

## Least Squares Techniques for Extracting Water Level Fluctuations in the Persian Gulf and Oman Sea

Farzaneh, S.<sup>1\*</sup>, Forootan, E.<sup>2</sup> and Parvazi, K.<sup>3</sup>

1. Assistant Professor, Department of Surveying and Geomatics Engineering, Faculty of Engineering, University of Tehran, Tehran, Iran

2. Lecturer, school of Earth and Ocean Sciences, Cardiff University, Cardiff, United Kingdom

3. Ph.D. Student Department of Surveying and Geomatics Engineering, Faculty of Engineering, University of Tehran, Tehran, Iran

(Received: 28 Nov 2018, Accepted: 1 Oct 2019)

### Abstract

Extracting the main cyclic fluctuations from sea level changes of the Persian Gulf and Oman Sea is vital for understanding the behavior of tides and isolating non-tidal impacts such as those related to climate and changes in the ocean-sea circulations. This study compares two spectral analysis methods including: Least Squares Spectral Analysis (LSSA) and Least Squares Harmonic Estimation (LSHE), to analyze satellite altimetry derived sea surface height changes of the Persian Gulf and Oman Sea. SSH data are derived from about 16 years of satellite altimetry observations (1992 to 2008), including the Topex/Poseidon and Jason-1 missions. By analyzing the real data, we extract significant tidal components in the spectrum of LSSA and LS-HE including those with the period of 62.07, 173.3, 58.71, 45.68, 88.86, 364.2 and 117.5 days, which are interpreted as Principal Lunar semi-diurnal, Luni-Solar Diurnal, Principal Solar Semi-diurnal, Principal Lunar Diurnal, GAM2, annual, Solar Diurnal periods are dominant in the level fluctuations. Moreover, some tidal components appear in the spectrum of LSSA and LS-HE, from which the Moon's semi-diurnal component  $M_2$  is dominant. Also, to evaluate the efficiency of these two techniques, we run three experiments in each extracted frequency from LSSA, LS-HE, and astronomical tide tables are separately used to predict the sea level in the Persian Gulf and Oman Sea for three years. The results of this prediction indicate that RMSE from LSSA, astronomical table, and LS-HE is 0.101 m, 0.093 m, and 0.086 m, respectively. According to the results LS-HE is found a more efficient technique to analyze cyclic fluctuations from altimetry measurements.

**Keywords:** Persian Gulf and Oman Sea, Least Square Spectral Analysis (LSSA), Least Square Harmonic Estimation (LS-HE), Satellite Altimetry.

### 1. Introduction

The description, understanding and quantitative determination of the tides has been an important research topic in geodesy oceanography. Ocean tides, resulting from the gravitational attractions of the moon and the sun, causes more than 80% of the total variability of the sea surface. Tides have strong influence on coastal environment and the protection of its ecosystem, and play a significant role in climate and also needed for the precise treatment of space observations (Fok, 2012).

Advances in satellite radar altimetry technology have enabled a globally sampled record of sea surface and has become an important tool for monitoring global and regional sea surface height (SSH) (Fu and Cazenave, 2001), measuring level fluctuations of inland water bodies (Khaki et

al., 2015), and even soil moisture (Frappart et al., 2015; Papa et al., 2003). Thereby, complementing traditional tide gauge (in situ) measurements, which despite their valuable utility in reflect local sea level fluctuations, they have limitations such as inhomogeneous spatial distribution, inconsistency between reference datum's, and a suboptimal reference to the moving coast (Chelton et al., 2001).

Tide data are frequently used for different applications such as safe navigation and hydrographic surveys. Their identity can be expressed by their frequencies and noise structure. Tidal analysis, which is focused in this paper, tries to condense a long-term record of observations into a brief collection of time-invariant constants. Due to the periodic behaviour of the tide-generating

\*Corresponding author:

farzaneh@ut.ac.ir

forces (e.g., those resulting from the relative [to Earth] motions of the Moon and Sun), periodicities contained within a tidal record is often extracted in order to describe the tidal displacement at a location as a sum of the associated harmonics (see, e.g., <https://tidesandcurrents.noaa.gov/predhist.html>). A reliable tidal analysis and prediction requires a reliable knowledge on the (main) tidal frequencies and noise structure. So far, all the proposed methods for determining tidal frequencies have been Theorized based; in other words, on these methods applied the ephemeris of Moon, Sun and other planets to extract tidal frequencies without any use of tidal observations, as mentioned by Doodson (1954), Tamura (1993), Parvazi et al. (2015), Xi and Hou(1987). These methods assume that the tidal frequencies are known, but their amplitudes are unknown.

To extract the tidal frequencies, many studies have analysed 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 (Boashash and Butland, 2003). 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 (Wu et al., 2009). Fourier analysis, however, exhibits some drawbacks in analysing time series, which are unequally sampled or those with data gaps (Rubin, 2002). 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 (Papa et al., 2003).

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, we compare the application of the Least Square Spectral Analysis (Vanicek, 1969, 1971) and the Least Squares Harmonic Estimation (LS-HE) developed by Amiri-Simkooei and Asgari (2014), Amiri-Simkooei (2014, 2012), and Amiri-Simkooei (2007). 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.

Therefore, the mathematical objectives (MOs) of this study include: (MO1) justifying the mathematical (dis-)similarity of these two techniques and their relationship with the commonly used (discrete) Fourier Analysis; (MO2) assessing the accuracy of the extracted frequencies, while evaluating the effect of the noise that contaminate the observations, effect of blunders, impact of missing values in time series; and (MO3) recognizing the accuracy of extraction of nearby frequencies and effect of the data length in extracting the nearby frequencies. A detailed investigation of these objectives has not been provided in previous studies, thus, this assessment complements the literature by evaluating the skill of LSSA and LS-HE techniques for tidal analysis studies.

To assess the ability of LSSA and LS-HE, on a real case study, we apply them at the SSH time series of the Persian Gulf and the Oman Sea derived from the Topex/Poseidon mission (1992-2002), Jason-1 (2003-2008), and Jason-2 (2009-2014). The Persian Gulf, located in the southwest of the Asian continent is a shallow, semi-enclosed basin in a typical arid zone and is an arm of the Indian Ocean. It is located between the longitude of 48–57° E and the latitude of 24–30° N (Figure 9). This Gulf is connected to the deep Gulf of Oman through the narrow Strait of Hormuz. The Persian Gulf covers an area of approximately 226,000 km<sup>2</sup> with a length of 990 km. Its width varies from 56 to 338 km, separating Iran from the Arabian Peninsula with the shortest distance of about 56 km in the Strait of Hormuz. This basin has an average depth of about 35 m, and the deepest water depth is approximately 107 m (Purser and Seibold, 1973).

The remaining part of this study is organized as follows: in Section 2, the datasets of the study are introduced, and the methodology of their analysis is explained in Section 3. The results are reported in Section 4, and finally, the study is summarized and concluded in Section 5.

## 2. Data

The first studies regarding the use of the satellite for sea level were raised in 1969 in the Williamstown Conference of Solid Earth and Ocean Physics. Four years later in spatial laboratory, the first measures of sea level were performed by altimetry radar S-193 with accuracy of about one meter. Later, various satellites were used for different space missions (Aviso and Podaac, 2008).

Topex/Poseidon generated SSH measurements with RMSE accuracy better than 5 cm for a single-pass and better than 2 cm at global scale. To eliminate atmospheric effects, this satellite was placed in an orbit with high altitude, 1336 km to perform exact orbit modelling. With the progress of distance with laser, JGM3 (Joint Gravity Model) and non-gravity models reduce the radial error of RMS to about 2 cm. For the first time, seasonal period and other time changes of which ocean were determined as globally with high accuracy.

Jason-1 and Jason-2 were placed in orbit in continuance of the Topex/Poseidon mission and the mission of Topex/Poseidon and Jason-1, respectively. The altimetry measurement accuracy of these two satellites is about 2.5 cm. With the investigation in the data of Topex/Poseidon, Jason-1 and Jason-2, when these satellites were in a similar orbit, had equal coverage and by intermission bias on data of Jason-1, Jason -2 (in data files of these satellites), they were combined with the data of Topex/Poseidon and achieved about 22 years of data on similar passes. The satellite data in this study include Topex/Poseidon satellite data during 1992-2002, Jason-1 satellite during 2003-2008 and Jason-2 satellite during 2009-2014 in the Persian Gulf and the Gulf of Oman. In fact, this data is data on which the effect of the wet Troposphere error ( $\Delta w$ ), dry troposphere error ( $\Delta d$ ), ionosphere error ( $\Delta I$ ), polar tide error ( $\Delta pT$ ), inverse pressure effect bias ( $\Delta IB$ ), sea state bias, electromagnetic bias ( $\Delta E$ ) and error of centre of gravity changes of altimeter antenna ( $\Delta c$ ); is applied. All of these corrections apply to observations. These data are available at <ftp://avisoftp.cnes.fr/AVISO/pub> (Aviso and Podaac, 2008).

## 3. Methodology

Consider  $\underline{f}$  containing sampled SSH observations, it can be written as a functional model of (Vaniček, 1969, 1971; Amiri-Simkooei, 2007):

$$\underline{f} = A\underline{x} + A_j\underline{x}_j \quad (1)$$

where  $\underline{f}$  is vector of observations containing both periodic and non-periodic terms;  $A\underline{x}$  and  $A_j\underline{x}_j$  refer to non-periodic and periodic part of the model, respectively. In a similar manner, both LSSA and LS-HE methods try to extract the periodical part of the model in Equation (1) by using a series of sinusoidal base-functions (Vanicek, 1969). Both LSSA and LS-HE use a least square approximation to find the amplitudes associated to the base-functions.

