

Multi-satellite Monitoring SST Data Fusion based on the Adaptive Threshold Clustering Algorithm

Hongmei Shi[1], Han Dong[2], Lingyu Xu[3], Cuicui Song[4], Fei Zhong[5], Ruidan Su[6]
 [1, 3-5]Department of Computer Engineering & Science, Shanghai University, Shanghai, China, 200072
 [2]The National Marine Information Centre, Tianjin, China, 300171
 [6] College of Information Science and Engineering, Northeastern University, Shenyang, China 110004
 Email: { [1] shmei1988@163.com , [3] xly@shu.edu.cn }

Abstract—This paper proposes a method which describes the information precision with a soft fusion model, instead of the traditional rigid fusion method. The method is divided into two steps, the pretreatment model and fusion center model. Each forms a relative independent model, and the two models have a progressive relationship. The former is used for consistency evaluation, data cleaning and invalid data eliminating, while the latter provides fusion results and variable precision fusion expression by the adaptive threshold clustering algorithm. Experimental results show that the fusion method can not only give every SST data a different precision, but also carry more information to describe precision multiple distribution, which make users get high-quality data and enjoy more rights.

Index Terms—information fusion, sea surface temperature (SST), variable precision, threshold, clustering, adaptive threshold

I. INTRODUCTION

Multi-source remote sensing information fusion, one of the intense topics of international remote sensing over several years, is becoming a key technology to process multi-source massive data fusion. The development is transforming from research into function, but the previous papers on the remote sensing sea surface temperature (SST) data fusion were only limited in some localities. In order to overcome the bottleneck of single technology, and achieve the complementary advantages, foreign scholars have started to comprehensively study on marine monitoring regarding information fusion since 2000. There are typical composite fusion test platforms which are Marine environment monitoring and information integrated system (SEAWATCH) developed by Norwegian, Marine environment remote-control measurement and detection system (MEAMAID) developed by Germany [1]. MISST [2] (Multi-sensor Improved Sea Surface Temperatures) can get improved, high resolution, global coverage, near real-time SST data by fusing the infrared and microwave sensor data. Besides, Bovith et al. [4] have detected SST elements through multi-remote sensing fusion image processing. Oesch et al. [5] used multi-scale remote sensing fusion to observe tepidity

situation in lake. Kozai et al. [6] used multi-satellite remote sensing data to fuse the sea surface wind.

In China, the state 2nd oceanic administration Ma [7] pointed out that it's difficult to practically operate the common assimilation method in the specific application due to their respective shortcomings, so he proposed to use fusion method to deal with the assimilation problems of the common Marine observation data. Guan [8] adopted the objective analysis to fuse TMI, AVHRR, TRMM/VIRS satellites. Shi [9] first proposed the concept of soft fusion quantitative precision in remote sensing and assimilation product fusion. Xu [10] first gave the fusion products and the precision analysis in East China seas by using multi-remote sensing fusion method.

However, the research results mentioned above are all used traditional rigid fusion method, which means the rigid processing method is used through the fusion process to filter out fuzzy and inconsistent information, and finally certain results are got, but it doesn't have a precise description to the accuracy and reliability of the fusion point, and user can't learn a conflict among the data sources just from those products, they have no idea of which points are fused in certain circumstances, and which are in hesitation. This paper proposes a more advanced fusion method with the adaptive threshold clustering algorithm, and it's divided into two steps, the pretreatment model and the fusion center model, each forms a relative independent model, and the two models have a progressive relationship. The former is used for consistency evaluation, data cleaning and eliminating invalid data, while the latter provides fusion results and variable precision fusion expression by the adaptive threshold clustering algorithm. The first part of this paper introduces the classic fixed threshold clustering algorithms; The second part proposes the adaptive threshold clustering algorithm to make up the insufficient of the fixed threshold clustering algorithm; The third part is the case analysis of the adaptive threshold clustering algorithm; The fourth part is fusion result comparison and analysis.

II. CLASSIC FIXED THRESHOLD CLUSTERING ALGORITHM

A. The Idea of Classic Algorithm

The basic process of classic fixed threshold clustering algorithm is determine the threshold which is fit for all data samples according to the priori knowledge, and then do clustering analysis to data samples under the threshold, and finally obtain the clustering results.

B. The Description of the Algorithm

Input: the $N(N \geq 2)$ unspecified data samples X_N , threshold T

Output: M class data results

Steps:

Step1 Sort the unspecified data samples in ascending order ($X_1 \leq X_2 \leq \dots \leq X_N$).

