin{cases} \dfrac{1}{P}\displaystyle\sum_{k=mP}^{(m+1)P-1}\mathcal{R}_{x_1,x_2}(k,\tau)e^{-i\alpha_p k}, & \alpha = \alpha_p \\ 0, & \text{others} \end{cases} \tag{13}$$

where $\alpha_p \in \mathcal{A}_P$, $\mathcal{A}_P = \left\{\alpha_p | \alpha_p = p\frac{2\pi}{P}, p = 0,1,\ldots,P-1\right\}$. It means that the cyclo-stationary relation function is a periodic function that has its harmonic terms at $\alpha_0, \alpha_1, \ldots, \alpha_{P-1}$. In the output over-sampling scheme, the relation functions associated with $u_\Delta(k)$ or $y_\Delta(k)$ are periodic. Compared with the stationary signals, the information of $\mathcal{C}_{x_1,x_2}(\alpha,\tau)$ at $\alpha = \alpha_p$ can be utilized to deal with the ill-conditioned problem [12, 14].

Moreover, the Fourier transform of (13) with respect to the time shift τ yields the cyclo-stationary spectral density function

$$\mathcal{S}_{x_1,x_2}(\alpha,\omega) = \sum_{\tau=-N/2+1}^{N/2}\mathcal{C}_{x_1,x_2}(\alpha,\tau)e^{-j\omega\tau}, \tag{14}$$

where N is the number of frequency grids, then the spectra of input, output and the noise term $\mathcal{S}_{u_\Delta,u_\Delta}(\alpha,\omega)$, $\mathcal{S}_{u_\Delta,y_\Delta}(\alpha,\omega)$, $\mathcal{S}_{u_\Delta,w_\Delta}(\alpha,\omega)$ are cyclo-stationary, and satisfy the following equation:

$$\mathcal{S}_{u_\Delta,y_\Delta}(\alpha,\omega) = G^\circ_\Delta\left(e^{-i(\alpha-\omega)}\right)\mathcal{S}_{u_\Delta,u_\Delta}(\alpha,\omega) + H^\circ_\Delta\left(e^{-i(\alpha-\omega)}\right)\mathcal{S}_{u_\Delta,w_\Delta}(\alpha,\omega). \tag{15}$$

Compared with the stationary relation functions, the cyclo-stationary correlation functions or spectra are quite different at angles $\alpha_1,\ldots,\alpha_{P-1}$, where the stationary relation functions and spectra are 0. Then, it is easy to distinguish the cyclo-stationarity by checking (12) or (14) rather than to evaluate the magnitude of stationary correlation functions or spectra in the conventional methods, especially under the severe numerical conditions.

3.2. **Estimation of cyclo-stationary relation functions and spectra.** It can be seen that if the cyclo-stationary spectral density function is estimated by their definitions, the estimation will be time consuming since the correlation functions $\mathcal{C}_{x_1,x_2}(k,\tau)$ should be estimated with respect to both the possible k and τ. Here a fast estimation algorithm presented in [16] is used in this work.

It is seen that the time average periodograms can be used for the relation analysis of the stationary signals. Then, by separating the data record of cyclo-stationary signals into several windows, the cyclo-stationary relation functions and spectra are estimated for determination of model structure.

Assume that the experimental data of the cyclo-stationary signals $x_1(k)$ and $x_2(k)$ are recorded, where the period of their correlation functions is P. Let $x_{1,p}^{(m_P)}(m) = x_1(K_{m_P} + mP+p)$, $x_{2,p}^{(m_P)}(m) = x_2(K_{m_P}+mP+p)$, where K_{m_P} is the start of the m_Pth data window, $p = 0, 1, \ldots, P-1$. Then, $x_{1,p}^{(m_P)}(m)$ and $x_{2,p}^{(m_P)}(m)$ can be treated as stationary signal respectively; hence, Fourier transform can be applied for the estimation of the correlation functions as follows:

$$\begin{aligned}\mathcal{X}_{p_1,p_2}^{(m_P)}(\tau_P) &= E\left\{x_{1,p_1}^{(m_P)}(m+\tau_P)x_{2,p_2}^{(m_P)}(m)\right\} \\ &\approx \frac{1}{M}\sum_{l=-M/2+1}^{M/2}\frac{1}{M}X_{1,p_1}^{(m_P)}\left(e^{-i\omega_l}\right)\left(X_{2,p_2}^{(m_P)}\left(e^{-i\omega_l}\right)\right)^* e^{i\omega_l\tau_P},\end{aligned} \tag{16}$$

