INTERNATIONAL JOURNAL OF MARITIME TECHNOLOGY

IJMT Vol.13/ Winter 2020 (41-49)

Available online at: www. ijmt.ir

Assessment of noise in time series analysis for Buoy tide observations

Saeed Farzaneh[1*], Mohammad Ali Sharifi[2], Kamal Parvazi[3], Bahareh Namazi[4]

[1] *Assistant Professor; School of Surveying and Geomatics Engineering Faculty of Engineering University of Tehran; farzaneh@ut.ac.ir.*
[2] *Associate Professor; School of Surveying and Geomatics Engineering Faculty of Engineering University of Tehran; sharifi@ut.ac.ir.*
[3] *PhD Student. School of Surveying and Geomatics Engineering Faculty of Engineering University of Tehran; kamal.parvazi@ut.ac.ir.*
[4] *MSc. School of Surveying and Geomatics Engineering Faculty of Engineering University of Tehran; bahare.namazi@ut.ac.ir.*

ARTICLE INFO

Article History:
Received: 22 Dec. 2019
Accepted: 25 Jun. 2020

Keywords:
Buoy station's time series
Least Square Estimation
Least Square Harmonic Estimation
tides observation's noise analysis
Maximum Likelihood Estimation

ABSTRACT

To extract valid results from time series analysis of tides observations, noise reduction is vital. This study aimed to use a precise statistical model to investigate noise types. Noise component amplitude of the proposed model was studied by Least Square Estimation (LS-VCE) through different statistical models: (1) white noise and auto-regressive noise, (2) white noise and Flicker noise, (3) white noise and random walk noise, (4) white noise and Flicker noise and random walk, and (5) auto-regressive noise and Flicker noise. Based on the values obtained for the Likelihood Function, it was concluded that the noise model that can be considered for observations of the Buoy time series includes two white and Flicker noises. In addition, tide forecasting for all stations was done by extracting important frequency calculated in two cases: (1) the first case in which matrix of observation weight matrix was considered as the unit matrix or the noise model was just a white noise (2) the case in which matrix of observation weight matrix was considered as a combination of white and Flicker noises. The results show that use of precise observation weight matrix resulted in 11 millimeter difference compared to the case in which observation with unit weight matrix was used.

1. Introduction

Tides data are used for wide variety of application such as hydrography or navigation. Accordingly, data analyses and forecasting as well as knowledge of data's structure and nature have been considered very important. The nature of tide's data can be expressed by their fundamental frequencies; tide's frequencies can be extracted by an observation-based method. To do this, a harmonic analysis of sea surface level by Fourier series expansion is utilized. Harmonic analysis is a very valuable method for tide data analysis. In this field Amiri-simkooie et al [1], has worked on extraction of tide's data frequencies for the United Kingdom gauges. Mousavian and Mashhadi Hossienali [2] used single-variable analysis for this purpose.

One of the most important part of tide's data is the noise structure which has not been studied properly yet. It's worth noting that without enough information related to the tide's data noises, frequency extraction will confront many problems. There are different approaches to study geodetic time series' noises, such as: spectral power method [3] , Maximum Likelihoods Estimation [4], Limited Maximum Likelihoods Estimator [5], Minimum norm quadratic unbiased estimator [6,7], Best Linear Unbiased invariant quadratic estimator [8,9], Helmert method [10,11], Bayesian Estimation [12], and Least Square Estimation [13].

Because of the noise significant effect on the time series and this fact that time series can always be accompanied by noise, existence of the noise in time series must be studied carefully. The ideal condition is when the observations are independent. Time series may have spatial and temporal correlation. If the observations do not have temporal correlation, they have white noise. But in the time series derived from

[DOR: 20.1001.1.23456000.2020.13.0.5.7]

buoy, temporal correlation is present, which means the existence of color noise in observations. Therefore, we must inevitably need to know the effects of other noise in observations. In the methods for estimating variance components, including MINQUE, BIQUE, Helmert, REML all relationships provide an estimate of the variance components based on the assumption that the observations have a normal distribution while LS-VCE avoids the assumption of normal distribution for observations. LS-VCE method is based on the least squares principle therefore it is very flexible and it is easy to apply the least squares theory in this method. Some of the features that make using of least squares method are as follow: In this method we can provide a general class of unbiased estimators that are independent of the distribution considered for observations, minimal variance estimators even for one class of observations with elliptical distribution can be presented. In this method, covariance matrix, the estimator variance is obtained directly and conveniently. LS-VCE provides a clearer geometric interpretation of least squares, the properties of matrix Normal and orthogonality are easily established, and unstable observations and statistical tests are presented easily by this method. Finally, we can study the structure of model as if the number of considered parameters is suitable for a model or not. In other words, other methods can be considered a special case of the LS-VCE method. Our observations may not follow the normal distribution and for this reason, in this paper, the LS-VCE method is used to estimate noise of time series derived from buoy observations. Next, in order to determine the noise structure of the data, the logarithm of the likelihood function which is based on the maximum likelihood method is used and the appropriate random model was studied using a valid statistical model [1,14,32].

