

Program VAV/2000 for Tidal Analysis of Unevenly Spaced Data with Irregular Drift and Colored Noise

Angel Venedikov^{1,2)*}, José Armoso¹⁾ and Ricardo Vieira¹⁾

1) Institute of Astronomy and Geodesy, University Complutense, Madrid 28040

2) Geophysical Institute, Bulgarian Academy of Sciences, Acad.G.Bonchev street, block 3, Sofia 1113

(Received September 30, 2000; Revised January 31, 2001; Accepted February 1, 2001)

Abstract

A general description of the algorithm used by a new computer program for tidal data processing is presented. The main stages of the algorithm are: (i) filtration of intervals without overlapping and (ii) application of the Method of the Least Squares on the filtered numbers. A new element is that all series of filtered numbers, related with different tidal species are processed all together. It is shown that this way of analysis may be equivalent to a direct processing of the data, without any filtration. However, the processing of the filtered numbers has the advantage to provide frequency dependent estimates of the precision. Another new element is that the program can be applied on totally unevenly spaced data.

1. Introduction

The program VAV is a further development of our algorithm whose last realization is made by the program NSV (Venedikov et al., 1977a). It consists in (i) filtration of independent (without overlapping) intervals and (ii) processing of the filtered numbers by the Method of the Least Squares (MLS). The stage (i) can be considered as a time/frequency window that is eliminating the drift and transforming the data in separate pairs of series, each pair corresponding to one of the main tidal species. In a way, the data from the time domain are transformed in a mixed time/frequency domain. In stage (ii) the series are processed independently, and the parameters for each species are determined individually. In such a way we get frequency dependent MLS estimates of the precision.

In VAV, following an idea of B. Ducarme, in stage (ii) MLS is applied on all series together. This eliminates any possibility of a leakage, not perfectly excluded by NSV for large series of very high precision. Correspondingly to this we had to use a new way for getting frequency dependent estimates of the precision.

NSV deals with hourly data, eventually with gaps between the intervals. VAV may accept totally unevenly spaced data (Figure 1). This new ability allows avoiding any interpolation of the data. Some special devices, e.g. the absolute gravity instruments may provide unevenly spaced data, if they will be used for the observation of the tides.

*E-mail: vened@geophys.bas.bg

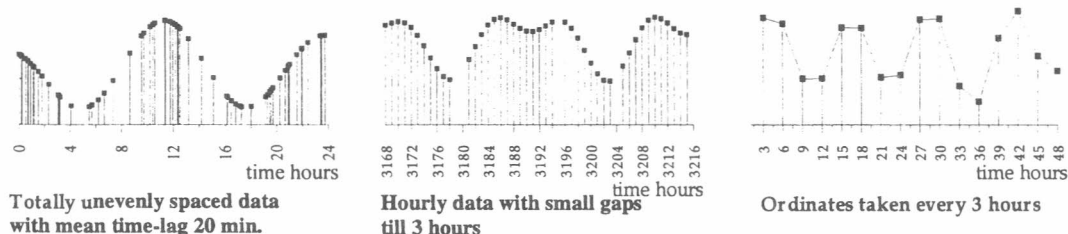


Fig. 1. Examples of data distribution in the filtered intervals, acceptable by the VAV program.

Other new elements of VAV are: separation of the tides, originated by the tidal potential 4th degree, determination of high frequency tidal and non-tidal waves, time and seasonal variations of the regression coefficients.

Within the present paper we shall restrict our attention to the main principles of the basic algorithm of VAV with a few results at its end.

2. Model Equations (ME) and a Solution by MLS

We shall deal with data partitioned in N intervals with central epochs $T = T_1, T_2, \dots, T_N$, containing n ordinates, whose total number is $M = nN$. The intervals and the whole set of data will be represented by the vectors (column vectors)

$$\mathbf{y}_i = \mathbf{y}(T_i) = \begin{bmatrix} y(T_i + t_1) \\ \dots \\ y(T_i + t_n) \end{bmatrix}, i = 1, 2, \dots, N \text{ and } \mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 \\ \dots \\ \mathbf{y}_N \end{bmatrix} = \begin{bmatrix} y(T_1 + t_1) \\ \dots \\ y(T_N + t_n) \end{bmatrix}. \quad (1)$$

In the case of unevenly spaced data the set t_1, \dots, t_n may vary in different \mathbf{y}_i . In the actual application of VAV the number n may also vary but, for an easier development of the theory in the following parts, we shall deal with $n = \text{const}$.

