Marine Geodesy, 28: 209–217, 2005

ISSN: 0149-0419 print / 1521-060X online
DOI: 10.1080/01490410500204546

Taylor & Francis
Taylor & Francis Group

Are Global Tide Gauge Data Stationary in Variance?

H. BÂKI IZ AND H. M. NG

Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China

Tide gauges distributed all over the world provide valuable information for monitoring mean sea level changes. The statistical models used in estimating sea level change from the tide gauge data assume implicitly that the random model components are stationary in variance. *We show that for a large number of global tide gauge data this is not the case for the seasonal part using a variate-differencing algorithm. This finding is important for assessing the reliability of the present estimates of mean sea level changes because nonstationarity of the data may have marked impact on the sea level rate estimates, especially, for the data from short records.*

Keywords PSMSL, tide gauge, stationary, variate-differencing, mean sea level, global warming

Tide gauges (TG) are among the oldest methods for monitoring variations in the mean sea level during the last two centuries (see Chapter 11 in IPCC Third Assessment Report, 2001 for a more detailed account of the mean sea level monitoring). TG measurements broadly consist of a systematic and a random part.

The systematic part involves variables due to wind stress, atmospheric pressure, precipitation, river discharge, currents, temperature and salinity of the water, and long periodic lunar tidal activities (Lambeck 1988), subseasonal to decadal periodic effects due to atmospheric pressure variations, sea surface winds, ocean circulation patterns causing the movement of sea level locally and regionally (Pugh 1987; Hannah 1990; Woodworth et al. 1999), a datum bias (Iz and Shum 2000), and a linear component to represent any secular trend in MSL and a nonlinear parameter for the deceleration or acceleration in the trend (Gornitz and Solow 1991).

The random part is composed of effects that cannot be modeled or replicated by empirical or physical means like the parts mentioned in the previous paragraph. They include random variations such as instrumental source or stochastic variations due to the transient choppiness of the sea level over a short or a long period of time because of climatic changes. They may also be due to, for instance, irregular and episodic discharges from a close-by river, or seasonal variations due to the atmospheric pressure or temperature variations that are not periodic in nature.

Received 26 February 2004; accepted 13 November 2004.
This study was supported by funds from The Hong Kong Polytechnic University. Helpful comments by the reviewers are gratefully acknowledged.
Address correspondence to H. M. Ng, Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China. E-mail: carolina.ng@polyu.edu.hk

Recent analyses of global, regional, and local sea level have tacitly used a stochastic model for the random part through the application of Ordinary Least-squares Estimation, OLE, in which the data is assumed to be stationary in variance. Nonetheless, nonrandom appearance of the residuals of some of the models has long been recognized (Pugh 1987) and identified as having nonstationary origins (Pugh 1987; Douglas et al. 2001).

This assumption can have a severe repercussion of the mean sea level estimates if the data sets are short, rendering them useless (Iz and Shum 2000). Even very long time series can be affected significantly should the nonstationary data be located at the end or beginning of the series or both (ibid.).

Regardless of their impact on the rate of change of the MSL, nonstationary properties of the series affect the construction of the confidence intervals for the estimated parameters using their estimated variances. Any hypothesis testing based on such nonoptimal confidence intervals will be misleading.

Stationarity can be understood as the statistical properties of a physical system that remains unchanged with time (Emery and Thomson 2001). Along these lines, we identify two types of stationary series: one having a *constant mean* and another fluctuating about that mean with a *constant variance* (Kendall 1973).

In this study, we choose to address the second requirement, that is, whether the stochastic component of the TG data has a constant variance. The first requirement is relatively easy to achieve, if not readily part of the time series (by proper modeling of the variations in the sea level due to known causes and by using descriptive models to account for the variations observed in the residuals such as polynomials, and modeling for intermittent changes in the expected values of the random values of the series using mean shift analyses, etc).

For this purpose, we look at all the tide-gauge data from 1862 stations located over the world from PSMSL (Permanent Service for Mean Sea Level). A simple variate-differencing method is used to remove all the systematic variations in the data without resorting to complex models to represent them. We then obtain estimates for the averaged random variations over yearly intervals. Two statistical methods are used to test the constancy of the averaged variances.

