Homogenization methods for the Sea Surface Temperature Data over the South China Seas

G S Wang[1], M Hou[2], Y Li[1, 3], H Wang[1], W J Fan[1], K X Liu[1], J Gao[1], C Li[1]

[1] National Marine Data and Information Service, Tianjin 300171, China;
[2] Meteorological Bureau of Binhai New Area, Tianjin 300457, China

E-mail address: ok8200@sina.com

Abstract. Based on the metadata, Monthly Sea Surface Temperature (SST) series from nine marine stations over the South China Sea (SCS) are homogeneity detection and correction by Penalized Maximum T Test (PMT) method. The reference stations are developed using surrounding meteorological stations. Correction results show that: (1) The homogeneity detection and correction of marine observation stations should be based on the metadata, meanwhile, fully consider the influence of regional climate change factors. (2) Correlation analysis found that, the air temperature series from the surrounding meteorological stations is currently the optimal reference series. (3) The marine stations has 1~2 change points average, among them, changes of instrumentation and changes of location environment has great impact on the discontinuities. (4) The trend of the SST over SCS have a more pronounced warming trend during the past 52 years. Correction results indicate that the homogenization research in the SCS has an important meaning for the study of the SCS coast SST changes and climate change.

1. Introduction

The South China Sea (SCS) is the most typical monsoon region in the world, and it is influenced by the local Hadley circulation [1-2]. The inter-annual variability of SST over the SCS has important significance for the climate change research [3-5]. The quality of climate data is of extreme relevance for environment analysis, geographic information and remote sensing. Due to the changes of observation environment, measuring instrumentation, observation method or the human habit , the observed climate data sequences often includes many non-natural irregularities, which could lead to errors and bias in the results of environment analysis. For instance, the conclusion of coastal environment change research may has certain uncertainty caused by the influence of inhomogeneity [6]. As a result, the SCS marine observation data need to be detection and correction of irregularities.

Many countries take the climatological data homogeneity as an important task, and has develop many homogeneity datasets, which as the primary reference for the IPCC report [7]. Numerous studies has found [8-9], climatological data homogeneity detection must use direct and indirect homogenisation methods in both, meanwhile, absolute and relative homogenization methods are also very popular homogenization methods for climatic variables. The application of metadata is very important in homogeneity detection [10]. The Standard Normal Homogeneity Test (SNHT), Multiple Analysis of Series for Homogenization (MASH), Penalised Maximal F-test (PMF) and Penalised Maximal T-test (PMT) have been widespread used for assessing the significance trend in hydrological time series. Stephenson [10] was systematic evaluation of the homogeneity of daily SST for the Caribbean and neighbouring regions by used RHtest software. Zhai Panmao [11] firstly analysed gross

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

errors and biases in China's historical radio sonde datasets. Li Qingxiang [12] combination statistical methods with station metadata, designed several tests and tested for inhomogeneity in all Chinese historical surface air temperature series from 1951 to 2001.

In summary, many domestic and scholars have made various researches on homogeneity detection and correction. However, previous studies focused on the meteorological data, and less for the SST data over the SCS. This paper will based on the metadata, surrounding meteorological stations as reference stations, and climate rationality analysis was carried out on the detection of change points and correction, monthly SST series of nine marine stations in coastal zone of the SCS are from 1960 to 2001 would be homogeneity detection and correction by PMT method.

2. Data and Methods

2.1. Data

The SST series form marine stations over the SCS used in this paper, which are obtained from the National Marine Data and Information Service (NMDIS), has been quality control checked , includes checks for physically unreasonable values, unreasonably long consecutive occurrences of the same values , illegal code, consistency inspection , and so on. The missing data has been interpolated by multiple linear regression method (Qinglan, Dongshan and Nanao marine station exit some missing data). SST series time range is from 1960 to 2011, the position distribution is shown in Figure 1.