The data \mathbf{Y} can be represented by the ME in the following general form

$$\mathbf{Y} = \mathbf{AX} + \mathbf{PZ} + \mathbf{E}. \quad (2)$$

Here \mathbf{AX} is the model of the tidal signal, \mathbf{A} being $(M \times m)$ known matrix and \mathbf{X} is $(m \times 1)$ vector of the tidal unknowns. The model is created by using the grouping of the tides defined in (Venedikov, 1961, 1966) and used by (Chojnicki, 1973, Schüller, 1977, Tamura et al., 1991, Wenzel, 1994). If regression on a side agent is used, e.g. air-pressure, some non-tidal terms should be included in \mathbf{AX} .

The term \mathbf{PZ} is the model of the drift, the same as in NSV. It uses independent representations of the drift in $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N$ by polynomials of given low power $k \geq 0$ of the time t . If \mathbf{d}_i is the drift component of \mathbf{y}_i , we set up $\mathbf{d}_i = \mathbf{p}_i \mathbf{z}_i$, where \mathbf{p}_i is $(n \times (k+1))$ known matrix, composed by polynomials of $t = t_1, \dots, t_n$ and \mathbf{z}_i is a $((k+1) \times 1)$ vector of drift unknowns, related with \mathbf{y}_i . Then \mathbf{P} and \mathbf{Z} in (2) are created as follows

$$\mathbf{P} = \begin{bmatrix} \mathbf{p}_1 & \mathbf{0} \\ \dots & \\ \mathbf{0} & \mathbf{p}_N \end{bmatrix}, \text{ and } \mathbf{Z} = \begin{bmatrix} \mathbf{z}_1 \\ \dots \\ \mathbf{z}_N \end{bmatrix}. \tag{3}$$

VAV transforms the polynomials so that $\mathbf{P}^T \mathbf{P} = \mathbf{I}$ (identity matrix). The last term \mathbf{E} is the noise of \mathbf{Y} . If \mathbf{E} is white noise (WN) we have

$$\text{cov}(\mathbf{E}) = \text{cov}(\mathbf{Y}) = \mathbf{I}\sigma^2. \tag{4}$$

Under this assumption \mathbf{E} has in both time and frequency domains a constant power. Actually (4) is not true, \mathbf{E} is correlated red noise and it has frequency dependent power, higher at lower frequencies.

Under the assumption (4) the MLS estimates $\tilde{\mathbf{X}}$ of the unknowns \mathbf{X} can be obtained through the following operations, known as algorithm for separation of the unknowns

$$\tilde{\mathbf{X}} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{Y} \text{ where } \mathbf{B} = \mathbf{A} - \mathbf{P} \mathbf{P}^T \mathbf{A}. \tag{5}$$

The assumption (4), although false, does not affect seriously the values of $\tilde{\mathbf{X}}$. It is, however, seriously deforming the estimate of the precision. Under this assumption we get a frequency independent estimate of the precision while the effect of the real colored noise is frequency dependent. In this relation we shall use an algorithm providing estimates $\tilde{\mathbf{X}}$ equivalent to (5), which, in the same time, will allow us to find reasonable, frequency dependent estimates of the precision.

3. Filters Applied on Independent Intervals

We shall use some frequencies (angular) $\omega_j, j = 1, 2, \dots, \mu$ and $(n \times 2)$ matrices

$$\mathbf{c}_{ij} = \begin{bmatrix} \cos \omega_j t_1 & \sin \omega_j t_1 \\ \dots & \\ \cos \omega_j t_n & \sin \omega_j t_n \end{bmatrix}, i = 1, 2, \dots, N. \tag{6}$$

The matrices \mathbf{c}_{ij} are related with \mathbf{y}_i through the corresponding set $t = t_1, t_2, \dots, t_n$. For evenly spaced data \mathbf{c}_{ij} are one and the same for all \mathbf{y}_i .

The ω_j are chosen in such a way that $[\mathbf{p}_i \mathbf{c}_{i1} \mathbf{c}_{i2} \dots \mathbf{c}_{i\mu}]$ is a square and nearly orthogonal matrix, due to which the last $\mathbf{c}_{i\mu}$ may have only one column. An example of ω_j for hourly data, $n = 24$ and drift power $k = 0$ is $\omega_j = 2\pi j / n$ or, in degrees per hour, $\omega_j = 15^\circ, 30^\circ, 45^\circ \dots 180^\circ$. In any case some ω_j are related with the main tidal frequencies, as in this example. In the following the expression "at frequency ω_j " should be understood as concerning a tidal band around ω_j .