PSMSL Tide Gauge Data Holdings

We used monthly tide gauge data available at the PSMSL (PSMSL 2001). The service has maintained a tide gauge data base from more than 1800 stations since 1933. About 200 national authorities around the world provide data to PSMSL in monthly and yearly formats. Until February 2001, PSMSL contained more than 47000 station-years of sea level data, and it receives approximately 2000 station-years of data each year (ibid.).

PSMSL offers *Metric and Revised Local Reference* (RLR) *data* (Permanent Service for Mean Sea Level 2001). Metric data is the raw data directly received from the authorities whereas RLR data contains monthly and annual mean sea level data (MSL) corrected so that the annual and monthly data refer to a common datum, which is not the case with the metric data. The RLR datum is 7 m below the global mean sea level to avoid negative monthly and annual mean sea level values. However, only two thirds of the stations in the PSMSL database have been adjusted to a common datum.

In this study, we used all the PSMSL metric data since the variate-differencing algorithm deployed to investigate constant variance of the series is invariant to year-to-year datum changes (i.e., first difference of the data eliminates the common constant offset). Most of the datum changes in the PSMSL metric data occur at the beginning of calendar years, therefore only a small number of stations with different datum within a given year had to be adjusted to refer to the same datum. Furthermore, any changes in data spikes larger than three times the rms of data within a yearly bin were removed as gross error and replaced by the average of the two neighboring data if the spike involves only one data point. If there is more than one point involved, consecutive spikes were replaced by the average of all data points.

The metric data span from 1 to 158 years for different stations, however, majority of the stations have records spanning less than 40 years.

We corrected the metric data to account for the missing monthly values using *Locally Weighted Least Squares* (LOESS) interpolation scheme in between adjacent data points with a 0.75 smoothing factor determined by trial-and-error. For more than a month of missing data, yearly averages are used for interpolated values.

Eliminating Systematic Variations: Variate-Differencing

A direct method for quantifying the random sea level variations at a TG station data is to model all the systematic variations in the series. Once the correlated effects are removed, what remains are random and uncorrelated. The variances of residuals can be scrutinized whether they are stationary or not. Such an approach however will require modeling each data set one-by-one which is a cumbersome task.

Alternatively, TG data, in this case monthly PSMSL data, can be grouped into subseries, such as yearly or decadal, and each subseries can be represented by a descriptive polynomial. Again, the residuals of the polynomial fit can be used to assess the random behavior of the data over time for each TG station. Such an approach, though simpler then the previous approach, will still require significant amount of interaction with the data to determine the order of each polynomial and to ensure that the fit is appropriate (polynomial fits are well known for causing ill-conditioned solutions). In this study we will use a variate-differencing algorithm to achieve the same results of a polynomial fit or modeling of the TG data but free from the computational complexity and drawbacks of either approach.

Suppose that a set of TG data can be represented by a polynomial of degree p and a random element. Note that the underlying phenomena can be more complex than to be represented by a simple polynomial but the data can be decomposed always into a level that it can be approximated by a polynomial. The random errors ε are assumed to be independent, (although smoothing time series using running medians or means can create serial correlation, filtering nonoverlapping data reduces such artifacts), and identically distributed with zero mean and variance σ^2, i.e.,

$$\varepsilon \sim (0,\ \sigma^2 I). \tag{1}$$

The first successive differences of the series elements result in a polynomial of degree $p - 1$ and residuals in increasing variance is given by,

$$E(\Delta\varepsilon_t) = E(\varepsilon_{t+1} - \varepsilon_t) = 0 \rightarrow \mathrm{var}(\Delta\varepsilon_t) = E(\varepsilon_{t+1} - \varepsilon_t)^2 = 2\ \mathrm{var}(\varepsilon) = 2\sigma^2. \tag{2}$$

Similarly, second differences (differences of the first differences) give

$$\mathrm{var}(\Delta^2\varepsilon_t) = \mathrm{var}(\varepsilon_t) + 4\mathrm{var}(\varepsilon_{t-1}) + \mathrm{var}(\varepsilon_{t-2}) = 6\mathrm{var}(\varepsilon) = 6\sigma^2. \tag{3}$$

For the rth differences we obtain,

$$\mathrm{var}(\Delta^r \varepsilon_t) = \binom{2r}{r}\mathrm{var}(\varepsilon) = \binom{2r}{r}\sigma^2, \tag{4}$$

where the expression within large parentheses is the number of combinations of $2r$ objects taken r at a time.