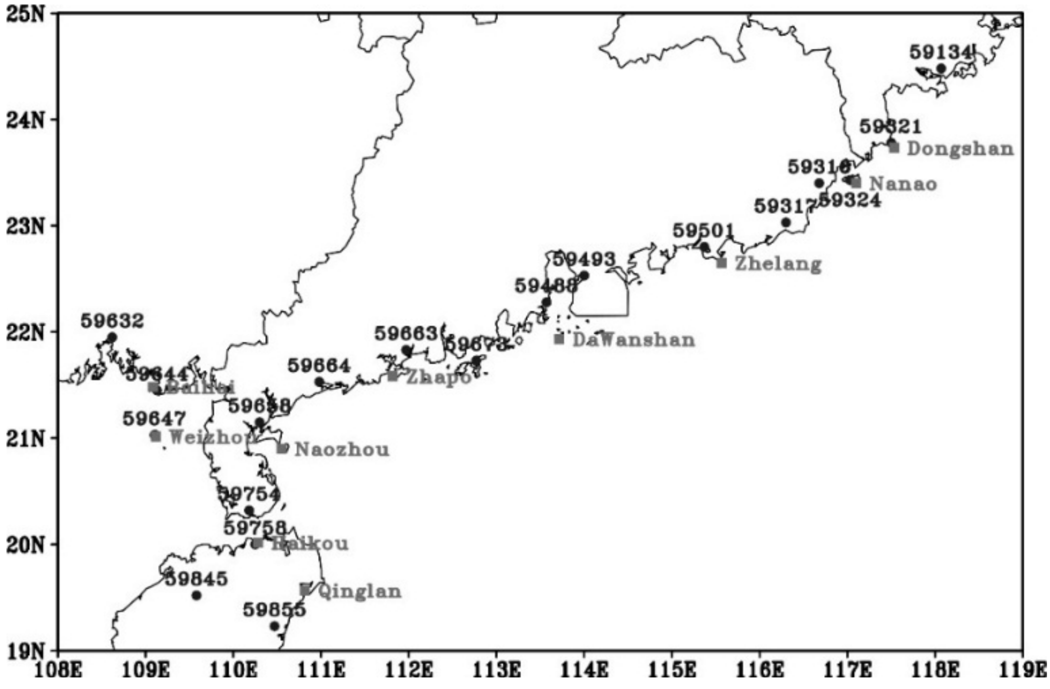

Figure 1. Locations of marine stations (block) and meteorological stations (circle) around the SCS.

The metadata used in this paper are obtained from the NMDIS, including relocation of monitoring stations, changes of measuring instrumentation, observing practices or environment change, and so on. In addition, the monthly mean surface air temperature (SAT) from meteorological observation stations over the SCS are obtained from National Meteorological Information Centre (NMIC). Surface air temperature series from 1960 to2011, has been quality controlled and procedures to detect and correct inconsistencies have been applied to homogenize the data. The position distribution is also shown in Figure 1.

The SST reanalysis data used in this paper includes: The Hadley Centre Global Sea Ice and Sea Surface Temperature (HadISST v1.1) data (1°×1°, 1961-2011) is provided by UK Met Office Hadley Centre. OISST (Optimum Interpolation sea Surface temperature, 1°×1°, 1981-2011,) monthly mean reanalysis data [13] and higher resolution daily average SST datasets (OISST.v2, 1/4°×1/4°, 1981-2011) [14] are both provided by US NOAA (National Oceanic and Atmospheric Administration) .(HadISST v1.1 data can be download from https://climatedataguide.ucar.edu/ climate-data/sst-data-hadisst-v11, OISST data can be download from http://www.ncdc.noaa.gov/oisst/data-access, OISST.v2 data can be download from ftp://eclipse.ncdc.noaa.gov/pub/OI-daily-v2/)

2.2. Methods

We use the RHtest version 4 software package to detect and adjust inhomogeneity in this study. This package includes the PMTred algorithm, which is based on the penalized maximum t (PMT) test. This algorithm deal with multiple change points using a recursive testing algorithm and account for the first order autocorrelation. The PMT test is greatly depend on the reference series. Since the much lower density of the SCS marine stations, it is difficult to choose the SST time series from the nearby marine stations to make the reference series. Compared to other homogenization methodology, PMT method can eliminate local trend signal from climate change, obtain better correction and results, but this method requires complete and high-quality reference comparisons. Rayner [14] adopted HadISST1 datasets (1°×1°) as a reference sequence to homogenise the candidate stations and has achieved good results. Since the SCS stations location of sparse, it is difficult to choose the right surrounding stations (reference stations) to homogenise the candidate station.

