A method for detecting and locating the gross error of tidal based on wavelet analysis

Jin Zhang1,2,3,*, ShuJun Li2,3, and FanJun Meng2,3

132023 Troop, Aerospace Photogrammetry, Dalian 116021, China
2Department of Military Oceanography and Hydrography & Cartography, Dalian Naval Academy, Dalian 116018, China
3Key Laboratory of Hydrographic Surveying and Mapping of PLA, Dalian Naval Academy, Dalian 116018, China

Abstract. Tidal harmonic constants are necessary data for the evaluation of tidal model, tidal prediction and chart datum. Compared with the wavelet multi-scale analysis model, a new method is put forward to detect and locate tidal discrete and continuous gross error. Based on the properties of the wavelets, the wavelet suitable for detecting the gross error of tidal is selected. And taking db6 wavelet as an example, the feasibility and effectiveness of this method are proved by experiments. The results show that the method can not only simultaneously detect and locate discrete gross error and continuous gross based on high frequency information, and can detect and locate the systematic deviation caused by the zero drift according to the low frequency information. Experimental result shows that the method is more simple, the efficiency and accuracy of detecting and locating gross error are improved.

1 Introduction

The application of wavelet analysis is very extensive, and wavelet analysis mainly uses the self-similarity, multi-resolution analysis of wavelet function and the function of mathematical microscope to process non-stationary signals [1][2]. The wavelet transform is used to detect and analyze the singularity of ECG signals [3]; the fault detection and diagnosis of machinery are performed; the image is compressed, merged, denoised and watermarked [4][5]. The wavelet is used to monitor the deformation date, forecast earthquakes, monitor GPS satellite faults and filter disturbances in aeronautical gravity measurements [6-9].

In this paper, using wavelet to study the tides and tidal. Like other artificial signals (signals in radar and sonar systems), most natural signals are non-stationary signals, such as hydrological data, meteorological data, and so on. The most striking feature of non-stationary signals is that the statistical properties of the signal change over time. The observed tides level of the tide station is changed with time. It is generally written as a function of time \( f(t) \). \( f(t) \) is a non-periodic function and the statistical characteristics change with time, so the tide level signal is also a non-stationary signal.

2 Wavelet multiscale analysis

2.1 Principle of analysis

Multiscale analysis was proposed by S. Mallat and Y. Meyer. On this basis, S.Mallat introduced a fast algorithm for calculating discrete raster wavelet transform, namely Mallat algorithm, which made a breakthrough in wavelet theory. The algorithm decomposes the signal into different frequency domain spaces through the high-pass filter \( G(w) \) and the low-pass filter \( H(w) \): the high frequency subspace \( W \) and the low frequency subspace \( V \). The low-frequency subspace obtained for each decomposition can be repeatedly and repeatedly decomposed using a similar process. The specific process of wavelet multi-scale decomposition is shown in Figure 1.

![Fig. 1 The process of wavelet multi-scale decomposition](image1)

Assuming that the total frequency band occupied by the original function \( f(t) \) is \( 0-\pi \) and the space is \( V_0 \), the frequency distribution of each high and low frequency subspace is shown in Figure 2.

![Fig. 2 Frequency space distribution map](image2)

During the decomposition process, the low pass filter \( H(w) \) of each stage is the same as the high pass filter \( G(w) \). After the first stage decomposition, \( V_0 \) is divided
into two subspaces: the low frequency subspace $V_1$ of $0$–$\pi/2$ and the high frequency subspace $W_1$ of $\pi/2$–$\pi$. After the first-stage decomposition, $W_1$ is decomposed into the low-frequency subspace $V_2$ of $0$–$\pi/4$ and the high-frequency subspace $W_2$ of $\pi/4$–$\pi/2$, and so on into the low-frequency subspace $V_j$ and high frequency. Subspace $W_j$. The low frequency subspace reflects the overview and trend of the signal, and the high frequency subspace reflects the details of the signal. The relationship between each subspace and $V_0$ is:

$$V_0 = W_1 \oplus V_1 = W_1 \oplus W_2 \oplus V_2 = \cdots = W_1 \oplus W_2 \oplus \cdots \oplus W_j \oplus V_j \quad (1)$$

