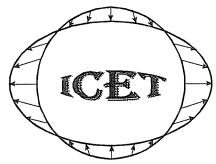
MAREES TERRESTRES

BULLETIN D'INFORMATIONS

INTERNATIONAL CENTER FOR EARTH TIDES
CENTRE INTERNATIONAL DES MAREES TERRESTRES



Federation of Astronomical and Geophysical Data Analysis Services
(FAGS)

International Association of Geodesy - International Gravity Field Service
(IAG – IGFS)

Publié avec le soutien de l'Observatoire Royal de Belgique

BIM
1 3 9

15 DECEMBRE 2004

Editeur: Dr. Bernard DUCARME
Observatoire Royal de Belgique
Avenue Circulaire 3
B-1180 Bruxelles

BIM 139

15 Décembre 2004

OBITUARY Baron Paul Melchior 11011

MIYAMA S..PING J., TSUBOKAWA Y., TAMURA Y., HEKI K., MATSUMOTO K., SATO T.
Estimating the fluid core resonance based on strain observation 11015

GUO J. Y., GREINER_MAI H., DIERKS O., BALLANI L., NEUMEYER J., SHUM C.K. .....
Application of the folding-averaging algorithm for the determination of the periods of the Earth's free oscillation using superconducting gravimeter data..... 11025

VENEDIKOV A.P., VIEIRA R.....
Guidebook for the practical use of the computer program VAV-Version 2003 11037

Obituary
Baron Paul Melchior

On the 15th of September 2004, Baron Paul Melchior passed away. He was 78 years old. Recently, he had hip replacement surgery and died painfully from complications.

Paul Melchior was an exceptional person. He contributed immensely to the development of geophysics not only as an outstanding scientist but also as a great leader. From 1973 to 1991, he served as Secretary General of International Union of Geodesy and Geophysics (IUGG), and was the Honorary Secretary General of IUGG until his death. His tenure lasted so long because he had an extraordinary ability to meet the expectations of his colleagues, and it was difficult to find someone who was willing to follow in his footsteps to manage such a demanding position. One of his successes as Secretary General was the adhesion of China as member of IUGG, for which he showed all his skill in diplomatic matters.

During his long and fruitful scientific career, Paul Melchior had been :
Director of the International Center for Earth Tides (1958-1995),
President of the Commission of Earth Rotation of the International
Astronomical Union (1967-1970), and President of CODATA (1974-1978).

Paul Melchior studied mathematics at the Free University of Brussels. His Doctorate in Mathematical Sciences was earned at the same University in 1951. His professional career started as assistant at the Royal Observatory of Belgium in 1949. He became the Director in 1981 and served in that capacity until his retirement in 1990. Paul Melchior had a profound attachment to the Observatory. His contributions to the development of geophysics and geodesy at the Observatory are countless and brought great renown to his grateful Institute. For his dedication for science and his international reputation, in 1993 King Baudouin bestowed upon him the title of Baron. Paul Melchior valued this mark of honour above all. He was very attached to his country and its monarchy.

Paul Melchior began his career as an astronomer. Between 1950-1957, he spent long nights observing at the Askania Great Meridian Circle in Uccle. After compiling the data, he published the most precise star catalogue of that time. Soon, he was interested in the Earth's

rotation. He then developed the complete theory of the motions of the Earth's rotational axis and its link to Earth's tides. This became his main subject of research. He began measuring Earth tides in 1957 with the Verbaandert-Melchior quartz tiltmeters. In 1958, he was the first to analyze Earth tide observations using an electronic computer, the famous IBM 650. In 1968, Paul Melchior founded, with Johnny Flick, the Underground Laboratory of Walferdange in Luxembourg. In 1969, with Prof. Manfred Bonatz, they installed tiltmeters, gravimeters and a satellite camera in Spitzbergen (Norway). Thanks to his growing scientific reputation, the US Navy supported him to set up the first permanent Transit Satellite Doppler recording station in Europe, which operated until 1993. In 1973, his skillful and accurate interpretation of Earth tide gravity observations led the US Air Force to entrust his team to carry out Trans World Tidal Gravity Profiles. A total of 127 stations
were observed worldwide for at least 6 months. This exceptional data set was used to assess the precision of the oceanic tidal models derived from Topex-Poseidon a few years ago. Always looking for more precise
observations in gravimetry, Paul Melchior succeeded in raising funds to install the first superconducting gravimeter in Europe in Uccle.

Paul Melchior was also a Professor at the Catholic University of Louvain. His lectures were gripping due to his tremendous experience and expertise. He directed several Ph. D. students who considered themselves lucky to have him as an advisor. Being absorbed by his scientific and administrative life, he chose to trust his students instead of directing their theses on a daily basis. His door was open at any time to discuss science and he always gave advice without imposing direction. Paul Melchior succeeded in hiring some of his students at the Royal Observatory of Belgium, and helped and supported the others to get positions in academic institutes. In his last years, he was deeply involved in the development of the European Center for Geodynamics and Seismology (ECGS) in Luxembourg. Paul Melchior, one of the founders of the ECGS, was the most active member of the scientific committee.

Although Paul Melchior was very busy, he found the time to write about 300 scientific papers as well as two books published by Pergamon Press. The first one, published a few decades ago, was devoted to Earth Tides. It is still the only one on the subject. He will be remembered as "the Father of Earth tides," and was bestowed during his lifetime with numerous international honours and distinctions. Notable among these (in addition to the title of Baron awarded by the King of Belgium) are his election to the title of "Fellow" of the American Geophysical Union in 1978; his nomination as honorary Professor of the Institute of Geodesy and Geophysics of the Chinese Academy of Sciences; and his nominations as foreign member to various scientific academies in Finland, the Netherlands, Spain, and Romania.

