Seawater Temperature Outlier Detection Strategy Based on Wavelet Analysis for Solving the Migration Problem of Fishery Companies

Yuan Liu¹, Zheyong Qiu², Zhongyue Wang³ and Jiaqi He⁴

¹School of Electronics and Information Engineering, Hangzhou Dianzi University, Hangzhou Zhejiang 310018, China
²School of Science, Hangzhou Dianzi University, Hangzhou Zhejiang 310018, China
³Zhuoyue Honors College, Hangzhou Dianzi University, Hangzhou Zhejiang 310018, China
⁴School of Intelligence, Hangzhou Dianzi University, Hangzhou Zhejiang 310018, China
Email: 17011419@hdu.edu.cn

Abstract. As the global ocean temperature rises, many marine organisms begin to find new habitats. The problem of fish migration has a profound impact on the development of fisheries. This paper uses wavelet anomaly analysis method to find the abnormal value of seawater temperature, uses discrete wavelet decomposition and reconstruction to de-noise the time series, and predicts the age of the fishing company. Firstly, we conducted continuous wavelet analysis of seawater temperature data to obtain the relationship between seawater temperature change rate and time. We performed Fourier transform to obtain its transform function, and then reconstructed the sequence to obtain wavelet coefficients of seawater temperature transformation speed. After that, we used wavelet coefficients to conduct simulation screening to obtain all outliers. In this way, the time for bankruptcy under the best and worst conditions of the fishing company can be obtained. Although the model comprehensively considers many factors, in order to establish a universal model, idealizing many influencing factors has certain limitations. The predicted bankruptcy time may have a certain error from the actual.

1. Introduction
The sea water temperature sequence can be regarded as a periodic and phased water temperature sequence[1]. Usually, when the seawater temperature sequence is abnormal, it will greatly affect the migration of fish[2]. Therefore, finding the abnormal value of water temperature plays a key role in the study of fish migration routes. The traditional signal analysis is based on the Fourier transform[3]. Because Fourier analysis uses a global transformation, either completely in the time domain or completely in the frequency domain[4], it is impossible to express the time-frequency local nature of the signal.

Wavelet transform[5] is a time-frequency analysis method of signals, which has the characteristics of multi-resolution analysis[6], and has the ability to characterize the local characteristics of signals at both time and frequency. The sea water temperature sequence[7] can be regarded as a digital signal sequence composed of different frequency components[8], and the outliers can be regarded as singular points of the signal. Therefore, the signal singularity theory of wavelet analysis can be used to detect the abnormal value of seawater temperature[9].
2. Problem Description
Changes in ocean temperature can cause changes in the habitat quality of some marine life. When the degree of temperature change exceeds the species’ ability to accept it, the species will migrate to areas where the temperature of the sea is suitable. And this kind of geographical species migration will affect the local marine fisheries.

In response to the problems of the Scottish North Atlantic fisheries, we need to consider the migration routes of Scottish herring and mackerel due to the temperature increase of the sea so as to help small fishing companies in Scotland overcome the difficulties.

The changing process of seawater temperature often produces complicated and irregular information data, and the singularity of water temperature data often contains rich dynamic system information. Finding out irregular and non-compliant temperature data plays a key role in the research of the whole seawater temperature system.

We establish a wavelet analysis model to predict the number of years that a small fishing company will be able to maintain its business in the worst, best, and most likely three cases without changing the fishing area without changing the fishing area.

Based on the above description, the following assumptions are given:
• We will not consider the impact of individual differences such as fish volume and age on its migration speed. We will not consider the impact of other conditions such as climate and natural enemies on the fish population on the migration speed of fish. The migration speed of fish of the same species is the same.
• Assuming that the small fishing company does not move the company's location, the company's fishing range is certain, and the company's revenue sources are all herring and mackerel. For the time being, the company's fishing of other fish is not considered.
• Assuming that the cost and profit of each fishing boat are the same, the difference between the fishing capacity of employees and the reduction in the number of fish caused by the fishing intensity are not considered.
• Assuming that there is enough food in the seawater, the target fish death due to interspecies competition is not considered.

3. Wavelet Analysis Theory
Wavelet analysis, also known as wavelet transform, originated from Fourier analysis. The traditional Fourier analysis is only an analysis tool about frequency. The frequency domain has the highest resolution, but there is no resolution in the time domain. The short-time Fourier transform (STFT) developed later has a certain resolution of the time-domain characteristics, but its window size does not change with the change of frequency, and it cannot take into account the requirements of frequency and time resolution. That is, the frequency resolution and time resolution of the short-time Fourier transform cannot be optimized at the same time. The wavelet analysis overcomes the shortcoming that the window size of the short-time Fourier transform does not change with the change of frequency by shifting and scaling the wavelet function. Wavelet analysis has become a mathematical tool that can handle non-steady-state signals well.

