UDC 519.246.8
V. Lyubchich, Research Assistant Professor, University of Maryland Center for Environmental Science
e-mail: lyubchic@umces.edu

DETECTING TIME SERIES TRENDS AND THEIR SYNCHRONIZATION IN CLIMATE DATA

Analysis of weather and climate dynamics largely depends on reliable assessment of trends and other complex nonlinear interrelationships among atmospheric variables. However, assumptions of many conventional parametric methods for statistical inference, particularly tests for trend detection, are not satisfied by environmental data. The goal of this paper is to provide an insight into utility of modern data-driven procedures that address many limitations of the conventional parametric trend tests. The discussed methods of trend testing employ local regression and local factor methods, hybrid bootstrap, and m-out-of-n subsampling. The results of two climatological case studies demonstrate the utility of the WAVK testing procedure in initial detection of regime shifts in a non-sequential change point analysis and applicability of the new synchronism test to intercomparison of climate trends. One of the possible directions for future research is to employ the discussed testing approaches for intercomparison and scoring of multiple global climate models.
Keywords: *bootstrap, resampling, inference, time series, temperature, limnology, CMIP*

1. Introduction

Modeling and forecasting of weather and climate is critical for understanding complex socio-environmental interactions, with applications ranging from tourism, agriculture, and food security to insurance and health studies. Reliable detection of changes and aligning the results within wider geographic areas or with some other known trends and processes are of particular importance in such studies of environmental dynamics. This has led to an ever-increasing demand for reliable statistical inference tools that are data-driven and can be robust against serial correlation in time series and misspecification of distributional assumptions.

This paper discusses utility of a new nonparametric methodology for trend analysis that is based on an artificial one-way analysis of variance (ANOVA) and can be used for detection of (non-)monotonic trends in atmospheric sciences [23, 24]. Based on the authors' names of the trend test statistic (Wang, Akritas and Van Keilegom), the trend test is called WAVK. The data-driven extension of the WAVK procedure [10] employed in this paper is shown to be powerful in detecting smooth non-monotonic trends and robust to different time dependence structures, conditional heteroscedasticity, and non-normality. The multivariate extension of this method [8, 9] demonstrates high power in detecting misalignments of trends embedded into linear noise of possibly infinite order while remains free from any distributional assumptions.

The testing techniques are demonstrated with two case studies of climate dynamics. The trend testing procedure is applied to ice phenology data. Ice phenology refers to ice-related processes, such as freeze-up, break-up, and ice cover duration. Long-term phenological observations provide valuable information about the climate change [13] and might serve as an insight into the patterns of atmospheric teleconnections [3, 7, 20], i.e., "relationships in the variability of large-scale features of the atmospheric circulation as well as tropical and extratropical precipitation and temperature relationship" [21]. The observations are subject to serial correlation and multiple change points, thus, the conventional Student's t-test and Mann–Kendall trend testing procedure may not be applicable.

The second case study deals with testing the similarity in multiple trends using an example of observed air temperatures and projected for the same time period by climate models. One of the problems addressed in this case study is the stochastic nature of the climate system's response to the forcings. The Earth's response to a particular forcing might be different each time that forcing is applied, however, this variability cannot be studied directly, since there is only one replication of the Earth's climate system. In the model world, a climate model can be run under the same forcings several times to assess the variability of response. Even if the model correctly represents the climate system, there is a chance that the observed climate appears as an outlier in the ensemble of the model runs, thus, the model would have to be renounced. In view of the fact that no model is known to give the ground truth representation of the climate system, it is common to create multi-model ensembles that incorporate uncertainty from a number of models. Therefore, in the ensemble case, there is a lower probability to consider the actual climate dynamics as an outlier. The study investigates whether two trends – of observed global mean temperature and of the multi-model average – are statistically similar.

The remainder of the paper is organized as follows. Section 2 describes the new nonparametric procedures for trend detection and testing trend synchronism. Section 3 and Section 4 provide the relevant case studies from climate research. The paper is concluded with discussion in Section 5.