Table 1. The correlation coefficients of SSTA. The most significant correlation coefficients at confidence 0.01 level are in bold.

Marine Stations (1960-2011)	Meteorological Stations (1960-2011)			HadISST1 (1960-2011) 1°×1°	OISST (1981-2011) 1°×1°	OISST.v2 (1981-2011) 1/4°×1/4°
Baihai	59644 **0.907**	59647 **0.931**	59632 **0.856**	0.395	0.335	0.023
Dongshan	59324 **0.790**	59134 **0.746**	59321 **0.833**	-0.124	0.058	-0.134
Da Wanshan	59488 **0.568**	59673 **0.569**	59493 **0.544**	0.022	-0.088	-0.120
Haikou	59758 **0.776**	59754 **0.788**	59855 **0.765**	0.420	**0.492**	**0.613**
Naozhou	59658 **0.84**	59754 **0.81**	59664 **0.83**	0.370	**0.558**	**0.704**
Qinglan	59855 **0.856**	59845 **0.836**	59758 **0.784**	-0.022	-0.341	-0.381
Weizhou	59647 **0.819**	59644 **0.761**	59754 **0.772**	**0.528**	**0.471**	**0.654**
Nanao	59324 **0.768**	59321 **0.810**	59316 **0.720**	-0.130	-0.043	-0.186
Zhelang	59493 **0.674**	59317 **0.748**	59501 **0.759**	**0.681**	**0.506**	**0.582**
Zhapo	59663 **0.876**	59664 **0.890**	59673 **0.898**	0.378	0.409	**0.665**
Average		0.784		0.252	0.225	0.347

Studies have found that [15], the correlations must be greater than 0.70, which is indicate that at least 50% of the correction sequence variability is captured by the reference series. Therefore, this article compared the SST anomaly of candidate station correlation between the SAT data (distance criterion), OISST and HadISST, respectively. The HadISST1 data sets largely span 1960-2011, while the OISST and OISST.v2 data span 1981-2011. Due to the different lengths of the SST time series from different sources, we calculate the correlation coefficient between SST time series from candidate stations and the SAT from meteorological stations and SST time series from the International gridded data sets on different period which is based on the common data set lengths, respectively. In our study, when n=52, rα=0.05=0.27, rα=0.01=0.35; when n=31, rα=0.05=0.35, rα=0.01=0.45, here α denotes the confidence level. The results show that the average correlation coefficient of the SAT data was 0.784, which is much higher than OISST, OISST.v2 and HadISST1 datasets (Table 1).

Based on the above analysis, three surrounding meteorological stations, which is the correlation coefficient is higher than 0.70, is selected as reference stations for each candidate station. Because of lack of metadata information, and the correlation coefficients is lower than 0.70, the Da Wanshan station is temporarily revised. The Zhelang station uses only two meteorological stations as a reference stations due to the low correlation coefficient with No.59493 meteorological station. For each tested station, the arithmetic average of the three or two SAT time series from the reference stations is defined as a reference series. Using the PMT method as well as the reference series, the SST time series of the candidate stations are detected and adjusted. Consequently, the statistical break points will be obtained.

3. Detection of change points over the SCS

Tan Jun [1] pointed out that the SST anomaly over SCS has close relation to the ENSO phenomenon. Wang Weiqiang [4] found that the SST over SCS has inverted relationship with the Southern Oscillation Index. Based on the preliminary analysis of the marine station SST series over the SCS (Figure 2), the interannual variability of SST over the SCS has closely related with the climate change. In the process of station data correction, the inhomogeneity may be superimposed on a true climate trend [16], so the breakpoints are generally more difficult to detect. Several statistical methods have been proposed for detection of “undocumented change points” [7]. It is need to be fully consider the impact of local climate factors in the detection of “undocumented change points” of SST over the SCS. In this paper, the SST over the SCS are conducted by using PMT test methods [17]. Change points that are significant at the 95% level are significant change points for further investigation. Except for Qinglan station and Nanao station, the remaining seven candidate stations have full and detail metadata. Here in Dongshan station and Nanao station stand two typical example, briefly explain the process of detection of “undocumented change points” and detection of “documented change points”. Then, the identified change points are adjusted using the quantile-matching (QM) adjustment method which is also included in the RHtest-v4 software.