Through a linear transformation of \mathbf{c}_{ij} we get another set of $(n \times 2)$ matrices \mathbf{f}_{ij} . The latter are arranged in the matrices

$$\mathbf{F}_j = \begin{bmatrix} \mathbf{f}_{1j} & \mathbf{0} \\ \dots & \\ \mathbf{0} & \mathbf{f}_{Nj} \end{bmatrix}, \mathbf{F} = [\mathbf{F}_1 \mathbf{F}_2 \dots \mathbf{F}_\mu] \text{ and } \mathbf{D} = [\mathbf{P} \mathbf{F}], \text{ Size}(\mathbf{D}) = (M \times M). \tag{7}$$

The transformation of \mathbf{c}_{ij} in \mathbf{f}_{ij} is made in such a way that

$$\mathbf{D}^T \mathbf{D} = \mathbf{D} \mathbf{D}^T = \mathbf{I} . \quad (8)$$

For evenly spaced data the $(n \times 2)$ blocks \mathbf{f}_{ij} in \mathbf{F}_j are the same for all i . In this sense we call \mathbf{F}_j a filter. The result of the application of this filter on the data is

$$\mathbf{u}_j = \mathbf{u}(\omega_j) = \mathbf{F}_j^T \mathbf{Y} = \begin{bmatrix} \mathbf{f}_{1j}^T \mathbf{y}_1 \\ \dots \\ \mathbf{f}_{Nj}^T \mathbf{y}_N \end{bmatrix}, \text{ Size}(\mathbf{f}_{ij}^T \mathbf{y}_i) = (2 \times 1), \text{ Size}(\mathbf{u}_j) = (2N \times 1) . \quad (9)$$

The result from the application of all filters is

$$\mathbf{U} = \mathbf{F}^T \mathbf{Y} = \begin{bmatrix} \mathbf{u}_1 \\ \dots \\ \mathbf{u}_\mu \end{bmatrix} = \begin{bmatrix} \mathbf{u}(\omega_1) \\ \dots \\ \mathbf{u}(\omega_\mu) \end{bmatrix} . \quad (10)$$

The elements of \mathbf{u}_j are obtained through filtration of intervals without overlapping and may be considered as disposed in a time domain. On the other side the elements of \mathbf{U} are vectors, which may be considered as disposed in a frequency domain. Thus we have a transformation of the data \mathbf{Y} from the time domain in a mixed time/frequency domain.

4. ME for the Filtered Numbers and Application of MLS

From (8) follows $\mathbf{F}^T \mathbf{P} = \mathbf{0}$ and $\mathbf{F}^T \mathbf{F} = \mathbf{I}$. By using this and replacing \mathbf{Y} in $\mathbf{U} = \mathbf{F}^T \mathbf{Y}$ through (2) we get

$$\mathbf{U} = \mathbf{G} \mathbf{X} + \mathbf{E}' , \quad (11)$$

where $\mathbf{G} = \mathbf{F}^T \mathbf{A} = \mathbf{F}^T (\mathbf{B} + \mathbf{P} \mathbf{P}^T \mathbf{A}) = \mathbf{F}^T \mathbf{B}$ and $\mathbf{E}' = \mathbf{F}^T \mathbf{E}$ is the noise of \mathbf{U} . Under the assumption (4) and due to $\mathbf{F}^T \mathbf{F} = \mathbf{I}$ the noise has $\text{cov}(\mathbf{E}') = \mathbf{I} \sigma^2$. Hence, as far as we are allowed to apply MLS on \mathbf{Y} we can apply MLS on \mathbf{U} by using (11) as ME. This application will provide new estimates of \mathbf{X}

$$\tilde{\mathbf{X}}' = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{U}' = (\mathbf{B}^T \mathbf{F} \mathbf{F}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{F} \mathbf{F}^T \mathbf{Y} . \quad (12)$$

From (5) follows $\mathbf{B}^T \mathbf{P} = \mathbf{0}$ and from (8) - $\mathbf{F} \mathbf{F}^T = \mathbf{I} - \mathbf{P} \mathbf{P}^T$. By using this

$$\tilde{\mathbf{X}}' = (\mathbf{B}^T (\mathbf{I} - \mathbf{P} \mathbf{P}^T) \mathbf{B})^{-1} \mathbf{B}^T (\mathbf{I} - \mathbf{P} \mathbf{P}^T) \mathbf{Y} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{Y} = \tilde{\mathbf{X}} . \quad (13)$$

I.e., the estimates obtained through MLS, applied on the filtered numbers \mathbf{U} are the same as those in (5), obtained by the direct application on the original data \mathbf{Y} .