In order to create a time-dependent observation vector, we create a vector ( $\underline{f} = \{f_1, \dots, f_n\}$ ) where we have an observation at any time ( $\underline{t} = \{t_1, \dots, t_n\}$ ). And finally, we come to a time series. Let  $M$  be vector space defined by an inner product ( $\sigma(\underline{a}_j^i, \underline{a}_j^i) = \langle \underline{a}_j^i | \underline{a}_j^i \rangle = \sum_{t_k \in M} \underline{a}_j^i(t_k) \underline{a}_j^i(t_k)$ ) and  $\underline{a}_j^i$  as its base-functions, i.e.  $\underline{a}_j^i$ ,  $i=1, \dots, m$  are column-vectors with the same dimension as  $\underline{f}$ . The aim of a spectral analysis (LSSA or LS-HE) is to look for unknown coefficients  $x_j^i$ ,  $i=1, \dots, m$ , which provide the best approximation of  $\underline{f}$  in the vector space  $M$ :

$$\underline{f} = \sum_{i=1}^m x_j^i \underline{a}_j^i = A_j \underline{x}_j \quad (2)$$

$\underline{x}_j$  is a vector containing the coefficients  $x_j^i$  and each base-function  $\underline{a}_j^i$  is  $i$ -th column of the matrix  $A_j$ . In a matrix form, the optimum  $x_j^i$  follows the minimization criterion of (Vaniček, 1969, 1971; Amiri-Simkooei, 2007):

$$\| \underline{f} - A_j \hat{x}_j \| = \| \hat{v} \| \rightarrow \min \quad (3)$$

Using the least squares adjustment (LSA) with the mean quadratic norm, coefficients  $x_j^i$  are therefore determined as:

$$\hat{x}_j = (A_j^T W A_j)^{-1} A_j^T W \underline{f} \quad (4)$$

Considering the periodic pattern of model in

Equation (1), one should a priory select appropriate base-functions to model the sampled observations. As mentioned in Vaníček (1969), LSSA thus uses general trigonometric polynomial of:

$$T_m(t) = \sum_{j=1}^{m-1} r_j \cos(\omega_j t - \theta_j) \quad (5)$$

where  $\omega_j$  is Frequency,  $r_j$  is Amplitude,  $\theta_j$  is Phase component of  $j$ .

Equation (5) consists of two main parts of a signal, the linear trend and the periodic portion. For representing the periodic term. Therefore, Equation (5) should be defined by considering the minimum quadratic distance from the observed vector  $f$ , while detecting the major periodic terms with frequencies  $\omega_j$ . Note that the frequencies are assumed to be known beforehand. Therefore, each general trigonometric polynomial  $T_j(t)$  expands as below:

$$T_m(t) = a_j \cos(\omega_j t) + b_j \sin(\omega_j t) \quad (6)$$

To achieve Equation (6), two equations of  $r_j = \sqrt{a_j^2 + b_j^2}$  and  $\theta_j = 2\arctan(\frac{a_j}{b_j+r_j})$ , are used.

For a specific frequency  $\omega_j$ , matrix  $A_j$  with the base-functions  $\underline{a}_j^i$ , can be written as:

$$A_j = \begin{bmatrix} \cos(\omega_j t_1) & \sin(\omega_j t_1) \\ \cos(\omega_j t_2) & \sin(\omega_j t_2) \\ \vdots & \vdots \\ \cos(\omega_j t_n) & \sin(\omega_j t_n) \end{bmatrix} \quad (7)$$

Each spectral value, corresponding to the frequency  $\omega_j$ , is defined as the ratio of the length of  $\underline{f}$  projected on the orthogonal base-functions  $\underline{a}_j^i$  and scaled by the length of observations as (Vaníček, 1969, 1971; Amiri-Simkooei, 2007):

$$s(\omega_j) = \frac{\underline{f}^T W A_j (A_j^T W A_j)^{-1} A_j^T W \underline{f}}{\underline{f}^T W \underline{f}} \quad (8)$$

Choosing  $W = I_{n \times n}$ ; least square spectrum simplifies as:

$$s(\omega_j) = \frac{\underline{f}^T A_j (A_j^T A_j)^{-1} A_j^T \underline{f}}{\underline{f}^T \underline{f}} \quad (9)$$

Equations (8) and (9) are related to the LSSA technique, while Equation (8) is the weighted

case and Equation (9) is a simple equally-weighted case.

The univariate harmonic estimation spectrum is defined as below (Vaníček, 1969, 1971; Amiri-Simkooei, 2007):

The univariate harmonic estimation power spectrum shown as below:

$$S(\omega_j) = \hat{e}_0^T W A_j (A_j^T W P_A^\perp A_j)^{-1} A_j^T W \hat{e}_0 \quad (10)$$

where

$$\hat{e}_0 = P_A^\perp \underline{f} \quad (11)$$

Represents least square residuals (Amiri-Simkooei, 2007)

$$P_A^\perp = I - A(A^T W A)^{-1} A^T W \quad (12)$$

Substituting Equations (11) and (12) into Equation (10), at first, the problem of non-inevitability is solved, and then the final formula is presented as follows:

$$S(\omega_j) = \left( (I - A(A^T W A)^{-1} A^T W) \underline{f} \right)^T W A_j (A_j^T W (I - A(A^T W A)^{-1} A^T W) A_j)^{-1} A_j^T W (I - A(A^T W A)^{-1} A^T W) \underline{f} \quad (13)$$

Is the orthogonal projector of the univariate model that projects along range space of  $A$ , i.e.  $R(A)$ , onto orthogonal component of range space of  $A$ , i.e.  $R(A)$ . For zero mean stationary random process containing only white noise, linear part of model vanishes ( $A = 0$ ) and the weight matrix of  $W = I_{n \times n}$ , the univariate LS-HE estimated spectrum simplifies as (Amiri-Simkooei and Asgari, 2012; Amiri-Simkooei, 2007):

$$S(\omega_j) = \underline{f}^T A_j (A_j^T A_j)^{-1} A_j^T \underline{f} \quad (14)$$

To obtain  $S(\omega_j)$  using several approaches are possible. One of these methods, using analytical relations for the matrix  $A_j$ , which is obtained for a maximum value of  $S(\omega_j)$ . However, this method due to the presence of large local amounts is very cumbersome and complicated. Therefore, numerical methods are used to solve the problem. For this purpose, a discrete relationship between  $\omega_j$  and  $S(\omega_j)$  is established and we can obtain the graph of the spectral values of  $S(\omega_j)$  in terms of the discrete values of  $\omega_j$ . In this way,  $\omega$  which has the maximum value of the spectral value  $S(\omega_j)$  it is chosen as the frequency of interest.

The difference between the LSSA-derived spectrum Equation (9) and the univariate LS\_HE derived spectrum Equation (14) is the scaling (normalization) factor of  $\underline{f}^T \underline{f}$ .

In case of existing several time series, the model is referred to as a multivariate linear model which uses all the time series simultaneously with  $\Sigma$  as the cross-correlation matrix of observations. Multivariate power spectrum obtained as (Vaníček, 1969, 1971; Amiri-Simkooei, 2007):

$$s(\omega_j) = \text{tr}(\hat{E}^T W^{-1} A_j (A_j^T W^{-1} P_A^\perp A_j)^{-1} A_j^T W^{-1} \hat{E} \Sigma^T) \quad (15)$$

where  $\hat{E}^T$  is least squares residual of observation matrix  $F$  with time series stored in its columns.

$$\hat{E} = P_A^\perp F \quad (16)$$

For zero-mean uncorrelated series choosing  $\Sigma = \text{diag}(\sigma_{11}, \sigma_{22}, \dots, \sigma_{rr})$ , spectrum simplifies as (Vaníček, 1969, 1971):

$$s(\omega_j) = \sum_{i=1}^r \hat{E}^T A_j (A_j^T A_j)^{-1} A_j^T \hat{E} / \sigma_{ii} \quad (17)$$

Spectrums in the Equations (8), (9), (14), (15) and (17) are functions of frequency  $\omega_j$ . These functions may include several local maximums. More significant frequencies will show up with greater spectral values and maximize the spectrum (Amiri-Simkooei and Asgari, 2012; Amiri-Simkooei, 2007; Vanicek, 1969, 1971). Analytical methods for extracting the local maximums might be complicated. Spectral values, therefore, can be computed for a set of primary frequencies to find the significant ones that appear with maximum spectral values in the spectrums. Equations (8) and (9) can respectively be used to compute spectrums for a weighted and equally-weighted LSSA; Equation (14) corresponds to the univariate LS-HE; and finally Equations (15) and (17) refer to the weighted and equally-weighted multivariate LS-HE approach, respectively.

### 3-1. Least square harmonic estimation for frequency extraction

Harmonic estimation method is used to introduce periodic patterns in functional model. For a time series, the simplest periodic behavior that can be added to

improve a functional model is to include:

$$y(t) = a_j \cos \omega_j t + b_j \sin \omega_j t \quad (18)$$

This relationship is in fact a sinusoidal wave with a primary phase,  $a_j$  and  $b_j$  signal amplitude and  $\omega_j$  frequency. Therefore, the functional model changes as follows:

$$E(\underline{y}) = A x + A_j x_j \quad (19)$$

After merging Equations 18 and 19, we will have Equation (20):

$$A_j = \begin{bmatrix} \cos \omega_j t_1 & \sin \omega_j t_1 \\ \cos \omega_j t_2 & \sin \omega_j t_2 \\ \vdots & \vdots \\ \cos \omega_j t_m & \sin \omega_j t_m \end{bmatrix}, \quad x_j = \begin{bmatrix} a_j \\ b_j \end{bmatrix} \quad (20)$$

Since in this model, in addition to  $a_j$  and  $b_j$ , the frequency  $\omega_j$  is also unknown; therefore, the least squares harmonic estimation method should be used to solve the problem. For this purpose, the assumptions of zero  $H_0$  and the opposite of  $H_a$  are defined as follows:

$$\begin{aligned} H_0: E(\underline{y}) &= A x \\ H_a: E(\underline{y}) &= A x + A_j x_j \end{aligned} \quad (21)$$

An important issue here is whether the theory or assumptions used in the model is correct or not. For this purpose, two assumptions are assumed to be zero and the assumption is opposite to. For example, an assumption is that the data are unmistakable. To examine whether these assumptions are valid or not, we oppose the zero hypothesis and consider the opposite hypothesis, so that the variables become clearer for us. In a model that is assumed to have error and error effects absent ( $H_0$ ),  $H_0$  test against  $H_a$  informs us that indicates whether additional variable should be added to the calculation or not. The main goal is to solve the problem of finding frequencies that can maximize the following. In fact, the goal is to find the frequencies that can be obtained for these frequencies by maximizing the amount of power spectrum. Then, based on this maximum power

spectrum, it is possible to analyze the frequency extraction of important tidal components.

