International Journal of Advanced Engineering Research and Science (IJAERS)
Peer-Reviewed Journal
ISSN: 2349-6495(P) | 2456-1908(O)
Vol-11, Issue-11; Nov, 2024
Journal Home Page Available: https://ijaers.com/
Article DOI: https://dx.doi.org/10.22161/ijaers.1111.6

WSIC Criterion for Decomposition Level Selection of Orthogonal Wavelet Transform

Yimei Zheng[1], Cheng Liu[*2], Yan Fang[3]

[1,2]School of Mathematics, Southwest Jiaotong University, Chengdu 611756, China
1554047784@qq.com
[3]School of Electrical Engineering, Southwest Jiaotong University, Chengdu 611756, China
*Corresponding author

Received: 10 Oct 2024,

Receive in revised form: 12 Nov 2024,

Accepted: 18 Nov 2024,

Available online: 24 Nov 2024

***Keywords*— Wavelet Transform, Decomposition Level, Model Selection, WSIC**

Abstract*— *Wavelet transform is a widely used method in the field of signal analysis, and its application effect depends largely on the selection of wavelet decomposition level. In view of the unknown original signal, this paper transforms the selection of wavelet decomposition level into a model selection problem through statistical modeling, and then analyzes and presents wavelet selection information criteria (WSIC) of orthogonal wavelet transform level selection from the perspective of model selection. Finally, simulation experiments are conducted to verify and compare the effect of the WSIC information criteria with AIC and BIC criteria on wavelet level selection. The results show that the level accuracy of WSIC is up to 15.1% higher than that of AIC criterion and up to 14.3% higher than that of BIC criterion, indicating that WSIC criterion has better stability in selecting wavelet decomposition level than that of AIC and BIC criterion in existing literatures.

I. INTRODUCTION

Wavelet transform is widely used in the field of signal analysis. It is a linear transform with wavelet function as its kernel. In the practical application of orthogonal wavelet transform, the signal containing noise is decomposed into different frequency components according to Mallat algorithm. In the signal reconstruction, some high frequency detail coefficients related to noise are set to zero, and the reconstructed signal after noise removal is obtained. The larger the level of wavelet decomposition, the more beneficial to noise removal, but will also lose more detailed signals, which may cause signal distortion. The smaller the level of wavelet decomposition, the noise can't be well eliminated. Therefore, in the process of using wavelet transform, choosing a suitable decomposition level is the key to improve the effect of wavelet denoising.

Under the condition that the original signal is known, the signal after wavelet transformation is usually compared with the original signal, and the optimal wavelet decomposition level is selected according to various evaluation indicators [1,2,3,4]. Under the condition that the original signal is unknown, only the observed signal containing noise can't be compared with the original signal after wavelet transform. Literature [5] uses blind estimation of SNR to analyze the changes of SNR and SNR gain by improving the index of SNR and SNR gain, and then evaluates the denoising effect of wavelet, thus solving the problem of selecting the optimal decomposition levels of wavelet transform when the original signal is unknown. Literature [6,7,8,9] takes energy and information entropy as evaluation indexes to determine the wavelet decomposition level, and uses the distribution characteristics of signals in different frequency bands to estimate the effectiveness of the wavelet, essentially selecting the wavelet that is most similar to the signal of interest. However, the selection result of this method is easily affected by the noise distribution, and it is more suitable for signal diagnosis and recognition. When denoising deformation detection signals, literature [10] proposed that the wavelet decomposition level should be selected based on the minimum information criterion. Literature [11] and literature [12] used traditional information criteria such as AIC and BIC when selecting wavelet decomposition level. When analyzing large-level non-stationary data in civil and mechanical engineering applications, literature [13] proposes a new framework that comprehensively considers multiple indexes for selecting the most suitable wavelet base and decomposition level, and systematically applies existing evaluation indexes for selecting wavelet base and decomposition level. However, because of its systematicness, the process of this method is relatively complicated in actual operation, and it needs to calculate and analyze the signal in many aspects. Therefore, it is necessary to find a simple, reasonable and efficient method for the level selection of wavelet transform.