### 2.2 Decomposition recursive model

The signal obtains the low frequency coefficient $Ca_j$ and the high frequency coefficient $Cd_j$ in the decomposition process. Each coefficient is reconstructed by a single layer to obtain low-frequency information $a_j$ and high-frequency information $d_j$ ($j = 1, 2, \cdots, j$) of the same physical quantity as the original signal. Let the tidal sequence $Ca_j^0 = f(n)$, then the decomposition recursive model for solving the high frequency coefficient and low frequency coefficient algorithm is [10]:

$$Ca_j^{j+1} = \sum_n h(n-2k)Ca_j^k$$

$$Cd_j^{j+1} = \sum_n g(n-2k)Ca_j^k$$

In the formula, $h$ is a low-pass filter, $g$ is a high-pass filter, $n$ is the number of data in the tidal sequence, and $j$ is the number of layers to be decomposed.

According to the decomposition process of the signal, among the high-frequency information $d_1 \sim d_j$, $d_1$ has the highest frequency range relative to other high-frequency information, mainly composed of noise, and the mutation information is most obvious in $d_1$. Among the low-frequency information $a_1 \sim a_j$, $a_j$ has the lowest frequency range relative to other low-frequency information, and the trend information is most obvious in $a_j$. Therefore, when multi-scale analysis of tidal level signals, it is best to select high-frequency information $d_1$ to identify and locate the catastrophic points, and select low-frequency information $a_j$ to understand the trend of tidal level[11].

### 3 Based on wavelet detection and positioning gross error

The diversity of wavelet functions determines the effectiveness and efficiency of wavelet analysis. Using different wavelet functions to analyze the same problem will produce different results [12]. Combined with the analysis of the properties of wavelet function, the vanishing moment order of db6 wavelet function is 6, which can effectively divide the tidal level data and ensure the tight support and time resolution of the wavelet function, which is beneficial to the mutation point. The detection ensures the efficiency of the operation. Selecting db6 as the wavelet function, and performing six-layer decomposition on the tidal level signal, and detecting and locating the discrete gross error and continuous gross error that may occur during the tidal water level observation process.

#### 3.1 Detection and localization of discrete gross errors

Taking db6 as a wavelet function, the tidal level signal is decomposed into six layers. The signal is decomposed by multi-scale to obtain low-frequency information and high-frequency information.

#### 3.2 Detection and localization of Continuous gross error

It mainly detects and locates the continuous gross error that may occur during the tidal water level observation process. The continuous error may be caused by random vibration of the instrument under the influence of severe factors such as strong waves. There is a trend change in systematic bias, and the low frequency information obtained by wavelet decomposition can reflect the trend change of water level.

#### 3.3 Detection and localization of systematic deviations

In the tidal level data, in addition to the gross errors, there may be systematic deviation caused by zero drift. There is a trend change in systematic deviations and the low-frequency information obtained by wavelet decomposition can reflect the trend change of water level. Therefore, while using high-frequency information to detect and locate the rough points, the application of wavelets can be extended to the detection and localization of systematic deviations so that the advantages of wavelet multi-scale analysis characteristics can be fully utilized.

### 4 Experiment and analysis

A total of 720 tidal height original tidal data were used, and the data were normal values, no gross error, and the sampling interval was $\Delta t = 1h$. It is subjected to wavelet multi-scale analysis. The low-frequency information ($a1$~$a6$) is shown in Figure 3, and the high-frequency information ($d1$~$d6$) is shown in Figure 4, and the high-frequency information ($d1$) is shown in Figure 5.
When the data does not contain gross errors, 720 data in the high-frequency information $d_1$ are distributed in the range of -5 cm to 5 cm, and there is no mutation value. On the basis of the original data, gross errors were artificially added to conduct experiments and analyze the detection and localization effect of $db_6$ wavelet.

4.1 Experiment of discrete gross errors

In the 114th, 250th, 266th, 492th, and 650th data of the tidal level signal, the gross error is added, which is 20cm, -15cm, -15cm, 20cm, and 15cm in order. The 6-layer decomposition is performed on the tidal level signal after the gross error is added. The low-frequency information $a_6$ obtained by the decomposition is shown in Figure 6, and the high-frequency information $d_1$ obtained by the decomposition is shown in Figure 7.

In Figure 6, the trend of the water level in the month can be seen from the curve. In Fig. 7, the values of the five mutations can be found intuitively, quickly, and accurately based on the high frequency information $d_1$ (Five mutations are indicated by arrows).