Now observe that the left hand side of the above equation can be estimated from the sample variance of the differences given by

$$\mathrm{var}(\Delta^r \varepsilon_t) \cong \sum_N \frac{(\Delta^r \varepsilon_t)^2}{N}. \tag{5}$$

Because the above expression can be computed from the differenced data set, the underlying variance of the random series is estimated using the following expression,

$$\hat{\sigma}^2 = \sum_N \frac{(\Delta^r \varepsilon_t)^2}{N} \bigg/ \binom{2r}{r}, \tag{6}$$

where, N is the number of elements in the differenced data set.

Note that the first differencing removes the common constant part of the data, the second differencing removes the linear trend, the third differencing removes the second degree of the polynomial, and so on. Finally, it comes to the point where how to determine the systematic parts has been eliminated to a considerable degree, and the remaining part is the approximate random part. Considering that a series only consists of random elements, then the variances of the successive differences are equal, which is corrected for the multiplication by a binomial coefficient. Hence, the variances of its first and the higher difference must remain the same as the variance of the original random series. Thus, the differencing continues until the difference between two subsequent estimates for the series variance remains constant (Figure 1).

In this study, we applied variate-differencing to monthly PSMSL metric data grouped into yearly series for each station. For instance, a TG station with 40 years of data will have 40 subseries with each having 12 data points (12 monthly averages). Variate differencing is applied to each one of the 40 subseries, generating 40 data points each representing the average yearly variance of the station data over time (Figure 2). We then scrutinize the sampled variance of the subseries to determine whether they remain constant from subseries to sub-series. This is the topic of the next section.

Test for the Constancy of Error Variances and Conclusion

We will use both Bartlett and Hartley tests for resting the constancy of the error variances over time (Neter and Wasserman 1974). Both methods are widely used in practice and assume that the elements of the series are stochastically independent and normally distributed.

Bartlett Test

The main idea behind the Bartlett test a comparison between the weighted arithmetic and weighted geometric averages (*GMSE*). If the two averages are equal, then the variances are equal, otherwise, the variances are unequal. Actually, this relation between these two

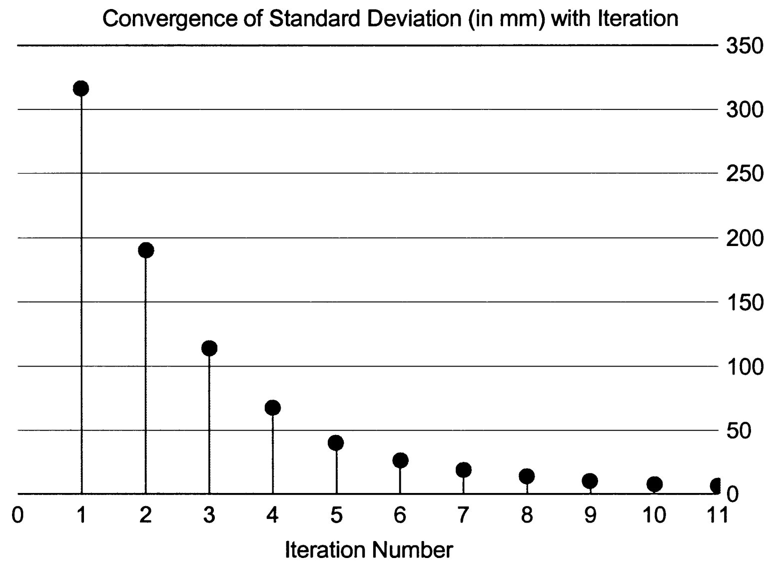

Figure 1. The changes in the estimated variance of the series are smaller than 5 cm after the 4th differencing.