where $X_{1,p_1}^{(m_P)}(e^{-i\omega_l})$ and $X_{2,p_2}^{(m_P)}(e^{-i\omega_l})$ are the Fourier transform in the m_Pth data window, M is the data window length, and fast Fourier transform (FFT) can be executed for Fourier transform if M is an integer of power of 2. Following the definitions of $x_{1,p}^{(m_P)}(m)$ and $x_{2,p}^{(m_P)}(m)$, the following equation of $\mathcal{X}_{p_1,p_2}^{(m_P)}(\tau_P)$ holds for $x_1(k)$ and $x_2(k)$

$$\begin{aligned}\mathcal{X}_{p_1,p_2}^{(m_P)}(\tau_P) &= E\left\{x_1\left((m_P+m+\tau_P)P+p_1\right)x_2\left((m_P+m)P+p_2\right)\right\} \\ &= \mathcal{R}_{x_1,x_2}((m_P+m)P+p_2, \tau_P P+p_1-p_2).\end{aligned} \tag{17}$$

Let $\mathcal{X}_{p_1,p_2}^{(m_P)}(\tau_P)$ be arranged in such a vector that $\tau = \tau_P P + p_1 - p_2$ is fixed

$$\begin{bmatrix} \mathcal{X}_{p_1,p_2}^{(0)}(\tau_P) \\ \mathcal{X}_{p_1+1,p_2+1}^{(0)}(\tau_P) \\ \vdots \\ \mathcal{X}_{\bmod(p_1+m),\ \bmod(p_2+m)}^{\left(\left\lfloor\frac{\max(p_1,p_2)+m}{P}\right\rfloor\right)}(\tau_P) \\ \vdots \\ \mathcal{X}_{\bmod(p_1+M-1),\ \bmod(p_2+M-1)}^{\left(\left\lfloor\frac{\max(p_1,p_2)+M-1}{P}\right\rfloor\right)}(\tau_P) \end{bmatrix} = \begin{bmatrix} \mathcal{R}_{x_1,x_2}(p_2,\tau) \\ \mathcal{R}_{x_1,x_2}(p_2+1,\tau) \\ \vdots \\ \mathcal{R}_{x_1,x_2}(p_2+m,\tau) \\ \vdots \\ \mathcal{R}_{x_1,x_2}(p_2+M-1,\tau) \end{bmatrix}. \tag{18}$$

Then performing Fourier transform of the vector in (18) with respect to $p_2, p_2+1, \ldots$ yields the cyclo-stationary functions at the angle grids $0, \frac{2\pi}{M}, \ldots, \frac{2(M-1)\pi}{M}$. Furthermore, for a given α, arrange the cyclo-stationary relation functions with respect to $\tau = \ldots, -1, 0, 1, 2, \ldots$ in vectors, then their Fourier transform with respect to τ yields the cyclo-stationary spectra.

It is seen that the estimated cyclo-stationary relation functions or spectra can be used to detect whether $x_1(k)$ or $x_2(k)$ contains the cyclo-stationary components by comparing the terms at α_p with the others.

4. **Determination of Model Structure.** Let the one-step prediction of the system output be given by

$$\hat{y}_\Delta(k,\hat{\boldsymbol{n}}) = \frac{\hat{G}_\Delta\big(q^{-1},\hat{\boldsymbol{n}}\big)}{\hat{H}_\Delta\big(q^{-1},\hat{\boldsymbol{n}}\big)}u_\Delta(k) + \frac{\hat{H}_\Delta\big(q^{-1},\hat{\boldsymbol{n}}\big)-1}{\hat{H}_\Delta\big(q^{-1},\hat{\boldsymbol{n}}\big)}y_\Delta(k), \tag{19}$$

and the model residuals be

$$\varepsilon_\Delta(k,\hat{\boldsymbol{n}}) \;=\; y_\Delta(k) - \hat{y}_\Delta(k,\hat{\boldsymbol{n}}) \tag{20}$$

in the output over-sampling scheme. It is deduced that if the model structure is suitable for the system model, the model residual will be close to the noise term. On the other hand, it is clear that if the model structure of $\hat{G}_\Delta\big(q^{-1},\hat{\boldsymbol{n}}\big)$ and $\hat{H}_\Delta\big(q^{-1},\hat{\boldsymbol{n}}\big)$ does not sufficiently approximate the true system characteristics, the terms associated with $u_\Delta(k)$ remain in the residuals, and lead to the cyclo-stationarity of $\varepsilon_\Delta(k,\hat{\boldsymbol{n}})$. Consequently, model structure will be determined through detecting the cyclo-stationarity of the residuals.

