

Article

Exploratory Method for Spatio-Temporal Feature Extraction and Clustering: An Integrated Multi-Scale Framework

Wen Luo ¹, Zhao-Yuan Yu ^{1,2,3,*}, Sheng-Jun Xiao ¹, A-Xing Zhu ^{1,2,3} and Lin-Wang Yuan ^{1,2,3}

¹ Key Lab of Virtual Geographic Environment, Ministry of Education, Nanjing Normal University, Nanjing 210023, China; E-Mails: luow1987@163.com (W.L.); xiaoshengjun@126.com (S.X.); axing@njnu.edu.cn (A.Z.); yuanlinwang@njnu.edu.cn (L.Y.)

² State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing 210023, China

³ Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

* Author to whom correspondence should be addressed; E-Mail: yuzhaoyuan@njnu.edu.cn; Tel./Fax: +86-25-8589-1049.

Academic Editor: Wolfgang Kainz

Received: 9 June 2015 / Accepted: 11 September 2015 / Published: 25 September 2015

Abstract: This paper presents an integrated framework for exploratory multi-scale spatio-temporal feature extraction and clustering of spatio-temporal data. The framework combines the multi-scale spatio-temporal decomposition, feature identification, feature enhancing and clustering in a unified process. The original data are firstly reorganized as multi-signal time series, and then decomposed by the multi-signal wavelet. Exploratory data analysis methods, such as histograms, are used for feature identification and enhancing. The spatio-temporal evolution process of the multi-scale features can then be tracked by the feature clusters based on the data adaptive Fuzzy C-Means Cluster. The approach was tested with the global 0.25° satellite altimeter data over a period of 21 years from 1993 to 2013. The tracking of the multi-scale spatio-temporal evolution characteristics of the 1997–98 strong El Niño were used as validation. The results show that our method can clearly reveal and track the spatio-temporal distribution and evolution of complex geographical phenomena. Our approach is efficient for global scale data analysis, and can be used to explore the multi-scale pattern of spatio-temporal processes.

Keywords: clustering; scale; tracking; oceanography; classification; data mining

1. Introduction

Most geographical phenomena have different characteristics at different scales in spatio-temporal domain. Various influences caused by many different factors simultaneously acting across scales lead the data on the phenomena to contain a mixture of multi-scale signals. The mixture of the multi-scale characteristics affects the data characteristics in both spatial and temporal domains. In the spatial domain, the characteristics of a given phenomenon should be represented with certain spatial range or resolution to support the complex remote sensing analysis including the object extraction, image segmentation, change detection, object tracking, and geostatistical analysis, *etc.* [1,2]. In the temporal domain, the multi-scale characteristics of the data are acting as periodical or quasi-periodical signal at frequency ranging from hours to several decades. The effects of the multi-scale characteristics have attracted much attention, have been discussed in many different disciplines, *etc.* [3–6].

Remote sensing data have been quickly accumulated continuously. More and more long-term remote sensing sequences (e.g., satellite and aerial images, SAR, *etc.* [7]), which cross the time span of more than several decades, are logged and used in various fields. These long-term data observations provide valuable information for detailed revelation of the evolution of the geographical processes [8]. Along with the wide application of multi-temporal remote sensing data in many different fields, researchers have also proposed several different kinds of advanced analysis methods for feature extraction, segmentation, *etc.* [9].

There have already been some techniques proposed to solve these problems. Multidimensional data modeling and databases [10,11] and Array Databases (e.g., rasdaman [12], SciDB [13]) have largely improved the performance of storing and querying of the massive multidimensional spatio-temporal data. MOLAP (Multidimensional On Line Analysis Process), commonly used in data warehouse, is rapidly improved for multidimensional data queries and aggregation [14]. Technologies, such as wavelets and hierarchical clustering of data are introduced to MOLAP to partly support the computation [15,16]. However, these analyses require accessing directly the most detailed data since no prior knowledge for the query exists. The spatio-temporal unified approaches (e.g., STIM [17], Clifford Algebra/Tensor approaches [18,19]) are also proposed for the unified analysis of such huge amount of spatio-temporal data. However, these approaches are still in an early stage and too complex to be used directly in the operational environment.

Despite the large progresses achieved in the developing of multi-temporal remote sensing analysis methods, two critical problems still exist: (1) Most of these methods treat spatial and temporal dimensions separately. Several approaches can do multi-scale analysis in either spatial or temporal domain (e.g., EMD, wavelet, and parallel adaptive kernel smoothing) [20–22]. Few methods can support for the spatio-temporal domain analysis in the spatio-temporal unified way [23,24]. With the separation of spatial and temporal signal in data analysis the spatio-temporal data can only be analyzed either spatially or temporally, which will lead to information missing, pattern inconsistency, and synchronizing difficulties. (2) Rarely methods integrate the multi-scale information during the exploratory data

analysis. Although there are several exploratory data analysis methods, which can well reveal certain characteristics (e.g., value distributions, spectrum characteristics, and feature variations), developed, large percent of these methods are only applied to the overall data rather than the features at each scale. It is still complicated to reveal the multi-scale characteristics from different perspectives and perform the exploratory multi-scale spatio-temporal feature extraction.

Feature clustering, one of the most typical and commonly used exploratory analysis methods for multi-temporal remote sensing data, is largely restricted by the two problems mentioned above. Integrating multi-scale information in feature clustering can produce rich information about the spatio-temporal process and increase the feature clustering performance for feature extraction and process investigation. For example, various signals, including ocean currents and tides, high frequency variations (e.g., Madden-Julian Oscillation (MJO)), inter-decade variations (e.g., El Niño-Southern Oscillation (ENSO)), *etc.* can be separated from the global ocean data according to their multi-scale characteristics differences [25,26]. With the multi-scale cluster in the spatio-temporal domain, the spatio-temporal evolution characteristics and processes of the above various signals can be revealed.

A fine performance multi-scale spatio-temporal feature clustering method should satisfy the following demands: (1) Since most geographical processes change continuously in the spatio-temporal domain, the analysis should be performed in the spatio-temporal domain with the consideration of the data continuity. (2) The feature clustering should be applied to components of different scales, which have different data distributions and different noise levels. Thus the feature clustering is better to be data adaptive and flexible. (3) Since the data volume of the original multi-temporal remote sensing data is usually large, and the multi-scale decomposition increases the data volume significantly, the feature clustering should be efficient for real world operational analysis. To achieve all the above requirements, new approaches should be developed to perform the multi-scale spatio-temporal feature clustering analysis.

This paper presents an integrated approach to perform the multi-scale spatio-temporal exploratory feature clustering. The paper is organized as follows: Section 2 presents the overall idea of the proposed method. Section 3 provides the details of the method. The case study with the satellite altimeter data is shown in Section 4. The technical and geographical effectiveness of the method is evaluated in Section 5. Discussions and conclusions are presented in Sections 6 and 7, respectively.

2. Overall Idea

The overall idea of the exploratory multi-scale spatio-temporal feature clustering of multi-temporal remote sensing data is depicted in Figure 1.

From the unified spatio-temporal perspective, the original multi-temporal remote sensing data can be considered as a unified spatio-temporal data cube. In this data cube, each location (pixels with same spatial coordinates) is a sampling with a time series indicating the temporal fluctuations through the temporal dimensions. Due to the lack of spatio-temporal unified multi-scale decomposition methods, the spatio-temporal multi-scale decomposition can only be performed in the spatial domain or the temporal domain. Therefore, the original spatio-temporal data, which are considered as a unified spatio-temporal data cube, should be unfolded as a multi-channel time series or spatial series. Under this notion, the entire multi-temporal remote sensing data over an area can be perceived as a set of time series with each location being a channel of time series. Therefore, the whole spatio-temporal data can be seen as a single

multi-signal time series that is continuously observed at a synchronized temporal sampling rate. The spatial distributions of the data are inherited in the channel-to-channel relations.

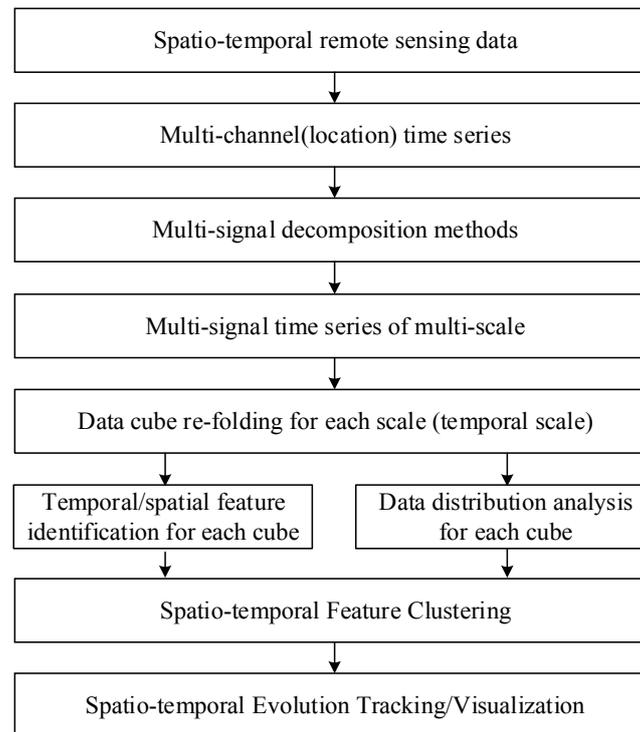


Figure 1. The overall framework of the exploratory multi-scale spatio-temporal feature clustering.

Multivariate multi-scale analysis methods, such as multi-signal wavelet, can be applied to the multi-signal time series to extract the multi-scale variation patterns. The multi-signal wavelet decomposition estimates the common multi-scale variation pattern of all channels synchronically using the full information of the spatio-temporal data. Both global and the local spatio-temporal variations at different spatio-temporal scales can be extracted synchronically [27]. With the multi-signal wavelet decomposition, the original multi-temporal remote sensing data can be decomposed into several multi-scale components. Each decomposed component, which is a multi-signal time series that indicates a certain feature of the multi-scale spatio-temporal variations, can be re-folded to a data cube with the same size of original data. Therefore, the connotative signals that only affect at certain significant scales can be extracted and revealed.