The goal is to find the frequency  $\omega_j$  by solving the following minimization problem:

$$\underline{\omega}_j = \underset{\omega_j}{\operatorname{argmax}} P(\underline{\omega}_j) \quad (22)$$

$$\omega_j = \underset{\omega_j}{\operatorname{argmin}} \| P_{[\bar{A}_j]}^\perp y \|_{Q_y^{-1}}^2 = \underset{\omega_j}{\operatorname{argmin}} \| \hat{e}_a \|_{Q_y^{-1}}^2$$

$$\omega_j = \| P_{\bar{A}_j}^\perp y \|_{Q_y^{-1}}^2 \cdot \bar{A}_j = P_{\bar{A}_j}^\perp A_j$$

$$P_{\bar{A}_j} = \bar{A}_j (\bar{A}_j^T Q_y^{-1} \bar{A}_j)^{-1} \bar{A}_j^T Q_y^{-1}$$

$$\omega_j$$

$$= \underset{\omega_j}{\operatorname{argmax}} \hat{e}_0^T Q_y^{-1} A_j (A_j^T Q_y^{-1} P_{\bar{A}_j}^\perp A_j)^{-1} A_j^T Q_y^{-1} \hat{e}_0$$

$$= \underset{\omega_j}{\operatorname{argmax}} P(\underline{\omega}_j)$$

$$\hat{e}_0 = P_{\bar{A}_j}^\perp y$$

$$P(\underline{\omega}_j) = \hat{e}_0^T Q_y^{-1} A_j (A_j^T Q_y^{-1} P_{\bar{A}_j}^\perp A_j)^{-1} A_j^T Q_y^{-1} \hat{e}_0 \quad (23)$$

$$\hat{e}_0 = P_{\bar{A}_j}^\perp y$$

$$P_{\bar{A}_j}^\perp = I - A_j (A_j^T Q_y^{-1} A_j)^{-1} A_j^T Q_y^{-1} \quad (24)$$

In the above relationship,  $\bar{A} = [A_1 \dots A_{k-1}] \cdot \|\cdot\|_{Q_y^{-1}}^2 = (\cdot)^T Q_y^{-1} (\cdot)$  and  $\hat{e}_a$

the remainder of the least squares is subject to the opposite.  $A_j$  is obtained using Equation (20). The first two columns of this matrix include a linear behavior of time series and the following columns for finding periodic behavior in time series is used.  $\hat{e}_0$

The least squares residue vector and  $P_{\bar{A}_j}^\perp$  is the orthogonal projector. These parameters are obtained under the assumption  $H_0$ .

In order to find the matrix  $A_j$  in (23), for different  $\underline{\omega}_j$  bands, different matrices are tested to find the maximum value of  $P(\underline{\omega}_j)$ . The matrixes  $A_j$  have the same matrix structure as in (20). The matrix  $A_j$ , the maximum value of  $P(\underline{\omega}_j)$ , obtained by it, and is chosen as the matrix  $A_j$ .

To obtain  $P(\omega_j)$  using several approaches are possible. One of these methods, using analytical relations for the matrix  $A_j$ , which is obtained for a maximum value of  $P(\omega_j)$ . However, this method due to the presence of large local amounts is very cumbersome and complicated. Therefore, numerical methods are used to solve the problem. For this

purpose, a discrete relationship between  $\omega_j$  and  $P(\omega_j)$  is established, and we can obtain the graph of the spectral values of  $P(\omega_j)$  in terms of the discrete values of  $\omega_j$ . In this way,  $\omega$  which has the maximum value of the spectral value  $P(\omega_j)$  is chosen as the frequency of interest.

The choice of different  $\underline{\omega}_j$  is done by Equation (25):

$$T_{j+1} = T_j (1 + \alpha T_j / T), \quad \alpha = 0.1, \quad j = 1, 2, \dots$$

$$\underline{\omega}_j = 2\pi / T_j \quad (25)$$

where  $T_1$  is the Nyquist period and  $T$  is the total length of the time series. The value of  $T_1$  in Equation (25) is equivalent to 1 hour. Considering the value of 0.1 for alpha is because, first, this coefficient must be chosen so that the algorithm can extract all the different frequencies. Second, this factor should be such that the volume of calculations does not last long. Therefore, for the above two goals, the appropriate value for this coefficient is 0.1.

After finding  $\omega_j$  and also the maximum value of  $P(\omega_j)$ , it is necessary to test the selected frequencies with statistical assumptions. The statistics used for this test are as follows:

$$\underline{T}_2 = \hat{e}_0^T Q_y^{-1} A_k (A_k^T Q_y^{-1} P_{\bar{A}_k}^\perp A_k)^{-1} A_k^T Q_y^{-1} \hat{e}_0 \quad (26)$$

The statistics  $T_2$  under the assumption  $H_0$ , has a Chi Squared distribution function with two degrees of freedom.

#### 4. Simulation

The JASON-1 and missions are T/P's follow-on that overfly almost the same reference T/P ground tracks from 2002 to 2008 (continued on the Topex satellite mission), respectively (Picot et al., 2003). The two satellites fly 254 ground tracks (passes) in 9.9156-day repeat cycle to measure topography changes of the sea surface (measure the height of the sea surface). The time series of observations are constructed according to satellite measurements when crossing a point (Sharifi et al., 2013). In fact, there are missing values, high amplitude noise, outliers, and blunders in the SSH time series caused by altimeter failures due to, e.g. a

raining weather, ice and land coverage etc. The mentioned conditions are considered within the simulation stage, when LSSA and LS-HE are applied to analyse the SSH time series.

In order to simulate the time series, we used a periodic signal with a constant trend which was simulated using Equation (27):

$$SSH(t) = y_0 + rt + \sum_{j=1}^K [a_j \cos(\omega_j t) + b_j \sin(\omega_j t)] \quad (27)$$

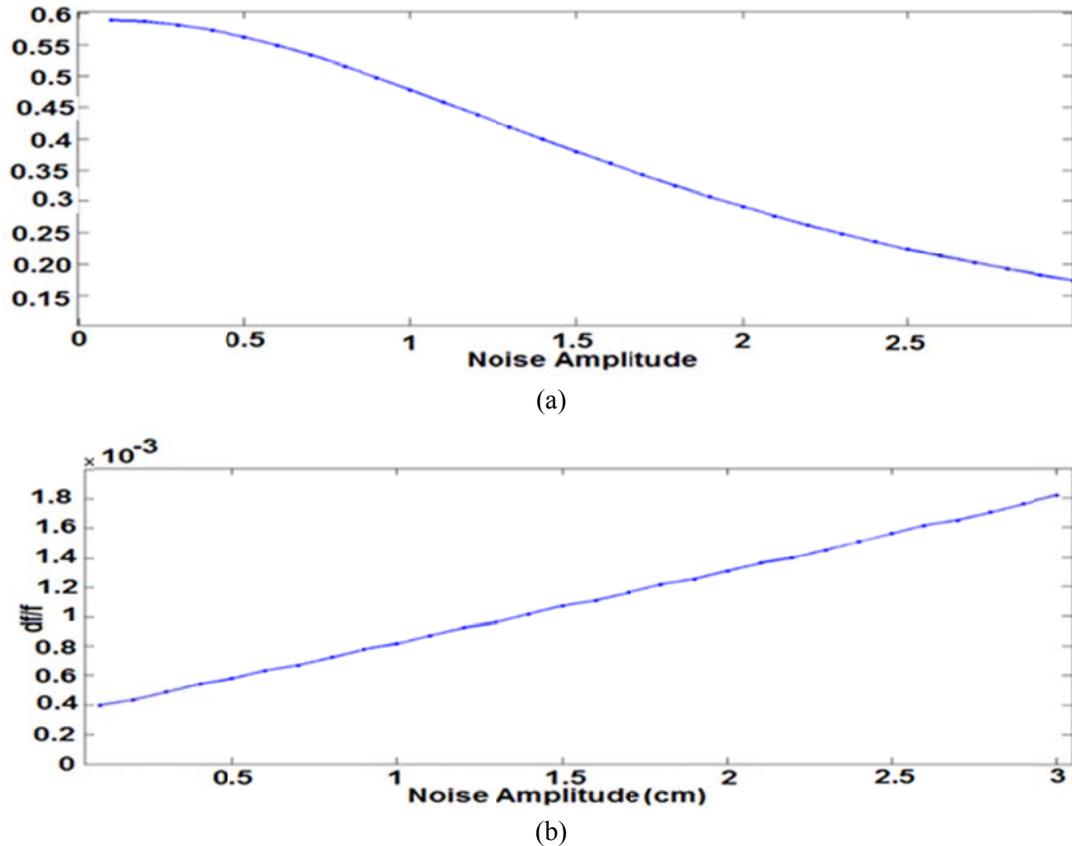
where  $SSH(t)$  is the water altitude at time moment of  $t$ ,  $y_0$  is the mean time series (water level),  $j=1,2,3,\dots,k$  of the components number,  $K$  is determined based on the number of applied components. and finally,  $r$  is the rate of change.

Here, we assumed that the simulated time series contain a superposition of seven different sinusoidal signals with the frequencies of  $62.07^{-1}$ ;  $173.3^{-1}$ ;  $58.71^{-1}$ ;  $45.68^{-1}$ ;  $88.86^{-1}$ ;  $364.2^{-1}$  and  $117.5^{-1}$  cycle/day with different amplitudes and phases. Sampling rate equals to that of TOPEX/POSIDON (T/P) and Jason-1 missions, i.e. 9.9156 days. The simulation was done by MATLAB software using Equation (27). Considering the seven main tidal components that play the most role in producing a periodic signal. White noise is also used to consider and add noise to the time series.

#### 4-1. Effect of noise

Similar to the most of spectral analysis methods, LSSA and LS-HE are also sensitive to noise. Therefore, when the time series are noisy, some undesirable effects would appear (consider the effect of noise on the desired time series), for instance, the peak behavior of the estimated spectrum.

To investigate a vast possibility of noisy behavior, temporally random noise with a standard normal distribution were added to the simulated time series with 16-year length. The simulated noise amplitudes differ from 1 to 30 mm. The results show that  $\frac{\Delta\omega}{\omega}$  ratio, in which  $\Delta\omega$  is the difference of the estimated frequency and its actual simulated value and  $\omega$  is the simulated frequencies, corresponding to the annual frequency became more than 5 times greater when the noise amplitude increased from 1 to 30 mm. This is also combined with decreasing the amplitude of the significant peaks in the spectrum. For instance, the effect of noise on the accuracy of the extracted annual cycle and its spectrum is illustrated in Figure (1). Figure (1-a) indicates that when the amplitude of the noise is greater than 14 mm, a bias of 9 hours appears in the estimation of the annual cycle. Figure (1-b) shows that the power of the simulated significant frequency, decreased when the noise presented. For those simulations containing noise amplitude greater than 15 mm, the power of the annual frequency reduced to 0.3, which might even fail a statistical significant test. The graph in Figure (1-b) indicates that the relationship between the noise amplitude and the estimated power of the peak is not linear. Note that the significance of the peaks, estimated by LSSA and LS-HE, can be tested statistically (Sharifi et al., 2013; Farzaneh and Parvazi, 2018). When a high magnitude noise is present in the time series, subsequently, the probability of the rejection of statistical test increases. Discussing a reliable testing method or a proper significant threshold is behind the scope of the current study.