For further judgment, $3\sigma$ is used as the criterion and $\sigma$ is the standard deviation which obtained from the 720 data in $d_1$. If the absolute value of the mutation value in $d_1$ exceeds $3\sigma$, the original data has a gross error at that position; if it is less than $3\sigma$, the original data is a normal value at that position. It was calculated that $\sigma = 2.2$ cm. Comparing each value in $d_1$ with $3\sigma$, the result is that there are only 5 values greater than $3\sigma$. These 5 values are exactly the 5 mutation values and their specific positions are consistent with the position where the artificial gross error is added. The detection results are shown in Table 1.

| Point number | $|\text{Mutation}|$ (cm) | $3\sigma$ | Conclusion |
|--------------|-----------------|----------|------------|
| 114          | 13.9            | $>3\sigma$ | Y          |
| 250          | 12.6            | $>3\sigma$ | Y          |
| 266          | 12.3            | $>3\sigma$ | Y          |
| 492          | 13.4            | $>3\sigma$ | Y          |
| 650          | 9.9             | $>3\sigma$ | Y          |

The point numbers of the five mutation values are 114, 250, 266, 492, and 650, corresponding to the point number that was previously added to the gross error, and
their absolute value is greater than $3\sigma$. According to the judgment criterion, the five measured tidal points corresponding to the mutation values contain gross errors. Therefore, in this example, after the tidal level signal is decomposed by 6 layers using the db6 wavelet, 5 coarse errors can be quickly and accurately detected based on the high frequency information $d_1$.

4.2 Experiment of Continuous gross error
In order to study the detection and location of wavelet on continuous gross error, adding 24 continuous gross errors to the original data for experimental analysis. The gross errors are 10cm, -10cm, 10cm, -10cm, 10cm, -10cm, 10cm, -10cm, 10cm, -10cm, 10cm, -10cm, 10cm, -10cm, 10cm, -10cm, 10cm, -10cm, 10cm, -10cm, 10cm, -10cm, 10cm, -10cm, 10cm, -10cm. The tidal level signal was decomposed into 6 layers, the high frequency information $d_1$ is obtained as shown in Figure 8, and the low frequency information $a_1$~$a_6$ is shown in Figure 9.

From the high-frequency information $d_1$, it can be clearly found that the value in a section is abrupt, and the value of the mutation is much larger than other normal values. Therefore, the segment where the continuous gross error occurs can be accurately located according to the high frequency information $d_1$, thereby providing a reliable guarantee for the later data processing.

4.3 Experiment of systematic deviations
In tidal data, there are not only gross errors, but also systematic deviations caused by zero drift. There is a trend change in systematic deviation, and the low frequency information obtained by wavelet decomposition can reflect the trend change of water level. Therefore, while using high-frequency information to detect and locate the gross error, the application of wavelets can be extended to the detection and localization of systematic deviations, and the advantages of wavelet multi-scale analysis characteristics can be fully utilized.

4.3.1 Wavelet decomposition
Using the monthly data of a long-term tidal station A, a systematic deviation of 10cm-80cm is added within 5 days to simulate the temporary tidal station B where the zero drift occurs. Using the db6 wavelet to perform 7-layer decomposition of the monthly data of the temporary tidal station B, the low-frequency information $a_1$–$a_7$ can be obtained. Using the median method to get the hourly average sea surface $MB_0$, $a_3$–$a_7$ and $MB_0$ were selected for comparison. The curves of $MB_0$, $a_7$, and $a_6$ are shown in Figure 10, and the curves of $MB_0$, $a_5$, and $a_4$ are as shown in Figure 11, and the curves of $MB_0$ and $a_3$ are as shown in Figure 12.
4.3.2 Wavelet selection

In Figure 15-17, the low-frequency information $a_4$–$a_7$ is extremely similar to the daily average sea surface, the similarity of $a_3$ is poor, but it can also reflect the general trend of the water level. Calculating the correlation coefficient between the low frequency information $a_3$–$a_7$ and $MB_0$. The correlation coefficient is shown in Table 2.

Table 2. Correlation coefficient table of $a_3$–$a_7$ and $MB$

| low-frequency information | $a_3$ | $a_4$ | $a_5$ | $a_6$ | $a_7$ |
|---------------------------|-------|-------|-------|-------|-------|
| Correlation coefficient   | 0.777 | 0.951 | 0.975 | 0.979 | 0.966 |

From Table 2, the similarity between $a_6$ and $MB_0$ is the highest, and the correlation coefficient between $a_4$, $a_5$, $a_7$ and $MB$ is also above 0.95, and $a_3$ and $MB$ have the lowest similarity. Therefore, the four layers of low frequency information $a_4$, $a_5$, $a_6$, and $a_7$ can be used to detect and locate the zero drift.