Step2 Select the first member X_1 as the first group data for the first class C_1 .

Step3 Calculate the distance d_{1i} between X_1 and X_i in dataset $X_i (i = 2, \dots, N)$ orderly, which $d_{1i} = |X_1 - X_i|$. Classify $X_1 \sim X_{k-1}$ as Class C_1 until a data point X_k that makes $d_{1k} \geq T$, and classify X_k to a new Class C_2 , and take X_k as the first element of Class C_2 as well.

Step4 Similarly, calculate the distance between X_k and the data after X_k . Get all the elements of the class and the first element of next class.

Step5 Repeat step3 and step4 operations, until finishing all the samples and get M kinds of data totally.

C. The Adaptability Analysis of the Algorithm to SST Observation

The advantage of the algorithm is its simple calculation, and if there is a priori knowledge of the sample distribution for the threshold value selection, we can get reasonable clustering results quickly. However, this algorithm also has a series of problems. In practice, it is difficult to obtain accurate priori knowledge for high dimensional model samples. Therefore, we can only choose different threshold to tempt, and thus the clustering results largely depend on the choice of threshold T. Different thresholds lead to different clustering results, large threshold will get smaller cluster number, and small threshold may get more cluster number relatively, but there is no criterion to determine whether the large threshold effect, or small threshold effect.

III. THE ADAPTIVE THRESHOLD CLUSTERING ALGORITHM

A. The Thought of the Algorithm

When the user lacks priori knowledge of clustering data, it is difficult to determine the appropriate threshold, and often needs to execute many times of experiment and compare the test results to find the most suitable threshold. For the fixed threshold clustering, once the user changes the threshold value, it often needs to restart the clustering computing. Each-time clustering would cost a lot of time, and sometimes the clustering efficiency is very poor. Therefore, on the basis of the existing fixed threshold, the adaptive threshold algorithm is proposed.

B. The Description of the Algorithm

Input: SST data of thirteen satellites

Output: inside-class data and outside-class data after clustering, variable precision fusion expression, and fusion reliability.

Steps:

1. Extract thirteen satellites' temperatures of a certain day in the same latitude and longitude.

2. Input the minimum threshold and maximum threshold.

3. Calculate the distance D between any two of thirteen temperature data (totally $13 * 13$ distance), discard those temperature data which the calculated distance are greater than the minimum threshold, and find out the largest frequency temperature data involved in the calculation among the remaining distances, remember the temperature and those temperatures which the distance D less than the threshold as inside-class temperature, the rest of the temperatures as outside-class temperature.

4. If the number of inside-class temperature less than the outside-class temperature, then increase step length (increase one-tenth of the difference of the maximum threshold and minimum threshold) and return to step 3. If it is still not satisfied conditions until the threshold value greater than the maximum threshold, then exit algorithm and the fusion is failed. Otherwise, if the number of inside-class temperature greater than the outside-class temperature, then jump to Step5.

5. The inside-class temperature data recorded as (1), outside-class temperature data noted for (0), then according to the inside-class temperature data, calculate the fusion expression, the weighted center temperature, the center temperature, and the fusion reliability.

6. Computing formula of fusion expression: fusion expression = fusion temperature \pm error.

Set: the inside-class maximum temperature recorded as C_Max, inside-class minimum temperature recorded as C_Min, fusion temperature recorded as C_fusion, the error record as Error, fusion expression denoted by E, then:

$$\text{Fusion temperature } C_{\text{fusion}} = \frac{C_{\text{Max}} + C_{\text{Min}}}{2} \tag{1}$$

$$\text{Error} = \frac{C_{\text{Max}} - C_{\text{Min}}}{2} \tag{2}$$

$$\text{Fusion expression } E = C_{\text{fusion}} \pm \text{Error} = \frac{C_{\text{Max}} + C_{\text{Min}}}{2} \pm \frac{C_{\text{Max}} - C_{\text{Min}}}{2} \tag{3}$$

7. The center temperature calculation: the center temperature is the average value of all the inside-class temperature. Then, $W' = \frac{W1 + W2 + \dots + Wn}{n}$ (4)

8. Calculate reliability C_reliability: the number of inside-class temperature number recorded as C_num, effective temperature denoted by C_num',

$$\text{Then, } C_{\text{reliability}} = \frac{C_{\text{num}}}{C_{\text{num}'}} \tag{5}$$