Compared with the Fourier transform, the wavelet transform is a localized analysis of time (spatial) frequency. It gradually refines the signal (function) at multiple scales through scaling and translation operations, and finally meets the requirements of time subdivision at high frequencies and frequency subdivision at low frequencies, and can automatically adapt to the analysis of time-frequency signals so that it can focus on any details of the signal.

The wavelet function is derived from multi-resolution analysis. Its basic idea is to express the expanded function f(t) as a series of successive approximation expressions, each of which is a smoothed form of f(t). They correspond to different resolutions. That is to say, for a given wavelet function, as the value of the scaling factor and the value of the translation factor change, the shape of the wavelet function continuously changes. It uses a local analysis method that changes both the frequency domain window and the time window for the entire original sequence, making the local characteristics of the time series more obvious.
4. Data Transformation under Wavelet Analysis

For solving the minimum bankruptcy time of small fishing companies, we need to get outliers in seawater temperature changes. The temperature of seawater plummets or rises, resulting in insufficient time for fish to cope with sudden temperature changes through migration, and eventually die, which makes it impossible for small fishing companies to make profits. This model uses wavelet analysis to analyse and detect abnormal data of seawater temperature changes, and performs a series of transformations such as scaling, translation, etc. on seawater temperature change data, and through multi-angle and multi-resolution processing of signals to obtain abnormal values.

When wavelet analysis is used to process seawater temperature change data, we assume that the speed of seawater temperature change is:

\[ \phi(t) \in L^2(R), \]  

in which \( L^2(R) \) is the space made up of all square integrable functions.

The Fourier transform of the speed of seawater temperature change is \( \hat{\phi}(\omega) \), which satisfies the complete reconstruction condition:

\[ C_\phi = \int_{-\infty}^{+\infty} \left| \frac{\hat{\phi}(\omega)}{|\omega|} \right|^2 d\omega < \infty, \]  

After expanding and translating the speed \( \phi(t) \) of seawater temperature change:

\[ \phi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \phi \left( \frac{t - b}{a} \right), a,b \in R, a \neq 0, \]  

in which, \( a \) is the scaling factor, \( b \) is the translation factor.

The continuous wavelet transform of \( f(t) \in L^2(R) \) for any function is:

\[ W_f(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} f(t) \phi \left( \frac{t - b}{a} \right) dt. \]  

Its reconstruction formula (inverse transformation) is:

\[ f(t) = \frac{1}{C_\phi} \int_0^\infty \int_{-\infty}^{+\infty} \frac{1}{a^2} W_f(a,b) \phi \left( \frac{t - b}{a} \right) dadb. \]  

And it should be noted that the speed of change in seawater temperature \( \phi \) should meet the constraints of the general function:

\[ \int_{-\infty}^{+\infty} |\phi(t)| dt < \infty. \]

Therefore, its Fourier transform function \( \hat{\phi}(\omega) \) is a continuous function, and:

\[ \hat{\phi}(0) = \int_{-\infty}^{+\infty} |\phi(t)| dt = 0. \]  

In order to ensure the stability of the reconstructed sequence, Fourier of \( \phi(t) \) satisfies the following formula:

\[ A \leq \sum_{-\infty}^{+\infty} \left| \frac{\hat{\phi}(\omega)}{2^j} \right|^2 \leq B, \]
in which $0 < A \leq B < \infty$.

Table 1. Mathematical properties of those wavelet functions used in signal processing

| Wavelet system | Abbreviation | Number | Compact support | Symmetric set | Vanishing Moments | Orthogonality | Biorthogonality |
|----------------|--------------|--------|----------------|---------------|------------------|---------------|----------------|
| Haar           | haar         | 1      | +              | -             | 1                | +             | +              |
| Daubechies     | dbN          | 9      | +              | -             | N                | +             | +              |
| Symlets        | symN         | 7      | +              | +             | N                | +             | +              |
| Coiflets       | coifN        | 5      | +              | +             | 2N               | +             | +              |
| Dmeyer         | dmeN         | 1      | +              | +             | /                | +             | +              |
| Bior splines   | biorMN       | 15     | +              | +             | M                | -             | +              |
| Gaussian       | gaus         | 8      | -              | +             | /                | -             | -              |
| Morlet         | moel         | 1      | -              | +             | /                | -             | -              |
| Mexican hat    | mexh         | 1      | -              | +             | /                | -             | -              |
| Meyer          | meyr         | 1      | -              | +             | /                | -             | -              |

Since the self-similarity problem is common when using continuous wavelet transform to process data, the phenomenon of information redundancy in continuous wavelet transform is mainly manifested in the following two aspects:

- The reconstruction fraction of continuous wavelets is not unique, that is, the sequence $f(t)$ after the wavelet transform does not correspond to the re-ditched sequence, and the Fourier transform and its inverse transform are consistent.
- The choice of wavelet functions $\phi_{a,b}(t)$ is not unique. The correlations between different $(a, b)$ are different, which increases the difficulty of the inspection and analysis process.