**Figure 1.** Impact of noise amplitude on estimation of the annual frequency. (a) shows the difference of an estimated annual frequency and its simulated value against the amplitude of noise. When the noise amplitude increases, the difference is also increased. (b) shows the power of the estimated annual cycle against the amplitude of noise. The graph indicates that the relation is non-linear.

It should be mentioned here that for time series with a high level of noise, e.g. containing a random noise with the standard deviation close to that of data, the performance of the LSSA and LS-HE is worsened significantly. Figure (2-a) shows a time series of SSH changes contaminated with a random noise of 50 mm amplitude. We applied LSSA on the data in Figure (2-a), when windows with the length of 3 years to 16 years were selected from the data. Figure (2-b) shows the accuracy of the estimated cycles with respect to the length of time series. The results show the reliable accuracy of the estimations derived from the time series that are longer than 14 years.

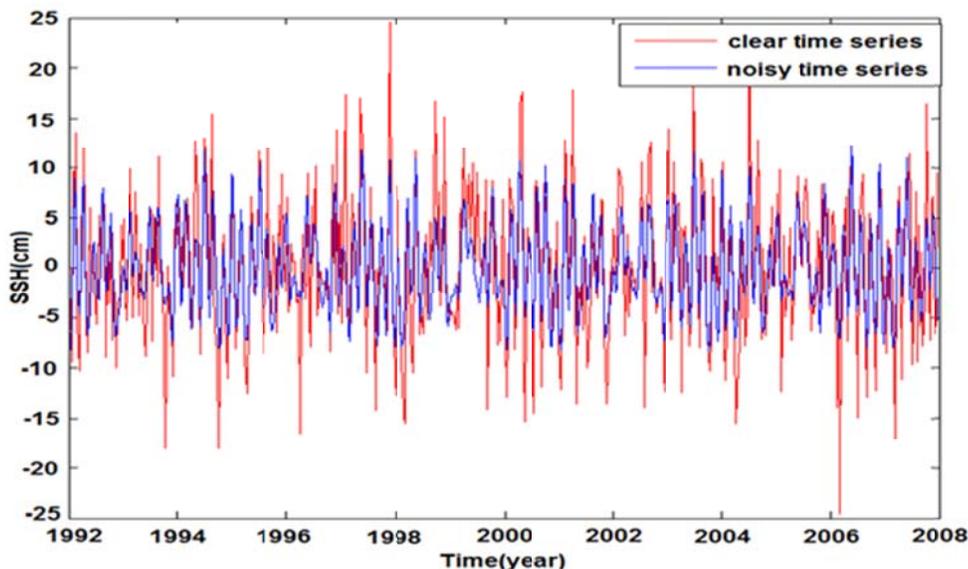
#### 4-2. Effect of blunders

Blunders are usually referred to the observations that are inconsistent with the other sampled values. Blunders, therefore,

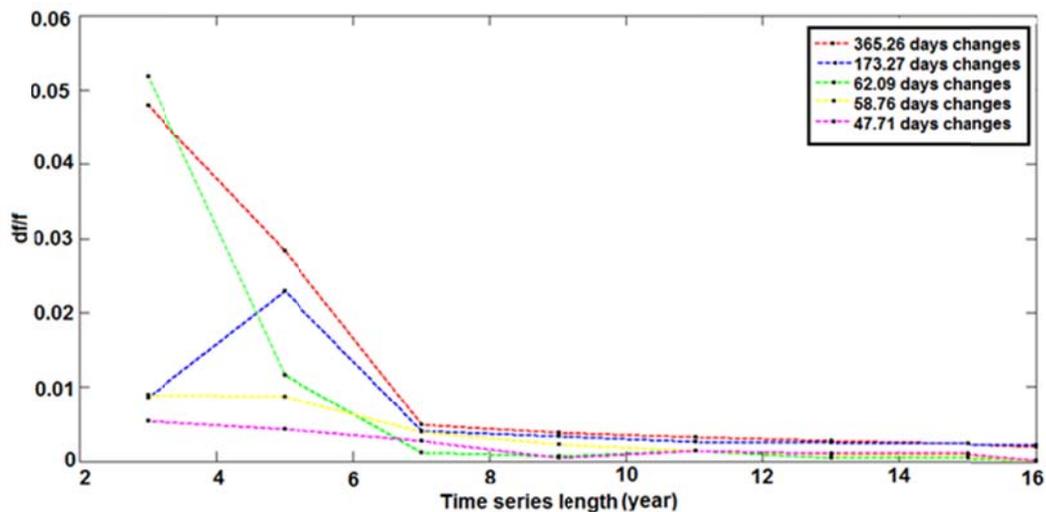
should be removed and/or, if it is necessary, be replaced by corrected values (Kern et al., 2005). However, the detection and replacement of these values might be ambiguous. Although, different methods have been developed to detect blunders, e.g. thresholding, Mahalanobis distance (Mahalanobis, 1936), Grubbs' test, Dixon test, Wavelet outlier detection algorithm etc. (see in details in Kern et al., 2005), there is no guarantee to find and replace all of them in a correct way. Therefore, some of them may remain in the final time series and as a result would affect the final results of spectrum estimation. To show how blunders affect the result of LSSA and LS-HE, we simulated a time series with blunders. For this purpose, we replace the value of a time series with an out-of-range number in some observation epochs. Number of blunders and their positions and values were selected randomly. Note that, very large outliers can

be detected before applying LSSA and LS-HE, e.g. simple thresholding method with  $3\sigma$  test. Figure (3-a) shows a simulated time series with and without blunders. Both time series of Figure (3-a) were sampled with the

length of 15 years and contain frequencies mentioned in section (3-1). Figure (3-b) indicates that all the introduced frequencies are extracted with an accuracy better than 0.004 cycle/day.

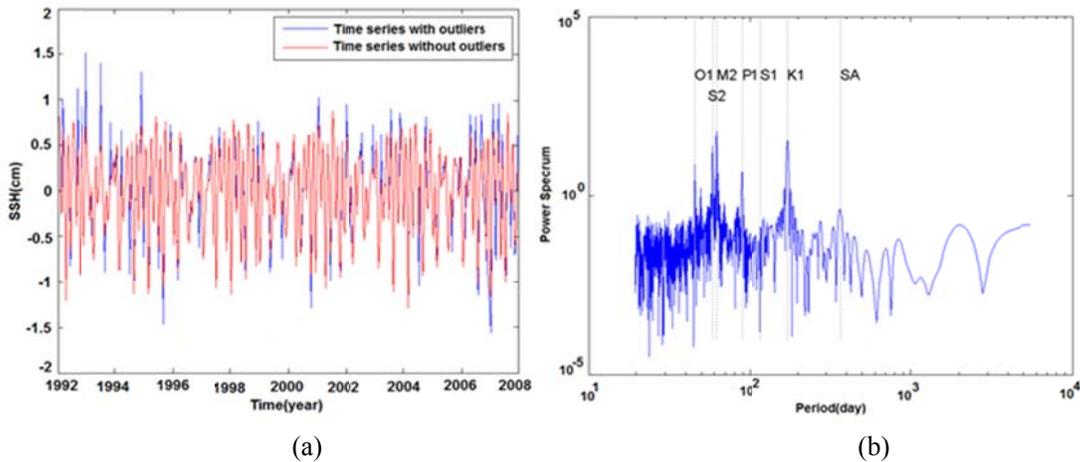


(a)



(b)

**Figure 2.** a) A simulated time series with a random noise of 50 mm amplitude. The blue bold line refers to the simulated series, containing seven frequencies mentioned in section (3-1) without noise. The red line refers to the superposition of the simulated data and a random noise with amplitude of 50 mm, b) Accuracy of Annual and Semi-annual terms, shown with respect to the length of time series. Generally, when the length of time series increases, more than 14 years in this case, the accuracy of estimation improves.



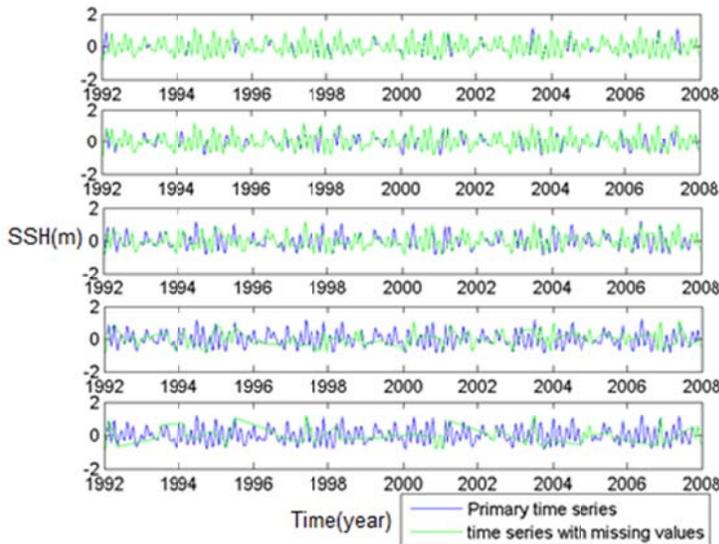
**Figure 3.** a) A simulated time series containing blunders; the red-line represents the simulation without blunders and the blue-lines are derived after adding blunders. There are 62 values replaced in primary time series that is equivalent to 11.48% of all time series values. b) Power of spectrum of the simulated time series in Figure (3-a). Four frequencies are separated in spectrum. All these frequencies are estimated with an accuracy better than 0.004 cycle/day.

**4-3. Effect of missing value**

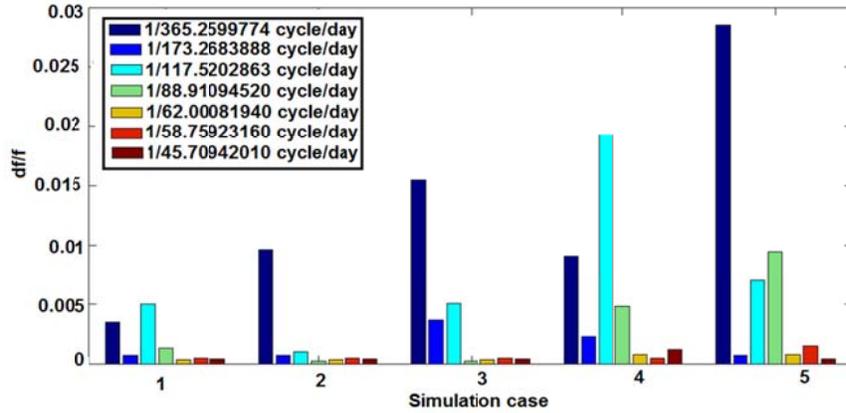
As mentioned before, one advantage of LSSA and LS-HE is that time series with unequally spaced data can be analyzed. To assess this property, we simulated time series in five cases with different number of missing value. Number of missing values increase until time series become sparse. In this case, sparse time series differs to one that missed more than 50% of possible observations. The positions of the missing values were selected randomly. Therefore, in the time series that are sampled with the rate of 9.9156 days for 16 years, there should exist 588 epochs of observations. For our

investigations, we assumed unequally spaced time series containing 529, 471, 353, 177 and 89 SSH values, which correspond to 10%, 20%, 40%, 70%, and 85% of missing samples, respectively. The mentioned five time series are shown in Figure 4.