4.3.3 Wavelet detection

The zero drift is detected by using the low frequency information $a_6$ with the highest correlation coefficient. Firstly, the low-frequency information $a_{6B}$ of the temporary tide station $B$ is obtained; then the low frequency information $a_{6A}$ of the synchronous observation period adjacent to the long-term tidal station $A$ is obtained; finally, the mutual difference $\Delta a_6$ of the two low-frequency information $a_6$ is obtained. The three curves are shown in Figure 13.

![Fig. 13 The curves of $a_{6A}$, $a_{6B}$, $\Delta a_6$](image)

Since the low-frequency information $a_{6A}$ and $a_{6B}$ are at a much higher water level than the $\Delta a_6$ water level, the curve is "compressed" to some extent. In order to more easily observe the change of $\Delta a_6$, the two curves of $a_6$ and $a_{6B}$ are shifted downward by a constant, as shown in Figure 14.

![Fig. 14 The curves of $a_{6A}$, $a_{6B}$, $a_{6A} - a_{6B}$](image)

According to the curve of the $\Delta a_6$, it can be judged that there is a significant zero drift phenomenon in station $B$. It may be that the submarine foundation of the tidal gauge is permanently settled or the location is affected by strong tide. Therefore, the zero drift can be detected by the tidal low-frequency information $a_6$.

5 Conclusion

This paper studies the characteristics of tidal level data and the principle of wavelet multi-scale analysis. The method of wavelet multi-scale analysis is used to detect and locate the gross error of tidal. Through experimental analysis, the method does not rely on the tidal model, only the wavelet decomposition of the original tidal level data is needed; the characteristic of the gross error of tidal is the mutability, and the wavelet is good at detecting and locating the mutational point. The validity and feasibility of the wavelet multi-scale analysis method for gross error detection and location are verified, and it can be extended to the effective detection of zero drift.

References

1. Heng Tong. Wavelet analysis and its Application[D]. Sichuan University, 2003.
2. Pan Quan, Meng Jinli, Zhang Lei. Wavelet Filtering method and its Application[J]. Journal of Electronics and Information Technology, 2007(01): 236-242.
3. Wang Ting. Research on EMD algorithm and its Application in signal denoising[D]. Harbin Engineering University, 2010.
4. Mohamed Rouis, Abdelkrim Ouafi, Salim Sbaa. Optimal level and order detection in wavelet decomposition for PCG signal denoising[J]. Biomedical Engineering and Biomedizinische Technik, 2019, 64(2).
5. J. Rabi, T. Balusamy, R. Raj Jawahar. Analysis of vibration signal responses on pre induced tunnel defects in friction stir welding using wavelet transform and empirical mode decomposition[J]. Defence Technology, 2019.
6. Wei Wang, Dunqiang Lu, Xin Zhou et al. Statistical wavelet-based anomaly detection in big data with compressive sensing[J]. EURASIP Journal on
7. Yanfang Sang, Dong Wang, Jichun Wu et al. Wavelet-Based Analysis on the Complexity of Hydrologic Series Data under Multi-Temporal Scales[J]. Entropy, 2011, 13(1).

8. D. Maraun, J. Kurths. Cross wavelet analysis: significance testing and pitfalls[J]. Nonlinear Processes in Geophysics, 2004, 11(4).

9. He Liwen, Li Xinguo. Application of wavelet analysis in GPS satellite fault detection[J]. Chinese Space Science and Technology, 2017, 37(06): 56-61.

10. Shi Yan, Li Yuqian, Cai Maolin, Zhang Xiaohua Douglas. A Lung Sound Category Recognition Method Based on Wavelet Decomposition and BP Neural Network[J]. International journal of biological sciences, 2019, 15(1).

11. Yanan Liu, Yinghua Liu, Zhangzhi Cen. Multi-scale Daubechies wavelet-based method for 2-D elastic problems[J]. Finite Elements in Analysis & Design, 2010, 47(4).

12. Yanan Liu, Fei Qin, Yinghua Liu, Zhangzhi Cen. A Daubechies wavelet-based method for elastic problems[J]. Engineering Analysis with Boundary Elements, 2009, 34(2).