To extract the tidal frequencies, many studies have analyzed sea level height with different methods such as the Fourier and wavelet. Historically, Fourier spectral analysis has been used to examine the global energy and frequency distributions of SSH time series [37]. Its popularity is due to the prowess of the method, as well as its simplicity of application. As a result, the term ’spectrum’ has become almost synonymous with the Fourier transform of time series [38]. Fourier analysis, however, exhibits some drawbacks in analyzing time series, which are unequally sampled or those with data gaps [39]. Filling the gaps with inverted data might be erroneous when large gaps present in the time series, or due to the approximation approach used for interpolation [40]. In this paper, we focus on time-invariant base-functions to detect tidal frequencies using tidal observation analysis without predefining these frequencies. For this purpose, the Least Squares Harmonic Estimation (LS-HE) method is used. Our motivation to select these techniques is: 1- they are not limited to evenly-spaced data nor to integer frequencies; 2- they allow us to detect common-modes of signals, in a least squares sense, and thus are very efficient in detecting cyclic patterns; and (3) they can be easily used for univariate and multivariate examples [1,43,32].

In this study, Maximum Likelihood function values was expanded in multi variable states. All these methods, except spectral power, define observation’s covariance matrix. On the other words, all these methods estimate the noise’s amplitude in the observation random model, thus they are called variance component estimator methods. Precise estimation of observation weight matrix using variance-covariance estimation method plays an important role in tides component phase and amplitude as well as estimation of Mean Sea Level- Sea Surface Height(MSL-SSH).

proposed noise’s model confirms white and flicker noise for observation. After determination of observation weight matrix, tide forecasting for two months is done and compared in two different states: weighted and un-weighted.

2. Methodology

2.1. Noises Least Square Estimation Principal

In this study, Least Square Estimation method is applied to estimate noise’s amplitude for tide’s data. Least Square Estimation method (In previous research, it was used for the noise assessment of GPS time series.) creates unbiased estimators with minimum variance. Different studies have used multi variable time series for noise’s amplitude analyzing [1, 14–29]. This method resulted in close estimation of real data in tide’s forecasting by considering dependency between different time series. As previously mentioned, we utilized Likelihood logarithm function based on Least Square Estimation method to determine the noise model that the equation has been expanded for multi-variable state.

2.2. Univariate Least Square Noise Estimation

Least Square Noise Estimation is presented at 1988 by Teunissen [13]. Equation (1) shows the linear model of observations.

$$E(\underline{y}) = Ax, \quad D(\underline{y}) = Q_y = \sum_{k=1}^{p} s_k Q_k \tag{1}$$

In equation 1, E and D are mathematical expectation and dispersion operators of observation respectively, A is design matrix of dimension $m \times n$, x is unknown vector of dimension n, y is observations vector of dimension m, s_k is variance of unknown unit weight, Q_y is covariance matrix of observations of dimension

$m \times m$ and $Q_k, (k=1,2,...,p)$ are known co-factor matrices of models. The covariance matrix Q_y is considered as a definite-positive matrix and is represented as an unknown linear combination of known co-factor matrices Q_k. Cofactor matrices are assumed to be symmetric so that sum of the series in equation (1) becomes a definite-positive matrix. The necessary condition for obtaining a regular solution in stochastic model, is that cofactor matrices are linearly independent. For more studies, the works of Amiri-Simkooei and Xu et al. [1,30] are recommended.

Least square estimation from p dimensional vector, consist of unknown covariance in $\underline{\hat{\sigma}}$ random observation model can drive from equation.(2):

$$\underline{\hat{\sigma}} = N^{-1}\underline{l} \tag{2}$$

where matrix N elements (with $p \times p$ dimension) and vector $\underline{l}$ (with dimension of p) can be obtained as follows:

$$n_{kl} = \frac{1}{2}\,\mathrm{tr}\,(Q_y^{-1} P_A^{\perp} Q_k Q_y^{-1} P_A^{\perp} Q_l) \tag{3}$$

$$\underline{l}_k = \frac{1}{2}\underline{\hat{e}}^T Q_y^{-1} Q_k Q_y^{-1} \underline{\hat{e}}, \qquad k,l = 1,2,...,p \tag{4}$$

Due to presence of Q_y in Eqs. (3) and (4), the elements of covariance matrix can be estimated by applying a recursive method. In this way, first we assign an initial value for covariance elements. Then new estimation value is replaced in each step. The repeat process continues till estimated values for covariance elements do not show significant changes. Since the $\underline{\hat{\sigma}}$ estimators are driven by the least square method, the covariance matrix of estimated values can be obtained from inverse of normal matrix N as follows:

$$Q_{\hat{\sigma}} = N^{-1} \tag{5}$$

Accordingly, the $\underline{\hat{\sigma}}$ estimators precision can be calculated.

2.3. Multivariate Least Square Noise Estimation

The time series used have both temporal and spatial dependence. For this reason, the multivariate method is used to analyze them [31]. If both time dependency and between series correlations were taken into account, the least square noise estimation method enables us to estimate all the parameters at once. In multi-variable method, several solutions were presented by Simkooie for random models [32].