2. Methods

Let a time series be represented as a sum of an unknown smooth parametric function $\mu(u)$ and error process ϵ_t that can be linear (and potentially of infinite order) or generalized autoregressive conditionally heteroscedastic (GARCH) [10]:

$$Y_t = \mu\left(t/T\right) + \epsilon_t \left(t = 1,\ldots,T\right). \quad (1)$$

The WAVK test statistic is employed for testing the null hypothesis of whether $\mu\left(u\right)$ can be represented by a function $f\left(\theta,u\right)$ from some known smooth parametric family (see [10] and references therein):

$$\mathrm{H}_0\text{: } \mu\left(u\right) = f\left(\theta,u\right) \text{ vs. } \mathrm{H}_1\text{: } \mu\left(u\right) \neq f\left(\theta,u\right); \quad (2)$$

$$WAVK\left(k_T\right) = MST - MSE, \quad (3)$$

where MST is mean square for treatments and MSE is mean square for errors in an artificial balanced one-way ANOVA with T groups of size k_T (also called the window size).

If N time series are considered

$$Y_{it} = \mu_i\left(t/T\right) + \epsilon_{it}\left(i = 1,\ldots,N; t = 1,\ldots,T\right), \quad (4)$$

the hypothesis of a common trend $f\left(\theta,u\right)$

H_0: $\mu_i\left(u\right) = c_i + f\left(\theta,u\right)$, (5)

H_1: there exists i, such that $\mu_i\left(u\right) \neq c_i + f\left(\theta,u\right)$ can be tested using the statistic [8]:

$$S_T = \sum_{i=1}^{N} k_{iT}^{-1/2} WAVK_i\left(k_{iT}\right). \quad (6)$$

The hypothesis (5) can be also referred to as the hypothesis of trend synchronism, i.e., alignment of all individual trends $\mu_i\left(u\right)$ with $f\left(\theta,u\right)$, which parameters θ are estimated from an averaged time series

$$\overline{Y}_{\cdot t} = N^{-1}\sum_{i=1}^{N} Y_{it}. \quad (7)$$

Convergence of the test statistics (3) and (6) to their asymptotic distributions is slow in small samples, thus, a hybrid bootstrap procedure is employed to enhance the finite sample performance of these tests. Particularly, B bootstrapped test statistics are obtained from normally distributed T-dimensional vectors scaled by the variance estimate s_i^2 ($i = 1,\ldots,N$). The variance is estimated from the residuals, which are obtained by extracting from the original time series the estimated trend $f(\hat{\theta},u)$ and applying a BIC-based autoregressive filtering to each time series.

Since simulation studies showed dependence of the test statistics (3) and (6) on the window size, an automatic m-out-of-n selection procedure was developed (see [8, 10, 12] and references therein). Figure 1 shows the steps of this procedure in the univariate case (1), using $k_T\left(j\right) = q^j T$; $j = 7,\ldots,11$; $q = 3/4$ [10]. Notice that the optimal window in Figure 1 is shown only as an example, as it can be different for each time series.

With the performance boost provided by the hybrid bootstrap and m-out-of-n selection procedure, the tests (3) and (6) were shown to be robust against various forms of time dependence, heteroscedasticity, and non-normality. Particularly, the WAVK test (3) has a high power in detecting smooth non-monotonic trends and outcompetes the conventional Student's t-test and Mann–Kendall test, which are limited to linear and monotonic trends [10]. At the same time, the synchronism test (6) exhibits a high power in identifying asynchronous trends and outcompetes the integrated square error (ISE) based test of [2, 22] (see [8] for more details). Both WAVK and synchronism tests are implemented in the **R** package *funtimes* [11].

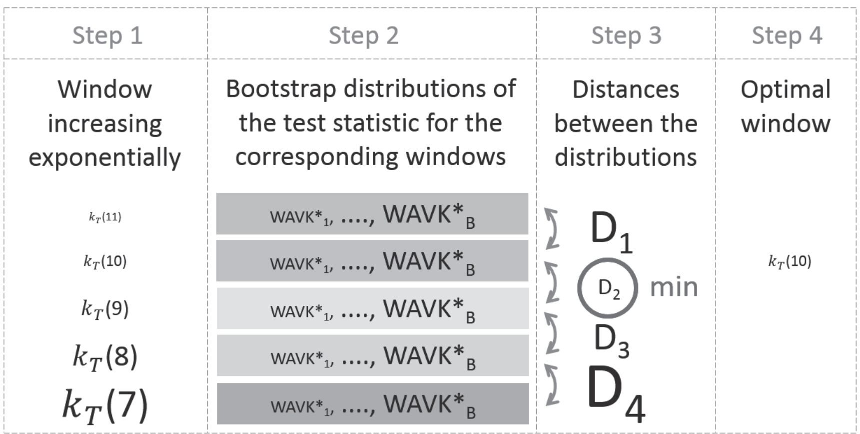

Figure 1. Visualization of the m-out-of-n subsampling algorithm for selecting the optimal window, which corresponds to the minimal distance between bootstrap distributions of the statistic $WAVK\left(k_T\right)$