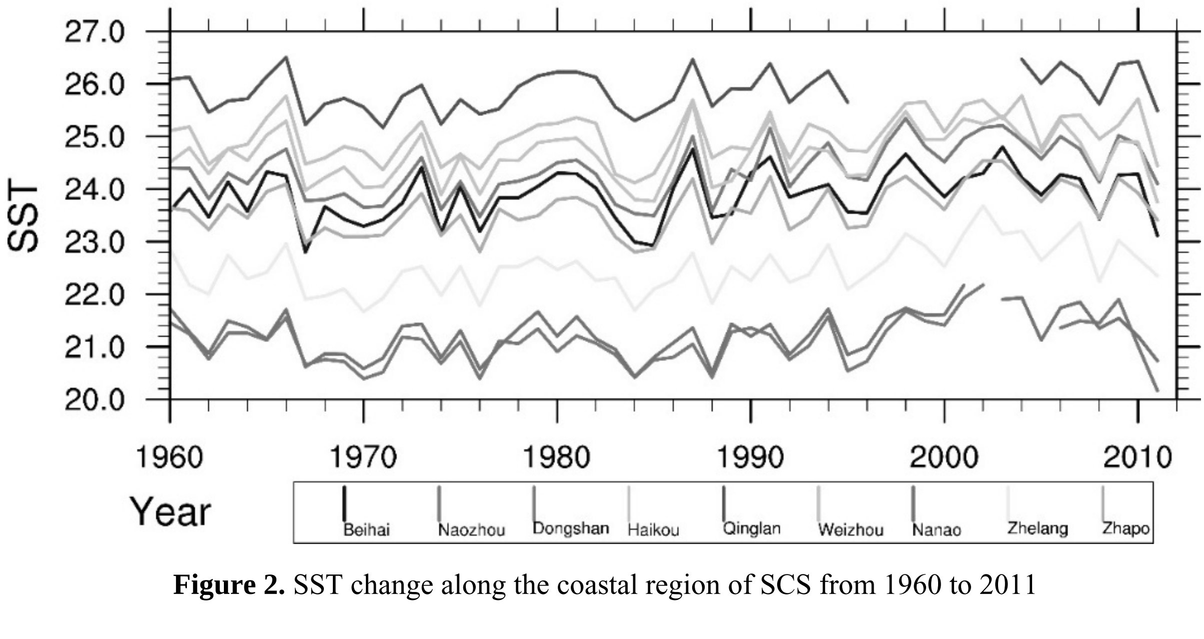

Figure 2. SST change along the coastal region of SCS from 1960 to 2011

3.1. Detection of documented change points

Based on metadata information of Dongshan station can be found, the temperature and salinity measuring site of Dongshan station was change in January 1969, the new station site of about 1km far away the old site. Since in September 1959, the temperature measurement instrument had changed many times (Table 2). Early adopt the SWL1-1 type SST thermometer measurements. The SST observation was cancelled in July 2002, and restore observation in April 2006. In January 2007 started using YSI-600-1 SST automation system for SST observation. According to the above metadata, Dongshan station has experienced a transformation from manual observation to automated observation in 2006-2007 period, observation instrument accuracy also exists a substantial change.

Figure 3 shows plots of the anomalies of the difference series between SST and SAT. The plots indicate that, the SST anomaly time series has a significant shifts in the 1969-70 period compare with SAT anomaly time series. This is the new location evident for the 1969-70 period for Dongshan station. Meanwhile, the SST anomaly time series has a significant shifts in the 2006-07 period compare with SAT anomaly time series, which is the new measuring instrumentation and new observing procedures evident for the 2006-07 period for Dongshan station.