To enhance the feature extraction performance and filter out possible uncertainties or pseudo-variations caused by scale mixture or noise, classical exploratory data analysis methods, such as principal components analysis, power spectrum analysis and histograms, can be used to characterize the spatial/temporal features (e.g., spatial shapes or temporal indexes) and the data value distributions of the multi-scale variations. This can be achieved by comparing the features extracted by principal components analysis and power spectrum analysis with existing geographical indexes or spatial patterns to reveal geographical meanings of the multi-scale components. To identify the spatio-temporal distribution characteristics of the geographical phenomena, feature cluster methods such as K-Means, Fuzzy C-Means Cluster (FCM), *etc.*, can be applied to the spatial/temporal histograms to achieve the

spatio-temporal feature clustering. To achieve the data adaptive clustering for each decomposed multi-scale component, data indexes, such as entropy, can be used to develop the data adaptive optimized cluster number selection procedure. To improve the computational performance, the histograms can be used for feature clustering. The histogram-based cluster of the decomposed components helps to enhance the features of the data, which makes the data more separable. Since the data value distributions in the histogram bins are interlinked with different regions of data in the spatio-temporal domain, the original spatio-temporal data can be clustered by separating the histogram with the value distribution corresponding to the desired regions. With the clustered features, the evolution of the features can be tracked and visualized crossing different time stamps.

3. Methods

According to the overall framework, several methods are assembled and steadily integrated in a processing workflow (Figure 2). The original spatio-temporal data are firstly reorganized as a multi-signal time series, which is then decomposed into several feature series cube and a noise cube by the multi-signal wavelet decomposition. The empirical orthogonal function analysis (EOF) and power spectrum analysis are applied to extract the spatio-temporal and frequency characteristics of each decomposed component. The integer approximation, which stretches the data, is used to enhance the data feature and reduce the computational cost. By reordering the data of integer approximation of each feature series cube in an ascend order and splitting them into several data bins, the histogram approximations can be constructed. Then the Fuzzy C-Means Cluster (FCM), with adaptively selection of cluster numbers, is applied to the stretched feature cubes. The segmentations of histograms are then reversely mapped to the spatio-temporal domain to achieve the spatio-temporal segmentation of original multi-temporal remote sensing data.

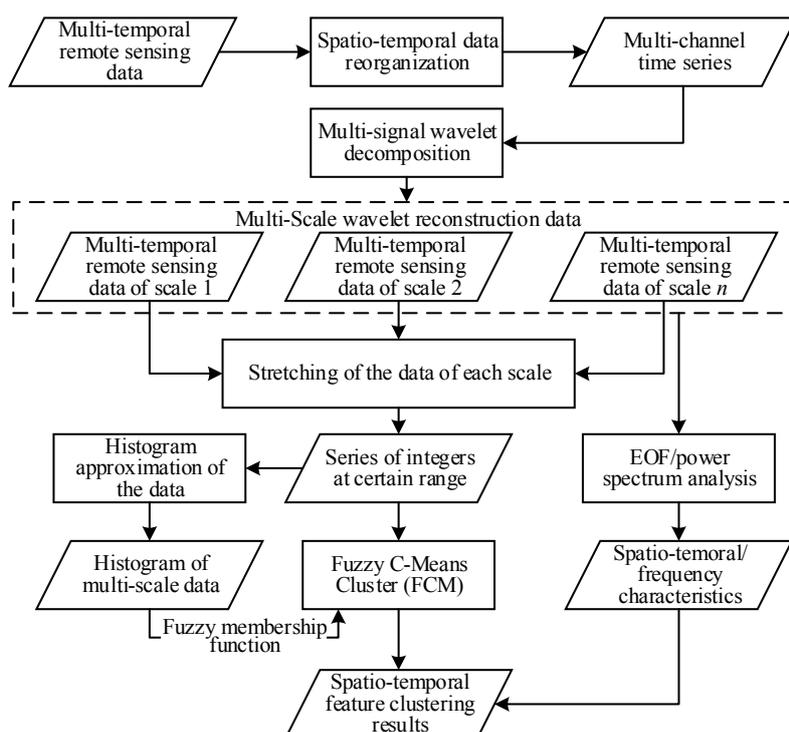


Figure 2. The processing workflow.

3.1. Spatio-Temporal Data Reorganization

In the beginning, we organize the multi-temporal remote sensing data as a unified multidimensional data cube [28]. The unfolding is then applied to the data cube to form the multi-signal time series. Since there are usually more than one coordinates in the spatial domain, unfolding the original data cube into a set of time series with each spatial coordinates a time series (*i.e.*, spatial unfolding) will make the data much simpler. Therefore, we reorganize original data as unified spatio-temporal cube and then apply spatial unfolding to form the multi-signal time series. Although, the unfolding process of the original spatio-temporal data into the multi-channel signal appears to separate the spatial and temporal dimensions, the pattern extraction is done through the spatio-temporal distribution of the data and the data synchronization problems are naturally solved during the analysis.

3.2. Multi-Signal Wavelet Decomposition

Multi-signal wavelet decomposition is applied to the reorganized multi-signal time series to retrieval the multi-scale spatio-temporal variations. Assuming $X(t) = \{x_1(t), x_2(t), \dots, x_M(t)\}$ is the M dimensional multi-signal time series unfolded from the multi-temporal remote sensing data, the multi-signal wavelet decomposition of $X(t)$ can be formulated as follows:

$$X(t) = \sum_k F_k(t) + \varepsilon(t) \quad (1)$$

where $F_k(t)$ is the deterministic feature series signal that has certain spatio-temporal scale characteristics, $\varepsilon(t)$ is the spatial correlated residuals, which can be seen as noise, and k is the level of features extracted.

Let $h(t) = \sum_n h(n) h(Mt - n) \in L^2(R)$ be the scale function. $M - 1$ wavelet functions (notated as $j^{(i)}(t)$) can be constructed. For each of the wavelet function that meet:

$$j^{(i)}(t) = \sum_n g^{(i)}(n) h(Mt - n), i = 1, 2, \dots, M - 1 \quad (2)$$

where $g^{(i)}$ is the coefficient function of the wavelet filter of wavelet function $j^{(i)}(t)$.

Rapid algorithm [29], which is constructed by parameterization of the mother wavelet and decomposition of the signal in the feature space, is used here for efficient decomposition. For each channel of the multi-signal, irrelevant elements involved in the signals are eliminated by a dyadic decimation, which reduces the original resolution into half of its length. This procedure is recursively applied on approximation coefficients to produce increasingly smoother versions of the original signals. Thus, we have the following decomposition:

$$\begin{cases} A_{j+1}^1 X = M^{1/2} \sum_k h(Mn-k) A_j^k X \\ D_{j+1,n}^{(i)} X = M^{1/2} \sum_k g^{(i)}(Mn-k) A_j^k X \end{cases} \quad (3)$$

where $A_{j+1}^1 X$ is the approximation coefficient of the first scale, $A_j^k X$ is the approximation coefficients of the k -th scale, $D_{j+1,n}^{(i)} X$ is the detailed coefficient of the $(k + 1)$ -th scale, and $g^{(i)}$ is the high pass filter of the scale function. Since the original data X are a signal that have total M channels, all the

approximation coefficients A_j^k is a $M \times p_k$ matrix, where p_k is the same length as the DWT approximation coefficients at the k -th scale.

The reconstruction of the multi-signal wavelet decomposition can be adopted in wavelet synthesis by up-sampling and inverse filtering of the decomposed coefficients. That is:

$$A_{j,n}^k X = M^{-1/2} \sum_k h(k - Mn) A_{j+1}^k X + M^{-1/2} \sum_k g^{(i)}(k - Mn) D_{j+1}^{(i)} X \tag{4}$$

According to Equation (4), the approximation component at level j can be reconstructed with the approximation and detail components at level $j + 1$.

Since all the time series are decomposed at the same time, they are at the same time scales. As the decomposed components have the same size of the original data, exploring methods, such as principle components, histograms and power spectrums, can be direct applied to identify the characteristics and geographical meaning of these spatio-temporal variations at each scale.

3.3. The Stretching of the Data

To enhance the feature pattern of each decomposed component and reduce the computational cost, the data are stretched to the range that best represents its features. A simple mechanism, integer approximation, which projects float point data into a series of integers at certain range, is used. Given a multi-signal time series $X(t) \in (a, b)$, which has a value range (a, b) , i.e., $X(t) \in (a, b)$, where a and b are float values. The mapping of the multi-signal time series $X(t)$ to the integer approximation in the range of (m, n) can be formulated as follows:

$$f: \mathbb{R}_{R(X(t))} \mapsto \mathbb{R}_{(m,n)}^+ \tag{5}$$

where \mathbb{R} is the real number set, $R(X(t))$ is the value range of the series $X(t)$, and the $\mathbb{Z}_{(m,n)}^+$ is the positive integer set with the value range of (m, n) . A typical solution of the mapping is:

$$x'_i = \left\{ m + (n-m) \frac{x_i - \text{Min}(X)}{\text{Max}(X) - \text{Min}(X)} \right\} \tag{6}$$

where (x) is the function which aims at finding the closest integer of real value x .

For each decomposed feature series cube, we can apply Equation (6) to project the whole data from the original value range (a, b) into the new value range (m, n) with their spatial and temporal distribution rarely changed. Since both the boundary value of the data range (m, n) and the level value i are all customizable, they can approximate the original signal at any given precision.

3.4. Histogram Approximation of the Data

Histogram is used to summarize the value distribution of the integer-approximated data to reduce the data value and data dimensions, which also helps to enhance the features of the data.

Let $f \in L^1(\mathbb{R}^n)$ is the integer approximation of multi-signal time series $X(t)$, the cumulative histogram operator H is defined by:

$$H(f)(y) = \mu\{x: f(x) \leq y\} \tag{7}$$

where μ is the Lebesgue measure. $H(f)$ is the real function that represents the cumulative histogram. With the integer approximation of the multi-signal series, the histogram m_i counts the number of observations that fall into each of the bins meets the conditions: $n = \sum_{i=1}^k m_i$, where n is the total number of observations and k is the total number of bins. The data value distributions in the histogram bins are interlinked with different regions of data in the spatio-temporal domain, clustering the histogram with the value distribution corresponding to the desired regions can result in a spatio-temporal cluster of original data according to their value distributions. Since the data volume of the histogram are much less than the original data, the cluster of the histogram is much more efficient than that of the original data.