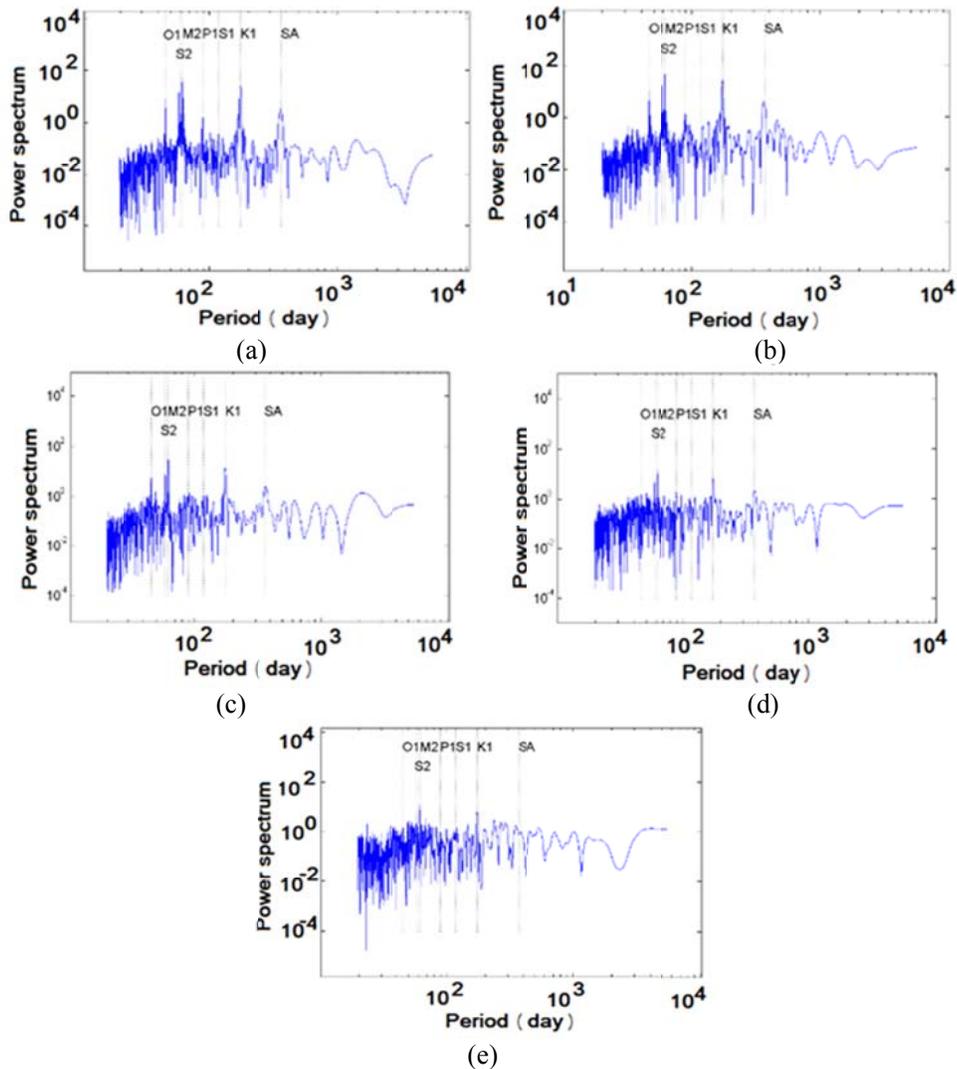
Computing the spectrum using LSSA to the sparse time series, those moderate ( $365.259^{-1}$ ;  $88.910^{-1}$ ; and  $117.520^{-1}$ ) became considerably lower for the time series with 177 and 89 samples. The  $O_1$ ,  $K_1$ ,  $M_2$  and  $S_2$  frequencies were, however, extracted with sufficient accuracy in all cases. Figure 5 shows the  $\frac{\Delta\omega}{\omega}$  ratio of frequencies in each case.



**Figure 4.** simulated time series with missing values. Graphs that are shown from top to the bottom of the figure represent the time series that contain 529, 471, 353, 177 and 89 samples, respectively.



**Figure 5.** An overview of the LSSA results when it is applied to five simulated time series with missing values, shown in Figure 6. The cases 1, 2, 3, 4 and 5 correspond to 10%, 20%, 40%, 70%, and 85% missing values. For those time series that are not sparse (case 1, 2, and 3), the desired frequencies were estimated accurately. For those with considerable number of missing values (case 4 and 5,) the estimated frequencies were less accurate.



**Figure 6.** a) Power spectrum of a time series with 529 observations (corresponding to case 1), b) time series with 471 observations (corresponding to case 2), c) Power spectrum of a time series with 353 observations (corresponding to case 3), d) Power spectrum of a time series with 177 observations (corresponding to case 4), e) Power spectrum of a time series with 89 observations (corresponding to case 5).

The corresponding power spectrums of all the cases in Figure 4 are displayed in Figure (6-a) to (6-e). In case 1, 2 and 3 spectrums, one can easily find six dominant frequencies, and therefore, the missing values do not cause a considerable impact on the estimation. In cases 4 and 5, only four dominant frequencies were detectable.

Figure 6 shows an overview of the estimated spectrums, corresponding to the time series of Figure 4. In each graph, the x-axis refers to the period of change in (days), and the y-axis refers to the spectral value.

Next simulation shows the effect of gap in the time series. We assumed five years of gap (i.e. 25% of the data missed) in the time series of case 1 to 5, shown in Figure 4.

The  $\frac{\Delta\omega}{\omega}$  ratio of frequencies is recomputed for the new time series with a prolonged gap. The results are summarized in Figure 7.

Comparing Figure 7 with Figure 5, it can be seen that a single gap, which is 5 years in our case, makes slight changes in the accuracy of the simulation in case 1, 2, and 3. In case 4 and case 5, moderated length changes are estimated less accurate in comparison with the results in Figure (5). The estimated spectrum of the time series with gaps were similar to those of Figure 6, therefore, they are not shown here. When the missing values occur with a periodic occurrence, the results of spectral analysis might completely change. To cover this aspect, we simulated a time series that exhibits 70 days of a year without any observation. Therefore, the mentioned gaps were considered for the time series of Figure 4. The results of spectral analysis indicated that the long periods can be extracted almost with the same accuracy of Figure 6. In the case 5 (with 85% missing values), however, the results of the LSSA became significantly biased. Moreover, there were several artificial frequencies emerging in the spectrum with high spectral values, making detection of the real periodic

component more difficult. For brevity, the results are not shown here.

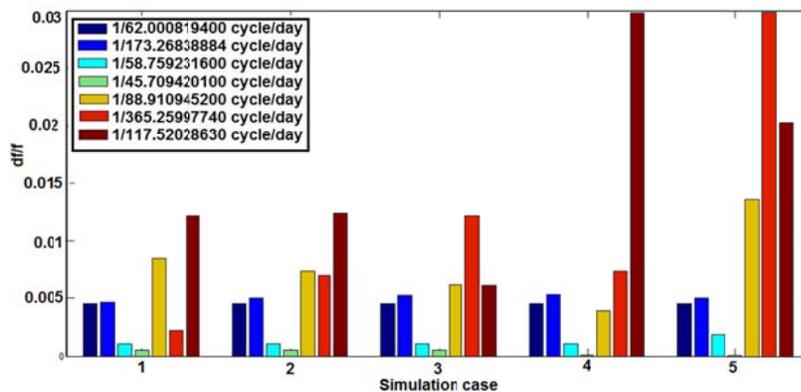
#### 4-4. Detection of nearby frequencies

Detection of the nearby frequencies is important for analyzing SSH time series since there are other different physical movements of water level at sea, which might be represented by close spectral structure, e.g. the annual tidal frequency and annual fluctuations caused by river flows. Our simulation results show that detecting the nearby frequencies depends strongly on both the frequency of interest as well as the length of time series. For instance, with a 16-year SSH time series, the two low frequency fluctuations of 0.00273786 and 0.00281690 cycle/day were not separable, however, one can separate them when the length of observation was 30 years. Detected numerical values of the frequencies were 0.002717 and 0.002838, respectively. For the nearby semi-annual and monthly frequencies, the situation was different. Using time series with the length of 16 years, the nearby semi-annual frequencies of  $182.6^{-1}$  and  $188^{-1}$  cycle/day, as well as those of monthly " $32^{-1}$ " and  $31.82^{-1}$ " cycle/day were successfully separated in the power spectrum. Estimated frequencies corresponding to the nearby semi-annual frequencies were  $181^{-1}$  and  $188.5^{-1}$  cycle/day, and corresponding to the nearby monthly frequencies, they were " $32^{-1}$ " and  $31.81^{-1}$ " cycle/day.

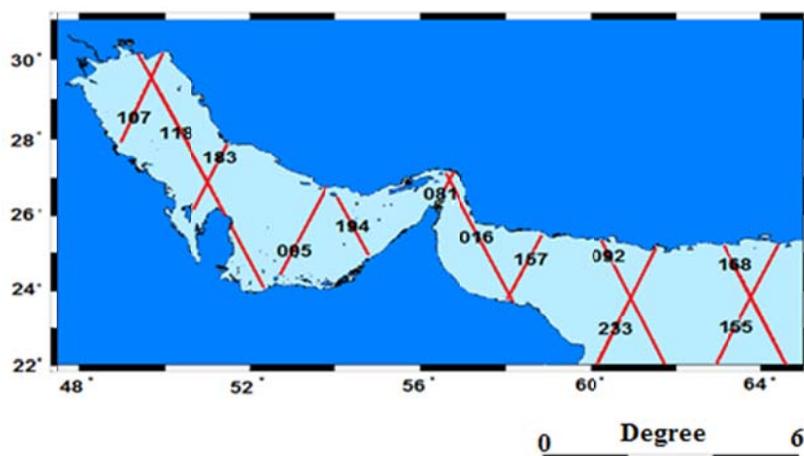
### 5. Implementation of LSSA and LS-HE on the observed SSH time series over the Persian Gulf and Oman Sea

#### 5-1. Building the SSH Time series

Considering the temporal coverage of the two satellite altimetry missions Topex/poseidon and Jason-1, SSH time series with the length of ~16 years (1992-2008) can be built. From 254 satellite passes, twelve ground tracks over the Persian Gulf and Oman Sea were selected in this study (see Figure 9).