Most of the existing methods directly select some indicators for the selection of wavelet transform decomposition level, but this paper starts with the unknown "uncertainty" of the original signal and chooses to analyze this problem from the perspective of statistics. Firstly, the wavelet transform problem of the signal is modeled statistically, and the choice of optimal decomposition level is transformed into a model selection problem. Then, the Wavelet Selection Information Criterion (WSIC) is derived by reasonably balancing the goodness of fit and complexity of the model. Finally, it is verified by simulation experiments that WSIC criterion is more effective in selecting orthogonal wavelet transform models than AIC and BIC criteria used in literature [11]. The research in this paper provides the method and theory support for determining the optimal decomposition level of wavelet transform in practical application.

The rest of this paper is organized as follows. Some necessary preliminaries are introduced in Section 2. In Section 3, the orthogonal wavelet transform model selection criteria is introduced in detail. Simulation experiment is provided in Section 4 to verify the effectiveness of the proposed WSIC criterion for the selection of wavelet decomposition level. Finally, conclusions are drawn in Section 5.

II. PRELIMINARIES

2.1 Orthogonal Wavelet Transform Theory

The implementation of orthogonal wavelet transform is based on multi-resolution analysis (also known as multi-scale analysis). According to the finite precision multi-scale analysis and approximation [14], the signal $x(t)$ can be expressed as

$$x(t)=\sum_{k\in Z}c_{J,k}\phi_{J,k}(t)+\sum_{j=1}^{J}\sum_{k\in Z}d_{j,k}\psi_{j,k}(t) \tag{1}$$

where $\phi_{J,k}(t)$ and $\psi_{j,k}(t)$ are the functions obtained by the scale function $\phi(t)$ and the wavelet function $\psi(t)$ after translation and expansion transformation, respectively, and they are the basis functions of the scale space and the wavelet space under the corresponding scale.

The scale coefficient $c_{J,k}$ and the wavelet coefficient $d_{j,k}$ are the inner products of the signal and the scale basis function $\phi_{J,k}(t)$ and the wavelet basis function $\psi_{j,k}(t)$, respectively:

$$c_{J,k}=<x(t),\phi_{J,k}(t)> \tag{2}$$

$$d_{j,k}=<x(t),\psi_{j,k}(t)>$$

There is a step by step derivation relationship between the scale coefficient and the wavelet coefficient, and the calculation of $\{c_{j,k}\}_{j,k\in\mathbf{Z}}$ and $\{d_{j,k}\}_{j,k\in\mathbf{Z}}$ has the following transfer relationship for j:

$$c_{j,k}=\sum_{n\in Z}h(n-2k)c_{j-1,n} \tag{3}$$

$$d_{j,k}=\sum_{n\in Z}g(n-2k)c_{j-1,n} \tag{4}$$

where h and g are low-pass and high-pass filters of wavelet respectively. Equations (3) and (4) are the basis of Mallat algorithm. As can be seen from the two equations, as long as $\{h(n)\}_{n\in Z}$ and $\{g(n)\}_{n\in Z}$ are known, the scale coefficient and wavelet coefficient under each level j can be calculated by observing the $\{c_{0,n}\}_{n\in Z}$ obtained by the signal.

2.2 Statistical Model of Orthogonal Wavelet Transform

It is assumed that the observed signal $y(t)$ of length n consists of the original signal $f(t)$ and Gaussian white noise interference $\varepsilon(t)$:

$$y(t)=f(t)+\varepsilon(t), t=1,\cdots,n \tag{5}$$

where $\varepsilon(t)$ meets the conditions:

$$\varepsilon(t) \overset{i.i.d}{\sim} N(0,\sigma^2), t=1,\cdots,n$$

After the discrete wavelet transform, the signal can be expressed in the form of equation (1), assuming that the original signal is composed of the approximate signal obtained by the wavelet transform, that is,

$$f(t)=\sum_{k\in Z} c_{J,k}\phi_{J,k}(t)$$

According to the above assumptions, the statistical model of orthogonal wavelet transform can be given:

$$\begin{cases} y(t)=\sum_{k\in Z} c_{J,k}\phi_{J,k}(t)+\varepsilon(t), \\ \varepsilon(t) \overset{i.i.d}{\sim} N(0,\sigma^2), t=1,\cdots,n. \end{cases} \quad (6)$$