For each decomposed feature series, we can calculate one histogram by recognizing the data cube as a one dimensional data vector. In the histogram, the data are organized in an ascending order, so the clustering is very simple. The multi-signal wavelet decomposition, which can be seen as a denoising procedure, largely reduces sudden changes in the data. Therefore, the data samples that have close spatial distances are highly possible to be classified into the same bins. This will further lead to large potential to keep the spatial continuity of the clustering. The feature cluster performance is customizable by controlling the number of bins, which make the feature cluster flexible to achieve the balance between performance and the approximation level.

3.5. Adaptive FCM on the Histogram for Feature Clustering

Original spatio-temporal clustering measures the distances between pixels in the spatio-temporal domain. This kind of spatio-temporal domain is computational intensive and time consuming. The histogram, which is a reordered index about the value distribution of the original spatio-temporal data, abstracts the value features into different bins. Since each bin of the histogram relates to certain pixels in the spatio-temporal domain, the clustering of the histogram bins is much simpler and can better indicate the value distribution of original data in the spatio-temporal domain. In our approach, the histogram used can be seen as a pre-process, which gathers the pixels into fewer groups according to their value data. Then the similarities between different bins are then clustered. In this way, not only the features are enhanced but also the computation can be largely reduced. Since for each bin, there are bidirectional mapping functions to map the data from the value domain to the spatio-temporal domain, the clustering between bins of the histogram in deed performs the clustering in the spatio-temporal domain. Therefore, the clustering is definitely a spatio-temporal domain clustering, which can be used to track the spatio-temporal evolution of the features.

To reduce the shape changes between the boundaries to achieve more smooth separation in the spatio-temporal domain, the Fuzzy C-Means Cluster (FCM), which has soft thresholds for data separation, is used. Here, FCM, which clusters the data sampler to two or more clusters according to the fuzzy membership function, is applied to the histogram. The computation of the FCM is an iterative method that moves the cluster centers and assigns the data points to each cluster centers recursively to optimize the data segmentation according to a certain object function [30]. The objective function [31] we have used is:

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2, 1 \leq m \leq \infty \quad (8)$$

where m is a real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j , x_i is the i -th of d -dimensional histogram of the observed data (original data or the decomposed feature series), c_j is the d -dimension center of the cluster, which is initially randomly selected and recursively updated. $\|\cdot\|$ is any norm expressing the similarity between any histogram data and the center.

The determination of the cluster numbers is a critical but also important issue for the feature clustering. Since the clustering is applied to the histogram, the difference between histogram bins (*i.e.*, the value distribution) are much smaller than the difference between the original data, lots of traditional cluster number determination methods are not efficient for such task. So the fuzzy membership between different histogram bins, which measures the relative variability and similarity between different value ranges, is used to index the affection of the cluster numbers to the whole data segmentation and their statistical distributions. The Classification Entropy (CE) [32], which measures the fuzziness of the cluster partition, is used to determine the optimized segmentation cluster number adaptively according to the data distribution. The definition of the CE is:

$$CE(c) = -\frac{1}{N} \sum_{i=1}^c \sum_{j=1}^N u_{ij} \log(u_{ij}) \quad (9)$$

where u_{ij} is the membership of data point j in cluster i . According to the definition of the CE, the larger CE indicates the large fuzziness between different clusters. When the cluster number is growing, the value of CE is going smaller, which suggests a more disorderly data distribution. Therefore, according to the change characteristics of the CE against to the cluster numbers, the data segmentation structures against value distributions can be identified. The change point of the CE suggests the significant change of the cluster distribution *i.e.*, change of the characteristics. Thus, the cluster number located at the significant change point of the CE can be selected as the optimal cluster number of each decomposed component.

4. Case Studies

4.1. Research Data and Experiment Configuration

The Monthly Global Delayed Time Mean Sea Level Anomaly (MSLA) data produced by SSAIto/Duacs, AVISO [33] were used as the experiment data. The data were produced by combining the satellites of T/P, Jason-1, Jason-2 and Envisat. The spatial resolution of the data is 0.25° and the time period of the data is from January 1993 to December 2013. The total data are $1440 \times 721 \times 252$. The data covers most of the global ocean containing the area between 81°S – 81°N , the land part and the ocean in the polar area are represented as missing values. The original data in the form of several individual NetCDF files is read recursively into the MATLAB 2011b environment on an Inspur NP 3560 server with two Intel Xeon E5645 (2.4G) processors and 48 GB DDR-3 ECC Memory. The operating system is Windows Server 2008 R2. All the test codes are written in the same software and hardware environment. The Multivariate ENSO Index (MEI) is used as an index of ENSO event [34].

The MSLA data are firstly represented as multi-signal time series, with the time series at each pixel as a channel. Then the multi-signal matrix is decomposed with the multi-signal wavelet with the nearly symmetric wavelet “Sym4” as mother wavelet. Symlet wavelet family (SymN) is a reversion of the commonly used Daubechies wavelet. It is an orthogonal and nearly symmetric wavelet with support

width $2N-1$ and vanishing moments N . Due to its nearly symmetric properties, it can reduce the phase shifts during the wavelet reconstruction [35]. Similar to the individual DWT analysis, the decomposition level in this work was restricted to the greatest integer contained in $\log_2(n)$, where “ n ” is the length of the time series. According to the length of the data, 6-level decomposition is applied to the whole data set.

With the decomposed multi-signal data, the minimum of the data is subtracted to keep all the data positive. According to Equations (5) and (6), the integer approximation of the data is applied. The data range parameters (m, n) are selected according to the decomposition value distribution of the data. To best distinguish the difference of all the decomposed components, all components are re-normalized to the value that ranging from 0–160. With the integer approximation, the histogram distribution of each decomposed component is calculated and the histogram-based FCM algorithm is then applied to each decomposed feature series. The cluster with different number of clusters is simulated and the classification entropy is computed for each simulation to determine the optimal cluster number of each decomposed component. The final data cluster at each scale is computed with the optimized cluster number for each decomposed component. The computation time and the memory occupations changing are logged and analyzed according to the cluster numbers. Finally, the distribution and the evolution of the cluster result are compared with typical indexes such as ENSO, *etc.* The evolution of the spatio-temporal pattern is also compared with existing literature.

4.2. Multi-Scale Decomposition Result

The reconstruction of the multi-scale components of the satellite altimeter data extracted from the multi-signal wavelet decomposition is depicted in Figure 3. Since the full representation of all the data is impossible, we only provide several time slices. From the spatial dimension, the A6 components are much smoother than the original data. This suggests the multi-signal wavelet decomposition not only is multi-scale filter in the temporal domain, but also affects the spatial domain. Although the information of the generalized spatial scale affection of the multi-scale decomposition is still limited, in our cases, the larger temporal scales will also result in a smoother spatial distribution.

To reveal the detailed spatio-temporal distribution of the decomposed components at different scales, we applied the empirical orthogonal function analysis (EOF) to each decomposed component. The EOF method can extract the significant and common pattern of the spatio-temporal data. The variance contribution and the principal components (PCs), which suggest the consistency of the components across the spatial distribution and the significant temporal change of certain patterns, are extracted from the EOF (Table 1). To further understand the temporal patterns of each decomposed component, the power spectrum (PSD) of each extracted PC is also calculated. The peak of the PSD suggests the significant period of the PC. The PCs and the PSD of the PCs are depicted in Figure 4. To reveal the detailed frequency characteristics of the PCs, the highest peak of PSD and its associated frequency and period are extracted and represented in Table 2.