**Figure 7.** an overview of the LSSA results when it is applied to five simulated time series with missing values and gap. The time series of case 1 to case 5 are similar to those of Figure 7; however, they also contain a prolonged gap of 5 years.



**Figure 8.** Topex/poseidon and Jason-1 ground tracks over the Persian Gulf and Oman Sea. 626 time series with reliable SSH measurements are built.

SSH time series are computed similar to that of (Sharifi et al., 2013), with also considering Jason-1 observations. The bias between different missions was removed by adjusting them as a simple shift function. Therefore, an ideal SSH time series, in our case, should contain more than 576 observations covering 1992 to 2008. However, due to the noise, outliers, blunders and missing values, shorter time series were derived. In this implementation, 626-time series with the sample points of more than 530 were built over the Persian Gulf and Oman Sea.

### 5-2. Spectral Analysis based on LSSA

From the available time series, four time series are selected to be presented in this paper (More than 600-time series were generated using observations of satellites altimetry in the Persian Gulf and Oman Sea. For example, four time series are selected at

several points in the Persian Gulf and the Oman Sea and are presented in the text of the article). The positions of the selected time series are shown in Figure 9 and summarized in Table 1. Figure 11 shows the selected time series ordered with respect to the positions of Table 1. In this section, we only focus on the LSSA-extracted tidal frequencies and their amplitude from the Persian Gulf and Oman Sea. This aim is motivated by the fact that there are fewer attempts that have been undertaken to estimate tidal amplitudes from satellite observations over the Persian Gulf and Oman Sea. The results corresponding to the tidal frequencies are reported in Tables 2 to 4.

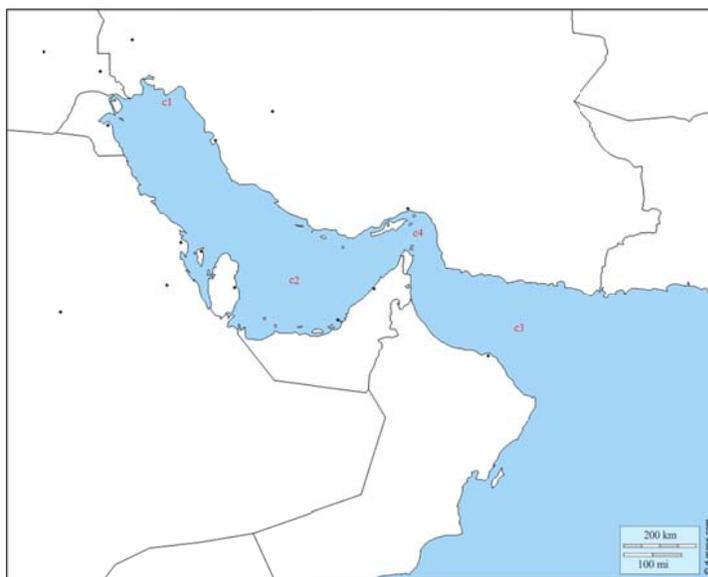
Table 2 summarizes the theoretical values of the main celestial tidal frequencies derived from astronomical studies (Cartwright, 1993). The aliased values of the frequencies in Table 2 are computed following the

empirical formula in (Sharifi et al., 2013) while considering the sampling rate of 9.915625 days. The main tidal frequencies and their amplitudes over the Persian Gulf and Oman Sea, computed from the time series of Figure 9, are summarized in Tables 3 and 4, reports the amplitude of tidal frequencies corresponding to the selected time series of Figure 11. Among 22 detected tidal components, mean amplitude of the annual, semiannual solar tides, along with  $M_2$ ,  $S_2$  (solar and lunar semi-diurnal

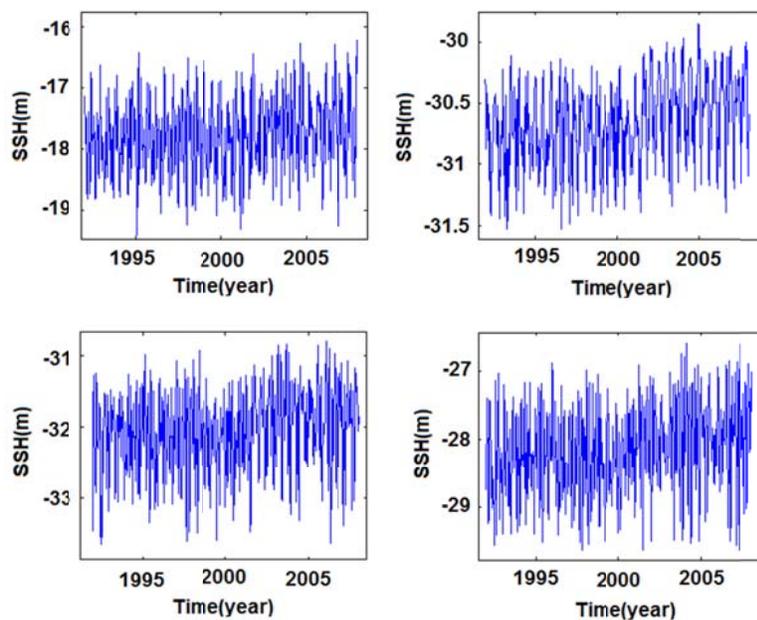
components) and  $k_1$  (luni-solar diurnal component) were greater than 1 cm, thus they should be considered as the main tidal frequencies of the Persian Gulf and Oman Sea.

**Table 1.** Positions of four selected time series in Figure 9, subjected to spectral analysis.

Station. no	Point-1	Point-2	Point-3	Point-4
Latitude	29.808	25.502	24.515	26.645
Longitude	49.651	53.203	58.416	56.580



**Figure 9.** Locations of the four selected sample time series, selected for spectral analysis.



**Figure 10.** Time series of the four selected sample points. Position of the time series are accordingly shown in Figure 8.

**Table 2.** Main tidal components and their aliased frequencies. The first column indicates the sign of tidal frequencies, the third column lists the real value of the tidal frequencies derived from astronomical studies (c.f. [5]), the fourth column shows the corresponding real period, fifth column lists the computed frequencies with respect to 9.915625 days sampling rate, and the sixth column lists their corresponding aliased periods.

Sign	Description	Frequency (cycle/day)	Period (day)	Aliased Frequency (cycle/day)	Aliased Period (day)
SA	Annual Solar Tide	0.0027378	365.2599774	0.0027378	365.2599774
SSA	Semi-annual Solar Tide	0.0054758	182.6211838	0.0054758	182.6211838
MSM	Solar monthly	0.0314347	31.8119340	0.0314347	31.8119340
MM	Lunar Monthly	0.0362916	27.5545492	0.0362916	27.5545492
MSF	Luni-solar Synodic Fortnightly	0.0677264	14.7652927	0.0331245	30.1891006
MF	Lunar Fortnightly	0.0732022	13.6607901	0.0276487	36.1680374
Q1	Larger Lunar Elliptic Diurnal	0.8932441	1.1195148	0.0144143	69.3755724
O1	Principal Lunar Diurnal	0.9295357	1.0758059	0.0218773	45.7094201
P1	Principal Sonar Diurnal	0.9972621	1.0027454	0.0112472	88.9109452
S1	Solar Diurnal	1.0000001	0.9999999	0.0085092	117.5202863
K1	Luni-Solar Diurnal	1.0027379	0.9972696	0.0057714	173.2683888
O2	Side, half-Daily Sun	1.8590714	0.5379030	0.0437547	22.8546975
N2	Larger Lunar Elliptic Semi-diurnal	1.8959820	0.5274312	0.0201857	49.5400487
M2	Principal Lunar semi-diurnal	1.9322736	0.5175250	0.0161060	62.0888194
S2	Principal Solar Semi-diurnal	2.0000000	0.5000000	0.0170186	58.7592316
K2	Luni-Solar Semi-diurnal	2.0054758	0.4986348	0.0115428	86.6341944
M3	Lunar ter-diurnal	2.8984104	0.3450167	0.0262665	38.0712527
M4	Shallow water over tides of principal lunar	3.8645472	0.2587625	0.0322119	31.0444328
S4	Shallow water over tides of principal solar	4.0000000	0.2500000	0.0340372	29.3796365
M6	Shallow water over tides of principal lunar	5.7968208	0.1725084	0.0483179	20.6962834
S6	Shallow water over tides of principal solar	6.0000000	0.1666667	0.0497951	20.0822785
M8	Shallow water eighth diurnal	7.7290945	0.1293813	0.0364271	27.4520769

**Table 3.** Extracted tidal frequencies from sample points of Figure 10. Frequencies are reported in (cycle/day).

Sign	Point-1	Point-2	Point-3	Point-4	Sign	Point-1	Point-2	Point-3	Point-4
SA	0.0027459	0.0027289	0.0027968	0.0027798	O2	0.0436363	0.0436363	0.0436533	0.0436363
SSA	0.0055495	0.0055325	0.0055325	0.0055155	N2	0.0201347	0.0201518	0.0201688	0.0201859
MSM	0.0312703	0.0313385	0.0314067	0.0314920	M2	0.0161113	0.0160943	0.0161113	0.0161113
MM	0.0363017	0.0360971	0.0363590	0.0363700	S2	0.0170318	0.0169636	0.0170489	0.0170318
MSF	0.0332487	0.0332487	0.0330099	0.0330781	K2	0.0115948	0.0114756	0.0115437	0.0115608
MF	0.0277059	0.0275012	0.0275865	0.0276377	M3	0.0261540	0.0258641	0.0261199	0.0263245
Q1	0.0144068	0.0144238	0.0143727	0.0143897	M4	0.0324130	0.0320719	0.0322083	0.0322083
O1	0.0218909	0.0218909	0.0218909	0.0218909	S4	0.0340503	0.0340674	0.0340674	0.0340844
P1	0.0112541	0.0112370	0.0112541	0.0112370	M6	0.0482761	0.0483443	0.0482931	0.0483443
S1	0.0085115	0.0085625	0.0086136	0.0085966	S6	0.0497261	0.0497772	0.0498796	0.0497431
K1	0.0057707	0.0057707	0.0057707	0.0057707	M8	0.0359342	0.0364723	0.0363529	0.0362368