In the decomposition of level J, the signal $y(t)\sim N(\mu_J(t),\sigma^2)$, where

$$\mu_J(t)=E\big(y(t)\big)=E\left(\sum_{k\in Z} c_{J,k}\phi_{J,k}(t)+\varepsilon(t)\right)=\sum_{k\in Z} c_{J,k}\phi_{J,k}(t)$$

So $y(t)\sim N(\sum_{k\in Z} c_{J,k}\phi_{J,k}(t),\sigma^2)$. The density function of this distribution is

$$p(y\,|\,c_{J,k},\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}\Big(y-\sum_{k\in Z} c_{J,k}\phi_{J,k}(t)\Big)^2\right\} \quad (7)$$

2.3 Monotone Property

The relevant properties of the orthogonal wavelet transform statistical model will be discussed in the following, which will serve as the basis for subsequent research.

Property 1 *The number of scale function* $\phi_{J,k}(t)$ *of the model in* J *level is* $q(J)$, *and for* $\forall J_1<J_2$, *there is* $q(J_1)>q(J_2)$.

Proof According to the uniform monotonicity of the closed subspace $\{V_j\}_{j\in \mathbf{Z}}$ in multi-resolution analysis, the scale space decreases with the increase of the decomposition level, and the number of basic functions of the scale space decreases with the decrease of the scale space. Because of the frequency domain decomposition of the signal, the sampling rate of each level is halved, and the length of the new sequence is reduced by half. Therefore, the number of scale function $q(J)$ decreases with the increase of decomposition level J .

Property 2 *The variance of the residual of the model on the* J *level is* $\hat{\sigma}^2(J)$, *and for* $\forall J_1<J_2$, *there is* $\hat{\sigma}^2(J_1)<\hat{\sigma}^2(J_2)$.

Proof According to equation (1), $\varepsilon(t)$ can be composed of a linear combination of wavelet functions, i.e.

$$\hat{\varepsilon}(t)=\sum_{j=1}^{J}\sum_{k\in Z} d_{j,k}\psi_{j,k}(t)=\sum_{j=1}^{J}\hat{\varepsilon}_j(t)$$

Since the wavelet function $\psi(t)$ is zero mean, let $\hat{\varepsilon}_j(t)\sim N(0,\hat{\sigma}_j^2)$, according to the mutually orthogonal property of the wavelet space at different levels, $\hat{\varepsilon}_j(t)$ can be independent of each other. From the additivity of the normal distribution, we know that

$$\hat{\varepsilon}(t)=\sum_{j=1}^{J}\hat{\varepsilon}_j(t)\sim N(0,\hat{\sigma}^2(J))$$

where $\hat{\sigma}^2(J)=\sum_{j=1}^{J}\hat{\sigma}_j^2$. It can be seen that the variance of model residuals increases with the increase of decomposition level.

III. ORTHOGONAL WAVELET TRANSFORM MODEL SELECTION CRITERIA

According to the principle of orthogonal wavelet transform, the trend information of the signal is mainly located in the scale space, and the detail information is mainly located in the wavelet space. If the scale space that completely contains the observation signal $y(t)$ is $\phi_{0,k}(t)$, that is, the space of scale 0, then the orthogonal wavelet transformation process of the signal is to decompose the space $\phi_{0,k}(t)$ again and again, and decompose the useful information such as the trend of the signal into the scale space, and decompose the interference information such as noise into the wavelet space. To determine the scale of wavelet decomposition is to determine the number of decomposition levels, so that the best decomposition effect can be achieved under the number of decomposition levels.

3.1 Measure of Model Goodness of Fit

Since the likelihood function is the most sensitive criterion for model parameters to deviate from the true value [15], the probability density function of the real model is $g(y)$, and it can be seen from equation (7) that the probability density function of the candidate submodel is $p(y\,|\,c_{J,k},\sigma^2)$. Suppose the parameter vector $\theta=(c_{J,k},\sigma^2)$, and consider the independent observed signal sequence $y_1,y_2,...,y_n$ of length n, then the mean logarithmic likelihood is

$$\frac{1}{n}\sum_{i=1}^{n}\ln p(y_i \mid \theta)$$

According to the strong law of numbers, we get

$$\frac{1}{n}\sum_{i=1}^{n}\ln p(y_i \mid \theta) \xrightarrow{a.s.} \int g(y)\ln p(y \mid \theta)dy = S(g; p(y \mid \theta)) \quad (8)$$

assuming that the integral exists.