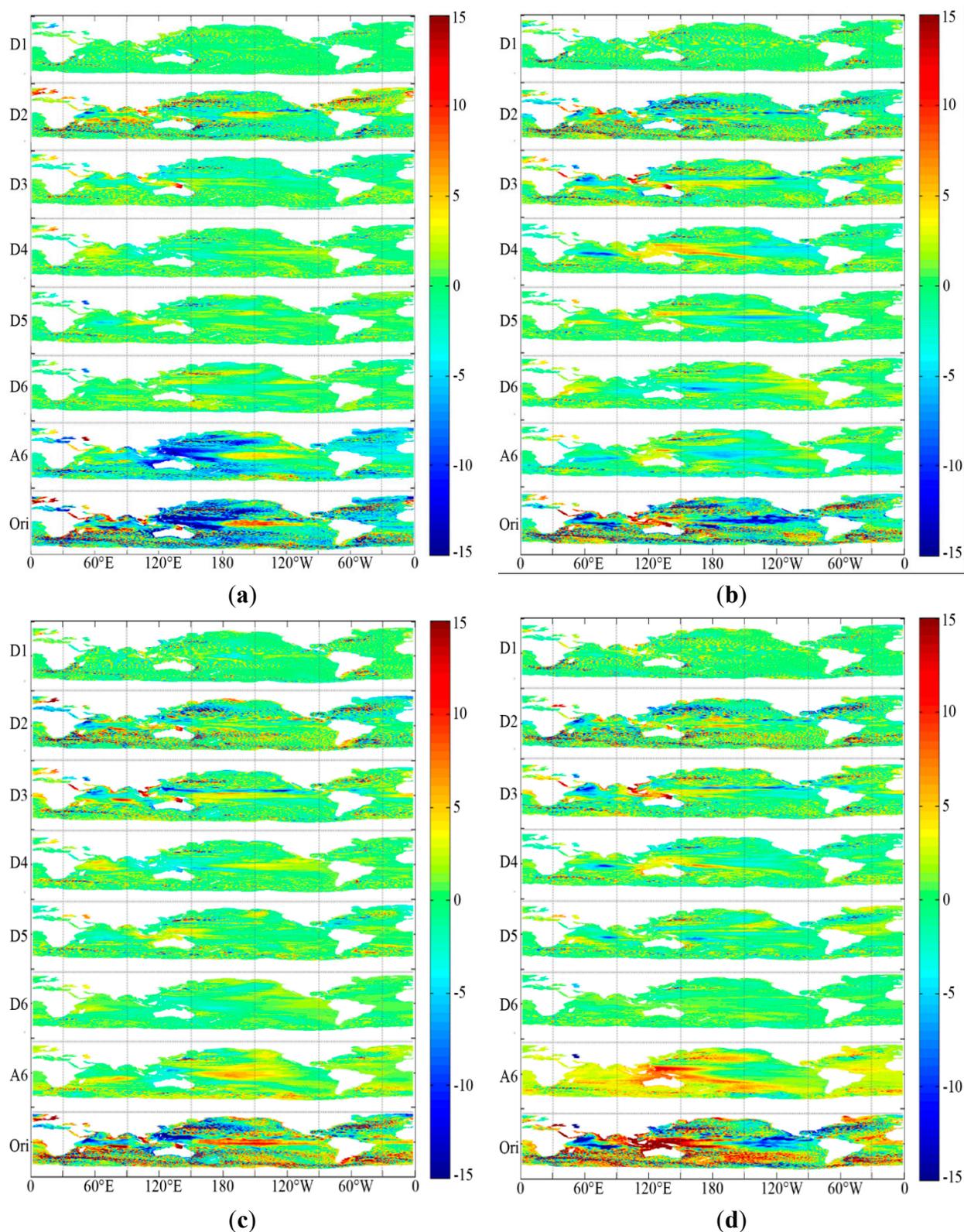


Figure 3. The multi-signal wavelet decomposition results. Four time slices are selected from the whole spatio-temporal data cubes of original data and reconstructed decomposed components ((a) January 1993; (b) January 1999; (c) January 2005; (d) January 2011). The color bar in each subgraph indicates anomaly range of the data. The multi-signal wavelet decomposition not only filters the data in the temporal domain, but also affects the spatial domain. Larger temporal scales will also result in a smoother spatial distribution.

Table 1. Explanation of the Variance from the EOF of the Decomposed Result (%).

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Oridata	31.2	27	16.2	10.5	5.5	4.9	4.3
A6	63.7	16.2	10.7	9.1	0	0	0
D6	31.7	23.7	22.9	19.8	1.7	0	0
D5	31.5	15.4	13	11.5	10.5	9.3	8.4
D4	31.2	17.8	11.8	11.2	10.4	8.8	8.4
D3	52.4	27.8	5.2	4.2	3.5	3.4	3.2
D2	32	22.8	9.8	9.6	8.8	8.5	8.3
D1	16.4	15.6	14.7	13.6	13.3	13	13

Table 2. Peak of PSD and corresponding frequency.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Oridata	0.0833 (12)/	0.0079 (126)/	0.0833 (12)/	0.0833 (12)/	0.0079 (126)/	0.0278 (36)/	0.0079 (126)/
	$1.01 \times 10^{9*}$	4.45×10^7	4.49×10^8	5.33×10^7	7.47×10^7	2.62×10^7	7.79×10^7
A6	0.0079 (126)/	0.0079 (126)/	0.004 (252#)/	0.0079 (126)/	0.0119 (84)/	0.0119 (84)/	0.0159 (63)/
	1.67×10^7	1.45×10^8	1.27×10^8	4.03×10^7	1.73×10^5	1.25×10^3	2.26×10^2
D6	0.0079 (126)/	0.0159 (63)/	0.0119 (84)/	0.0119 (84)/	0.004 (252#)/	0.0198 (50)/	0.004 (252#)/
	9.34×10^7	4.84×10^7	4.52×10^7	6.18×10^7	3.24×10^6	1.65×10^4	1.74×10^2
D5	0.0198 (50)/	0.0278 (36)/	0.0198 (50)/	0.0238 (42)/	0.0159 (63)/	0.0278 (36)/	0.0238 (42)/
	1.06×10^8	5.35×10^7	5.73×10^7	3.12×10^7	1.59×10^7	2.61×10^7	2.35×10^7
D4	0.0357 (28)/	0.0437 (23)/	0.0556 (18)/	0.0397 (25)/	0.0476 (21)/	0.0357 (28)/	0.0437 (23)/
	8.45×10^7	5.17×10^7	2.02×10^7	2.42×10^7	2.07×10^7	2.71×10^7	1.55×10^7
D3	0.0833 (12)/	0.0833 (12)/	0.0714 (14)/	0.0635 (16)/	0.0595 (17)/	0.0714 (14)/	0.0754 (13)/
	7.18×10^8	3.90×10^8	1.85×10^7	1.08×10^7	1.22×10^7	1.34×10^7	1.52×10^7
D2	0.1667 (6)/	0.1667 (6)/	0.1429 (7)/	0.1389 (7)/	0.1389 (7)/	0.127 (8)/	0.1349 (7)/
	1.24×10^8	8.81×10^7	5.47×10^6	1.12×10^7	8.33×10^6	6.87×10^6	4.71×10^6
D1	0.25 (4)/	0.25 (4)/	0.254 (4)/	0.2619 (4)/	0.2579 (4)/	0.3135 (3)/	0.2659 (4)/
	3.11×10^6	1.44×10^7	2.87×10^6	2.89×10^6	2.26×10^6	1.41×10^6	2.75×10^6

XX (YY)/ZZ in the table means principle frequency (principle period)/peak PSD value. The unit of the frequency and principle period is Hz and month, respectively. # in the table means principle frequency (principle period) is not trustable due to the period is larger than the half of the data length. For these three components, no other peaks of the PSD can be founded. This is caused by the limitation of the base theory of FFT. Since all the three components are not the first principle components, their impacts on the frequency characteristics of original data is small.

From Figure 4, we can find that the PCs of the original data, which are not easy to separate, are very similar. The PSD suggests that most of the PCs of original data have strong annual cycle. Most of the differences between different PCs are the phase difference caused by the spatial distribution lag. However, the PCs and the PSD of the PCs suggest the decomposed components are much more separable. The dominant periods of the PSD of each component are significantly different. The major period of the components from A6 to D6 is changing from inter-decades to bi-month. In addition, the PSD of each EOF at each scale can also be classified. This indicates that the patterns of the spatial different are also clearly extracted. The multi-scale decomposition can clearly improve the pattern classification of the data.

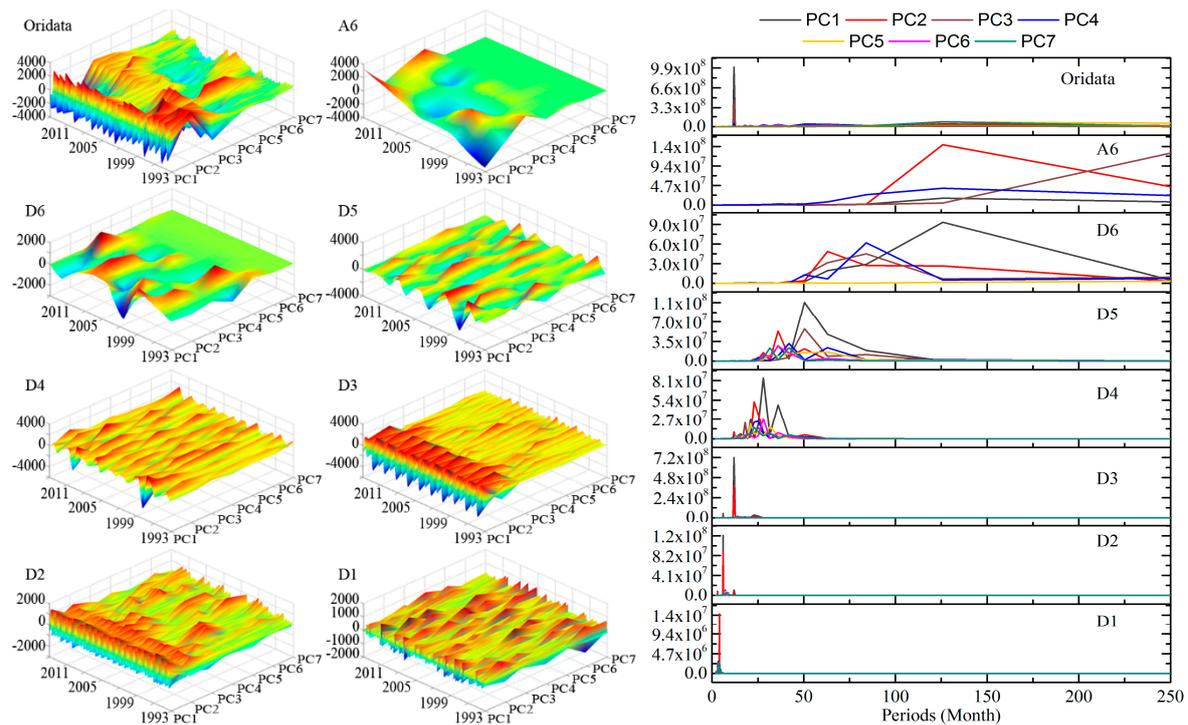


Figure 4. The principal components of original data and each decomposed components and their power spectrums. The 3D plot of the principal components of original data and each decomposed components (**left**); and the power spectrums of each decompose data of each principal component of original data and each decomposed component (**right**). The frequency is changed into period for better illustration.

From the explanation of the variance from the EOF of the decomposed result, the A6 and D6 components are much simpler than the other components, which have fewer principal components. This suggests large-scale variations are more consistent and smooth. Although, the PSD for the time series that are longer than half periods of the whole data length are not very accurate, the combining investigation from the PCs also suggest that both A6 and D6 are inter-decade variations. The A6 components have a period of ten years, which can be seen as the long-term trend of the data. Meanwhile, the D6 components have a quasi-period of 5–7 years. The strong spatial variation in the Equatorial Pacific and Indian Ocean suggests D6 may have close relations to the significant ENSO variations. The D5 and D4 components are also inter-decade variations, which have spatial distribution around the global oceans. The spatial distribution also suggests these two components may also have close relations with the ENSO events. D3 component, which has very high variance contribution of the first PC, has a significant annual cycle. Since the yearly period is the most common and significant period all around the global ocean, the significance of the yearly cycle also proves the consistency of the EOF. The D2 and D1 components have periods of half-a-year and two to four months, respectively. Some irregular high frequency noise is also contained in these components.

From the above discussion, the multi-signal wavelet decomposition can well extract the multi-scale temporal variation from the original multi-temporal remote sensing data. The decomposition is performed in the time domain but also affects the spatial distribution. The decomposed components have clear temporal variation with significant temporal period. By the multi-scale decomposition, the features

of original data are much more separable. Thus, it will be very useful and helpful for feature extraction and data segmentation at different scale levels.