The cyclo-stationary relation function $\mathcal{C}_{\varepsilon_\Delta,\varepsilon_\Delta}(\alpha,\tau)$ and spectrum $\mathcal{S}_{\varepsilon_\Delta,\varepsilon_\Delta}(\alpha,\omega)$ are estimated by the approach given in Section 3.2, and the cyclo-stationarity is detected by checking whether the magnitude of the cyclo-stationary relation functions and spectra at the angles $\alpha \in \mathcal{A}_{P_1} = \{\alpha_1,\ldots,\alpha_{P-1}\}$ are apparently different from those at $\alpha \in \mathcal{A}_{\overline{P}} = \{\alpha | \alpha \neq \alpha_0,\ldots,\alpha_{P-1}\}$. Compared with evaluating the magnitude of the mean square residual, the cyclo-stationarity based approaches are easier to distinguish the cyclo-stationarity from the stationary ones.

4.1. **Indices of cyclo-stationarity.** The information indices are detected from the over-sampled data, and they will be used in determination of the model structure.

4.1.1. *Mean square residual.* The conventional determination of model structure uses the mean square σ_ε^2 of the residual $\varepsilon_\Delta(k,\hat{\boldsymbol{n}})$

$$\sigma_\varepsilon^2 = \frac{1}{K}\sum_{k=1}^{K}\varepsilon_\Delta^2(k,\hat{\boldsymbol{n}}). \tag{21}$$

It is treated as an approximated indicator of the model accuracy in the conventional criteria. Nevertheless, due to the insufficient excitation or the strong noise terms, (21) has local minima that largely degrade the performance of both the determination of model structure and estimation of model parameters; hence, the cyclo-stationarity is introduced into the proposed approach to complement the necessary information.

4.1.2. *Magnitude of cyclo-stationary components.* If the residuals have cyclo-stationary components, the magnitude of both $\mathcal{C}_{\varepsilon_\Delta,\varepsilon_\Delta}(\alpha,\tau)$ and $\mathcal{S}_{\varepsilon_\Delta,\varepsilon_\Delta}(\alpha,\omega)$ at the angles $\alpha \in \mathcal{A}_{P_1}$ is quite larger than that at $\alpha \in \mathcal{A}_{\overline{P}}$. Let the maximal magnitude of $\mathcal{C}_{\varepsilon_\Delta,\varepsilon_\Delta}(\alpha,\tau)$ at all the angles $\alpha \in \mathcal{A}_{P_1}$ be denoted as $\mathcal{C}_{\max}$, and the maximal magnitude of $\mathcal{S}_{\varepsilon_\Delta,\varepsilon_\Delta}(\alpha,\omega)$ as $\mathcal{S}_{\max}$, while the mean magnitude $\mathcal{C}_{\mathrm{mean}}$ and $\mathcal{S}_{\mathrm{mean}}$ are defined by

$$\begin{aligned}\mathcal{C}_{\mathrm{mean}} &= \frac{1}{N(P-1)}\sum_{p=1}^{P-1}\sum_{\tau=-N/2+1}^{N/2}|\mathcal{C}_{\varepsilon_\Delta,\varepsilon_\Delta}(\alpha_p,\tau)|,\\ \mathcal{S}_{\mathrm{mean}} &= \frac{1}{N(P-1)}\sum_{p=1}^{P-1}\sum_{\omega=-\pi}^{\pi}|\mathcal{S}_{\varepsilon_\Delta,\varepsilon_\Delta}(\alpha_p,\omega)|.\end{aligned} \tag{22}$$

The maximal magnitude and mean magnitude indicate how strong cyclo-stationary components contained at the angles of $\mathcal{A}_{\overline{P}}$.