Synthetic analysis indicates that, the SST anomaly time series for Dongshan station has been detection of change points following variance comparisons with SAT anomaly time series. This approach is considered to be proper and reasonable because the homogeneous SAT series works to detection of change points of SST anomaly time series.

Table 2. Changes of measuring instrumentation.

No.	Observation period	Measuring instrumentation	Accuracy of observation	Specification
1	1959.9.1-1959.9.30	SWL1-1	0.2°C	GB/T 14914-94
2	1959.10.1-2002.6.30	SWL1-1	0.2°C	GB/T 14914-94
3	2006.4.1-2006.12	SWL1-1	0.2°C	GB/T 14914-94 GB/T 14914-2006
4	2007.1－now	YSI-600-1	0.05°C	GB/T 14914-94 GB/T 14914-2006

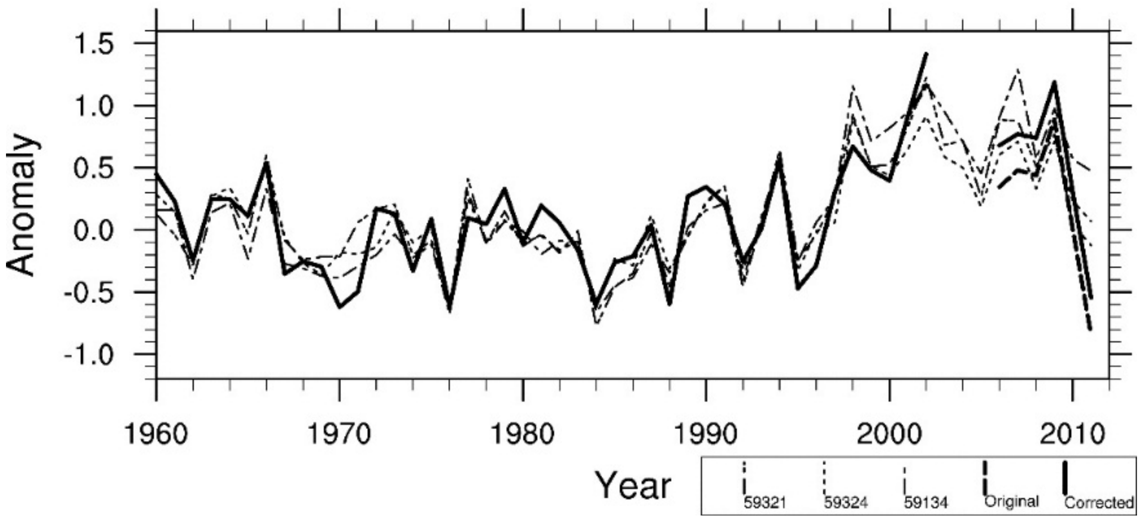

Figure 3. Plots of annually difference anomalies (°C) for Dongshan station (Original and Corrected, bold line) and three surrounding meteorological stations (59324, 59134 and 59321, dashed line).

Dongshan station, for example, shows the process of detection of documented change points. First, select three surrounding meteorological stations (59324, 59134 and 59321, the anomaly correlation coefficient are over 0.7) as reference stations to homogenise the Dongshan station. Second, with in significance at the 95% level , the homogeneity tests show that, only three breakpoints (January 1969 , April 2006 and January 2007) is consistent with the metadata (Figure 3). Third, eventually determine the three breakpoints and corrected.

3.2. Detection of undocumented change points

By analysing the marine station metadata has found, in addition to changes of location, changes of instrumentation, the effects of urbanization, land use of the surroundings, and other anthropogenic emissions of heat also caused SST data inhomogeneity. Nanao Station from October 1959 to begin SST observations, relying on temperature and salinity pier reef construction, conditions are suitable for SST observations. Because of the refrigeration factory is completed in 2007, SST measurement was crashed many times by the fishing boat. As a result of the fishing boat drainage, heat removal etc., can affect the accuracy of observation. Due to construction, strengthening pier, put the rocky drift into measurement area, causing the measurement area becomes shallow (Figure 4). Because the metadata not clear, based on the research of Peterson [18], in case of the metadata is unknown, after a significant test, and requires further climate rationality analysis of discontinuities. In this paper, in view of the stations are greatly influenced by local environmental change, based on the ENSO decadal distribution [19-20] and significant feature to detection of undocumented change points.