The Kullback-Leibler (KL) distance is a measure of the degree of difference between two probability distributions. If the density function of the candidate model is $p(\cdot \mid \theta)$ and the density function of the real model is g, then the KL distance between the distribution of the real model and the distribution of the candidate model is defined as [16]

$$\mathrm{KL}(g, p(\cdot \mid \theta)) = \int g(y)\ln\frac{g(y)}{p(y \mid \theta)}dy \quad (9)$$

By observing equations (8) and (9), it can be found that there is the following relationship between mean logarithmic likelihood and KL distance

$$\mathrm{KL}(g, p(\cdot \mid \theta)) = S(g; g) - S(g; p(\cdot \mid \theta)) \quad (10)$$

The KL distance represented by equation (10) is always greater than or equal to 0, and the KL distance is equal to 0 if and only if $g(y) = p(y \mid \theta)$. This shows that $S(g; p(\cdot \mid \theta))$ is a reasonable measure of the goodness of fit of a model, and by maximizing $S(g; p(\cdot \mid \theta))$, or minimizing $-S(g; p(\cdot \mid \theta)$, candidate submodels that are closest to the real model can be found. If the maximum likelihood estimator of θ is θ_0, then θ_0 is the parameter that minimizes the KL distance.

3.2 Deviation Correction

In the model selection problem, there are usually multiple candidate submodels, that is, there are multiple $p(y \mid \theta)$, and the parameter vector θ of different submodels is different. At this time, the maximum likelihood principle cannot provide a useful solution for this kind of problem, and a solution can be obtained by combining the basic idea of statistics with the maximum likelihood principle.

Let's consider the case where KL is equal to 0, which is $g(y) = p(y \mid \theta_0)$. $\mathrm{KL}(g, p(\cdot \mid \theta))$ and $S(g; p(\cdot \mid \theta))$ are respectively $\mathrm{KL}(\theta_0, \theta)$ and $S(\theta_0; \theta)$, and when θ is close enough to θ_0, $\mathrm{KL}(\theta_0, \theta)$ is approximately

$$\mathrm{KL}(\theta_0; \theta_0 + \vartriangle\theta) = \frac{1}{2}\|\vartriangle\theta\|_{J^2}$$

where $\|\vartriangle\theta\|_{J^2} = \vartriangle\theta' J \vartriangle\theta$, J is the Fisher information matrix, which is positive definite and defined as

$$J_{ij} = E\left\{\frac{\partial^2 \ln p(y \mid \theta)}{\partial\theta_i \partial\theta_j}\right\}$$

where J_{ij} represents the (*i,j*)th element of J, θ_i and θ_j are the *i*th and *j*th elements of θ, respectively. Thus, when θ is very close to the maximum likelihood estimator θ_0, the change of the distribution defined by $p(y \mid \theta)$ with the true distribution $p(y \mid \theta_0)$ over $S(\theta_0; \theta)$ can be measured by $\frac{1}{2}\|\theta - \theta_0\|_{J^2}$. Consider the case where the variation of θ that maximizes the likelihood function is confined to a low-dimensional subspace Θ that does not contain θ_0. For the maximum likelihood estimator $\hat{\theta}$ of θ_0 limited to in Θ, if the maximum $S(\theta_0; \theta)$ of θ is close enough to θ_0, then for a sufficiently large n, the distribution of $n\|\hat{\theta} - \theta\|_{J^2}$ is approximately the Chi-square distribution of degrees of freedom equal to the dimension of the restricted parameter space under certain regularity conditions [15]. So we have

$$E(X)2n\mathrm{KL}(\theta_0; \hat{\theta}) = n\|\theta - \theta_0\|_{J^2} + k \quad (11)$$

where the distribution of the variable X is the same as the approximate distribution of $n\|\hat{\theta} - \theta\|_{J^2}$, and k is the dimension of the space Θ or the number of parameters independently adjusted to maximize the likelihood function. Equation (11) describes the prediction error based on statistical data.