Furthermore, let the ratio of the cyclo-stationary relation function and spectrum at $\alpha \in \mathcal{A}_{P_1}$ to that at $\alpha \in \mathcal{A}_{\overline{P}}$ as

$$\mathcal{C}_{\text{rat}} = \frac{\frac{1}{P-1}\sum_{p=1}^{P-1}\sum_{\tau=-N/2+1}^{N/2}\left|\mathcal{C}_{\varepsilon_\Delta,\varepsilon_\Delta}(\alpha_p,\tau)\right|}{\frac{1}{M-P}\sum_{\alpha_{\overline{p}}\in\mathcal{A}_{\overline{P}}}\sum_{\tau=-N/2+1}^{N/2}\left|\mathcal{C}_{\varepsilon_\Delta,\varepsilon_\Delta}(\alpha_{\overline{p}},\tau)\right|},\ \mathcal{S}_{\text{rat}} = \frac{\frac{1}{P-1}\sum_{p=1}^{P-1}\sum_{\omega=-\pi}^{\pi}\left|\mathcal{S}_{\varepsilon_\Delta,\varepsilon_\Delta}(\alpha_p,\omega)\right|}{\frac{1}{M-P}\sum_{\alpha_{\overline{p}}\in\mathcal{A}_{\overline{P}}}\sum_{\omega=-\pi}^{\pi}\left|\mathcal{S}_{\varepsilon_\Delta,\varepsilon_\Delta}(\alpha_{\overline{p}},\omega)\right|}. \quad (23)$$

As shown in Section 3, if the residuals have cyclo-stationary components, the cyclo-stationary relation functions and spectra at $\alpha \in \mathcal{A}_P$ are apparently different from those at $\alpha \in \mathcal{A}_{\overline{P}}$. Then, $\mathcal{C}_{\text{rat}}$ and $\mathcal{S}_{\text{rat}}$ indicate whether the cyclo-stationary components are contained in the residuals or not.

4.1.3. *Distribution of cyclo-stationary components.* When the noise signals are treated as stochastic ones, the estimation errors at $\alpha \in \mathcal{A}_{\overline{P}}$ between theoretical values and their estimates obtained from the finite data number are approximated by normally distributed random numbers with zero mean. The same property also holds for $\alpha \in \mathcal{A}_{P_1}$ when the model structure can be used to describe the dominant characteristics of the true system; otherwise, the model error is subject to more complicated distribution. Therefore, the distribution information, which helps the criterion detect the cyclo-stationary components, can also be used as one of the indices for model structure determination. For the simplicity of computation, here the variance ratio $\mathcal{C}_{\text{cov}}$ of the magnitude of relation functions, and $\mathcal{S}_{\text{cov}}$ of spectra are used in the indices as follows:

$$\mathcal{C}_{\text{cov}} = \frac{\frac{1}{P-1}\sum_{p=1}^{P-1}\text{cov}(\mathcal{C}_{\varepsilon_\Delta,\varepsilon_\Delta}(\alpha_p,\tau))}{\frac{1}{M-P}\sum_{\alpha_{\overline{p}}\in\mathcal{A}_{\overline{P}}}\text{cov}(\mathcal{C}_{\varepsilon_\Delta,\varepsilon_\Delta}(\alpha_{\overline{p}},\tau))},\ \mathcal{S}_{\text{cov}} = \frac{\frac{1}{P-1}\sum_{p=1}^{P-1}\text{cov}(\mathcal{S}_{\varepsilon_\Delta,\varepsilon_\Delta}(\alpha_p,\omega))}{\frac{1}{M-P}\sum_{\alpha_{\overline{p}}\in\mathcal{A}_{\overline{P}}}\text{cov}(\mathcal{S}_{\varepsilon_\Delta,\varepsilon_\Delta}(\alpha_{\overline{p}},\omega))}, \quad (24)$$

where $\text{cov}(\cdot)$ indicates the variance function.