3. Analysis of Ice Phenology Trends in Lake Baikal

Lake Baikal is the world's largest freshwater lake in terms of volume. Its ice cover is highly linked to physical, chemical, and biological processes within the lake and has a great impact on the local transport, fishing activities, and tourism (see [6, 7, 15, 20], and references therein). Teleconnections possibly affecting Lake Baikal ice cover are:

– January–March North Atlantic Oscillation ($\mathrm{NAO_{JFM}}$) [7];

– Atlantic Oscillation (AO) [20];

– Pacific Decadal Oscillation (PDO) [5].

Historical ice break-up data for 1869–1996 were collected at Listvyanka limnological station that is located on the west coast of Southern Baikal [1, 7]

(Figure 2). Listvyanka is far from rivers reaching the lake and, hence, is suitable for a reliable ice phenology analysis [6].

Besides of some outliers in the right tail (Figure 3), the observed break-up dates do not deviate substantially from the normal curve. However, as discussed by [15], the break-up dates are serially correlated and their dependence structure can be approximated, for example, by an autoregressive model of sixth order, AR(6). The sieve bootstrapped modifications of the Student's *t*-test and Mann–Kendall test fail to reject the null hypothesis of no trend in favor of a linear or monotonic trend, with *p*-values of 0.277 and 0.104, respectively [15]. Moreover, the bootstrap-based change point study of [15] under a weakly dependent error structure suggests the year of 1924 as a potential regime shift. These results are also supported by the previous findings of [6, 7, 20] who doubt the existence of a linear trend in the break-up dates of the Lake Baikal but raise an issue of presence of multiple regime shifts (change points) in the observed ice phenology data.

If such regime shifts are present, the potential trend functions necessarily need to be of a non-monotonic form. In particular, the local minimum of the loess smoothing curve (Figure 2) corresponds to the year 1924 suggested by [6, 7] as a potential regime shift, while the local maximum around the year of 1970 corresponds to the PDO phase shift, affecting the Lake Baikal [19].

Remarkably, the WAVK trend test of [10] supports these findings and rejects the null hypothesis (2) of no trend (H_0: $\mu(u) \equiv 0$) as well as the null hypothesis of a linear trend (H_0: $\mu(u) = \theta_0 + \theta_1 u$), with highly statistically significant *p*-values of less than 0.001. Hence, a non-monotonic trend is likely present in the ice break-up dates, with local minima and maxima corresponding to the potential change points. Thus, the WAVK lack-of-fit procedure can be employed for initial detection of regime shift existence in a non-sequential change point analysis.