4.3. Feature Clustering Result

The integer approximation and the histogram representation of each decomposed component are performed for the feature clustering. The histogram of the original data and the decomposed multi-scale components are shown in Figure 5. The original histogram only has a wide band peak, because various spatial features are blended together. For this case, it is not very easy to classify different patterns directly from the original histogram. The histogram of the decomposed components is much more separable. Each histogram of the decomposed component has narrow bands and clear single peaks. The values of A6, D6 and D5 are smaller than those of D4–D1 component. This corresponds to the situation where the variation amplitude of the inter-decade cycles is lower than that of the inner-decade cycles in most areas. The segmentation based on the histograms of all the decomposed components is much more structurally clear than segmentation based on the histogram of original data. The cluster distribution of the histogram bins also suggests that the data are more significant and separable than the original data.

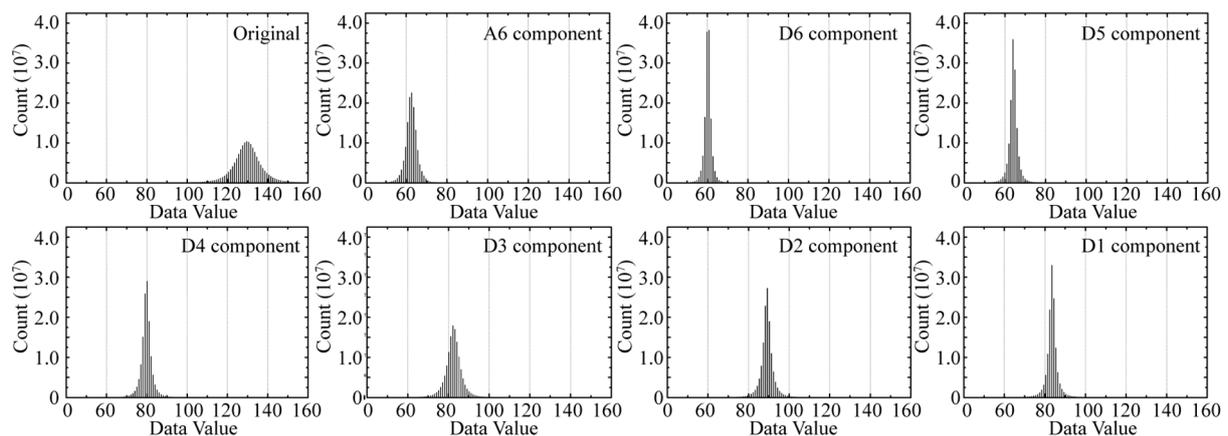


Figure 5. The histogram of the original data and the decomposed components.

Before the final feature clustering at each decomposed component, the optimized cluster number of each decomposed component is calculated from the simulation of the CE index changing according to the cluster number. The simulations are constructed by changing the number of the Clusters from 10 to 40. For each cluster number, the CE is calculated and logged. Figure 6 suggests how the CE changes according to the cluster number at each scale. For all decomposed components, the value of CE grows lower when the cluster number increases. The CE of the original data is significantly higher than that of the decomposed components. And the D1–D4 components have very similar CE. The CE of D5 and D6 components is also very similar, and they all have very low CE values. Unlike other components, the CE of A6 does not change smoothly, which suggests that the complexity degrees of the A6 component may be different. These suggest that the complexity of the inter-decade variations is lower than that of the inner-decade variations. This is partially because the inter-decade variations are much smoother and change more slowly and partially because fewer noises are contained in the long-term variations. We

will notice that in the lower sections the D5 and D6 are mainly influenced by the ENSO event, which is an irregular phenomenon in both spatial and temporal dimensions. Thus this kind of irregularity in CE also demonstrates the characteristics of the signal.

Most of the CE at different scales is changing very slowly after the cluster number gets bigger than 30. This suggests that the clustering with the cluster number bigger than 30 will not produce lots of different information. Therefore, we can consider that the minimal number of 30 will well represent the distribution and feature of original data. Thus, we select the optimized cluster number of 30 for all the decomposed components to produce the final segmentation. Although, it is optimal to select the different optimal cluster number for each level, we select the optimized cluster number of 30 for all the decomposed components according to CE for simplicity.

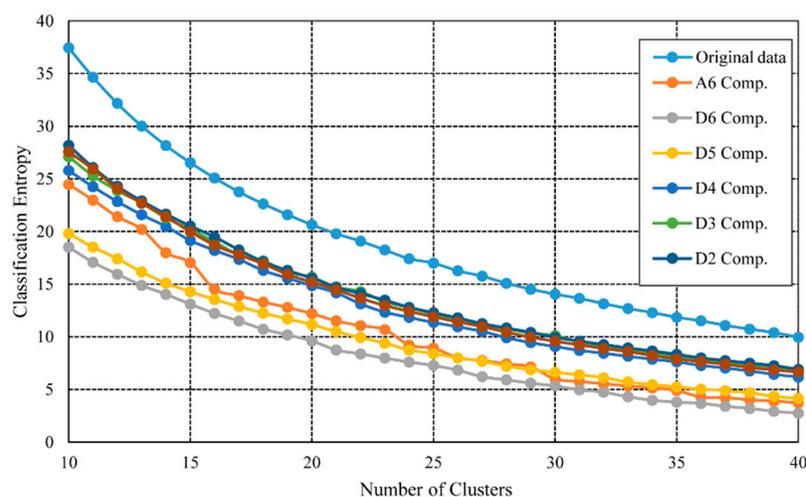


Figure 6. Determination of Cluster Numbers.

The feature clustering with the cluster number 30 is performed to each histogram of the decomposed components. The spatio-temporal distribution of the segmented result is shown in Figure 7. The cluster of the original data does not have a clear significant spatial structure. Lots of small patches can be found in the cluster result, which suggests the original segmentation cannot well separate the variations at different temporal and spatial scales. The cluster result of the decomposed components has much clearer spatial structures. The cluster result of the A6 component suggests clearly two stages before and after the year 2003. This is consistent with the trend change of the sea level height. The cluster result of the D6–D4 components is much more significant in the time dimension. The major changes are the spatial pattern changes crossing the global. Since these components mainly reveal the ENSO signal, detailed comparison of the spatio-temporal changes during ENSO events can be achieved. From the temporal cluster index changes, the evolution process, including the starting, growing, peaking, weakening and vanishing, is much clearer than the original data. The clustering of the D3–D1 components mainly changes periodically through the time which suggests that our cluster can well represent both the spatial and temporal changes of the data. All the above suggest our approach can be used as an exploring tool to find patterns in the spatio-temporal data. Unlike tensor methods, our approach can support irregular data that contain missing data at any spatial locations.

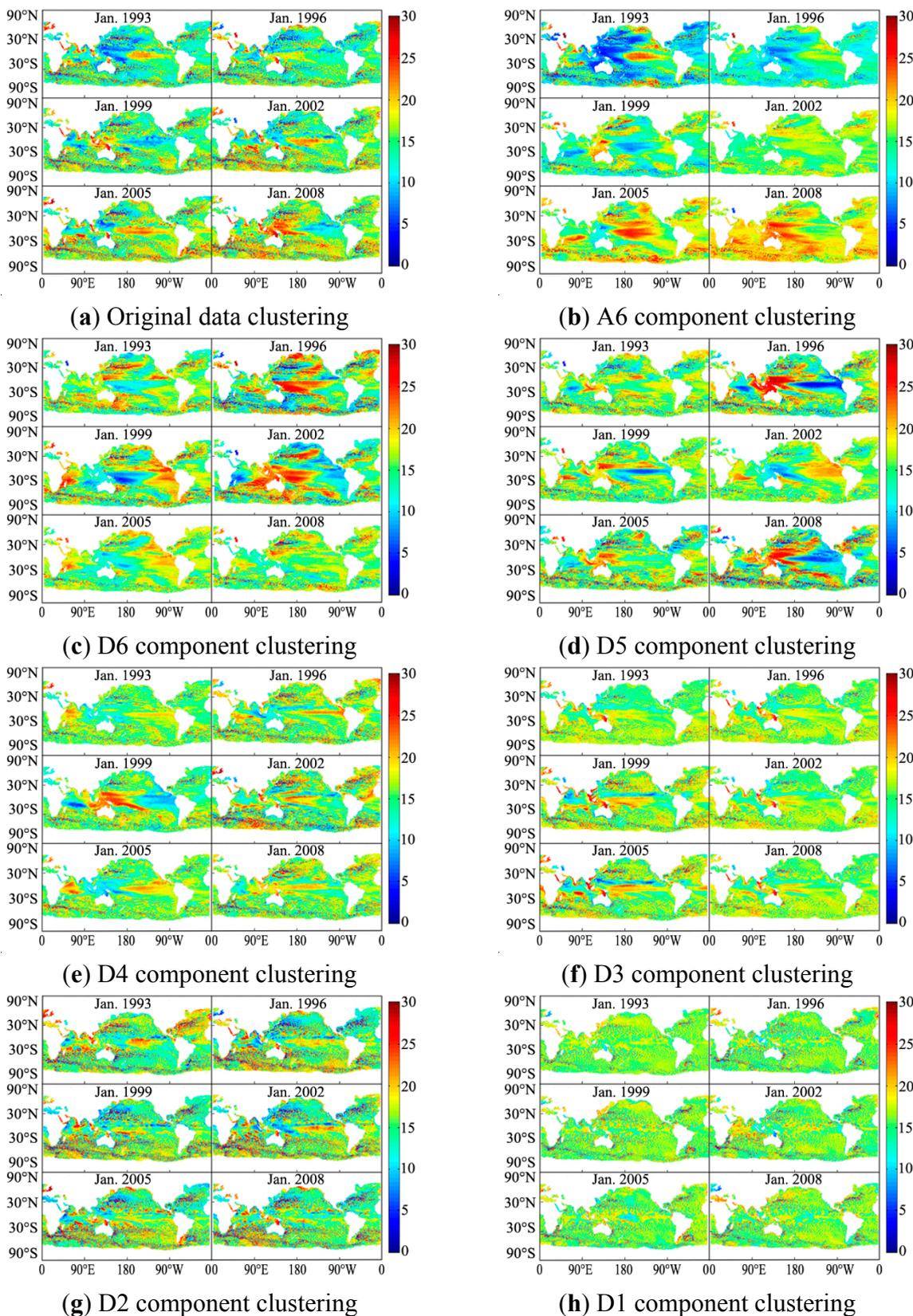


Figure 7. The cluster results of the original data and the decomposed components: (a) cluster result of the original data; and (b–h) cluster results of A6, D6, D5, D4, D3, D2, and D1 components, respectively. The color bar in each subgraph indicates the classification type index of the data. Due to the space limits, only six time slices are selected for demonstration.