averages could be replaced by $GMSE \leq MSE$ where, MSE is the mean square error, and is given by,

$$MSE = \frac{1}{n_T - r} \sum (n_j - 1)\hat{\sigma}_j^2. \tag{7}$$

Thus, the ratio of *MSE/GMSE* should be close to 1 to show that the variances are equal. On the contrary, if the *MSE/GMSE* is large, it indicates that the variances are unequal. To obtain the same conclusion, The test statistic is as follows, $\log(MSE/GMSE) = \log MSE - \log GMSE$ is the indicator for

$$B = \frac{2.302585}{C}[(n_T - r)\log_{10} MSE - \log_{10} GMSE],$$

which reformulates to

$$B = \frac{2.302585}{C}\left[(n_T - r)\log_{10} MSE - \sum_{j=1}^{r}(n_j - 1)\log_{10}\hat{\sigma}_j^2\right], \tag{8}$$

where r is the number of the sample variances, $\hat{\sigma}_j^2$, and n_j is the number of the observations in each sample (note that the Bartlett test does not require equal sample sizes). The constant 2.302585 is a result of converting logarithms of base 10 to logarithms of base e, and C is given by

$$C = 1 + \frac{1}{3(r-1)}\left[\sum_{j=1}^{r}\frac{1}{n_j - 1} - \frac{1}{n_T - r}\right]. \tag{9}$$

The null-hypothesis for the *Bartlett* test states

H_0: $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_r^2$, and
H_a: not all σ_j^2 are equal.

The test statistic B is Chi-square distributed with $r - 1$ degrees of freedom, hence if $B \leq \chi^2(1-\alpha; r-1)$ then H_0 is accepted.

In our case, $\hat{\sigma}_j^2$ is the yearly variance of a TG station data estimated by the variate-differencing method. If the yearly variances obtained from variate-differencing from a station agree with each other, we conclude that the random variations in the data set are