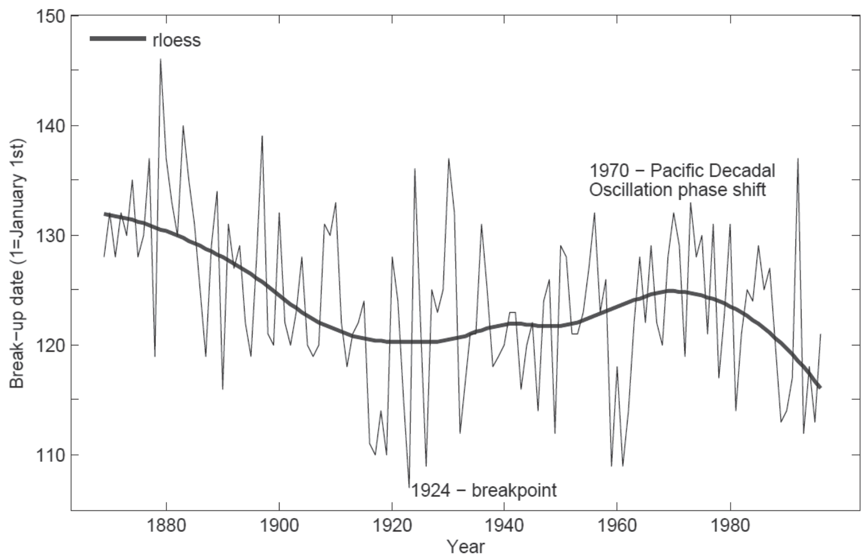

Figure 2. Ice break-up dates, Lake Baikal, Russia, 1869–1996

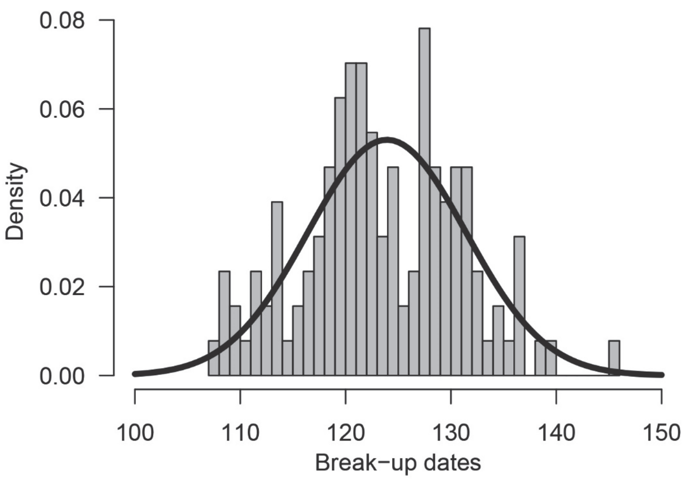

Figure 3. Histogram of the ice break-up dates for Lake Baikal, Russia, for the period 1869–1996, with a superimposed curve of fitted normal distribution

4. Testing for Similarities of Climate Trends

In this section, the synchronism test is employed to detect discrepancies in projected climate dynamics vs. observed data. Currently, global climate models is the main tool to forecast the future climate under assumed scenario of land use, solar irradiance, ozone concentrations, greenhouse gases, aerosol emissions, etc. Global climate models may incorporate different combinations of forcings (e.g., account for ozone depletion and solar changes but disregard other factors) and, supplied with various levels of these inputs, produce abundant model outputs. Thus, not only the differences between the models, but also the large number of possible scenarios yield a wide range of climate projections.

The Coupled Model Intercomparison Project, which is currently in its 5th phase (CMIP5), promotes coordinated climate model experiments, when different models are run under the same forcing scenarios. This project also aims to determine why similarly forced models produce various responses, and to evaluate how realistic the models are in simulating the recent past. To accomplish the later task, one can evaluate aligning of trends in the model output and observed data using the synchronism test of [8]. Notice that in this context the focus is not on the metric of individual discrepancies (such as mean square error) between the two time series, but on a more general comparison of dynamics of two mean functions. Following [4], trends in the global mean temperature and in the model output are considered (Figure 4). Based on the WAVK test of [10], both temperature time series have highly statistically significant individual trends in the analyzed period 1948–2013 (see the *p*-values for H_0 of no trend in Table 1). Now turn to testing H_0 of linear trend. While given the delivered *p*-value of 0.089 the hypothesis of linear trend in the observed series cannot be rejected, WAVK test *p*-value of 0.012 for the multi-model average is on the border of being highly statistically significant (Table 1).

How high should be the confidence in conducting such statistical tests is an open question. If the significance level of 0.01 is selected and, hence, (similarly to [17, 25]) linear trends in these temperature time series are hypothesized, the synchronism test can be applied to test the alignment of these linear trends. The synchronism test of [8] yields a *p*-value of 0.012 for testing the null hypothesis that the linear trends in observed and projected global mean temperature are synchronous. At the significance level of 0.01, the hypothesis of a common linear trend cannot be rejected, which implies that the multi-model average replicates the recent trends in the global mean temperature reasonably well. These results concur with [4] and [17] who reported that the discrepancies between data on global average warming and climate models vanished after the errors in satellite and radiosonde data have been corrected. However, the output of synchronism test is just on the border of statistical significance. Therefore, in view of the earlier findings on dynamics of multi-model average (Table 1), some deviations from the linear trend in the temperature time series can be suspected (e.g., see [18]), as well as an implication that trends in observed an modeled time series are not aligned sufficiently, which supports the phenomenon discussed in the recent climatological studies (e.g., see the discussion by [14, 16]).