Therefore, we must pay attention to the above two points when processing data, to avoid data processing errors due to self-similarity, and finally the detected outliers are not real outliers.

5. Outlier Detection

After the continuous wavelet transform of the seawater temperature change $X_t$, we need to obtain the wavelet coefficient of the seawater temperature change for detecting outliers. It is worth noting that the wavelet coefficients must meet the following properties:

- The wavelet coefficients at the same layer are stationary sequences with equal variance.
- The autocorrelation of each layer of wavelet will rapidly decay.
- The wavelet coefficients belonging to different layers do not exist or have weak correlation.

For the stationary time series of the same layer, the residual sequence $X(t)$ of the wavelet coefficients should obey the standard normal distribution.

First, a continuous wavelet analysis is performed on the seawater temperature change $X(t)$ to obtain the wavelet approximation coefficient:

$$A = \{a_1, a_2, ..., a_n\},$$

and wavelet detail coefficient:

$$D_{-1} = \{d_1, d_2, ..., d_n\}.$$  \hspace{1cm} (8)

Then, we use Monte Carlo to simulate 10,000 times and filter the maximum value of each simulation result, and take the average value as the threshold $k$.

Then, by sieving the wavelet detail coefficients, we find:

$$d_{max} = \max_{a < i < \frac{n}{2}} \{|d_i | > k^{0.05}\}.$$
And we set \(d_{\text{max}}\) in order to avoid the masking effect, assign the value at \(s_1\) in the sequence \(D_1\) to a suitable value, the remaining values remain unchanged, that is, let \(\tilde{D}_1\) be \(\{d_1, ..., d_{s_1-1}, 0, d_{s_1+1}, ..., d_2\}\). Afterwards, we use a continuous wavelet transform to reconstruct the seawater temperature change sequence. Repeat the above process until the absolute values of \(D_1\) are all less than the threshold, and we obtain a set \(S\) of all positions corresponding to \(d_{\text{max}}\). Locate the position of the outlier through \(S\), \(s\) represents any element in \(S\), and calculate the sample average of the residual sequence \(\tilde{x}_s\) after removing the values at the corresponding positions of \(2s\) and \(2s-1\):

\[
\text{If } |\tilde{x}_{2s}-\tilde{x}_{n-2}| > |\tilde{x}_{2s-1}-\tilde{x}_{n-2}|, \text{ then } 2s \text{ is an outlier, else } 2s-1 \text{ is an outlier.}
\]

We can obtain the year in which the abnormal value is located by comparing the calculated abnormal value with the annual seawater temperature change obtained in the previous question, so it is not difficult to conclude that the small fishing company maintain the status quo, comprehensive long-term data shows that seawater temperatures generally rise, showing a trend of first decline and then rise. In the worst case, the company may fail after 28 to 32 years due to the failure to catch fish, and in the best case, it will close after 40 to 44 years. Finally, based on the results of the previous question, we can know that the small fishing company is most likely to close after 34-36 years.

6. Acknowledgements
This study was supported by Zhejiang Xinmiao Talent Project (No. 2019R407054) and the National Natural Science Foundation of China (11871185).

7. References
[1] Hansen, & Mark. (1998). Applied wavelet analysis with s-plus. Technometrics, 40(2), p159-159.
[2] Flandrin, & P. (1992). Wavelet analysis and synthesis of fractional brownian motion. IEEE Transactions on Information Theory, 38(2), p910-917.
[3] Wickerhauser, Mladen Victor. (2010). Adapted wavelet analysis from theory to software. A.K. Peters.
[4] Wilkinson, W. A., & Cox, M.D. (1996). Discrete wavelet analysis of power system transients. IEEE Transactions on Power Systems, 11(4), P.2038-2044.
[5] Wang, Bin, Wang, & Yan. (1996). Temporal structure of the southern oscillation as revealed by waveform and wavelet analysis. Journal of Climate.
[6] Millet, P., Vicente Ibez, Delforge, J., Pappata, S., & José Guimón. (2000). Wavelet analysis of dynamic pet data: application to the parametric imaging of benzodiazepine receptor concentration. Neuroimage, 11(5), 458-472.
[7] François Anctil and Paulin Coulibaly. 2004: Wavelet Analysis of the Interannual Variability in Southern Québec Streamflow. Journal of Climate: Vol. 17, No. 1, pp. 163–173
[8] Paulin Coulibaly and Donald H. Burn. 2005: Spatial and Temporal Variability of Canadian Seasonal Streamflows. Journal of Climate: Vol. 18, No. 1, pp. 191–210.
[9] Shoshiro Minobe, Teruko Manabe and Akiko Shouji. 2002: Maximal Wavelet Filter and Its Application to Bidecadal Oscillation over the Northern Hemisphere through the Twentieth Century. Journal of Climate: Vol.15, No. 9, pp. 1064–1075.