5. Evaluations of the Tracked Spatio-Temporal Patterns

5.1. Evaluations by Tracking the 1997–1998 El Niño Evolution

The 1997–1998 El Niño is the strongest observed El Niño event that happened during the research time. The sea level anomalies have been largely affected by this El Niño event. Extensive studies have been performed concerning the pattern of sea level response to this El Niño event, especially in the tropical Pacific area [24]. Most of these response patterns are performed with certain pattern extraction methods (e.g., EOF and ICA) directly from the original data [36]. However, the ENSO signal, which is also a multi-scale phenomenon, often has quasi-periods from two to seven years, and may affect the sea level at several different scales since the interactions between ENSO events and sea level are complex [24–26]. Extracting and tracking the spatio-temporal evolution with the data cluster from the multi-scale decomposition will be helpful to reveal the possible mechanisms about how the ENSO events affects the sea level in the spatio-temporal scale domain.

The primary step of extracting the multiscale interactions between ENSO and sea level is to extract possible ENSO signal from the original SSHA data. Since there is no quantitative simple index in the spatio-temporal domain that can clearly indicate the evolution of ENSO, we use the MEI index in the temporal domain to identify the ENSO signals at different decomposed components. The MEI is compared to the first two principle components of all the decomposed feature series (Figure 8). The correlation coefficients between the principle components and the MEI index are presented in Table 3. From Figure 8, the first principle components (PC1) of D5 and D4 correspond to the MEI index well. The correlations between the PC1s of D5 and D4 and MEI are -0.534 and -0.571 , respectively. Since D6–D4 components represents the sea level variations of inter-decade time scales, it has close correspondence between the ENSO scales. In the second principle components, although the principle components of the original data have high correlations between MEI, and all the decomposed components have weak correlations between MEI. However, from the Figure 8, the PC2s of D5 and D4 also have high correspondence with MEI. Other components, e.g., A6, D3, D2 and D1 only have weak correlations between ENSO signals. Therefore, we can conclude that the D5, D4 may reveal the multi-scale evolution characteristics of ENSO.

Table 3. Time and memory cost of the feature cluster procedure.

Data	Time Cost(s) *		Memory Cost(MB)	
	Our Methods	K-Means	Our Methods	K-Means
Original Data	122.10	742.57	330.71	958.45
A6 Component	116.24	906.87	321.84	962.61
D6 Component	115.39	921.96	337.62	966.98
D5 Component	116.93	894.73	331.56	962.59
D4 Component	116.44	866.97	335.71	966.84
D3 Component	119.57	896.69	311.54	961.58
D2 Component	116.54	829.32	332.74	989.68
D1 Component	119.44	766.87	339.84	969.84

* The time/memory cost of our methods including the time/memory cost of wavelet decomposition, histogram computation and FCM clustering.

The multi-temporal decomposed components D5 and D4 and their cluster results from the May 1997 to August 1998, are selected to validate whether the feature extracted are affected by the 1997–1998 El Niño event. To well represent the data, only the data between 40°N and 40°S are shown in Figures 9 and 10. The selected components have the inter-decade cycle, which contains the El Niño signal, and the selected time range contains all the evolution stages including the starting, growing, peaking, weakening and vanishing stage of this El Niño event. From the decomposed result, both D5 and D4 components have clear patterns that can reveal the existence of the El Niño event. However, the detailed tracking of this El Niño event is still difficult directly from the decomposed data. So the feature clustering using the FCM is further applied to extract the multiscale spatio-temporal evolution of the El Niño event.

The cluster of the decomposed components shows how the El Niño changes in the spatio-temporal domain much clearer than original components data. The D5 components suggest a significant El Niño pattern from May 1997 to February 1998 in Tropic Pacific Ocean. The tongue of the high sea level has been moving from 180° to 120°W. After March 1998, the high sea level in the Tropic Pacific is growing weaker. However, peak heights are located in the West Tropic Indian Ocean. The locations and boundaries are comparable with the existing literature [37,38]. The significance of the Indian Ocean changes in this El Niño event, which cannot be directly seen and extracted from the original and decomposed components, has already been proved in the literature [39]. The D4 components are slightly different from the D5 components. The westward high sea level starts from September 1997 and ends in April 1998. The mixture of two significant peaks of the westward trade wind at different scales is one of the causes why this El Niño becomes the strongest El Niño [40,41].

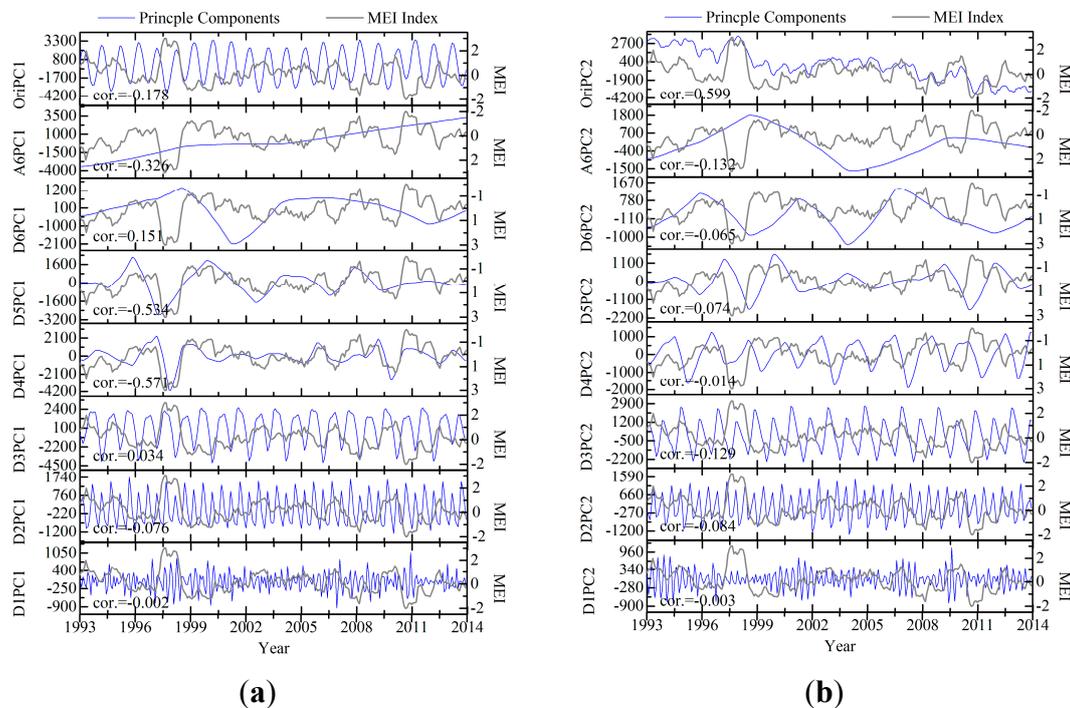


Figure 8. The comparison between the principle components of the decomposed components and the MEI index: (a) comparison between the first principle components of the decomposed components and the MEI index; and (b) comparison between the second principle components of the decomposed components and the MEI index.

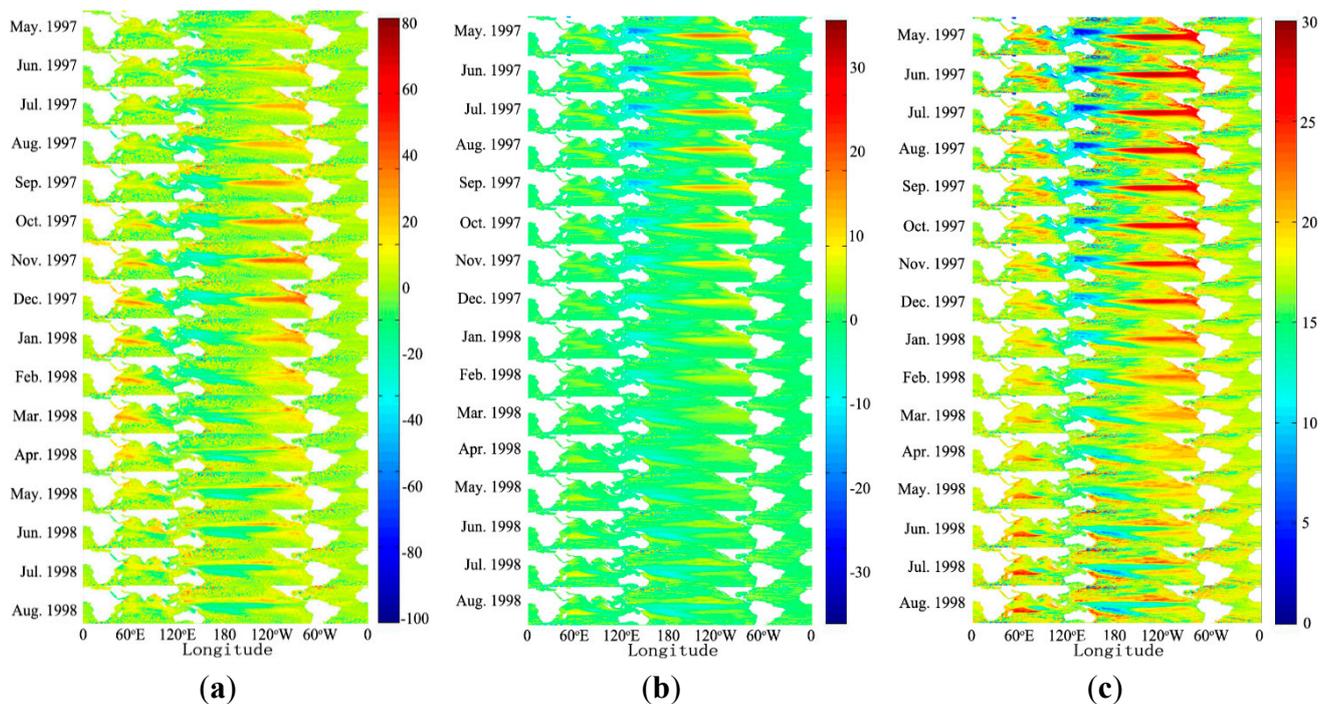


Figure 9. The spatio-temporal evolution of the D5 component during the 1997–1998 El Niño: (a) the original data; (b) the D5 components; and (c) the cluster of D5 components.