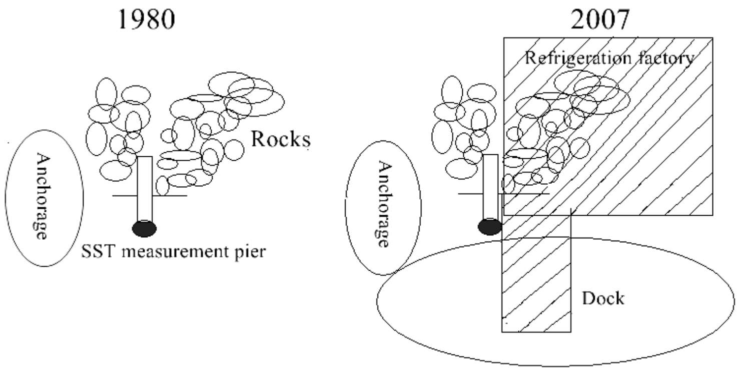

Figure 4. Changes of local environmental in Nanao station

In Nanao station as an example, explain the process of undocumented change points. Firstly, select three surrounding meteorological stations (59324, 59321 and 59316, the anomaly correlation coefficient are over 0.7) as reference stations to homogenise the Nanao station. Second, use PMT methods to detection the undocumented change points, then base on the metadata, there are six undocumented change points in 2005-2011 period (Table 3). On the basis of decadal distribution of ENSO, found that in August 2006-January 2007 and June 2009-April 2010 are weak and moderate warming events respectively, on June 2010-March 2011 is medium strength cold events. Significance test show that, there are three discontinuous points are the significant change points, that is, November 2006, April 2007 and October 2009.

Table 3. Change points of Nanao station.

No.	Change points	Significant test	Significant test range	ENSO	Significant change points	Rationality analysis
1	20050200	3.2442	(-2.8389-3.2816)	/	NO	Unreasonable
2	20060500	2.4962	(-2.5917-2.9981)	/	NO	Unreasonable
3	20061100	3.4407	(-2.2759-2.6289)	El Nino	YES	Unreasonable
4	20070400	3.8755	(-2.8369-3.2811)	/	YES	Reasonable
5	20091000	3.7217	(-2.8469-3.2935)	El Nino	YES	Unreasonable
6	20110100	2.3412	(-2.8172-3.2536)	La Nina	NO	Unreasonable

According to Peterson [18] and T.S.Stephenson [10] correction methods, climate rationality analysis of the above change points, at the same time satisfy the significant change points and the non-ENSO events both, established the change point of April 2007 of Nanao station and corrected.

3.3. statistics of change points

The statistics of nine marine stations change points in the SCS found (Table 4), the homogeneity tests produced a total of 14 change points. There are 5, 5 and 4 number of change points ,which is caused by changes of measuring instrumentation, changes of location or changes of local environmental of the surroundings, respectively, and the average absolute value of the corrected was 0.321 °C , 0.414 °C and 0.272 °C.

Correction of the four marine stations which is changes of location is about 1 time, moving station distance between 2 km and 8 km away. Six change points of instrument changes occurred after 2002, which is the observation to transmission and mechanisation processes.

Time distribution from the change points can be found, anthropogenic observation to automated observation is the main reasons for SST inhomogeneity in the nearly two decades. Shallow measuring area, port congestion, human emissions has a greater influence on the marine stations SST data. Time distribution of change points is scattered, did not focus in a certain time period (except for 2002), indicating that the process of detection between individual stations are independent.

Table 4. Statistics of change points.