These solutions are: general model, specific (particular) model and more practical model. In this study, the more practical model is chosen where Q matrix is:

$$Q = \sum_{k=1}^{p} s_k Q_k \tag{6}$$

In this case matrix Q_k and s_k coefficients are unknown and should be estimated by the least square noise estimation method. For this purpose, it is first assumed that matrix Q_k is known, then s_k coefficients is calculated using the following equation [32]:

$$\underline{\hat{s}} = N^{-1}\underline{l} \tag{7}$$

Where matrix N elements (with $p \times p$ dimension) and vector $\underline{l}$ (with dimension of p) can be obtained as follows:

$$n_{kl} = \frac{r}{2}\,\mathrm{tr}\,(Q^{-1} P_A^{\perp} Q_k Q^{-1} P_A^{\perp} Q_l) \tag{8}$$

$$\underline{l}_k = \frac{1}{2}\mathrm{tr}(\underline{\hat{E}}^T Q^{-1} Q_k Q^{-1} \underline{\hat{E}} \Sigma^{-1}), \quad k,l = 1,2,...,p \tag{9}$$

If matrix Σ is unknown, the unknown model's variable can be calculated in two steps: 1) calculating an estimation of matrix Σ, 2) estimating s_k coefficients by previous process. Matrix Σ estimation can be driven from the following equation:

$$\underline{\hat{\Sigma}} = \frac{\underline{\hat{\Sigma}}^T Q^{-1} \underline{\hat{\Sigma}}}{m-n} \tag{10}$$

Therefore, equation (9) will be rewritten as follows [32]:

$$\underline{l}_k = \frac{m-n}{2}\mathrm{tr}(\underline{\hat{E}}^T Q^{-1} Q_k Q^{-1} \underline{\hat{E}} (\underline{\hat{E}}^T Q^{-1} \underline{\hat{E}})^{-1}) \tag{11}$$

2.4. Random Model Determination for Univariate Time

Before utilizing the least square noise estimation, it is necessary to specify the observation noise's structure. An appropriate noise structure for random model can be achieved by employing w-test statistic and maximum likelihood logarithm function.

In order to investigate the existence of different noises in the random model related to the time series, two assumptions are considered, the null hypothesis (H_0) and alternative hypothesis (H_a) can be defines as follows [1]:

$$H_0 : Q_y = \sigma_1^2 \; Q_0 \tag{12}$$

$$H_a : Q_y = \sigma_1^2 \; Q_0 + C_y \delta \tag{13}$$

where Q_0 and C_y are known cofactor matrixes, σ_1^2 is covariance's parameter for the cofactor matrix Q_0, and δ is the unknown covariance parameter. The w-test statistic equation for evaluating H_0 versus H_a hypotheses is as follows [1]:

[DOR: 20.1001.1.23456000.2020.13.0.5.7]

$$\underline{w} = \frac{\hat{\underline{e}}^T Q_y^{-1}\,[\tfrac{1}{2} C_y - \frac{tr(C_y Q_{\hat{e}}^-)}{2b} Q_y]\, Q_y^{-1} \hat{\underline{e}}}{[\tfrac{1}{2} tr(C_y Q_{\hat{e}}^- C_y Q_{\hat{e}}^-) - \tfrac{1}{2b} tr(C_y Q_{\hat{e}}^-)\, tr(C_y Q_{\hat{e}}^-)]^{1/2}} \quad (14)$$

In this equation, $b = m - n$ is the degree of freedom of the functional model, and Q_y , $P_A^{\perp}$ and $\hat{\underline{e}}$ are calculated under null hypothesis. Matrix $Q_{\hat{e}}^-$ is reflex inverse matrix of residual vectors covariance for $Q_{\hat{e}} = P_A^{\perp} Q_y$ driven from $Q_{\hat{e}}^- = Q_y^{-1}\, Q_{\hat{e}}\, Q_y^{-1}$. In special cases that null and alternative hypotheses are defined using equations (15) and equations (16), equation (14) changes to equations (17) [1,2]:

$$H_0: Q_y = \sigma_w^2 I \quad (15)$$

$$H_a: Q_y = \sigma_w^2 I + C_y \delta \quad (16)$$

$$\underline{w} = \frac{b\, \hat{\underline{e}}^T C_y \hat{\underline{e}} - tr(C_y P_A^{\perp})\, \hat{\underline{e}}^T \hat{\underline{e}}}{\sigma_w^2 [2b^2\, tr(C_y P_A^{\perp} C_y P_A^{\perp}) - 2b\, tr(C_y P_A^{\perp})\, tr(C_y P_A^{\perp})]^{1/2}} \quad (17)$$

The mathematical expectation and variances of this statistic under null hypothesis are equal to zero and one, respectively. To test H_0 versus H_a hypotheses, Chebyshev's inequality can be used which is independent of distribution function [1].

By implying cofactor matrix C_y for different noise's types in equations (14) or (17), the presence possibility of them in dataset can be examined. The noise which is corresponded to maximum absolute vale of w-test statistic is chosen as current noise [1,33]. In fact, w-test statistic values under H_0 hypothesis has great values that for any confidence interval in Chebyshev's inequality is greater than critical value. Accordingly, the alternative hypothesis is considered as noise model which has maximum deviation from H_0 hypothesis, and thus has the maximum w-test statistic value.