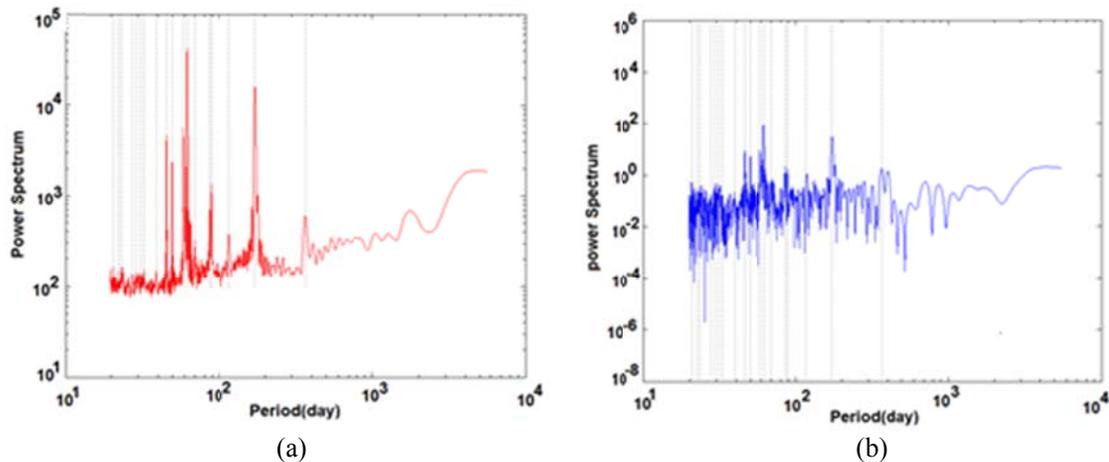
### 5-3. Spectral Analysis based on LS-HE

Before discussing the results of LS-HE, we should note that, since the LSSA and univariate LS-HE are equivalent. Our investigation of multivariate LS-HE in this section considers all the 626 virtual time series of SSH (Meaning Pseudo-Tide Gauge: This means that the Altimeter time series is a pseudo-tide. With the difference, we have observations similar to the Tide-Gauge observations in offshore.). Existing of those low frequency variations might cause a leakage to the LS-HE extracted frequencies. The magnitude of the leakage is however negligible for tidal frequencies as well as annual, semi-annual and seasonal frequencies (the effect of leakage phenomenon on the annual and semi-annual period is negligible. However, this phenomenon has a significant effect on small periods). Before implementing the multivariate LS-HE, the time series were temporally centered, i.e. temporal mean value was reduced from time series. The centered time series were then scaled by the standard deviation of time series to be standardized and unit-less. Under these circumstances, multivariate LS-HE spectrum would simplify as Equation 16, which is a linear combination of the univariate LS-HE spectrums. The result of the multivariate LS-HE and univariate LS-HE is displayed in Figures (11-a) and (11-b). For a better comparison between the two signals, a magnification signal daily-half, daily, annual and annual-half is displayed in Figure 12.

Comparing frequency extraction using two univariate and multivariate methods can only use one time series to extract the frequency. However, sometimes the number of frequencies may be too high or the frequencies may be small enough that only a series of times can be detected. Thus, by combining several series when the corresponding frequency occurs in the same epochs, one can increase the power of frequency detection. Therefore, the main purpose of this research is frequency extraction using univariate and multivariate analysis.

In the LSSA and the univariate LS-HE, results refer to the single observation point. In fact, all of time series demonstrate SSH changes, which have originated from common sources. The fundamental assumption in the multivariate LS-HE method is that in several time series, the same frequencies present with different amplitudes and phases. Therefore, the results can be generalized to the whole Persian Gulf and Oman Sea. Results are also statistically tested in the multivariate case. For implementing the statistical tests for both univariate and multivariate LS-HE, see Amiri-Simkooei and Asgari (2012), Amiri-Simkooei (2007), Amiri-Simkooei et al. (2017).

The main tidal frequencies, based on the 626 Time Series in the Persian Gulf and Oman Sea are presented the derived spectrum indicated 60 significant frequencies, which are summarized in Table 5.



**Figure 11.** a) Spectrum of frequencies of multivariate LS-HE applied on SSH time series and tested statistically. b) Spectrum of frequencies of univariate LS-HE applied on SSH time series and tested statistically.

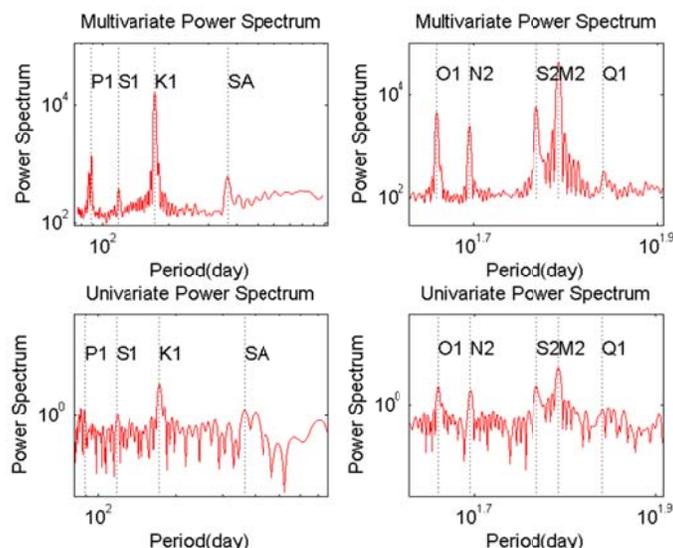


Figure 12. magnification signal of spectrum of frequencies of LS-HE applied on SSH time series and tested statistically.

Table 4. Extracted tidal frequencies of multivariate LS-HE applied on SSH time series and tested statistically. Frequencies are reported in (cycle/day).

Sign	Aliased Frequency (cycle/day)						
SA	0.0027459	O2	0.0443698	Q1	0.0143897	M4	0.0322424
SSA	0.0055325	N2	0.0201688	O1	0.0218909	S4	0.0340844
MSM	0.0313555	M2	0.0161113	P1	0.0112541	M6	0.0484296
MM	0.0362506	S2	0.0170318	S1	0.0085115	S6	0.0497261
MSF	0.0332316	K2	0.0115608	K1	0.0057707	M8	0.0363032
MF	0.0274160	M3	0.0264268				

Table 5. 60 significant frequencies extracted of multivariate LS-HE for 626 time series in Persian Gulf and Oman Sea.

period of changes [day]	Frequency of changes [cycle/day]	period of changes [day]	Frequency of changes [cycle/day]	period of changes [day]	Frequency of changes [cycle/day]
1750.8065641	0.0005712	88.8568160	0.0112540	39.1019403	0.0255741
1269.6996009	0.0007876	86.4994894	0.0115608	36.4978205	0.0273988
1047.1476005	0.0009550	77.1718132	0.0129581	35.3748777	0.0282686
677.2367535	0.0014766	74.2425116	0.0134693	32.7289862	0.0305539
601.5809809	0.0016623	69.4940375	0.0143897	31.0150451	0.0322425
541.0684803	0.0018482	65.1710508	0.0153442	29.3536008	0.0340669
491.5759325	0.0020343	64.3842958	0.0155318	27.5857826	0.0362503
446.9415718	0.0022374	63.6853601	0.0157023	26.2400202	0.0381098
409.7146016	0.0024407	63.0014284	0.0158728	25.8466994	0.0386892
364.1820676	0.0027459	62.0682245	0.0161114	23.8005614	0.0420150
262.0998441	0.0038153	61.1622492	0.0163500	23.4483384	0.0426476
239.6834004	0.0041722	60.5311415	0.0165205	22.5378643	0.0443695
192.5925023	0.0051923	58.7135803	0.0170317	21.6154456	0.0462642
186.4834825	0.0053624	51.9991749	0.0192311	21.4415156	0.0466374
180.7495450	0.0055325	50.1753457	0.0199302	21.2626469	0.0470301
173.2891677	0.0057707	49.5815178	0.0201686	20.9885916	0.0476449
166.4194269	0.0060089	48.9606698	0.0204244	20.8097787	0.0480538
117.4887903	0.0085114	46.5882494	0.0214648	20.6557978	0.0484121
112.7509923	0.0088691	45.6810504	0.0218909	20.3333910	0.0491811
96.9274052	0.0103170	45.1882794	0.0221298	20.1101790	0.0497265

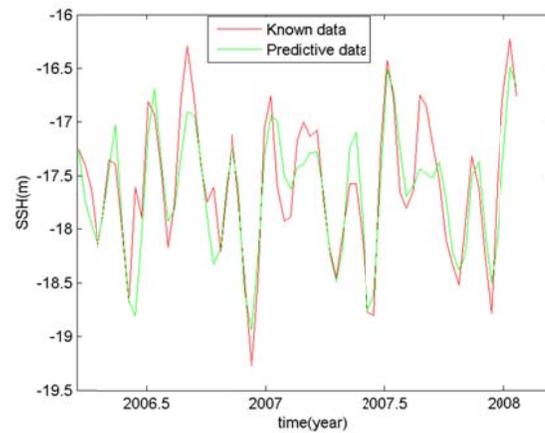
#### 5-4. Predictions based on three samples extracted frequency

In the following, to check the effectiveness of three samples extracted frequency (Astronomical-LSSA-Multivariate LSHE), the tide is predicted for a period of time. In this research, two methods of LSSA and LSHE have been used to extract the frequencies in the time series of satellite altimetry observations. After extracting the frequencies through the above two methods, a list of important frequencies is considered based on each method. On the other hand, principal frequencies are also provided through astronomical methods. Therefore, based on the output frequencies of the two above-mentioned methods and the frequencies of the astronomical method, there are three frequency lists available.

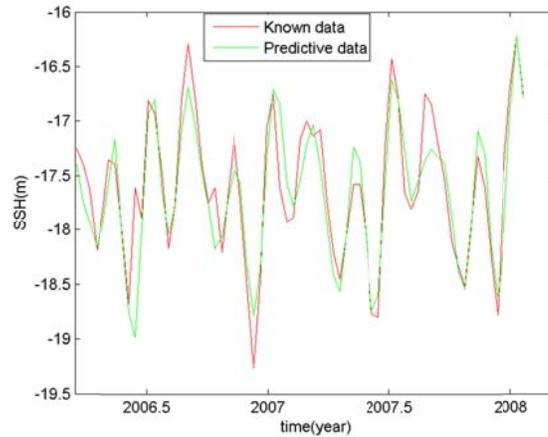
The total time interval of observation of sea surface in this research is from 1992 to 2008.

In order to predict the sea surface for the period 2006 to 2008, the prediction for sea surface observations in the two-year period, using three frequency list mentioned above was conducted. Accordingly, the height of the sea surface is calculated and compared with actual data (Amiri-Simkooei et al., 2017).