Marine Stations	Change points	Changes of location	Changes of measuring instrumentation	Changes of local environmental
Beihai	1	/	200505	/
Dongshan	4	196901,	200604,200701	/
Haikou	1	196601	/	/
Naozhou	3	199001,199606	200206	/
Qinglan	1	/	/	198307
Weizhou	1	200301	/	/
Nanao	1	/	/	200704
Zhelang	1	/	200201	/
Zhapo	2	/	/	197912,198803

4. Result analysis

To verify the correction of PMT methods, the OISST datasets and HadISST datasets were calculated the correlation coefficient respectively with the SST anomalies after correction (Table 5). After the corrected, the average correlation coefficient were 0.492, 0.482 and 0.594, respectively. Compared with the original, correlation coefficient increased about more than doubled, which Dongshan station, Nanao station and Qinglan station improvement is most evident. The linear trend of annual mean SST over the SCS can be found (Table 6), after the correction, the linear warming trend of the nine marine stations average annual SST over the SCS has obvious increase, this is consistent with the previous research [6]. The correlation coefficient of the homogenous SST series have significantly improved.

Table 5. The correlation coefficient of the homogenous SST series and the candidate SST time series. The most significant correlation coefficients at confidence 0.01 level are in bold.

	Beihai	Dongshan	Haikou	Naozhou	Qinglan	Weizhou	Nanao	Zhelang	Zhapo	Average
HadISST1	**0.400**	**0.608**	**0.421**	**0.366**	**0.458**	**0.531**	**0.616**	**0.652**	**0.373**	**0.492**
OISST	0.303	0.432	**0.507**	**0.575**	**0.632**	**0.461**	**0.496**	**0.521**	0.413	**0.482**
OISST.V2	0.031	**0.732**	**0.631**	**0.713**	**0.527**	**0.670**	**0.761**	**0.602**	**0.682**	**0.594**

Table 6. The linear trend of annual mean SST over the SCS.

Marine	Linear trend（°C/century）		Significantly test	
Stations	Original	Corrected	Original	Corrected
Beihai	0.94	1.45	0.45	0.67
Dongshan	0.81	1.27	0.56	0.72
Haikou	0.96	1.44	0.44	0.77
Naozhou	1.70	1.60	0.38	0.59
Qinglan	0.78	1.62	0.41	0.64
Weizhou	0.84	1.01	0.48	0.78
Nanao	0.85	1.14	0.51	0.65
Zhelang	1.43	1.53	0.32	0.72
Zhapo	1.28	1.57	0.52	0.69
Average	1.04	1.37	0.45	0.69

Figure 5 shows the corrected (dashed line) and original (solid line) mean annual SST curves for the all SST series around the SCS. This result shows that the inhomogeneous time series appears to underestimates the warming trend during the past 52 years. The average warming rate based on the homogenous SST series over the SCS is about 1.37 °C/century, and increased by a total of 0.71 °C in the last 52 years (1960- 2011). The warming trend in the central SCS is most prominent, and coastal SST changes in Guangdong province is the largest

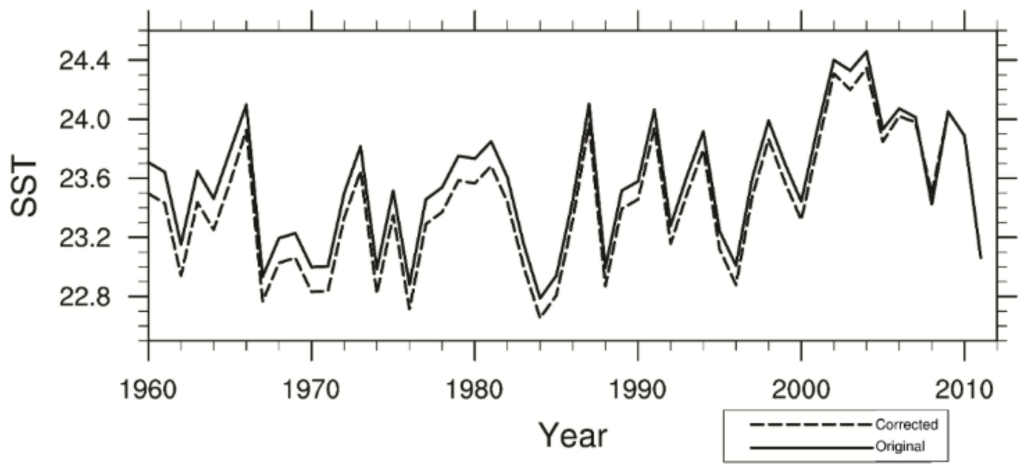

Figure 5. Corrected (dashed line) and original (solid line) mean annual SST curves for the all SST series around the SCS.