5. Conclusion

The discussed data-driven nonparametric methods for testing parametric trends in time series, including trend detection and synchronism, can be considered as competitive tools for reliable statistical inference on climate and weather data. The proposed procedures based on local regression are robust to various temporal dependence structures and deliver a competitive finite

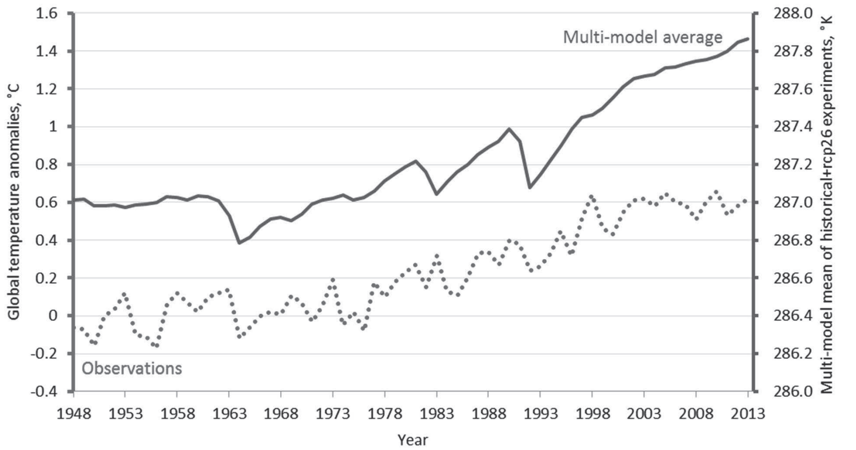

Figure 4. Global annual mean temperature in 1948–2013: observed anomalies with respect to the 1981–2010 base period and CMIP5 multi-model average. (*Data source:* National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center website accessed on April 2, 2014; Climate Explorer website accessed on April 22, 2014)

Table 1. Results of the local factor test for individual trends in global annual mean temperature in 1948–2013 (observed anomalies with respect to the 1981–2010 base period), and projected multi-model average. Number of bootstrap samples is 10000

Data	*p*-value	
	H_0: no trend	H_0: linear trend
Observations	$< 10^{-4}$	0.089
Multi-model average	$< 10^{-4}$	0.012

sample performance compared with other available methods for trend analysis. The case studies of ice break-up dates and global annual mean temperatures unveil additional aspects in the structure of climate dynamics.

Acknowledgements

The author would like to thank Yulia R. Gel for valuable discussions. This work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET, www.sharcnet.ca).

References

1. Benson, B., Magnuson, J. 2000, updated 2012. Global lake and river ice phenology database, Version 1. Boulder, Colorado, USA. NSIDC: National Snow and Ice Data Center. – Accessed May 10, 2012.

2. Degras, D., Xu, Z., Zhang, T., Wu, W.B. Testing for parallelism among trends in multiple time series. IEEE Transactions on Signal Processing. – 2012. – Vol. 60. – pp. 1087–1097.

3. Ebert-Uphoff, I., Deng, Y. Causal discovery for climate research using graphical models. Journal of Climate. – 2012. – Vol. 25. – pp. 5648–5665.

4. Karl, T.R., Hassol, S.J., Miller, C.D., Murray, W.L. Temperature trends in the lower atmosphere. Steps for understanding and reconciling differences. Report by the U.S. Climate Change Science Program and the Subcommittee on Global Change Research, Asheville, NC. – 2006. – 180 pp.

5. Katz, S.L., Hampton, S.E., Izmest'eva, L.R., Moore, M.V. Influence of long-distance climate teleconnection on seasonality of water temperature in the world's largest lake - Lake Baikal, Siberia. PloS ONE. – 2011. – Vol. 6 (2):e14688. – 10 p.

6. Kouraev, A.V., Semovski, S.V., Shimaraev, M.N., Mognard, N.M., Legrésy, B., Rémy, F. The ice regime of Lake Baikal from historical and satellite data: Relationship to air temperature, dynamical, and other factors. Limnology and Oceanography. – 2007. – Vol. 52. – pp. 1268–1286.

7. Livingstone, D.M. Ice break-up on southern Lake Baikal and its relationship to local and regional air temperatures in Siberia and to the North Atlantic Oscillation. Limnology and Oceanography. – 1999. – Vol. 44. – pp. 1486–1497.

8. Lyubchich, V., Gel, Y.R. A local factor nonparametric test for trend synchronism in multiple time series. Journal of Multivariate Analysis. – 2016. – Vol. 150. – pp. 91–104.

9. Lyubchich, V., Gel, Y.R., El-Shaarawi, A. Detecting non-monotonic trends and testing for synchronism: application to environmental time series. In: The 3rd International Workshop on Climate Informatics: CI2013. Boulder, CO. – 2013. – pp. 91–92.