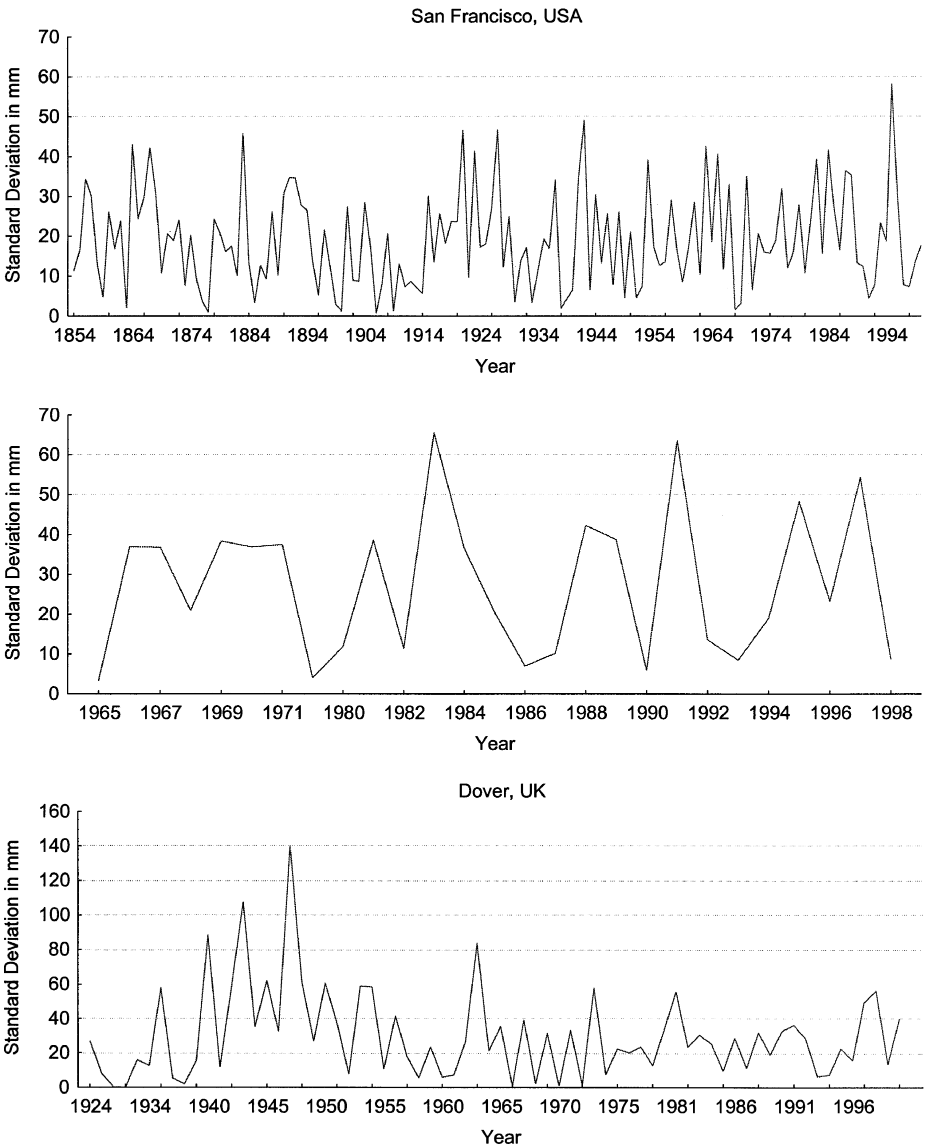

Figure 2. Sample graphs of the yearly standard deviation of the TG data (PSMSL) after the variate-differencing which removes the systematic variations. *(Continued on next page)*

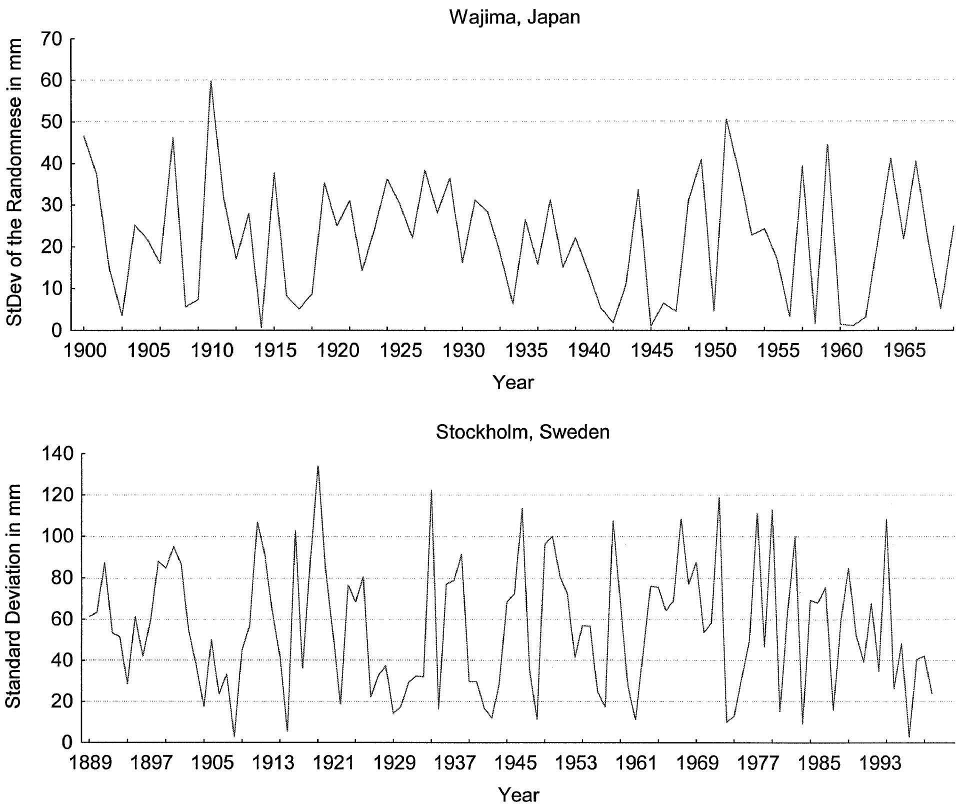

Figure 2. Continued

stationary in variance. Table 1 gives the test results for all stations. It shows that most of the stations failed in the test. Only 20 percent of all stations passed the test at 0.01 significance level. Although the *Bartlett* test is known to be quite sensitive for departures from normality, we used an alternative test to verify the outcome.