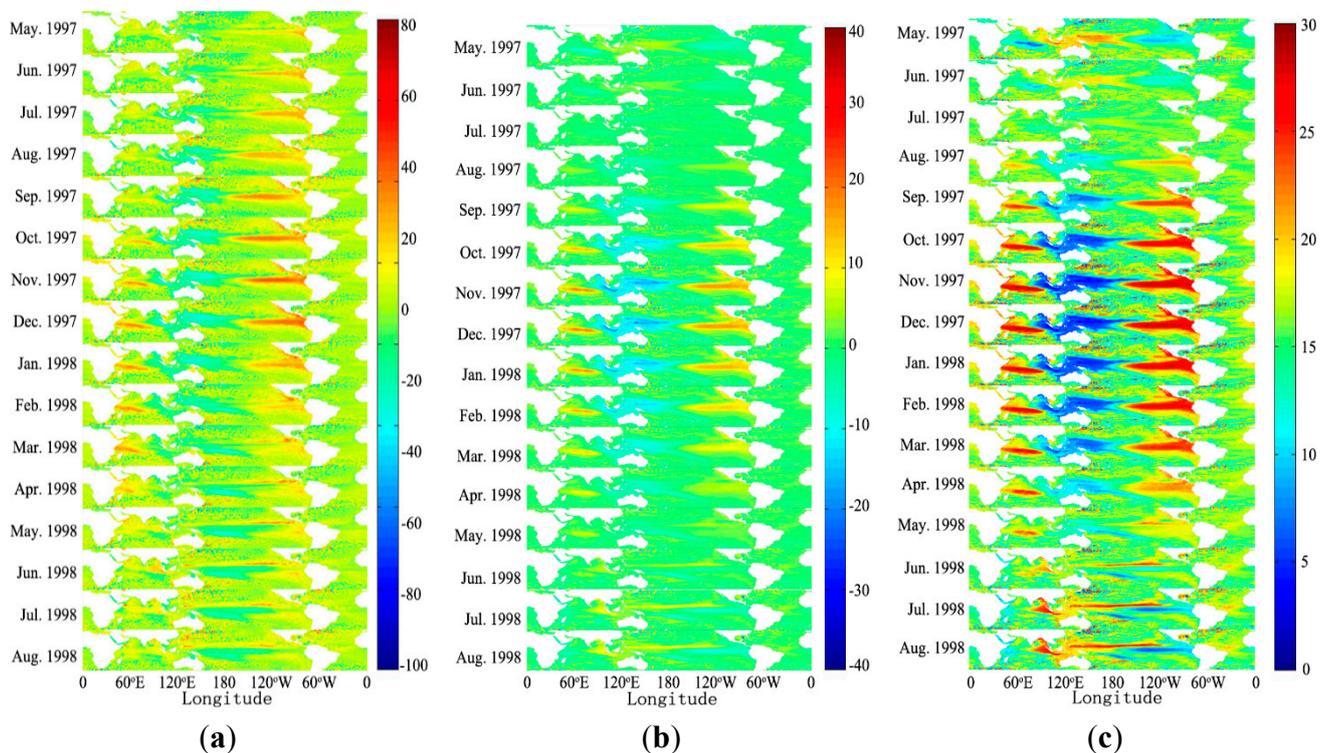


Figure 10. The spatio-temporal evolution of the D4 component during the 1997–1998 El Niño: (a) the original data; (b) the D4 components; and (c) the cluster of D4 components.

In most recent studies, there are two different El Niño events, the East Pacific El Niño (EP) and Central Pacific El Niño (CP), which may dominate the whole ENSO cycle [42]. The two different types of El Niño events can be distinguished by the starting location of the event in either East Pacific or Central Pacific.

However, the two different events are interlinked and are not easy to clearly separate. There are researches suggest the 1997 strong El Niño are caused by both two different El Niño events [43,44]. In our result, the spatial pattern difference between the D4 and D5 partly support such results. The cluster of the D5 component has its highest center in the Central Pacific while the D4 component has its center in only East Pacific. The start time of the D4 component agrees well with the time that the bloom of this El Niño event happens [45]. This suggests the two signals have different affections in spatio-temporal domain, which can be clearly revealed using the feature clustering of the spatio-temporal decomposed components at different scales. Since this pattern can not be observed clearly in both original data and the decomposed components, the cluster can reveal the multi-scale pattern from the multi-temporal remote sensing data and greatly enhance the representation. In other words, the proposed method can reveal the spatio-temporal evolution of certain kinds of patterns or events more clearly.

5.2. Computational Performance Evaluation

For the computational complexity, both theoretical computational complexity and experiment results are computed. For our method, the theoretical computational complexity is composed of the following parts: the complexity of the multi-scale decomposition, the integer representation, the histogram construction, the FCM and the cluster reverse mapping to the original data. For the multi-scale decomposition part, the theoretical complexity of the multi-signal wavelet decomposition is about $kO(n \log(n))$, where k is the number of decomposed levels; the complexity of the integer representation, histogram construction and the reverse mapping are all $O(n)$, where n is the total data count. The complexity of the FCM runs asymptotically in $O(Nc^2p)$ time, where N is the number of p -dimensional observations (*i.e.*, sum of number of histogram bins of each levels.) and c is the number of clusters [46]. Therefore, the overall complexity of the method is $kO(n \log(n)) + O(n) + O(Nc^2p)$. This computational complexity is much smaller than that of the original FCM, and it is customizable according to the number of the histogram bins.

The computational efficiency and the spatial continuity of the feature cluster are evaluated in the experiments. The computational time and the memory cost of the feature cluster procedure of the original data and decomposed components are logged in Table 3. From Table 3, we can clearly see that the time cost and memory cost of different components are very similar, which suggests the computational performance of our method is stable. The computation time and the memory cost are about 1/7 and 1/3, respectively, which are less than that of classical K-Means cluster, since the computational time and memory cost are affected greatly by the histogram representation. More detailed technologies that can enhance the patterns and reduce the bins of the histogram can reduce the computational time and memory cost even more. In conclusion, our method is efficient to deal with large amount of multi-temporal remote sensing data.

6. Discussion

In our approach, the multi-signal wavelet decomposition was only used to decompose the multi-temporal remote sensing data along the time dimension to extract the temporal multi-scale variations. However, the channel-to-channel (location to location) relations at different scales during the decomposition, which also reveal the spatial interaction and distribution of the signal at certain temporal scales, are not fully studied. Revealing the detailed channel-to-channel relations during the

decomposition and further integrate them in spatio-temporal feature cluster may also very helpful. In the integer representation, the adaptive selection of the bins to construct the histogram is important for data that have closed value in a very narrow band of the integer. Selecting the threshold adaptively according to the data value distribution to construct the histogram can solve this. The information or morphology criteria (e.g., entropy) can be used for the adaptive histogram construction. However, special attention should be paid to avoid the pseudo mode that may be produced by the equalizer. Since both the integer representation and the histogram representation are very fundamental representations, more advanced approximation in the time-frequency space (e.g., the representation of the data with Fourier basis or wavelet basis) may produce much better results. With the spectrum representation theory, the characteristics of the data may be extracted with much more details at a much lower computational cost.

The histogram re-organize similar data values into a single histogram bin, it interlinks the value distribution and the spatio-temporal distribution of the data. The integer representation, which is simple and efficient, not only enhances the feature pattern but also reduces the computational complexity. The soft threshold of the FCM results applied to the histogram can results in more consistency and continuity of the spatio-temporal feature cluster. In addition, the cluster is data adaptive, and a prior knowledge about numbers of objects/features and the distribution of the original data is not required. However, the cluster performance of the method can still be improved in several aspects. In the FCM, the selection of the optimal cluster numbers is still experimental and the efficiency of the selection procedure can also be optimized. In the reversed mapping procedure from the cluster of the histogram to the original data type, the constraints of spatial continuity can be integrated, which can further control the fine level of details of the cluster.

7. Conclusions

In this paper, we propose a new approach that integrates the multi-scale temporal decomposition into the feature cluster of the multi-temporal remote sensing data. The multi-signal wavelet decomposition, the integer approximation and histogram representation and FCM methods are combined to extract the segments of the significant features at different time scales. Our proposed method is efficient and stable, and can be used for extracting and tracking complex geographical phenomena such as El Niño events from the noisy multi-temporal remote sensing data. Experimental results demonstrate that our method can be used as an exploring analysis tool and a foundation to construct methods for more efficient and accurate spatio-temporal feature extraction and feature cluster of multi-temporal remote sensing or other spatio-temporal data.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 41471319, 41231173, 41201377), NCET program (Grant No. NCET-12-0735), and the PAPD project. Support from the Vilas Associate Award, the Hammel Faculty Fellow Award, the Manasse Chair Professorship from the University of Wisconsin-Madison, and the “One-Thousand Talents” Program of China are greatly appreciated. The authors also wish to acknowledge Ming-Song Xu for his constructive comments on this paper.