5. Conclusion

The SST series over the SCS is inhomogeneity, which has influence on the research of climate change and marine. Based on the metadata, the SST series of nine marine stations over the SCS are homogeneity detection and correction by PMT method. Select the surrounding meteorological stations as reference stations to homogenise the candidate station, and climate rationality analysis was carried out on the detection of change points and correction.

Correction results show that: (1) The homogeneity detection and correction of marine observation stations should be based on the metadata, meanwhile, fully consider the influence of regional climate change factors. (2) Correlation analysis found that, the surrounding meteorological station temperature data is the optimal reference data. (3) The average marine stations has 1~2 change points, among them, changes of instrumentation and changes of location environment is the most common. The anthropogenic observation change to automated observation is the main reasons of SST inhomogeneity in the last two decades. (4) The trend of the SST over SCS have a more pronounced warming trend, increased by an average 0.0137 °C each year. Correction results indicate that the homogenization research in the SCS has an important meaning for the study of the SCS coast SST changes and climate change.

References

[1] Tan J, Zhou F X and Hu D X 1995 *Oceanology Et Limnology Sinica* **26(4)** 377-82
[2] Liu X and Lin X P. 2013 *Periodical Of Ocean University Of China* **43(12)** 001-6
[3] Yu S Y, Zhou F X, Fu G, Wang D X. 1994 *Oceanologia Et Limnologia Sinica* 25(5) 546-51
[4] Wang W Q, Wang D X, Qi Y Q. 2000 *Acta Oceanologica Sinica*. 22(4) 8-16
[5] Arguez, A., Huang, B., Mcmahon, J. R., Menne, M. J., and Zhang, H. 2015 *Science* 348(6242) 1469-72
[6] Li Q X, Li W. 2007 Acta Meteorologica *Sinica* **65(2):**293-300
[7] Cao L J, Ju X H, and Liu X N 2010 *Meteorological Monthly* **10** 52-6
[8] Ren G Y, Guo J and Xu M Z 2005 *Acta Meteorologica Sinica* **63(6)** 948-52
[9] Li Q X, Liu X N, Zhang H Z, Tu Q Q 2003 *Meteorological Science And Technology* **31(1)** 3-10
[10] Stephenson T S, Goodess C M and Haylock M R 2008 *Journal of Geophysical Research Atmospheres* (1984–2012) **113(D21)** 23-31
[11] Zhai P M 1997 *Acta Meteorologica Sinica* **55(5)** 563-72
[12] Li Q, Liu X, Zhang H 2004 *Advances in Atmospheric Sciences* **21(2)** 260-8
[13] Reynolds R W, Rayner N A, Smith T M and Wang W 2002 *Journal of climate* **15(13)** 1609-25
[14] Reynolds R W, Smith T M, Liu, C, Chelton D B, Casey K S, and Schlax M G 2007 *Journal of Climate* **20(22)** 5473-96
[15] Lund R and Reeves J 2002 *Journal of Climate* **15(17)** 2547-54
[16] Rayner N A, Parker D E, Horton E B, Folland C K, Alexander L V, Rowell D P, and Kaplan A 2003 *Journal of Geophysical Research* Atmospheres (1984–2012) **108(D14)**
[17] Chapin F S, Berlow E L, Bloomfield J, Dirzo R, Hubersanwald E, Huenneke L F, and Wall D H 2000 *Science* **287(5459)** 1770-74
[18] Easterling D R, Groisman P Y, Nicholls N, Plummer N, Torok S, Auer I and Parker D E 1998 *International Journal of Climatology* **18(13)** 1493-517
[19] Malcher J, Schönwiese C 1987 *Theoretical and applied climatolog*y **38(3)** 157-66
[20] Schönwiese C 1987 *Beiträge zur Physik der Atmosphäre* **60(1)** 49-64