As mentioned above, maximum likelihood logarithm function is another method to examine data noise structure which is based on maximum likelihood. In this method, if observation vector $\underline{y}$ has multi-variable normal distribution as follows:

$$\underline{y} \sim N(Ax, \sum_{k=1}^{p} \sigma_k Q_k) \quad (18)$$

Then likelihood function logarithm related to vector $\underline{y}$ is [1]:

$$\begin{aligned} \ln L(y; x, \sigma) &= -\frac{m}{2}\ln 2\pi - \frac{1}{2}\ln\det(Q_y) - \frac{1}{2}(y-Ax)^T Q_y^{-1} (y-Ax) \otimes \\ &= -\frac{m}{2}\ln 2\pi - \frac{1}{2}\ln\det(Q_y) - \frac{1}{2}\hat{e}^T Q_y^{-1} \hat{e} \end{aligned} \quad (19)$$

Unlike w-test method where observation noise's type determination is done before amplitude estimation, this method estimates the different noise's amplitude at the first step. Then their presence possibility in observation is evaluated. After estimation of different Q_y values for different C_y values, they are replaced in equation. (19). Finally, a Q_y covariance matrix is chosen that provides the maximum value for likelihood function logarithm.

2.5. Random model determination for multi-variable time series

If we aimed to determine noise's structure by w-test statistic for multi-variable linear model in equation.(14), then H_0 and H_a hypotheses are defined as follows:

$$H_0: Q_y = \Sigma \otimes I_m \quad (20)$$

$$H_a: Q_y = \Sigma \otimes (I_m + C_y \nabla) \quad (21)$$

Where ∇ is an unknown parameter.

Here, we expand w-test statistical equation from single-variable to multi-variable. To do this, it is enough to replace this multi-variable phrase in equation (14):

$$\hat{\underline{e}} \leftarrow vec(\hat{\underline{E}}) \;,\; Q_y \leftarrow \Sigma \otimes I_m \;,\; C_y \leftarrow \Sigma \otimes C_y$$
$$P_A^{\perp} \leftarrow I_r \otimes P_A^{\perp} \;,\; m \leftarrow m\,r \;,\; n \leftarrow n\,r$$

By this replacement and Kronecker multiple properties, equation (14) will be changed to equation (22):

$$\underline{w} = \frac{b\, tr(\hat{\underline{E}}^T C_y \hat{\underline{E}} \Sigma^{-1}) - r\, tr(C_y P_A^{\perp})\, tr(\hat{\underline{E}}^T \hat{\underline{E}} \Sigma^{-1})}{[2b^2 r\, tr(C_y P_A^{\perp} C_y P_A^{\perp}) - 2r^2 b\, tr(C_y P_A^{\perp})\, tr(C_y P_A^{\perp})]^{1/2}} \quad (22)$$

In this equation, $b = (m - n)r$ is the degree of freedom.

By use of likelihood function logarithm to determine multi-variable random model's structure and use of the abovementioned replacements in equation (19), equation (23) can be driven:

$$\ln L(y; x, \sigma) = -\frac{mr}{2}\ln 2\pi - \frac{1}{2}\{m \ln\det(\Sigma) + r \ln\det(Q)\} - \frac{r(m-n)}{2} \quad (23)$$

Where

$$Q = I_m + C_y \nabla \quad (24)$$

3. Numerical results

This study aims to determine tide frequencies and the noise's structure. For this purpose, we used 57 Buoy stations. The data has been selected for a time period of 2005 to 2017 with 15-minute rate. The stations positions are shown in figure 1:

[DOR: 20.1001.1.23456000.2020.13.0.5.7]

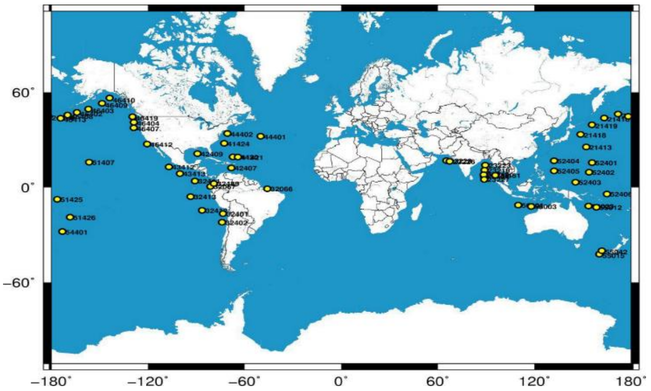


Figure 1. Buoy stations used in this study

In the following, first, numerical results for tide frequencies detection by harmonic least square estimation is presented. Next, tide data noise's structure is determined by the least square noise estimation.