When there are multiple candidate submodels, it is natural to choose the model that minimizes the $E\mathrm{KL}(\theta_0; \hat{\theta})$. Therefore, it is necessary to consider estimating $n\|\theta - \theta_0\|_{J^2}$ in equation (11). The asymptotic distribution of $\sqrt{n}(\hat{\theta} - \theta)$ is approximately a normal distribution with a mean of 0 and a variance matrix of J^{-1}. If

$$2\left(\sum_{i=1}^{n}\ln p(y_i \mid \theta_0) - \sum_{i=1}^{n}\ln p(y_i \mid \hat{\theta})\right) \quad (12)$$

is used as an estimate of $n\|\theta - \theta_0\|_{J^2}$, then the downward bias introduced by replacing θ with $\hat{\theta}$ needs to be corrected. This can be done by adding k to equation (12). When selecting the model, it is only necessary to compare the $E\mathrm{KL}(\theta_0; \hat{\theta})$ estimates of different candidate submodels.

Since the public items do not affect the comparison, the public items containing θ_0 can be discarded.

3.3 Model Parameter Estimation

For parameter vector $\theta = (c_{J,k}, \sigma^2)$, the estimator of $c_{J,k}(k \in Z)$ can be obtained as

$$\hat{c}_{J,k} = < y(t), \phi_{J,k}(t) >, (k \in Z)$$

according to equation (2).

The maximum likelihood estimator of σ^2 can be obtained by calculation as

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}\left(y(t) - \sum_{k \in Z}\hat{c}_{J,k}\phi_{J,k}(t)\right)^2 \tag{13}$$

Using equation (13), it can be obtained that the maximum value of log-likelihood function $\ln L\left(y; c_{J,k}, \sigma^2\right)$ is

$$\ln L\left(y; c_{J,k}, \sigma^2\right) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\hat{\sigma}^2 - \frac{n}{2} \tag{14}$$

3.4 Determination of Model Selection Criteria

In order to calculate the estimates of $E\mathrm{KL}(\theta_0; \hat{\theta})$ for different candidate submodels, equation (14) is brought into equation (12), and the deviation correction term k is introduced, where k can be regarded as the dimension of scale space. According to property 1 of the orthogonal wavelet transform model, the number of scale function $\phi_{J,k}(t)$ on the J -level is $q(J)$, then the dimension of the scale space is also $q(J)$. Then discard the common term and the irrelevant constant term, and finally get the orthogonal wavelet transformation model selection criterion is

$$n\ln\hat{\sigma}^2 + 2q \tag{15}$$

where n is the signal length, according to the property 2 of the orthogonal wavelet transform model , $\hat{\sigma}^2$ is the variance of the cumulant of all detail noise signals obtained after J decomposition, and q is the dimension of the scale space under the level J , that is, the number of scale functions or the number of scale coefficients.

By observing equation (15), it can be found that this criterion is very similar to AIC criterion used in literature [11] in form, but the meanings expressed by variables are different. In order to distinguish, the criterion represented by equation (15) given in this paper is named the wavelet selection information criterion (WSIC).

According to the property of orthogonal wavelet transform model, q decreases with the increase of decomposition level, and $\hat{\sigma}^2$ increases with the increase of decomposition level, so a suitable J can always be found to make WSIC reach the minimum value, and the J satisfying this condition is the optimal decomposition level of the corresponding wavelet.

IV. SIMULATION EXPERIMENT

4.1 Experimental Setup

In order to verify the effectiveness of the WSIC criterion for the selection of wavelet decomposition level given in this paper, WSIC criterion, AIC criterion and BIC criterion are used to process the same data respectively, and the effects of different criteria for the selection of wavelet decomposition level are compared. In order to avoid the useful high-frequency detail signal as noise removal, stein unbiased risk estimation is selected to determine the threshold and soft threshold function is used to process the coefficient. The simulation experiment steps are as follows:

Step 1: Selecting a wavelet and generating the original signal $f(t)$ using the scale function of the wavelet obtained by 7 iterations.

Step 2: Adding Gaussian white noise interference $\varepsilon(t)$ to $f(t)$, to obtain a noise signal with a specific SNR.

Step 3: Decomposing the signal to the largest level with the selected wavelet and store the detail coefficients of each level.