10. Lyubchich, V., Gel, Y.R., El-Shaarawi, A. On detecting non -monotonic trends in environmental time series: a fusion of local regression and bootstrap. Environmetrics. – 2013. – Vol. 24. – pp. 209–226.

11. Lyubchich, V., Gel, Y.R., Chu, C., Wang, X. Funtimes: Functions for Time Series Analysis. R package. – 2017. – Vol. 3.0. – 25 p.

12. Lyubchich, V., Wang, X., Heyes, A., Gel, Y.R. A distribution-free m-out-of-n bootstrap approach to testing symmetry about an unknown median. Computational Statistics & Data Analysis. – 2016. – Vol. 104. – pp. 1–9.

13. Magnuson, J.J., Robertson, D.M., Benson, B.J., Wynne, R.H., Livingstone, D.M., Arai, T., Assel, R.A., Barry, R.G., Card, V., Kuusisto, E., Granin, N.G., Prowse, T.D., Stewart, K.M., Vuglinski, V.S. Historical trends in lake and river ice cover in the Northern Hemisphere. Science. – 2000. – Vol. 289. – pp. 1743–1746.

14. McKitrick, R.R., Vogelsang, T.J. HAC robust trend comparisons among climate series with possible level shifts. Environmetrics. – 2014. – Vol. 25. – pp. 528–547.

15. Noguchi, K., Gel,Y.R., Duguay, C.R. Bootstrap-based tests for trends in hydrological time series, with application to ice phenology data. Journal of Hydrology. – 2011. – Vol. 410. – pp. 150–161.

16. Santer, B.D., Mears, C., Doutriaux, C., Caldwell, P., Gleckler, P.J., Wigley, T.M.L., Solomon, S., Gillett, N.P., Ivanova, D., Karl, T.R., Lanzante, J.R., Meehl, G.A., Stott, P.A., Taylor, K.E., Thorne, P.W., Wehner, M.F., Wentz, F.J. Separating signal and noise in atmospheric temperature changes: The importance of timescale. Journal of Geophysical Research: Atmospheres. – 2011. – Vol. 116: D22105. – 19 p.

17. Santer, B.D., Thorne, P.W., Haimberger, L., Taylor, K.E., Wigley, T.M.L., Lanzante, J.R., Solomon, S., Free, M., Gleckler, P.J., Jones, P.D., Karl, T.R., Klein, S.A., Mears, C., Nychka, D., Schmidt, G.A., Sherwood, S.C., Wentz, F.J. Consistency of modelled and observed temperature trends in the tropical troposphere. International Journal of Climatology. – 2008. – Vol. 28. – pp. 1703–1722.

18. Seidel, D.J., Lanzante, J.R. An assessment of three alternatives to linear trends for characterizing global atmospheric temperature changes. Journal of Geophysical Research: Atmospheres. – 2004. – Vol. 109: D14108. – 10 p.

19. Shen, C., Wang, W.C., Gong, W., Hao, Z. A Pacific Decadal Oscillation record since 1470 AD reconstructed from proxy data of summer rainfall over eastern China. Geophysical Research Letters. – 2006. – Vol. 33: L03702. – 4 p.

20. Todd, M.C., Mackay, A.W. Large-scale climatic controls on Lake Baikal ice cover. Journal of Climate. – 2003. – Vol. 16. – pp. 3186–3199.

21. Trenberth, K.E., Branstator, G.W., Karoly, D., Kumar, A., Lau, N.C., Ropelewski, C. Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. Journal of Geophysical Research: Oceans (1978–2012). – 1998. – Vol. 103. – pp. 14291–14324.

22. Vilar-Fernández, J., González-Manteiga, W. Nonparametric comparison of curves with dependent errors. Statistics. – 2004. – Vol. 38. – pp. 81–99.

23. Wang, L., Akritas, M.G., Van Keilegom, I. An ANOVA-type nonparametric diagnostic test for heteroscedastic regression models. Journal of Nonparametric Statistics. – 2008. – Vol. 20. – pp. 365–382.

24. Wang, L., Van Keilegom, I. Nonparametric test for the form of parametric regression with time series errors. Statistica Sinica. – 2007. – Vol. 17. – pp. 369–386.

25. Zhou, J., Tung, K.-K. Deducing multidecadal anthropogenic global warming trends using multiple regression analysis. Journal of the Atmospheric Sciences. – 2013. – Vol. 70. – pp. 3–8.