4.2. **Criterion approximated by neural network.** The conventional methods that use a single type of information index are largely affected by the experiment conditions such as the data quality and noise terms; hence, the information indices are compounded in the new criterion to deal with the severe experiment conditions. On the other hand, the simple linear combination cannot obtain satisfactory result since the relation between the different indices has nonlinear characteristics. Consequently, a neural network is used to compound the information indices detected from the cyclo-stationary relation functions and spectra to yield a nonlinear criterion

$$V(\hat{\boldsymbol{n}}) = f\left(\hat{\boldsymbol{n}}, \sigma_\varepsilon^2, \mathcal{C}_{\text{max}}, \mathcal{S}_{\text{max}}, \mathcal{C}_{\text{mean}}, \mathcal{S}_{\text{mean}}, \mathcal{C}_{\text{rat}}, \mathcal{S}_{\text{rat}}, \mathcal{C}_{\text{cov}}, \mathcal{S}_{\text{cov}}\right), \quad (25)$$

for model structure determination, where the function $f(\cdot)$ is nonlinear. Besides the mean square residual σ_ε^2 that is used in the conventional criteria to evaluate the model accuracy, the compounded information indices $\mathcal{C}_{\text{max}}$, $\mathcal{C}_{\text{mean}}$, $\mathcal{S}_{\text{max}}$, $\mathcal{S}_{\text{mean}}$ also include the maximal, mean magnitude of the cyclo-stationary relation functions and spectra, which indicate how strong cyclo-stationary components remained in the signals, the magnitude ratio $\mathcal{C}_{\text{rat}}$, $\mathcal{S}_{\text{rat}}$ at the angles $\alpha \in \mathcal{A}_{P_1}$ to that at $\alpha \in \mathcal{A}_{\overline{P}}$, and the distribution indices $\mathcal{C}_{\text{cov}}$, $\mathcal{S}_{\text{cov}}$. Compared with the conventional methods, the cyclo-stationarity helps criterion detect the residual components, and the compounded information indices improve the performance under severe experiment conditions.

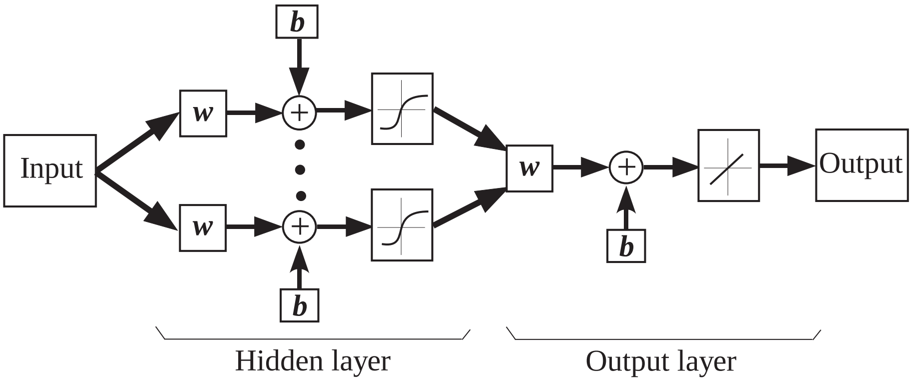

FIGURE 1. Illustration of neural network to approximate $V(\hat{\boldsymbol{n}})$

The illustration of neural network is shown in Figure 1, where $\boldsymbol{w}$ is the weight coefficient vector, and $\boldsymbol{b}$ is the bias vector. The coefficients are determined by the training data of the estimated information indices and the nominal model.

5. **Numerical Example.** An experiment of the magnetic levitation process operated in closed-loop is considered for the proposed approach. A steel ball is kept at a reference position by the electromagnetic force that is regulated by a digital PID controller with a control interval $T = 0.0024$s; correspondingly, the process of magnetic levitation is approximated by a discrete-time transfer function model. The data are sampled at the interval $\Delta = 0.0008$s, which is 3 times shorter than T.

The nominal theoretical poles of $G_{\Delta}^{o}\left(q^{-1}\right)$ are 1.0282, 0.9906, 0.9718, and the nominal steady gain is 0.69×10^{-8}, which is so small that the numerical problems easily occur in the numerical computation [14]. Moreover, several closed-loop poles are close to the unit circle, and the small steady gain also makes the closed-loop sensitivity to the external exciting signals be very low in the high frequency band, whereas the reference is a constant which does offer less excitations for model structure determination; hence, the numerical conditions are very poor for the conventional methods. On the other hand, the levitation process is disturbed by the interferences such as measurement noise, the disturbance of the ball's rotation, and the air floating. They are assumed to be compounded as a stationary stochastic process, whose structure of noise model is unknown and required to be determined from the experimental data.