Author Contributions

All four authors have contributed to the work presented in this paper. Wen Luo, Zhao-Yuan Yu and Lin-Wang Yuan formed the original idea. Wen Luo, Zhao-Yuan Yu and Sheng-Jun Xiao worked on the experiment and data analysis. Wen Luo, Zhao-Yuan Yu and A-Xing Zhu worked on the adaptive FCM algorithm. Wen Luo and Lin-Wang Yuan worked for the computation complexity analysis and evaluation. All authors worked collaboratively on writing the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Zhang, X.; Xiao, P.; Feng, X.; Wang, J.; Wang, Z. Hybrid region merging method for segmentation of high-resolution remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2014**, *98*, 19–28.
2. Ouma, Y.O.; Josaphat, S.S.; Tateishi, R. Multiscale remote sensing data segmentation and post-segmentation change detection based on logical modeling: Theoretical exposition and experimental results for forestland cover change analysis. *Comput. Geosci.* **2008**, *34*, 715–737.
3. Kalbermatten, M.; van de Ville, D.; Turberg, P.; Tuia, D.; Joost, S. Multiscale analysis of geomorphological and geological features in high resolution digital elevation models using the wavelet transform. *Geomorphology* **2012**, *138*, 352–363.
4. Gray, J.; Song, C. Mapping leaf area index using spatial, spectral, and temporal information from multiple sensors. *Remote Sens. Environ.* **2012**, *119*, 173–183.
5. Colditz, R.R.; Schmidt, M.; Conrad, C.; Hansen, M.C.; Dech, S. Land cover classification with coarse spatial resolution data to derive continuous and discrete maps for complex regions. *Remote Sens. Environ.* **2011**, *115*, 3264–3275.
6. Zhang, C.; Lin, H.; Chen, M.; Li, R.; Zeng, Z. Scale compatibility analysis in geographic process research: A case study of a meteorological simulation in Hong Kong. *Appl. Geogr.* **2014**, *52*, 135–143.
7. Del Ventisette, C.; Righini, G.; Moretti, S.; Casagli, N. Multitemporal landslides inventory map updating using spaceborne SAR analysis. *Int. J. Appl. Earth Observ. Geoinform.* **2014**, *30*, 238–246.
8. Corbane, C.; Jacob, F.; Raclot, D.; Albergel, J.; Andrieux, P. Multitemporal analysis of hydrological soil surface characteristics using aerial photos: A case study on a Mediterranean vineyard. *Int. J. Appl. Earth Observ. Geoinform.* **2012**, *18*, 356–367.
9. Ardila, J.P.; Bijker, W.; Tolpekin, V.A.; Stein, A. Multitemporal change detection of urban trees using localized region-based active contours in VHR images. *Remote Sens. Environ.* **2012**, *124*, 413–426.
10. Pedersen, T.B.; Jensen, C.S. Multidimensional database technology. *Computer* **2001**, *34*, 40–46.
11. Jensen, C.S.; Kligys, A.; Pedersen, T.B.; Timko, I. Multidimensional data modeling for location-based services. *VLDB J.* **2004**, *13*, 1–21.
12. Baumann, P.; Dumitru, A.M.; Merticariu, V. The array database that is not a database: File based array query answering in Rasdaman. In Proceedings of the 13th International Symposium on Advances in Spatial and Temporal Databases, Munich, Germany, 21–23 August 2013.

13. Soroush, E.; Balazinska, M.; Krughoff, S.; Connolly, A. Efficient iterative processing in the SciDB parallel array engine. In Proceedings of the 27th International Conference on Scientific and Statistical Database Management (SSDBM 2015), San Diego, CA, USA, 29 June–1 July 2005.
14. Cheung, D.W.; Zhou, B.; Kao, B.; Kan, H.; Lee, S.D. Towards the building of a dense-region-based OLAP system. *Data Knowl. Eng.* **2001**, *36*, 1–27.
15. Karayannidis, N.; Sellis, T. Hierarchical clustering for OLAP: The CUBE file approach. *VLDB J.* **2008**, *17*, 621–655.
16. Smith, J.R.; Li C.; Jhingran, A. A wavelet framework for adapting data cube views for OLAP. *IEEE Trans. Knowl. Data Eng.* **2004**, *16*, 552–565.
17. Scotese, C.R.; Rees, P.M. The Spatial-Temporal Information Matrix (STIM) cube: An efficient way to store geological information. In Proceedings of the 2004 Annual Meeting of the Geological Society of America, Denver, CO, USA, 7–10 November 2004.
18. Yuan, L.; Yu, Z.; Chen, S.; Luo, W.; Wang, Y.; Lv, G. CAUSTA: Clifford algebra based unified spatial-temporal analysis. *Trans. GIS* **2010**, *14*, 59–83.
19. Yuan, L.; Yu, Z.; Luo, W.; Hu, Y.; Feng, L.; Zhu, A. A Hierarchical tensor-based approach to compressing, updating and querying geospatial data. *IEEE Trans. Knowl. Data Eng.* **2015**, *27*, 312–325.
20. Hu, Z.Q.; Liu, L.P. Applications of wavelet analysis in differential propagation phase shift data de-noising. *Adv. Atmos. Sci.* **2014**, *31*, 825–835.
21. Kaplan, N.H.; Erer, I. Bilateral filtering-based enhanced pansharpening of multispectral satellite images. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 1941–1945.
22. Yuan, L.; Yu, Z.; Yi, L.; Luo, W.; Chen, S. Multiscale spatial decomposition for skew-distributed data with parallel spatial kernel smoothing. *Environ. Plan. B Plan. Des.* **2014**, *41*, 613–636.
23. Xue, C.; Song, W.; Qin, L.; Dong, Q.; Wen, X. A spatiotemporal mining framework for abnormal association patterns in marine environments with a time series of remote sensing images. *Int. J. Appl. Earth Observ.* **2015**, *38*, 105–114.
24. Yu, Z.Y.; Yuan, L.W.; Lu, G.N.; Luo, W.; Xie, Z.R. Coupling characteristics of zonal and meridional sea level change revealed by satellite altimetry data and their response to ENSO events. *Chin. J. Geophys.* **2011**, *54*, 1972–1982. (in Chinese)
25. Hoell, A.; Barlow, M.; Wheeler, M.C.; Funk, C. Disruptions of el nino-southern oscillation teleconnections by the maddenjulian oscillation. *Geophys. Res. Lett.* **2014**, *41*, 998–1004.
26. Oliver, E.C.J. Intraseasonal variability of sea level and circulation in the Gulf of Thailand: The role of the Madden-Julian Oscillation. *Clim. Dyn.* **2014**, *42*, 401–416.
27. Luisier, F.; Blu, T. Sure-let multichannel image denoising: Interscale orthonormal wavelet thresholding. *IEEE Trans. Image Process.* **2008**, *17*, 482–492.
28. Yuan, L.W.; Yu, Z.Y.; Chen, S.F.; Luo, W.; Wang, Y.J.; Lu, G.N. Causta: Clifford algebra-based unified spatio-temporal analysis. *Trans. GIS* **2010**, *14*, 59–83.
29. Sindhumol, S.; Kumar, A.; Balakrishnan, K. Automated brain tissue classification by multisignal wavelet decomposition and independent component analysis. *ISRN Biomed. Imaging* **2013**, *2013*, doi:10.1155/2013/473437.

30. Chen, S.C.; Zhang, D.Q. Robust image segmentation using fcm with spatial constraints based on new kernel-induced distance measure. *IEEE Trans. Syst. Man Cybernet. B Cybernet.* **2004**, *34*, 1907–1916.
31. Cai, W.L.; Chen, S.C.; Zhang, D.Q. Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation. *Pattern Recognit.* **2007**, *40*, 825–838.
32. Baldi, P.; Brunak, S.; Chauvin, Y.; Andersen, C.A.F.; Nielsen, H. Assessing the accuracy of prediction algorithms for classification: An overview. *Bioinformatics* **2000**, *16*, 412–424.
33. AVISO. The Altimeter Products were Produced by Ssalto/Duacs and Distributed by AVISO, with Support from CNES. Available online: <http://www.aviso.altimetry.fr/duacs/> (accessed on 3 November 2014).
34. Wolter, K.; Timlin, M.S. El niño/southern oscillation behaviour since 1871 as diagnosed in an extended multivariate enso index (mei.Ext). *Int. J. Climatol.* **2011**, *31*, 1074–1087.
35. Percival, D.B.; Walden, A.T. *Wavelet Methods for Time Series Analysis*; Cambridge University Press: Cambridge, UK, 2000.
36. Westra, S.; Brown, C.; Lall, U.; Koch, I.; Sharma, A. Interpreting variability in global SST data using independent component analysis and principal component analysis. *Int. J. Climatol.* **2010**, *30*, 333–346.
37. Jena, B.; Sahu, S.; Avinash, K.; Swain, D. Observation of oligotrophic gyre variability in the South Indian Ocean: Environmental forcing and biological response. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **2013**, *80*, 1–10.
38. Macías, D.; Franks, P.J.S.; Ohman, M.D.; Landry, M.R. Modeling the effects of coastal wind- and wind-stress curl-driven upwellings on plankton dynamics in the Southern California current system. *J. Mar. Syst.* **2012**, *94*, 107–119.
39. Lu, Q.; Zuo, J.; Li, Y.; Chen, M. Interannual sea level variability in the tropical pacific ocean from 1993 to 2006. *Glob. Planet. Change* **2013**, *107*, 70–81.
40. Chen, D.; Cane, M.A. El Niño prediction and predictability. *J. Comput. Phys.* **2008**, *227*, 3625–3640.
41. Yu, Z.; Luo, W.; Yi, L.; Hu, Y.; Yuan L. Clifford algebra-based structure filtering analysis for geophysical vector fields. *Nonlin. Process. Geophys.* **2013**, *20*, 563–570.
42. Ashok, K.; Yamagata, T. Climate change: The El Niño with a difference. *Nature* **2009**, *461*, 481–484.
43. Grinsted, A.; Moore, J.C.; Jevrejeva, S. Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlin. Process. Geophys.* **2004**, *11*, 561–566.
44. Liu, Q.; Liu, Z.; Pan, A. Conceptual model about the interaction between El Niño/Southern Oscillation and Quasi-Biennial Oscillation in far west equatorial Pacific. *Sci. China Ser. D Earth Sci.* **2006**, *49*, 889–896.
45. Wen, L.; Lin, Y.; Yu, Z.; Sun, H.; Yuan, L. Barycenter reflected equatorial Pacific sea level structure evolution and its indication of ENSO events. *Oceanologia* **2015**, *57*, 107–115.
46. Kolen, J.F.; Hutcheson, T. Reducing the time complexity of the fuzzy c-means algorithm. *IEEE Trans. Fuzzy Syst.* **2002**, *10*, 263–267.