9. Weighted center calculation: set central temperature as C', the weight P, the distance between all inside-class temperature and center temperature is $D = |C_1 - C'| | C_2 - C'| | \dots | C_n - C'|$ (6)

$P_i = 1 - (|C_i - C'| / D)$ as to ensure the smaller distance from the center temperature C' , the greater the weight. Calculate weighted center:

$$W = \frac{P_1 * W_1 + P_2 * W_2 + \dots + P_n * W_n}{W_1 + W_2 + \dots + W_n} \tag{7}$$

C. Advantages

Compared to fixed threshold algorithm, the adaptive threshold clustering algorithm has many advantages. First, this algorithm can get satisfied clustering results, because the results of the fixed threshold clustering largely depend on the choice of threshold T, different threshold leads to different clustering result and clustering effect is not ideal. Sometimes there's only a few or no data within the class, making the clustering without any meaning. While the adaptive threshold clustering algorithm guaranteed the clustering results reasonable. Second, the clustering speed is fast, because we define a threshold interval, it only needs to traverse the data space one time to achieve clustering. Third, using the proposed adaptive threshold clustering algorithm, user needs less prior knowledge but two input threshold parameters. For the set size of threshold, the user can test several times to determine it, and finally obtain the satisfactory clustering results. When the user changes threshold to re-clustering, this algorithm can get all kinds of different particle size of clustering.

IV. EXPERIMENT AND ANALYSIS

A. Algorithm Example

According to the proposed adaptive threshold clustering algorithm, fuse the SST data on the 9th day of 2006. SST data products including AVHRR, MODIS, TMI, MCSST of 13 types.

Example procedure:

1. Read all the data of thirteen satellites on the 9th day of 2006. Take the temperature of the point (38.875 °C, 128.125 °C) as example, the SST data is expressed in TABLE I.

TABLE I.
THE SST DATA IN (38.875 °C, 128.125 °C)

Satellite name	SST data
Aavhrr0.25	-
Amsre-A0.25	-
Amsre-D0.25	-
Clim0.25	5.712
Davhrr0.25	4.275
Modisasst40.25	3.979
Modisasstd0.25	3.978
Modisasstn0.25	3.936
Modistsst40.25	4.269
Modistsstd0.25	-
Modistsstn0.25	4.292
Tmi-A0.25	-
Tmi-D0.25	-

2. Let the minimum threshold value be 0.15, and the maximum threshold value be 1.0.

3. According to Step3 of the algorithm, calculate the distance between any two of the seven valid temperature data, and write down the distance following in the corresponding satellites in order, which is shown in TABLE II.

TABLE II.
ALL THE DISTANCES OF THE SEVEN VALID SST DATA

Clim0.25	Davhrr0.25	Modisasst40.25	Modisasstd0.25	Modisasstn0.25	Modistsst40.25	Modistsstn0.25
5.712	4.275	3.979	3.978	3.936	4.269	4.292
0	1.437	1.733	1.734	1.776	1.443	1.42
1.437	0	0.296	0.297	0/339	0.006	0.017
1.733	0.296	0	0.001	0.043	0.29	0.313
1.734	0.297	0.001	0	0.042	0.291	0.314
1.776	0.339	0.043	0.042	0	0.333	0.356
1.443	0.006	0.29	0.291	0.333	0	0.023
1.42	0.017	0.313	0.314	0.356	0.023	0

When threshold T = 0.15, according to Step3 of the algorithm, there are only three inside-class temperature while the outside has four, the result can be seen in TABLE III.

TABLE III.
THE NUMBER OF INSIDE-CLASS SST DATA WHEN THRESHOLD IS 0.15

Clim0.25	Davhrr0.25	Modisast40.2	Modisastd0.2	Modisastn0.2	Modistsst40.2	Modistsstn0.2	Inside-class data number when T=0.15
5.712	4.27	3.97	3.97	3.93	4.26	4.29	
	5	9	8	6	9	2	
0	0	0	0.00	0.04	0.00	0.01	1
		0.00	1	3	6	7	3
		1	0	0.04			3
	0.00	0.04	0.04	2			3
	6	3	2	0	0	0.02	3
	0.01				0.02	3	3
	7				3	0	

According to step4 of the algorithm, a fixed step size (0.085) increased the threshold, then T = 0.235, empathy in step 3, obtained three data inside the class, and four outside the class; Then add another step, so T = 0.32, get six data within the class, and one outside the class, thus, the number of inside-class data greater than the outside-class, which fit the termination conditions. At this time, (1) is used to represent the six inside-class data, and the rest one is expressed by (0), as shown below in TABLE IV.