Since detection of spectral power for common frequencies between time series by multi-variable harmonic analysis is much easier than single variable, the results from least square multi-variable spectral power of frequencies from 57 Buoy stations data are presented in figure 2:

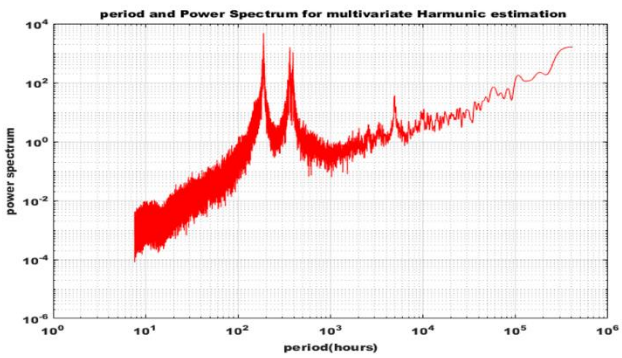


Figure 2. least square multi-variable spectral power of frequencies from 57 Buoy stations data

Then, frequencies with greater spectral power are extracted. The maximum spectral power specified in Figure 2 is related to the main components of tides. These components include M_2, S_2, K_1, and O_1. M_2, the tidal component lunar half−day. Due to the proximity of the moon to the earth is the most important factor in creating tides. This component has a period of 12.42 hours and a speed of 28.948 degrees per hour. The range of these components will be considered as an equal basis and other parameters are measured with respect to these components. S_2, the side component solar half−day. This component is the most influential component after M_2 and it's about the sun. The main features of this component are a 12-hour period, a speed of 30 degrees per hour and a range of 0.46. K_1 (the tidal component solar daily) and O_1 (the tidal component lunar daily) are the most important components of the day. The range of daily components depends on the angle of inclination of the moon and with the change of the inclination of the moon, the amplitude also changes. When the moon passes through the equator, that is, at the angle of zero inclination, the amplitude of the daily components has its lowest value, and when the angle of inclination of the moon has its maximum value, the range of these components also increases. K_1 component has a period of 23.93 hours and a speed of 15.041 degrees per hour. And O_1 component has a period of 25.82 hours and a speed of 13.943 degrees per hour

3.1 Tide observations noise's model

In this section, we will determine tide data noise's structure by multivariable time series least square noise estimation. According to equations (8), (9), we need to calculate the inverse of matrix Q for estimation of tide data noises by least square estimation. Due to lack of memory to store the information, calculation was not possible. Therefore, we used 2 hourly averaged data. Cofactor matrixes, dimensioned $m \times m$, can be obtained from the following equations where m is number of observations in time series. White noise cofactor matrix:

$$Q = I_{m\times m} \tag{25}$$

Flicker's cofactor noise matrix is defined as [34,35,36]:

$$Q_{ij}^{f} = \begin{cases} \frac{9}{8} & \text{if} \quad \tau \neq 0 \\ \frac{9}{8}\left(1 - \frac{\log\tau/\log 2 + 2}{24}\right) & \text{if} \quad \tau \neq 0 \end{cases} \tag{26}$$

Random walk cofactor matrix defined as below, where m is number of observations, T is the whole time of observation per year and f_s is sample frequency[34,35,36]:

$$Q_{rw} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 2 & \dots & 2 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & \dots & m \end{bmatrix}; f_s = \frac{m-1}{T} \tag{27}$$

C_y cofactor matrix for auto-regressive noise is:

$$C_y = \exp(-\alpha\tau) \tag{28}$$

where $\tau = |t_j - t_i|\tau$

As it is obvious, in auto-regressive cofactor matrix C_y, parameter α can have different values.

3.2 Line fitting method

To determine suitable α for autoregressive noise, there are other solutions than w-test statistic. Because of limitation in memory in w-test method and since the noise's effect is more obvious on time series spectra with longer lengths, line fitting solution based on the least square method for power spectra chart is used to determine suitable value for α. In figure 2, if a line be fitted to middle and lower frequencies, line's slope shows suitable value for α for autoregressive noise. To do this, spectra power values and frequencies logarithm should be first calculated. Moreover, in interval of long term frequencies to semi-day frequencies, a line be fitted to power spectra chart; the line's slope shows suitable value for α in autoregressive noise for long period frequencies.

After data normalization, as shown in figure 3, the slope of fitted line to power spectra chart in middle and low frequencies equals to -1857614002477414 and in long to semi-day frequencies equals to -0.960163273277107.

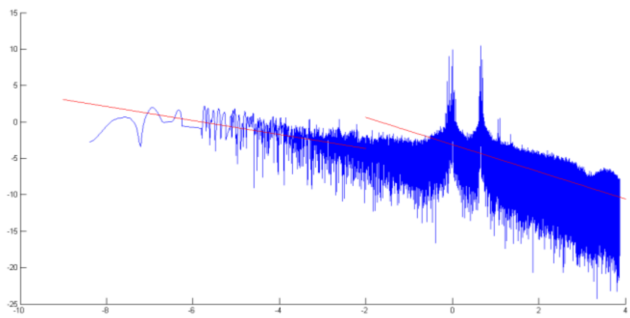
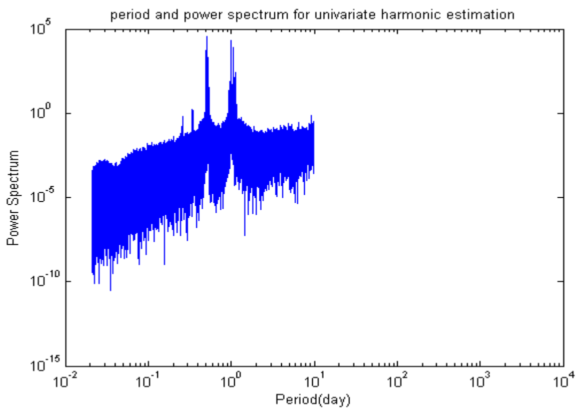
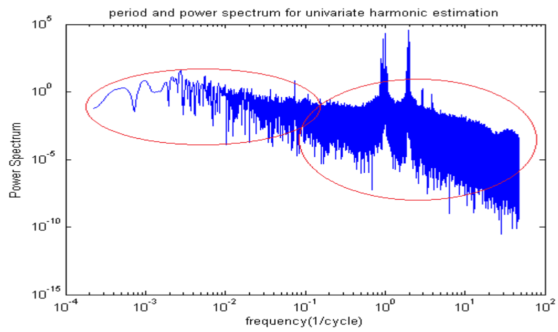