The ME (11) can be written separately for the vectors (9) as

$$\mathbf{u}_j = \mathbf{F}_j^T \mathbf{A} \mathbf{X} + \mathbf{E}'_j, j = 1, 2, \dots, N, \quad (14)$$

where $\mathbf{E}'_j = \mathbf{F}_j^T \mathbf{E}$ is the noise of \mathbf{u}_j with variance σ_j^2 .

The filter \mathbf{F}_j is built in such a way that it amplifies the corresponding frequency ω_j and retains all other $\omega_l, l \neq j$. Hence the variance σ_j^2 is actually the power of the noise at frequency

ω_j . Provided ω_j are growing with j and the noise is a red one, with higher power at the lower frequencies, we have $\sigma_1^2 > \sigma_2^2 > \dots > \sigma_\mu^2$.

We can represent the tidal signal decomposed into components

$$\mathbf{AX} = \mathbf{A}_1\mathbf{x}_1 + \mathbf{A}_2\mathbf{x}_2 + \dots + \mathbf{A}_\mu\mathbf{x}_\mu, \tag{15}$$

where the term $\mathbf{A}_j\mathbf{x}_j$ is the signal at frequency ω_j .

Due to the relation of \mathbf{F}_j with ω_j considered above, $\mathbf{F}_j^T\mathbf{AX} \cong \mathbf{F}_j^T\mathbf{A}_j\mathbf{x}_j$. It follows that through the solution of (11) or (14) the tidal unknowns \mathbf{x}_j are mainly determined through \mathbf{u}_j and affected by the noise $\mathbf{E}'_j = \mathbf{F}_j^T\mathbf{E}$ having variance σ_j^2 . This is namely the effect of the colored noise, which is not seen if a direct application of MLS on \mathbf{Y} through (5) is proceeded. The direct application actually accepts $\sigma_1^2 = \dots = \sigma_\mu^2 = \sigma^2$ and the precision of \mathbf{X} is estimated by using $\text{cov}(\mathbf{X}) = (\mathbf{B}^T\mathbf{B})^{-1}\sigma^2$. Our idea is to get individual estimates of σ_j^2 and, when the parameters \mathbf{x}_j are estimated, to replace σ^2 by σ_j^2 in the above expression.

The variance σ_j^2 is estimated by VAV in the following way.

We compute the residual vector $\mathbf{r}_j = \mathbf{u}_j - \mathbf{F}_j^T\mathbf{A}_j\tilde{\mathbf{X}}$ that involves $2N$ elements, as many as \mathbf{u}_j . Let the number of the unknowns in \mathbf{x}_j is m_j . Hence, due to $\mathbf{F}_j^T\mathbf{AX} \cong \mathbf{F}_j^T\mathbf{A}_j\mathbf{x}_j$, we can accept that \mathbf{r}_j has $2N - m_j$ degrees of freedom. Therefore VAV accepts the quantity

$$\tilde{\sigma}_j^2 = \mathbf{r}_j^T\mathbf{r}_j / (2N - m_j) \tag{16}$$

as an estimate of the variance $\tilde{\sigma}_j^2$ and $\tilde{\sigma}_j = \tilde{\sigma}(\omega_j)$ as the mean square deviation (MSD) at frequency ω_j (Figure 2). It is actually a discrete amplitude spectrum of the noise.

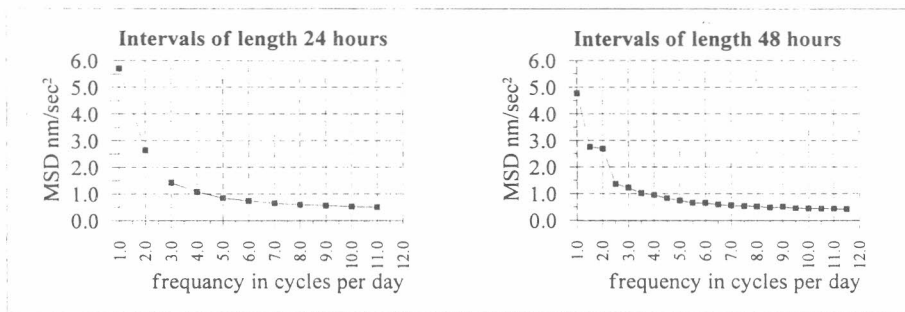


Fig. 2. Frequency dependent estimates of the precision (spectrum of the noise); SG data Brussels.