Step 4: Calculating the stein unbiased risk estimation threshold for each level of detail coefficients.

Step 5: Processing the detail coefficients of each level by using the soft threshold function.

Step 6: Reconstructing the signal with the coefficient after threshold processing, to obtain the estimated signal.

Step 7: Calculating the values of WSIC, AIC and BIC under each level respectively, and selecting the level corresponding to the minimum value of each index as the optimal level under the index.

Step 8: Increasing the signal-to-noise ratio from -5dB to 35dB with an interval of 5dB. Repeating Step 2-Step 7 1000 times for each signal-to-noise ratio, and recording the results of each experiment.

4.2 Analysis of Experimental Results

In order to avoid the chance of the experiment, db6, sym7 and coif2 wavelets-three different types and representative of the experiment were selected at the same time. Among them, the signal length generated by db6 and coif2 wavelet is 1409, and the signal length generated by sym7 wavelet is 1665. The above simulation experiments were carried out by matlab, and the level selected by each criterion and the number of experiments selected for this level were recorded. The frequency obtained by dividing the number of experiments of this level by the total number of experiments is taken as the probability of selecting this level. The level with the greatest probability is chosen as the optimal level of the criterion. The probability corresponding to the optimal level is denoted as the level accuracy, and the higher the value, the greater the probability that the selected level is the optimal level. The optimal level and the accuracy of level were compared between WSIC criterion, AIC criterion and BIC criterion. The experimental results are shown in TABLE 1 - TABLE 3.

Table.1: The selection level and corresponding probability of the three criteria under different SNR (db6 wavelet)

SNR/dB	-5	0	5	10	15	20	25	30	35
WSIC	**84.9%** (7)	**67.6%** (6)	**98.8%** (6)	**92.9%** (6)	53.6% (6)	**98.1%** (5)	**62.8%** (4)	99.9% (4)	**99.8%** (4)
	15.1% (6)	32.3% (7)	1.2% (5)	7.1% (5)	45.9% (5)	1.9% (4)	37.2% (5)	0.1% (3)	0.2% (3)
	/	0.1% (5)	/	/	0.5% (4)	/	/	/	/
AIC	**78.6%** (7)	**56.4%** (6)	**94.5%** (6)	**90.6%** (6)	63% (6)	**88.4%** (5)	**51.4%** (4)	99.9% (4)	**99.6%** (4)
	21.3% (6)	42.9% (7)	3.2% (5)	9.4% (5)	37% (5)	7.9% (6)	48.6% (5)	0.1% (3)	0.4% (3)
	0.1% (5)	0.7% (5)	2.3% (7)	/	/	3.7% (4)	/	/	/
BIC	**78.5%** (7)	**56.7%** (6)	**94.6%** (6)	**90.2%** (6)	62.7% (6)	**88.5%** (5)	**51.7%** (4)	99.9% (4)	**99.6%** (4)
	21.4% (6)	42.5% (7)	3.2% (5)	9.8% (5)	37.3% (5)	7.6% (6)	48.3% (5)	0.1% (3)	0.4% (3)
	0.1% (5)	0.8% (5)	2.2% (7)	/	/	3.9% (4)	/	/	/

Table.2: The selection level and corresponding probability of the three criteria under different SNR (sym7 wavelet)

SNR/dB	-5	0	5	10	15	20	25	30	35
WSIC	**79.8%** (7)	**80.8%** (6)	**99.3%** (6)	**97.3%** (6)	71% (6)	**98.4%** (5)	**90.3%** (5)	**98.3%** (4)	100% (4)
	20.2% (6)	18.8% (7)	0.7% (5)	2.7% (5)	29% (5)	1.5% (6)	9.7% (4)	1.7% (5)	/
	/	0.4% (5)	/	/	/	0.1% (4)	/	/	/
AIC	**71%** (7)	**75.7%** (6)	**96.6%** (6)	**93.3%** (6)	72.9% (6)	**89.5%** (5)	**88.7%** (5)	**92%** (4)	100% (4)
	29% (6)	23.5% (7)	3.3% (5)	6.7% (5)	27.1% (5)	10.3% (6)	11.3% (4)	8% (5)	/
	/	0.8%	0.1%	/	/	0.2%	/	/	/