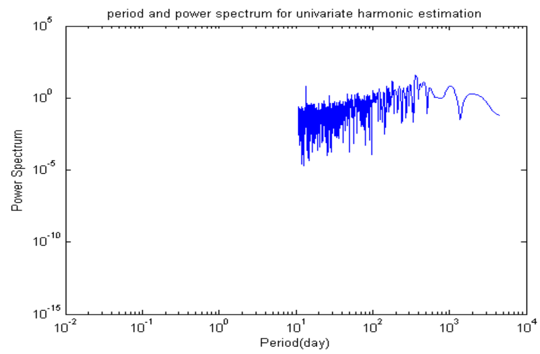
Figure 3. multivariable least square power spectra chart for 57 Buoy stations with sample rate of 15 minutes. Red lines: fitted lines to the chart by least square method



(a)



(b)



(c)

Figure 4. a, b and c, **multivariable least square power spectra chart for 57 Buoy stations with sample rate 15 minutes in low and high periods intervals**

To increase the accuracy of data results, dataset is divided in two parts; low and middle frequencies from .020833333 to 3.9133367 days and long to half-day period from 3.817232 to 4.3837606e+03; the high and low interval is shown in figure 4. To determine noise's model, likelihood function logarithm in equation (23) for different random model is calculated for these two parts. For this purpose, five different random models are considered and their likelihood function logarithm values are shown in Table 2. As can be seen in Table 1, likelihood function logarithm maximum value is related to combined Flicker and white noise models in random model.

Table 1. likelihood function logarithm normalized values for 5 random models for tide multivariable time series

Noise model	Likelihood function logarithm normalized values for low and middle frequencies	Likelihood function logarithm normalized values for high to semi-days frequencies
White + Flicker	2338.2285	218.9269
White + Autoregressive	-353.35441	-6027.3356
White + Random Walk	-3591.1898	-6091.2739
White + Flicker + Random Walk	-3455.9613	-5275.1757
Flicker + Autoregressive	-34.8323	106.220

According to the results in table 1, it can be concluded that tide date noises consist of: white noise + Flicker noise. If random model is considered as a combination of white and Flicker noises, their amplitude can be calculated by multivariable least square noise estimation method. s_k values for white and Flicker noises are estimated as $s_w = 0.114262$ and $s_w = 1.0496366$, respectively.

The calculated multivariable power spectra time series based on this random model is shown in figure 5.

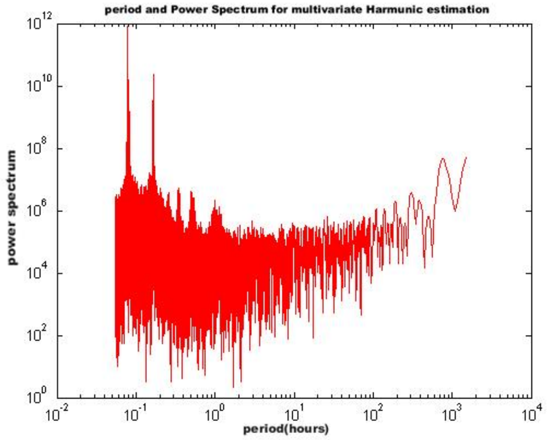


Figure 5. multivariable least square power spectra for weighted state

In order to evaluate precision of Buoy stations data, we predicted tides for 1-month interval. The results of this prediction in for the following two conditions are presented in Table (2), 1) assuming unit matrix for observation weight matrix or assuming white model as noise model and 2) combination of White and Flicker noises for weight matrix. As it can be seen, the latter condition (i.e. combination of White and Flicker noises for weight matrix) provided higher accuracy than the former one (i.e. assuming white model as noise model).

Table 2. RMSE residuals value in weighted and non-weighted for 1-month prediction of tide Buoy station

Station	Non-weighted	Weighted (Flicker+white)
RMSE(m)	0.0934888	0.0829430

Figures 6 and 7 present charts for weighted and non-weighted of 1-month prediction of tide Buoy station. Red diagram is related to real data and green diagram is related to the predicted data.