The results of predicted tides based on three samples of extracted frequency are presented in Figure 13. Two frequency samples listed in Figure 13. Two frequency samples listed were extracted from both methods LSSA and multivariate LSHE. Astronomical Frequency List is also used. Based on the prediction taken for these three frequency samples, the first method has RMSE 0.101 m, the second method has RMSE 0.086 m and use astronomical frequencies has RMSE 0.093 m. the results showed that to obtain the frequency of multivariate LSHE methods, the accuracy and reliability are higher.



(a)



(b)

**Figure 13.** a) A comparison between the data known and predicted data for the point-1, based on samples of extracted frequency of LSSA. b) A comparison between the data known and predicted data for the point-1, based on samples of extracted frequency of astronomical, based on samples of extracted frequency of LSHE.

## 6. Conclusion

In this contribution, first simulated data is used to investigate the effect of different phenomena on time series. However, actual observations have been used to extract frequencies and predict data. The total time interval of observation of sea surface in this research is from 1992 to 2008. In order to predict, predicted sea surface for the period 2006 to 2008. The prediction for sea surface observations in the two-year period, using three frequency list mentioned above was conducted. Accordingly, the height of the sea surface is calculated and compared with actual data. Particularly, we found that both the LS-HE and LSSA methods perform equivalently for analyzing the SSH time series of the Persian Gulf and Oman Sea. Generally, the longer time series of observations give the guarantee of more accurate results when time series contain noise. In the light of our results from the simulation, the methods were used to extract the main frequencies of SSH changes over the Persian Gulf and Oman Sea. Our main outcomes are summarized as follows:

- Impact of noise: for a 16-years length SSH time series: the impact of noise with the amplitude less than 10 mm is negligible and the accuracy of the extracted frequencies remains better than 0.003 cycles per day. Our LSSA results also indicated that for the time series that contain noise with an amplitude of greater than 50 mm, the extracted frequencies would be inaccurate. In this case, longer time series cannot compensate the noise effect.
- Impact of outliers: outliers of medium scales (i.e. the amplitude of outlier value is smaller than  $3\sigma$ , where  $\sigma$  is the standard deviation of SSH time series) presents a little influence on the extracted frequencies and the accuracy remains better than 0.004 cycle per day. When the number of outliers increases, e.g. more than 10% of the observations was contaminated with outliers, the estimated spectrums will be affected similar to that of noisy time series.
- Impact of missing values: for the case of randomly missing values and the length of time series being 16-years, we found that one needs at least half of the possible observations for an accurate estimation of the simulated frequencies. In case of sparse SSH time series, the accuracy of the spectral

analysis methods significantly decreased, specifically, those moderate length cycles were evaluated inaccurate. Existing up to 5-year concentrated missing values (or a gap) in a 16-year time series of SSH did not affect the accuracy of the extracted Principal Lunar Diurnal, Principal Solar Semi-diurnal and seasonal frequencies. In case of missing values with periodic pattern, methods were found highly vulnerable.

- Recognizing nearby frequencies: we found that the ability of recognizing nearby frequencies strongly depends on the magnitude of the frequencies themselves and how close they are; length of the time series; and the accuracy of observations. Our results showed that when an accurate 30-year SSH time series exists, the nearby low frequencies (with a difference of greater than 0.001 cycle per day) were detectable. Note that the closer frequencies, need longer time series to be separated. We also found that the existence of noise and missing values make the separation of the nearby frequencies more difficult.

In the light of the simulated results, both LSSA and LS-HE methods were applied to analyze 626 satellite altimetry-derived SSH time series over the Persian Gulf and Oman Sea. Particularly, we found that:

- In spite of missing values and noise that affects the performance of spectral analysis, LSSA and LS-HE successfully extracted dominant frequencies of the Persian Gulf and Oman Sea level changes. Nearby aliased frequencies of tidal components (e.g.  $M_2$  and  $S_2$ ,  $M_4$  and  $S_4$ ,  $M_6$  and  $S_6$ ) were detected with suitable accuracy. Using the multivariate LS-HE, we detected 60 frequencies using all 626 available time series over the Persian Gulf and Oman Sea. Long term fluctuations, (e.g. 0.2086, 0.2877, 0.3488, 0.5393, 0.6751, 0.7430 cycle/year respectively correspond to 4.8, 3.5, 2.9, 1.9, 1.5, and 1.3 years for one cycle of changes), moderate length frequencies (e.g. seasonal and monthly changes) and tidal components (e.g.  $M_2$ ,  $S_2$ ,  $N_2$ ,  $K_2$ ,  $K_1$ ) were detected. These frequencies can be generalized to the whole Persian Gulf and Oman Sea.

- In the following, to check the effectiveness of three samples extracted frequency (Astronomical-LSSA-Multivariate LSHE), the tide is predicted for a period of time. The

results of predicted tides is presented, the amount of RMSE based on extracted frequency of LSSA method 0.101 m, astronomical tables 0.093 m and frequency of Multivariate LSHE method 0.086 m, according to the results conclusion was that to obtain the frequency of multivariate LSHE methods the accuracy and reliability are higher.

As the future work, the impact of external variability on level fluctuations of the Persian Gulf and Oman Sea and the derived significant frequencies. This can be done by considering the rainfall data, river discharges, measurements of the surface evaporation rate, temperature, and pressure data sets, which enhance the interpretation of the results.

### References

- Amiri-Simkooei, A.R. and Asgari, J., 2012, Harmonic analysis of total electron contents time series: methodology and results. *GPS solutions*, 16(1), 77-88.
- Amiri-Simkooei, A., 2007, Least-squares variance component estimation: theory and GPS applications (Doctoral dissertation, TU Delft, Delft University of Technology).
- Amiri-Simkooei, A.R., Parvazi, K. and Asgari, J., 2017, Extracting tidal frequencies of the Persian Gulf and Oman Sea using multivariate least square harmonic estimation of sea level coastal height observations time series. *Journal of the Earth and Space Physics*, 43(1).
- Amiri-Simkooei, A.R., Zaminpardaz, S. and Sharifi, M. A., 2014, Extracting tidal frequencies using multivariate harmonic analysis of sea level height time series, *J. Geod.*, 88, 975-988, doi: 10.1007/s00190-014-0737-5.
- Aviso and Podaac, 2008, Aviso and Podaac Users Handbook Igrd and Gdr Jason Products. Jpl D-21352, October. Smm-Mu-M5-Op-13184-Cn.
- Boashash, B. and Putland, G., 2003, Polynomial Wigner-Ville Distributions and Design of High-Resolution Quadratic TFDs with Separable Kernels. In *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference* (pp. 3-27), Elsevier Ltd.
- Cartwright, D. E., 1993, *Theory of ocean tides with application to altimetry*. In *Satellite altimetry in geodesy and oceanography* (pp. 100-141), Springer, Berlin, Heidelberg.
- Chelton, D.B., Ries, J.C., Haines, B.J., Fu, L.L. and Callahan, P.S., 2001, *Satellite altimetry*. In *International geophysics* (Vol. 69, pp. 1-ii), Academic Press.
- Doodson, A.T., 1954, The analysis of tidal observations for 29 days. *The International Hydrographic Review*, (1).
- Fok, H.S., 2012, *Ocean tides modeling using satellite altimetry* (Doctoral dissertation, The Ohio State University).
- Fu, L.L. and Cazenave, A., 2001, *Satellite altimetry and earth sciences*. Vol. *International Geophysical Series* 69.
- Frappart, F., Fatras, C., Mougou, E., Marieu, V., Diepkilé, A. T., Blarel, F. and Borderies, P., 2015, Radar altimetry backscattering signatures at Ka, Ku, C, and S bands over West Africa. *Physics and Chemistry of the Earth, Parts A/B/C*, 83, 96-110.
- Farzaneh, S. and Parvazi, K., 2018, Noise Analysis of Satellites Altimetry Observations for Improving Chart Datum within the Persian Gulf and Oman Sea. *Annals of Geophysics*, 61, p.42.
- Khaki, M., Forootan, E., Sharifi, M. A., Awange, J. and Kuhn, M., 2015, Improved gravity anomaly fields from retracked multitemission satellite radar altimetry observations over the Persian Gulf and the Caspian Sea. *Geophysical Journal International*, 202(3), 1522-1534.
- Mahalanobis, P.C., 1936, On the generalized distance in statistics. *National Institute of Science of India*.
- Parvazi, K., Asgari, J., Amirisimkooei, A.R. and Tajfirooz, B., 2015, Determination of difference between datum and reference ellipsoid by using of altimetry datas of Topex/Poseidon Jason-1 and observations of coastal tide gauges. *Journal of Geomatics Science and Technology*, 5(1), 257-269.
- Papa, F., Legrésy, B. and Rémy, F., 2003, Use of the Topex-Poseidon dual-frequency radar altimeter over land surfaces. *Remote sensing of Environment*, 87(2-3), 136-147.
- Purser, B. H. and Seibold, E., 1973, The principal environmental factors

- influencing Holocene sedimentation and diagenesis in the Persian Gulf. In *The Persian Gulf* (pp. 1-9), Springer, Berlin, Heidelberg.
- Picot, N., Case, K., Desai, S. and Vincent, P., 2003, AVISO and PODAAC User Handbook. IGDR and GDR Jason Products. SMM-MU-M5-OP-13184-CN (AVISO), JPL D-21352 (PODAAC).
- Rubin, D. B., 2002, *Statistical Analysis with Missing Data*. ISBN: 978-0-471-18386-0.
- Sharifi, M.A., Forootan, E., Nikkhoo, M., Awange, J.L. and Najafi-Alamdari, M., 2013, A point-wise least squares spectral analysis (LSSA) of the Caspian Sea level fluctuations, using Topex/Poseidon and Jason-1 observations. *Advances in Space Research*, 51(5), 858-873.
- Tamura, Y., 1993, Additional terms to the tidal harmonic tables. In *Proc. 12th Int. Symp. Earth Tides*. Science Press, Beijing, 345-350.
- Vaniček, P., 1969, Approximate spectral analysis by least-squares fit. *Astrophysics and Space Science*, 4(4), 387-391.
- Vaniček, P., 1971, Further development and properties of the spectral analysis by least-squares. *Astrophysics and Space Science*, 12(1), 10-33.
- 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.
- Xi, Q. W. and HOU, T., 1987, A new complete development of the tide-generating potential for the epoch J2000.0. *Acta Geophysica Sinica*, 30(4), 349-362.
- Kern, M., Preimesberger, T., Allesch, M., Pail, R., Bouman, J. and Koop, R., 2005, Outlier detection algorithms and their performance in GOCE gravity field processing. *Journal of Geodesy*, 78(9), 509-519.
- PO. DAAC., 1993, *PO.DAAC Merged Geophysical Data Record Users Handbook* JPL D-11007, November 1996.