Due to the extremely ill-conditioned numerical conditions, the conventional methods are hard to obtain satisfactory results. In the proposed approach, the experimental data are recorded through output over-sampling, where the sampling rate is $P = 3$. The experimental data are collected for 100 seconds in one experiment. In the experiment, the model structure of the noise model is determined to make the output $V(\hat{\boldsymbol{n}})$ of neural network be close to $V_{\text{th}} = \left|e^{-\sigma_G^2} - 1\right|$, where σ_G^2 is the variance of parameter error between the estimates and the nominal values. The better model structure is, the closer V_{th} is to 0. In the neural network, the hidden layer size is 18, and the coefficients are trained by scaled conjugate gradient algorithm. Figure 2 shows the output of the trained neural network where the noise orders (n_c, n_d) are chosen form $(1, 3) \sim (4, 8)$, and Table 1 illustrates some examples of the indices vs. (n_c, n_d). The model orders are determined as $(n_c, n_d) = (2, 6)$, which can describe the system dynamics well in the experiment. From Figure 2 and Table 1, it is seen that the criterion has complicated nonlinearity, and severe experiment conditions lead to that any information index does not dominate the criterion characteristics. Consequently, compounding information indices is necessary. It

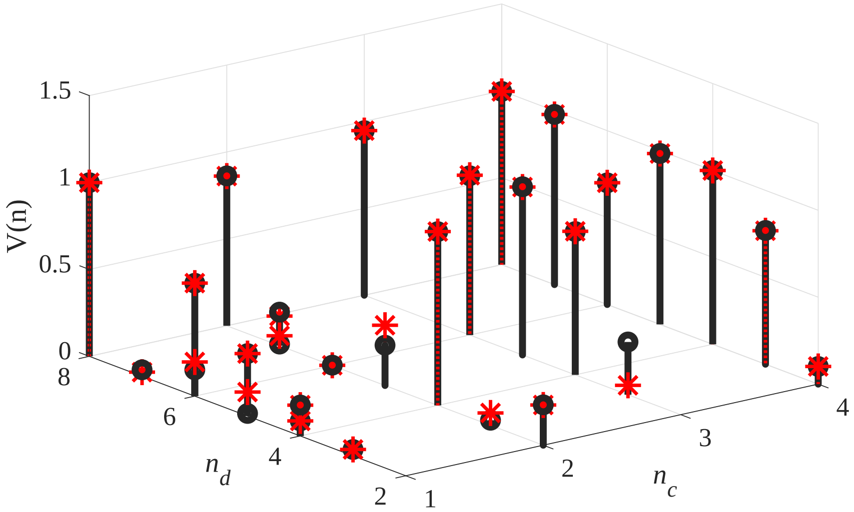

FIGURE 2. Values of V_{th} vs. output of neural network $V(\hat{\boldsymbol{n}})$. $*$: theoretical value; $\circ$: output of neural network.

TABLE 1. Examples of cyclo-stationary indices vs. model orders

n_c, n_d	σ_ε^2	$\mathcal{C}_{\max}$	$\mathcal{S}_{\max}$	$\mathcal{C}_{\text{mean}}$	$\mathcal{S}_{\text{mean}}$	$\mathcal{C}_{\text{cov}}$	$\mathcal{S}_{\text{cov}}$	$V(\hat{\boldsymbol{n}})$
1, 3	2.9772	5.7082	24.116	0.9657	0.9650	0.9550	0.9546	0.0365
1, 4	2.9981	5.9448	13.293	1.0893	1.1779	2.4543	1.7331	0.1780
1, 5	3.1020	5.8914	24.752	0.9646	0.9735	0.9777	0.9205	0.3593
1, 6	2.9446	5.2190	22.599	0.9859	0.9936	0.9473	0.9029	0.1512
1, 7	2.9769	5.4502	23.987	101.02	0.9771	0.9730	0.9562	0.0371
1, 8	5.7394	23.620	15.339	0.9818	0.9535	1.5319	1.4157	0.9989
2, 2	2.8636	6.4518	27.336	1.1496	1.1145	1.4485	1.4406	0.2322
2, 3	2.8316	5.7172	25.707	1.0871	1.1000	1.2845	1.1868	0.0292
2, 4	12.375	30.085	66.241	1.1352	1.3203	3.8145	2.0194	0.9996
2, 5	2.7721	3.9699	16.058	1.1110	1.1108	1.2799	1.2819	0.2316
2, 6	2.8287	5.6156	22.495	1.0859	1.0912	1.2589	1.2146	0.0026
2, 7	2.7537	3.8292	16.390	1.1041	1.0865	1.1727	1.3095	0.1932
2, 8	3.1726	8.0995	37.982	0.7763	0.7291	0.5202	0.6892	0.8622

is shown that the proposed approach can determine the appropriate orders even under severe experiment conditions, so it has application potential for practical systems.