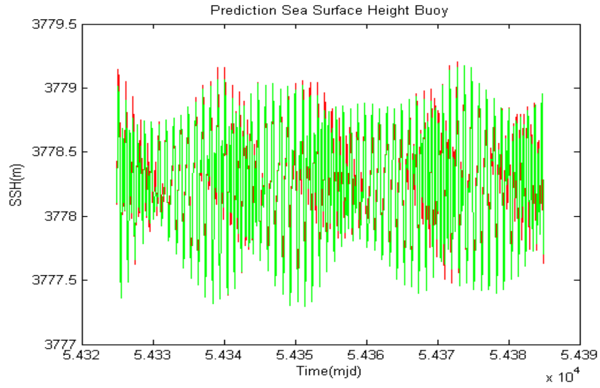


Figure 6. comparison between known and predicted data, in non-weighted mode for 1-month interval (red diagram: known data, Green diagram: predicted data)

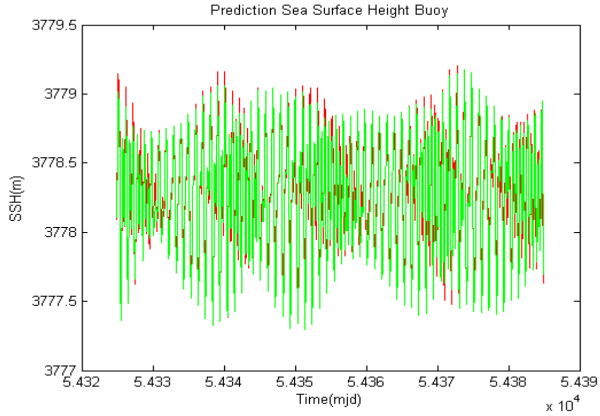


Figure 7. comparison between known and predicted data, in weighted mode for 1-month interval (Red diagram: known data, Green diagram: predicted data)

4. Conclusions

In this study, the tide observation noise's structure was initially defined by the least square estimation. For this purpose, Buoy data obtained from 57 tide stations during the period of 2005 to 2017 was analyzed. The proper noise model for tide's data was determined by likelihood function logarithm, and combination of white and Flicker's noises. Next, the noise component's amplitude was calculated using the least square estimation method.

Then, tide prediction for a duration of 1 month was conducted in two conditions: 1) Assuming unit matrix for observation weight matrix or assuming white model as noise model and 2) combination of White and Flicker noises for weight matrix. The results show that use of precise observation weight matrix resulted in 1 millimeter difference compared to the case in which observation with unit weight matrix was used.

5. References

1- Amiri-Simkooei, A. (2007). *Least-squares variance component estimation: theory and GPS applications*", PhD thesis, Mathematical Geodesy and Positioning,

Faculty of Aerospace Engineering, Delft University of Technology, Delft, Netherlands, 2007.
2- Mousavian, R., & Hossainali, M. M. (2012). *Detection of main tidal frequencies using least squares harmonic estimation method*. Journal of Geodetic Science, 2(3), 224-233.
3- Mandelbrot, B. B., & Van Ness, J. W. (1968). *Fractional Brownian motions, fractional noises and applications*. SIAM review, 10(4), 422-437.
4- Kubik, K. (1970). *The estimation of the weights of measured quantities within the method of least squares*. Bulletin Géodésique (1946-1975), 95(1), 21-40.
5- Koch, K. R. (1986). *Maximum likelihood estimate of variance components*. Bulletin Gæodésique, 60(4), 329-338.
6- Rao, C. R. (1971). *Estimation of variance and covariance components—MINQUE theory*. Journal of multivariate analysis, 1(3), 257-275.
7- Rao, C. R., Rao, C. R., Statistiker, M., Rao, C. R., & Rao, C. R. (1973). *Linear statistical inference and its applications* (Vol. 2, pp. 263-270). New York: Wiley.
8- Koch, K.R. (1978). *Schätzung von varianzkomponenten. Allgemeine Vermessungs Nachrichten*, 85, pp.264-269.
9- Koch, K. R. (1999). *Parameter estimation and hypothesis testing in linear models*. Springer Science & Business Media.
10- Helmert, F. R. (1872). *Die Ausgleichungsrechnung nach der Methode der kleinsten Quadrate: mit Anwendungen auf die Geodäsie*, die Physik und die Theorie der Messinstrumente (Vol. 1). Teubner.
11- Helmert FR, *Die ausgleichungsrechnung nach der methode der kleinsten quadrate*. 3. AUFL., Leipzig/Berlin,1924.
12- Koch, K. R. (1987). *Bayesian inference for variance components. Manuscripta geodaetica*, 12(4), 309-313.
13- Teunissen, P. (2004). *Towards a least-squares framework for adjusting and testing of both functional and stochastic models*. Technical report, Delft University of Technology. A reprint of original 1988 report is also available in 2004, No. 26, http://www.lr.tudelft.nl/mgp. Internal research memo, Geodetic Computing Centre.
14- Amiri-Simkooei, A. R., & Tiberius, C. C. J. M. (2007). *Assessing receiver noise using GPS short baseline time series. GPS solutions*, 11(1), 21-35.
15- Barnes, J. B. (2000). *Real time kinematic GPS and multipath: characterisation and improved least squares modelling* (Doctoral dissertation, University of Newcastle upon Tyne).
16- Bischoff, W., Heck, B., Howind, J., & Teusch, A. (2005). *A procedure for testing the assumption of homoscedasticity in least squares residuals: a case study of GPS carrier-phase observations*. Journal of Geodesy, 78(7-8), 397-404.
17- Bischoff, W., Heck, B., Howind, J., & Teusch, A. (2006). *A procedure for estimating the variance function of linear models and for checking the appropriateness of estimated variances: a case study of GPS carrier-phase observations*. Journal of Geodesy, 79(12), 694-704.
18- Bona, P. (2000). *Precision, cross correlation, and time correlation of GPS phase and code observations*. GPS solutions, 4(2), 3-13.
19- Chen, Y. Q. (1990). *Assessments of observations using minimum norm quadratic unbiased estimation (MINQUE)*. CISM J. ACSGC, 44, 36-49.
20- Fotopoulos, G. (2005). *Calibration of geoid error models via a combined adjustment of ellipsoidal, orthometric and gravimetric geoid height data*. Journal of Geodesy, 79(1-3), 111-123.
21- Kusche, J. (2003). *A Monte-Carlo technique for weight estimation in satellite geodesy*. Journal of Geodesy, 76(11-12), 641-652.
22- Kusche, J. (2003). *Noise variance estimation and optimal weight determination for GOCE gravity recovery*. Advances in Geosciences, 1, 81-85.
23- Satirapod, C., Wang, J., & Rizos, C. (2002). *A simplified MINQUE procedure for the estimation of variance-covariance components of GPS observables*. Survey Review, 36(286), 582-590.
24- Schön, S., & Brunner, F. K. (2008). *Atmospheric turbulence theory applied to GPS carrier-phase data*. Journal of Geodesy, 82(1), 47-57.
25- Schön, S., & Brunner, F. K. (2008). *A proposal for modelling physical correlations of GPS phase observations*. Journal of Geodesy, 82(10), 601-612.
26- Teunissen, P. J., Jonkman, N. F., & Tiberius, C. C. J. M. (1998). *Weighting GPS dual frequency observations: bearing the cross of cross-correlation*. GPS Solutions, 2(2), 28-37.
27- Tiberius, C. C. J. M., & Kenselaar, F. (2000). *Estimation of the stochastic model for GPS code and phase observables*. Survey Review, 35(277), 441-454.
28- Wang, J., Stewart, M. P., & Tsakiri, M. (1998). *Stochastic modeling for static GPS baseline data processing*. Journal of Surveying Engineering, 124(4), 171-181.
29- Xu, P., Shen, Y., Fukuda, Y., & Liu, Y. (2006). *Variance component estimation in linear inverse ill-posed models*. Journal of Geodesy, 80(2), 69-81.
30- Xu, P., Liu, Y., Shen, Y. and Fukuda, Y.(2007). *Estimability analysis of variance and covariance components*. Journal of Geodesy, 81(9), pp.593-602.
31- Williams, S.D., Bock, Y., Fang, P., Jamason, P., Nikolaidis, R.M., Prawirodirdjo, L., Miller, M. and Johnson, D.J.(2004). *Error analysis of continuous GPS position time series*. Journal of Geophysical Research: Solid Earth, 109(B3).
32- Amiri-Simkooei, A. R. (2009). *Noise in multivariate GPS position time-series*. Journal of Geodesy, 83(2), 175-187.