TABLE IV.
THE NUMBER OF INSIDE-CLASS SST DATA WHEN THRESHOLD IS 0.32

Cli m0. 25	Dav hrr0. 25	Mod isass t40.2 5	Mod isass td0.2 5	Mod isass tn0.2 5	Mod istsst 40.2 5	Mod istsst n0.2 5	Inside- class data number when T=0.32
5.71 2 (0)	4.27 5 (1)	3.97 9 (1)	3.97 8 (1)	3.93 6 (1)	4.26 (1)	4.29 2 (1)	
0	0	0.29	0.29			0.00	0.01
0	0.29	6	0.00		0.04	6	7
	6	0	1	0.04	0.29	0.31	5
	0.29	0.00	0	3	0.29	3	6
	7	1	0	0.04	1	0.31	6
		0.04	0.04	2		4	3
	0.00	3	0.29	0	0		5
	6	0.29	1		0.02	0.02	5
	0.01	0.31	0.31		3	3	
	7	3	4		3	0	

4. Calculate fusion expression, fusion expression = fusion temperature ± error.

By the formula (1),

$$C_{fusion} = \frac{C_{Sat1} + C_{Sat2}}{2} = \frac{4.262 + 3.936}{2} = 4.114,$$

By the formula (2),

$$Error = \frac{C_{Sat1} - C_{Sat2}}{2} = \frac{4.262 - 3.936}{2} = 0.178,$$

Therefore, from formula (3),

$$Fusion \quad expression \quad E = C_{fusion} \pm Error = \frac{C_{Sat1} + C_{Sat2}}{2} \pm \frac{C_{Sat1} - C_{Sat2}}{2} = 4.114 \pm 0.178.$$

Reliability calculation by formula (5),
 $C_reliability = \frac{C_num}{C_num + 7} = \frac{6}{7} = 85.7\%.$

B. The Experimental Analysis of Adaptive Threshold Clustering Algorithm

The proposed adaptive threshold clustering algorithm can apply to the project of multi-satellites monitoring SST data fusion. This project mainly studies the soft fusion method, which is divided into pretreatment and fusion center models, each forms a relative independent model, and two models have a progressive relationship. The former is used for consistency evaluation and data cleaning, while the latter provides fusion results and variable precision fusion expression by the adaptive threshold value clustering algorithm.

The pretreatment model is mainly used for the basic data collection and preliminary quality control. Process the multi-format, multi-precision, multi-source remote sensing data in east China sea with AVHRR, MODIS, TMI, MISST products and etc. Considering the data loss of the multi-resolution, single monitor products and other factors, we preliminary sort the SST data and eliminate the noise.

According to above adaptive threshold clustering algorithm, read the SST data on the first day of 2006 as experimental data. Satellite temperature data including 13 satellites, which are AVHRR, MODIS, AMSR-E, TMI and etc. TABLE V. shows the pretreatment results.

TABLE V.
THE PRETREATMENT RESULTS ON THE 1st DAY OF 2006

Latitude	Longitude	SST data												
20.625	118.875	-	24.9	25.2	24.74	-	-	24.511	-	25.251	24.405	24.886	-	24.9
20.625	119.125	-	24.45	25.05	24.712	-	-	24.453	-	24.462	24.173	24.108	-	24.75
20.625	119.375	-	24.15	24.9	24.962	-	-	24.185	-	24.554	24.076	24.185	-	24.9
20.625	119.625	-	24.45	25.5	25.017	-	-	24.126	-	24.335	-	23.86	-	25.5
20.625	119.875	-	24.9	26.55	25.065	-	-	24.302	-	24.467	-	23.956	-	25.95
20.625	120.125	-	25.5	26.85	25.267	-	-	24.345	-	-	-	-	-	26.4

The pretreatment module mainly picks up the satellite data, eliminates the invalid data files of each satellite, finally generates the pretreatment files to make the following fusion operation more conveniently, and improves fusion efficiency as well.

In order to overcome the shortcomings of the fusion results ever express less information, and the poor transparency, this paper establishes a soft fusion model, explored with multiple precision/reliability distribution of continuous valued fusion algorithm to solve the multiple source conflict problem of three-dimensional monitoring. It can improve the quality as well as quantify precision further, and provide the user more right to know.