		(5)	(7)			(4)			
BIC	**70.8%** (7)	**76.1%** (6)	**96.6%** (6)	**93%** (6)	72.3% (6)	**89.6%** (5)	**88.4%** (5)	**92.1%** (4)	100% (4)
	29.2% (6)	23.1% (7)	3.3% (5)	7% (5)	27.7% (5)	10.2% (6)	11.6% (4)	7.9% (5)	/
	/	0.8% (5)	0.1% (7)	/	/	0.2% (4)	/	/	/

Table.3: The selection level and corresponding probability of the three criteria under different SNR (coif2 wavelet)

SNR/dB	-5	0	5	10	15	20	25	30	35
WSIC	**83.1%** (7)	**68.2%** (6)	**92%** (6)	53.9% (5)	**74.4%** (5)	**99.6%** (4)	97.2% (4)	**93.6%** (3)	100% (3)
	16.2% (6)	30.3% (7)	7.6% (5)	45.8% (6)	25.4% (4)	0.3% (5)	2.8% (3)	6.4% (4)	/
	0.7% (5)	1.5% (5)	0.3% (7)	0.3% (4)	0.2% (6)	0.1% (3)	/	/	/
	/	/	0.1% (4)	/	/	/	/	/	/
AIC	**75.7%** (7)	**57.3%** (6)	**85.9%** (6)	71.3% (6)	**67.5%** (5)	**94.7%** (4)	97.9% (4)	**78.5%** (3)	100% (3)
	24.2% (6)	40.7% (7)	10.6% (5)	28.6% (5)	17.5% (4)	5.2% (5)	2.1% (3)	21.5% (4)	/
	0.1% (5)	2% (5)	3.5% (7)	0.1% (4)	15% (6)	0.1% (3)	/	/	/
BIC	**75.7%** (7)	**57.4%** (6)	**85.8%** (6)	70.7% (6)	**67.7%** (5)	**94.8%** (4)	97.9% (4)	**79.3%** (3)	100% (3)
	24.2% (6)	40.6% (7)	10.8% (5)	29.2% (5)	17.8% (4)	5.1% (5)	2.1% (3)	20.7% (4)	/
	0.1% (5)	2% (5)	3.4% (7)	0.1% (4)	14.5% (6)	0.1% (3)	/	/	/

In TABLE 1 to TABLE 3, the first line of each criterion is the optimal level and level accuracy selected by the criterion. TABLE 1 is taken as an example for analysis. The first column of data shows that the signal generated by db6 wavelet is interfered with by Gaussian white noise to obtain a noisy signal with a signal-to-noise ratio of -5dB. WSIC, AIC and BIC criteria are used to select the optimal decomposition level of the signal, and the optimal decomposition level selected by the three criteria is 7. However, the accuracy of WSIC, AIC and BIC criteria were 84.9%, 78.6% and 78.5%, respectively. It can be found that the level accuracy of WSIC criteria is higher than that of AIC and BIC criteria. Similarly, the level accuracy of the three criteria under other SNR and the experimental situation of other wavelet can be compared.

It can be seen from the results that the optimal level selected by WSIC criteria is basically the same as that selected by AIC and BIC criteria, but the level accuracy of WSIC criteria is greater than that of AIC and BIC criteria under most SNR conditions. The level accuracy of WSIC criterion is up to 15.1% higher than that of AIC criterion and up to 14.3% higher than that of BIC criterion. This shows that WSIC criterion has higher stability than AIC and BIC criterion when selecting the optimal level.

V. CONCLUSION

The choice of wavelet transform decomposition level is the key to the effect of wavelet noise reduction. In the case that the original signal is unknown, this paper looks at the problem of wavelet level selection from the perspective of model selection from a statistical point of view, gives the WSIC criterion, and compares the application effect of WSIC criterion with AIC and BIC criteria through simulation experiments. The simulation results show that the level accuracy of WSIC criterion is up to 15.1% higher than that of AIC criterion and up to 14.3% higher than that of BIC criterion. The results show that WSIC criterion is more stable than AIC and BIC criterion in the selection of orthogonal wavelet transform model. It shows that WSIC criterion has higher level accuracy when the optimal level selected by the three criteria is the same. This criterion can reasonably balance the goodness of fit and the complexity of the model in the process of wavelet transformation, and can more accurately select the optimal decomposition level under different signal-to-noise ratio, so that the wavelet noise reduction can achieve better results. The research in this paper provides the method and theory support for determining the optimal decomposition level of wavelet transform in practical application. Further study can apply the WSIC criterion in this paper to practical problems.