As a comparison, the possible model orders determined by the conventional method such as AIC under the same experiment conditions are $(n_c, n_d) = (2, 7)$ or $(2, 5)$. Though their values of σ_ε^2 are considerably small, with these model structures the estimated models' steady gain has estimation error; hence, the residuals contain a little of cyclo-stationary components. It implies that the conventional methods should introduce additional information to reduce the local minima's influences on information criterion under the severe experiment conditions.

6. **Conclusions.** The approach to determination of model structure under the severe numerical conditions is investigated in this paper. It is illustrated that conventional methods are affected by the severe numerical conditions due to their criteria depending

on the mean square residual. By detecting the cyclo-stationary indices from the experimental data collected in the output over-sampling scheme, and compounding them into a neural network based criterion, the determination performance has been improved, and can provide an effective model structure for system identification and system design. The improvement of numerical stability of neural network especially for the case of small training sample number and the evaluation of model error by using the cyclo-stationarity will be investigated in the future work.

REFERENCES

[1] L. Ljung, *System Identification – Theory for the User*, Prentice Hall, Englewood Cliffs, NJ, 1999.
[2] H. Akaike, Model selection: AIC, BIC and MDL, *Communications of the Operations Research Society of Japan*, vol.41, no.7, pp.375-378, 1996.
[3] S. Misra and M. Nikolaou, Adaptive design of experiments for model order estimation in subspace identification, *Computers and Chemical Engineering*, vol.100, pp.119-138, 2017.
[4] T. Söderström, *The Errors-in-Variables Problem for Dynamic Systems*, Springer, 2018.
[5] W. A. Gardner, *Cyclostationarity in Communications and Signal Processing*, IEEE Press, 1994.
[6] C. Tontiruttananon and J. K. Tugnait, Parametric identification of closed-loop linear systems using cyclic-spectral analysis, *American Control Conference 1998*, Philadelphia, USA, pp.3597-3601, 1998.
[7] L. Tong, G. Xu and T. Kailath, Blind identification and equalization based on second-order statistics: A time domain approach, *IEEE Trans. Information Theory*, vol.40, no.2, pp.340-349, 1994.
[8] E. Moulines, P. Duhamel, J. Cardoso and S. Mayrargue, Subspace methods for the blind identification of multi-channel FIR filters, *IEEE Trans. Signal Processing*, vol.43, pp.516-526, 1995.
[9] L. Sun, H. Ohmori and A. Sano, Output inter-sampling approach to closed-loop identification, *IEEE Trans. Automatic Control*, vol.46, no.12, pp.1936-1941, 2001.
[10] X. Liu, L. Sun, J. Meng and H. Lin, Intelligent directional system based on microphone array, *ICIC Express Letters*, vol.11, no.2, pp.475-481, 2017.
[11] J. Du, H. Yu, D. Ang, D. Zhang and G. Liu, Unified subspace fitting framework and its performance analysis for direct position determination in presence of multipath propagation, *IEEE Access*, vol.7, pp.6889-6914, 2019.
[12] L. Sun and A. Sano, Temporal-spatial information based approach to direct closed-loop identification, *Trans. of the Society of Instrument and Control Engineers*, vol.53, pp.346-354, 2017.
[13] M. Fang and Y. Zhu, Asymptotic variance expression in output over-sampling based closed-loop identification, *Proc. of the 17th IFAC Symposium on System Identification*, Beijing, pp.110-115, 2015.
[14] L. Sun, X. Liu and A. Sano, Direct closed-loop identification approach to magnetic levitation system, *Proc. of the 18th IFAC Symposium on System Identification*, Stockholm, Sweden, pp.610-615, 2018.
[15] L. Sun, C. Zhou and A. Sano, New prediction error method based on cyclostationarity for system identification, *ICIC Express Letters*, vol.11, no.1, pp.189-196, 2017.
[16] L. Sun, X. Liu and A. Sano, Model structure identification and parameter estimation for unstable process in closed-loop, *ICIC Express Letters*, vol.13, no.7, pp.625-633, 2019.