In order to overcome the poor effect by using the fixed threshold cluster method, we introduce the adaptive

threshold clustering method. First, according to the user's needs, input two thresholds, the minimum threshold and maximum threshold, to form a threshold interval. And make the threshold interval divided equally in a step length, compare the distance between every two satellite data circularly, it stopped until the number of the inside-class data more than half of effective data, and thus the clustering finished.

Since the threshold of each observation point is different, the error is also different, which eventually leads to different precision, resulting for each observation point a variable precision fusion expression, which can be expressed as the measured value ± measurement errors.

Fusion thirteen satellite data on January 1, 2006, fusion results stored in fusion result file (TABLE VI.)

TABLE VI.
FUSION RESULTS ON JANUARY 1,2006

Latitude	Longitude	Final threshold	Cluster result						Variable precision fusion result	reliability
29.375	128.125	0.15	22.35 (1)	22.233 (1)	22.209 (1)	22.333 (1)	22.5 (1)		22.355±0.146	100%
26.875	123.375	0.15	22.05 (1)	21.9 (1)	20.691 (0)	21.9 (1)	21.75 (1)		21.9±0.15	80%
24.375	126.125	0.405	23.4 (1)	24.358 (0)	23.744 (1)	23.772 (1)	23.395 (1)	23.775(1)	23.585±0.190	83%
21.375	119.375	0.405	25.2 (1)	25.2 (1)	24.855 (1)	24.857 (1)	24.826 (1)	24.45(1)	24.825±0.375	100%
23.875	122.875	0.825	23.85 (1)	23.85 (1)	24.635 (0)	22.955 (1)	23.028 (1)	22.8 (1) 23.85 (1)	23.325±0.525	85%

In order to more directly see the advantages of the adaptive threshold clustering algorithm than classic fixed threshold algorithm, Figure 1 and Figure 2 show all the SST data of the two algorithms in the first day of 2006 respectively. Fig. 1 set the threshold interval [0.15,1.0],and from which get that the threshold is variable, different locations have different threshold. Finally, this algorithm obtained more accurate and reliable fusion results which based on point-to-point precision. While Fig. 2 only has a threshold value (0.15), and the result is equal precision that we can't determine which point is relatively reliable from the figure, and the transparency is very poor.

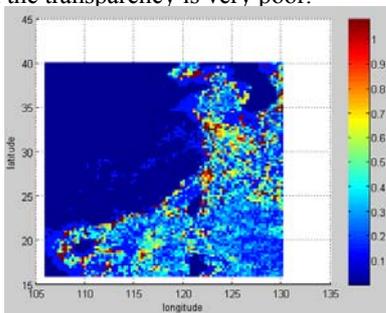


Figure 1. Threshold figure of adaptive clustering algorithm

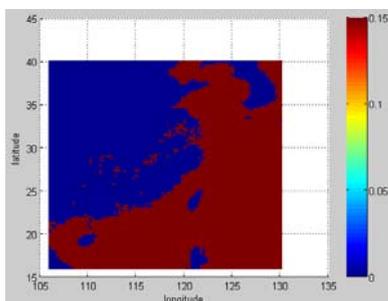


Figure 2. Threshold figure of classic fixed clustering algorithm

We use formula (7) calculate the weighted center value of all inside-class data in different latitude and longitude, to

describe the variable precision fusion result, and the result shows in Fig. 3.

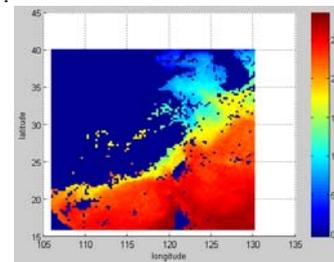


Figure 3. Variable precision fusion result of the weighted center

Similarly, Fig. 4 and Fig. 5 show the number of the inside-class data of the two algorithms, we can see that in some sea area, the number of inside-class data in Fig. 4 were significantly more than that in Fig. 5, which means that the reliability of the adaptive clustering algorithm has significantly increased than fixed threshold algorithm. The experiment shows that the algorithm proposed in this paper is more reliable.