Hartley Test

The Hartley test is applicable only to cases with equal sample sizes. The test is also more sensitive to the variations in the variances from sample-to-sample than the *Bartlett* test. The same null hypothesis is tested as before, with the following test statistic (Neter and Wasserman 1974),

$$H = \frac{\max(\hat{\sigma}_j^2)}{\min(\hat{\sigma}_j^2)}, \tag{10}$$

where, $\max(\hat{\sigma}_j^2)$ is the maximum sample variance and $\min(\hat{\sigma}_j^2)$ is the smallest sample variance. If the value of H is close to 1, H_0 is accepted. A critical value is obtained from an H-table. The distribution of Hdepends on the number of samples r and the fixed sample size n. Again, if, $H \leq H(1-\alpha;\ r-1,\ n)$ then H_0 is accepted.

Table 1
Heterogeneity of the variances by Bartlett and Hartley tests

Significance level	% of stations accepted in	
	Bartlett test	Hartley test
0.01	17	15
0.05	13	11
0.10	11	NA

Two different significance levels, 5% and 1%, are chosen for this test. The critical values are obtained from an existing H-table, but since the size of table is limited to 12, numerical computations, it cannot be easily extended to cover large number of data points. Nevertheless, we observed that the critical values converge rapidly with increasing sample size. Hence we can safely assume that for those stations containing more than 12 year-records, we can use the critical value for 12 data points in the H-table. To establish a safer measure we assumed that the test results are reliable only if the difference between the critical value and the testing value is larger than a predefined value (we used 2). Under these circumstances only 15 and 32 stations' test results, out of over 1800 stations, turned out to be unreliable at 5% and 1% significance levels, respectively. Thus, more than 98% of stations are used safely for hypothesis testing.

Table 1 shows the hypothesis test results using both tests. Only one-tenth of the stations are accepted at 5% significance level and about 17% of stations are accepted at 1% significance level.

Hence, the Bartlett and Hartley tests both show that the TG data are not stationary in variance. Out of over 1800 TG stations, 83% of the stations failed to pass the Bartlett test at 99% confidence interval. The results of Hartley test were similar. About 85% of the stations fail the test, confirming the results of the Bartlett test.

References

Douglas, B. C., M. S. Kearney, and S. P. Leatherman, eds. 2001. *Sea level rise: History and consequences.* Academic Press, San Diego, California, USA.

Emery, W. J., and E. Thomson. 2001. *Data analysis methods in physical oceanography*, second and revised edition, pp. 373–74. Elsevier.

Gornitz, V., and A. Solow. 1991. Observations of long-term tide-gauge records for indications of accelerated sea level rise. In *Greenhouse-gas-induced climatic change: A critical appraisal of simulations and observations*, M. E. Schlesingered, ed., pp. 347–367. Elsevier, Amsterdam.

Hannah, J. 1990. Analysis of mean sea level data from Zealand for the Period 1899–1988. *Journal of Geophy. Res.* 95(B8):12399–12405.

IPCC Third Assessment Report. 2001. Climate change 2001: The scientific basis. On line report http://www.ipcc.ch (accessed on 24th September 2002).

Iz, H. B., and C. K. Shum. 2000. Mean sea level variatlon in the South China Sea from four decades of tidal records in Hong Kong. *Mar. Geod.* 23(4):221–233.

Kendall, M. 1973. *Time-Series*, second edition, pp. 70–85. Charles Griffin and Company Ltd.

Lambeck, K. 1988. *Geophysical geodesy, The slow deformations of the earth*, pp. 108–168. Clerandon Press.

Neter, and Wasserman, 1974. *Applied linear statistical models*, pp. 763–766. Richard D. Irwin, Illinois, USA.

Permanent Service for Mean Sea Level (PSMSL). 2001. Home page, http://www.pol.ac.uk/psmsl/ (accessed on 24th September 2002).

Pugh, D. T. 1987. *Tides, surges and mean sea level*, pp. 300–330. John Wiley & Sons Ltd.

Sarkar, D., and R. G. Laha. 1955. A modification of the variate-difference method. *Econometrica* 23(1):67–75.

Woodworth, P. L., M. N. Tsimplis, R. A. Flather, and I. Shennan. 1999. A review of the trends observed in Bristish Isles. mean sea level data measured by tide gauges. *Geophys. J. Int.* 136:651–670.