REFERENCES

[1] Zhu, J. J., Zhang, Z. T., Kuang, C. L., & Pan, J.B. (2015). A reliable evaluation indicator of wavelet de-noising. *Geomatics and Information Science of Wuhan University*, 40(5), 688-694. doi: 10.13203/j. whugis20130417

[2] Rouis, M., Ouafi, A., & Sbaa, S. (2019). Optimal level and order detection in wavelet decomposition for PCG signal denoising. *Biomedical Engineering/ Biomedizinische Technik*, 64(2),163-176. doi: 10.1515/bmt-2018-0001

[3] Gao, Y. D., Wang, H. L., You, S. H., Feng, L., He, Y. Y., & Liu, K. (2020). A pulsar signal denoising algorithm based on wavelet basis function selection and improved threshold function. *Electronics Optics & Control*, 27(4),15-19. doi:10.3969/j.issn.1671-637X.2020.04.004

[4] Jang, Y. I., Sim, J. Y., Yang, J. -R., & Kwon, N. K. (2021). The optimal selection of mother wavelet function and decomposition level for denoising of DCG signal. *Sensors*, 21(5), 1851. doi: 10.3390/s21051851

[5] Zheng, Y., Chen, X. H., & Zhu, R. (2019). Research on improved optimal wavelet base selection method and frequency hopping signal detection. *Fire Control & Command Control*, 44(5), 102-106. doi: 10.3969/j.issn.1002-0640.2019.05.022

[6] Yan, R., & Gao, R. X. (2009). Base wavelet selection for bearing vibration signal analysis. *International Journal of Wavelets, Multiresolution and Information Processing*, 7(04), 411-426. doi: 10.1142/S0219691309002994

[7] Hafiz, F., Swain, A., Naik, C., Abecrombie, S., & Eaton, A. (2019). Identification of power quality events: selection of optimum base wavelet and machine learning algorithm. *IET Science, Measurement & Technology*, 13(2), 260-271. doi: 10.1049/iet-smt.2018.5044

[8] Kaplun, D., Voznesenskiy, A., Romanov, S., Nepomuceno, E., & Butusov, D. (2019). Optimal estimation of wavelet decomposition level for a matching pursuit algorithm. *Entropy*, 21(9), 843. doi: 10.3390/e21090843

[9] Freire, P. K. M. M., & Santos, C. A. G. (2020). Optimal level of wavelet decomposition for daily inflow forecasting. *Earth Science Informatics*, 13, 1163-1173. doi: 10.1007/s12145-020-00496-z

[10] Zhao, Q., & Huang, S. X. (2009). Determination of deformation-monitoring precision using wavelet transformation based on AIC criterions. *Journal of Geomatics*, 34(5), 46-47. doi: 10.14188/j.2095-6045.2009.05.012

[11] Xiang, D., & Gong, J. B. (2014). To determine the quantitative index for the optimal decomposition scale in wavelet de-noising of deformation series. *Geomatics and Information Science of Wuhan University*, 39(4), 467-470. doi: 10.13203/j. whugis20120216

[12] Zhao, F. Z. (2018). Wavelet denoising quality evaluation method and its application in deformation monitoring. *Shandong University of Science and Technology*. Shandong.

[13] Silik, A., Noori, M., Altabey, W. A., Dang, J., Ghiasi, R., Wu, Z. (2022). Optimum wavelet selection for nonparametric analysis toward structural health monitoring for processing big data from sensor network: A comparative study. *Structural Health Monitoring*, 21(3), 803-825. doi: 10.1177/14759217211010261

[14] Wang, H. Q. (2011). Wavelet Analysis and Application. *Beijing University of Posts and Telecommunications Press*, Beijing.

[15] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723.

[16] Claeskens, G., & Hjort, N. L. (2008). Model selection and model averaging. *Cambridge University Press*, Cambridge.