33- Amiri-Simkooei, A. R., Tiberius, C. C. J. M., & Teunissen, S. P. (2007). *Assessment of noise in GPS coordinate time series: methodology and results*. Journal of Geophysical Research: Solid Earth, 112(B7).

34- Zhang J, Bock Y, Johnson H, Fang P, Williams S, Genrich J, Wdowinski S, Behr J (1997). *Southern California Permanent GPS Geodetic Array: Error analysis of daily position estimates and site velocitties*. Journal of Geophysical Research, 102: 18035–18055.

35- Johnson HO, Wyatt FK (1994). *Geodetic network design for fault-mechanics studies*. Manuscripta Geodaetica, 19: 309–323.

36- Mao A, Harrison CGA, Dixon TH (1999). *Noise in GPS coordinate time series*. Journal of Geophysical Research, 104(B2): 2797–2816.

37- Boashash, B. and Putland, G., 2003, *Polynomial Wigner-Ville Distributions and Design of High-Resolution Quadratic TFDs with Separable Kernals. In TIme-Frequency Signal Analysis and Processing: A Comprehensive Reference* (pp. 3-27), Elsevier Ltd.

38- Wu, Z., Huang, N.E. and Chen, X., 2009, *The multi-dimensional ensemble empirical mode decomposition method*. Advances in Adaptive Data Analysis, 1(03), 339-372.

39- Rubin, D. B., 2002, *Statistical Analysis with Missing Data*. ISBN: 978-0-471-18386-0.

40- Papa, F., Legrésy, B. and Rémy, F., 2003, *Use of the Topex–Poseidon dualfrequencyradar altimeter over land surfaces*. Remote sensing of Environment, 87(2-3), 136-147.

[41] Tomczak, M. (2000). *Lecture Notes in Oceanography*. Flinders University, Adelaide, Australia School of Chemistry, Physics & Earth Sciences.

[42] Fitzpatrick, R. (2010). *Newtonian dynamics*. Austin, Tex: The University of Texas, 2011 [2012-5-14]: 201-217. http://farside. ph. utexas, edu/teaching/336k/Newton.

[43] Pirooznia, M., Raoofian Naeeni, M. and Amerian, Y., 2019. *A Comparative Study Between Least Square and Total Least Square Methods for Time–Series Analysis and Quality Control of Sea Level Observations*. Marine Geodesy, 42(2), pp.104-129.