5. Some Examples

The following result

$$\delta_4 = \delta(4\text{DSDTD}) = 0.6364 \pm 0.0384, \quad \kappa_4 = \kappa(4\text{DSDTD}) = -1.99^\circ \pm 3.46^\circ \tag{17}$$

is the first determination, as far as we know, of the tidal parameters of the D, SD and TD tides generated by the tidal potential 4th degree. It is obtained by the application of VAV on the largest SG series of data from station Brussels, 21.04.1982 – 22.03.2000, kindly provided to us by B.Ducarme. The result is actually obtained for a group named 4DSDTD, created according to a

new way of grouping (Venedikov et al., 1977b). This group includes namely all D, SD and TD tides coming from this potential.

For comparison here are the parameters of the usually determined group M4 whose origin is the same potential 4th degree

$$\delta(M4) = 0.7993 \pm 0.1218, \kappa(M4) = 56.06^\circ \pm 8.74^\circ. \quad (18)$$

The precision of (18) is considerably lower which is, probably, the reason for an abnormal phase.

Table 1 is a demonstration of the possibility of VAV to process data with different unit of time. Somewhat surprisingly we have a practically perfect coincidence between the two cases, although 3 times different quantity of data are used. The explanation is that the correlation between the noise of the hourly ordinates is so strong that 2/3 of the data do not provide almost any information.

Table 1. Analysis by the VAV program of SG Data Brussels, 21.04.1982 – 22.03.2000

Tidal Groups	Data every 3 hours		Data every 1 hour	
	δ	MSD	δ	MSD
Mf	1.14138	± 0.00334	1.14170	± 0.00343
Q1	1.15106	± 0.00042	1.15122	± 0.00041
O1	1.15328	± 0.00008	1.15331	± 0.00008
S1	1.66406	± 0.01190	1.65841	± 0.01154
2N2	1.16492	± 0.00118	1.16497	± 0.00106
M2	1.18379	± 0.00005	1.18380	± 0.00004
L2	1.17672	± 0.00158	1.17622	± 0.00140
3D	1.08397	± 0.00296	1.08341	± 0.00287
3SD	1.07424	± 0.00178	1.07408	± 0.00158
M3	1.05931	± 0.00256	1.05967	± 0.00194

The groups 3D and 3SD are groups involving all D and SD tides respectively, originated by the tidal potential of 3rd degree (Venedikov et al., 1977), as well as M3.

References

- Chojnicki, T. (1973): Ein Verfahren zur Erdzeitenanalyse in Anlehnung an das Prinzip der kleinsten Quadrate, *Mitt. Inst. Theor. Geod. Univ. Bonn*, **15**.
- Schüller, K. (1977): Tidal Analysis by the Hybrid least Squares Frequency Domain Convolution Method, 8th Int. Symp. Earth Tides, Bonn.
- Tamura, Y., T. Sato, M. Ooe and M. Ishiguro (1991): A Procedure for Tidal Analysis with a Bayesian Information Criterion. *Geoph. J.Int.*, **104**, 507-516.
- Venedikov, A.P. (1961): Application à l'Analyse Harmonique des Observations des Marées Terrestres de la Méthode des Moindres Carrées, *C. R. Acad. Bulg. Sci.*, **14**, 671-674.
- Venedikov, A.P. (1966): Une Méthode d'Analyse des Marées Terrestres à Partir d'Enregistrements de Longueurs Arbitraires, *Bull. d'Information Marées Terrestres*, **43**, 1687-1687.
- Venedikov, A.P., R. Vieira, C. de Toro and J. Arnosó (1997a): A New Program Developed in Madrid for Tidal Data Processing, *Bull. d'Information Marées Terrestres*, **126**, 9669-9704.
- Venedikov, A.P., R. Vieira and C. de Toro (1997b): On the Determination of the D and SD Earth Tides Generated by the Tidal Potential of the Third Order, *Bull. d'Information Marées Terrestres*, **126**, 9669-9704.
- Wenzel, H.-G., (1994): Earth Tide Analysis Package ETERNA 3.0. *Bull. d'Information Marées Terrestres*, **118**, 8719-8721.