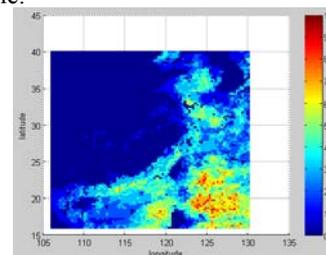


Figure 4. The number of the inside-class data using adaptive threshold clustering algorithm

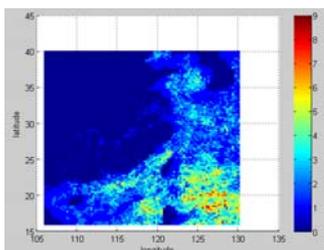


Figure 5. The number of the inside-class data using fixed threshold clustering algorithm

V. CONCLUSION

Adaptive threshold clustering algorithm are proposed on the basis of the existing fixed threshold clustering algorithm, the results of the fixed threshold clustering largely depend on the choice of threshold T , different threshold leads to different clustering result and clustering effect is not ideal. However, adaptive threshold clustering algorithm doesn't have this kind of problem, the threshold of different longitude and latitude is different, and then the final fusion expression is also different, thus realize the soft fusion which is put forward in this paper very well.

This paper proposes a soft fusion model which is divided into two models, pretreatment model and fusion center model, each forms a relative independent model, and two models have a progressive relationship. The former is used for consistency evaluation, data cleaning and invalid data eliminating, while the latter provides fusion results and variable precision fusion expression through the adaptive threshold clustering algorithm.

ACKNOWLEDGMENT

This work is supported by The National Natural Science Foundation of China. (No.40976108)

This work is supported by Shanghai Leading Academic Discipline Project in China. (NO. J50103)

The Ocean Public Welfare Project of The Ministry of Science and Technology under Grant No.201105033.

This work is supported by Innovation Action Plan supported by Science and Technology Commission of Shanghai Municipality (No.11511500200)

REFERENCES

- [1] Shi Suixiang, Study on Soft Fusion Strategy of Multi-Channel Incertitude Information in Digital
- [2] ocean[D]. Northeastern University, 2005:7-10.
- [3] <http://www.misst.org/>
- [4] Zhang hui. Merging AVHRR and AMSR-E Sea Surface Temperature Data Based on Wavelet Transform[D].qingdao: Ocean University of China, 2006:4-9.
- [5] Bovith, Thomas, Nielsen, Allan Aasbjerg, Hansen, Lars Kai, Overgaard, Soren, Gill, Rashpal S.. Detecting weather radar clutter by information fusion with satellite images and numerical weather prediction model output. In proc: 2006 IEEE International Geoscience and Remote Sensing Symposium, 2006, 6: 511-514.
- [6] Oesch, D., Jaquet, J. M., Klaus, R. Schenker.. Multi-scale thermal pattern monitoring of a large lake (Lake Geneva) using a multi-sensor approach. International Journal of Remote Sensing, 2008, Vol.29, No.31, pp: 5785-5808.
- [7] Kozai K., Ishida K., Shiozaki T., Okada Y.. Wind-induced upwelling in the western equatorial Pacific Ocean observed by multi-satellite sensors, Advances in Space Research, 2004, Vol.33, No.7, pp: 1189-1194.
- [8] Ma Jianpu, Huang Daji, Zhang Benzao. Blending method and its application in data assimilation[J]. Acta Oceanologica Sinica, 2003, Vol.25, No.2, pp: 33-41.
- [9] Guan lei, Chen Rui, He mingxia. Validation of Sea Surface Temperature from ERS-1/ATSR in the Tropical and Northwest Pacific[J]. Journal of Remote Sensing, 01, 63-69 (2007).
- [10] Shi Suixiang, Xia Dengwen, Yu ge. Method of fusing the data reversed from satellite remote-sensing with the assimilation data for sea surface temperature based on D-S theory[J]. Acta Oceanologica Sinica, 2005, Vol.27, No.4, pp: 31-37.
- [11] Zheng Jinwu, Xu Dongfeng, Xu Mingquan. A review of merging methods of all covered high resolution SST. Journal of Tropical Oceanography, 2008, Vol.27 No.4, pp: 77-81.
- [12] Sun Ji-Gui, Liu Jie, Zhao Lian-Yu, "Clustering Algorithms Research", Journal of Software, Vol.19, No.1, January 2008, pp.48-61
- [13] Lingyu Xu, Jun Xu, "Satellite Remote Sensing SST Quality Evaluation Based on Consensus Measurement", 2009 International Conference on Web Information Systems and Mining, pp807-810.
- [14] Zhong fei, Liu na, Liu yang, Xu Lingyu. New SST correction method from multi-satellite based on the coefficient of variation. Journal of Shanghai University (English Edition), 2011, 15(5):463-466.