Paul Melchior had an extremely rich personality. In addition to his scientific career, he was a humanist passionate about history and literature. He also enjoyed national and international politics, which
were favourite subjects of discussion. Anyone who had the chance to have dinner with Paul Melchior will never forget an enjoyable time listening to his incredible stories so well told. He is remembered for his taste
for good foods, especially Italian cooking.

As a friend, Paul Melchior was trusting and loyal. He enjoyed writing to his friends and his correspondence could easily fill the entire library of the Royal Observatory. Being a

stamp collector, one suspected that each time he sent a letter, he was anxious to get a nice stamp on the returning letter.

A man of his calibre could count on a wife of exception. Madame Melchior was literally at the service of the career of her husband. She prepared dinner for so many visiting scientists. She also learned Russian to translate the huge Russian bibliography on Earth Tides for her husband but also for all the readers of the Bulletin d'Informations des Marées Terrestres.

Paul Melchior will be missed not only by his family and friends, but also by his many colleagues, worldwide. His motto was: "In Omnibus Terris Amicus".

Prepared by V. Dehant and O. Francis

Estimating the Fluid Core Resonance Based on Strain Observation

J. Ping[1,2], T. Tsubokawa[1] Y. Tamura[1], K. Heki[3], K. Matsumoto[1], and T. Sato[1]
(jsping@miz.nao.ac.jp)
1. National Astronomical Observatory, 2-12 Hoshigaoka Mizusawa Iwate, 023-0861, Japan
2. Astrogeodynamic Center, Shanghai Astronomical Observatory, Shanghai, 200030, China
3. Division of Earth and Planetary Sciences, Hokkaido University, Sapporo, 060-0810 Japan

[Abstract] Esashi Earth Tides Station has been operated for tidal observations from 1979. Among the observations, a long term extensometer data set for 3 components of both free end and middle point transducers are calibrated and corrected to solve the tidal drift and tidal admittances. Based on the estimated admittances of diurnal tidal constituents, the fluid core resonance parameters are obtained as 419.9+/-1.3 sidereal days for the eigenperiod, and 5,900-7,440 for the quality factor. This result is precise enough to compare with the results obtained from other independent methods.

1. Introduction

Since 1979, the Earth tides and the secular crustal motion are observed at the Esashi Earth Tides Station of the Mizusawa Astrogeodynamics Observatory of National Astronomical Observatory by collocating three kind instruments of the gravimeter, the quartz tube extensometers and the water tube tiltmeters. Here after this site is referred to the Esashi station. The data obtained from these instruments clearly indicate that the Esashi station is very stable. For example, the average strain rate in the EW direction is to be about -3×10^{-8} per year in the sense of the contraction as a mean rate over the past 17 years.

The data obtained at the Esashi station have an high potential to study not only the detailed features of the Earth tides but also to study the tectonic phenomena around this area that should be related to the plate motion. We have compiled the strain data obtained from 1980. Based on this data set, we started the researches related to the Earth tide and the secular crustal motion. As a report using the newly compiled data, we will introduce here the analysis results for the fluid core resonance (FCR) due to the resonant motion of the Earth's core introduced by the diurnal tidal forces of the Sun and the Moon. The tidal factors of the strain tides are represented with a pure combination of the Love and Shida number h and l. There is a possibility that the strain tides reflect the time variations inside the Earth's elastic structural parameters.

The Esashi station (141°20'7"E, 39°8'53"N, +393m) is located on the north side of the Mt. Abara, about 16 km east of the Mizusawa Astrogeodynamical Observatory of NAOJ. The station consists of three observation tunnels and a gravity measurement room. (See Figure 1) The 150 m long tunnel is dug into granite bedrock and offers very stable environment for reliable measurements of the Earth tides and crustal deformation. Three quartz tube

extensometers along north-south (NS), east-west (EW) and north-east (NE) directions, respectively and two water tube tiltmeters have regularly been used for observation since June 1979. The observation with a borehole strainmeter started in January 1985. In addition, atmospheric pressure, rainfall and air temperature are also measured. Log-file of the entrances into the observational tunnel, offset of the signals, and occurrences of earthquakes, etc., is also recorded and available in machine-readable forms. A new type of absolute gravimeter (AG) is under construction. A cryogenic superconducting gravimeter (SG) was installed in 1988 for detecting tiny signals from the core of the Earth. Since 2001, the observation of comparing the AG and the SG is carried out once a year. The data observed from the extensometers, the tiltmeters and the borehole strain meter are sampled at every 1-minute interval, then, the data are sent to the Mizusawa campus by the telemeter system.

2. Strain data obtained by the extensometers and data preprocessing

The extensometer consists of a quartz-glass tube as the standard for measurement, supporting frames and a transducer. One end of each tube is fixed on bedrock and the other is kept free. The displacements of the pedestals at the free end (indexed by F) and the middle point (indexed by M) relative to the fixed-end are measured with differential transformers with primary exciting signal of 5 kHz and 2.5 Vp-p. (Tsubokawa & Asari, 1979). The resolution is better than 10^{-10} strain units for each component of the extensometers. To calibrate the transducer and to adjust the zero point of the output signal, each transducer is mounted on sliding stage with a differential micrometer. The routine strain observation at the free ends was started from June 1979. The observation at the middle points began in the beginning of 1980 to check the reliability of the observed strain steps caused by earthquakes (Sato et al., 1980)

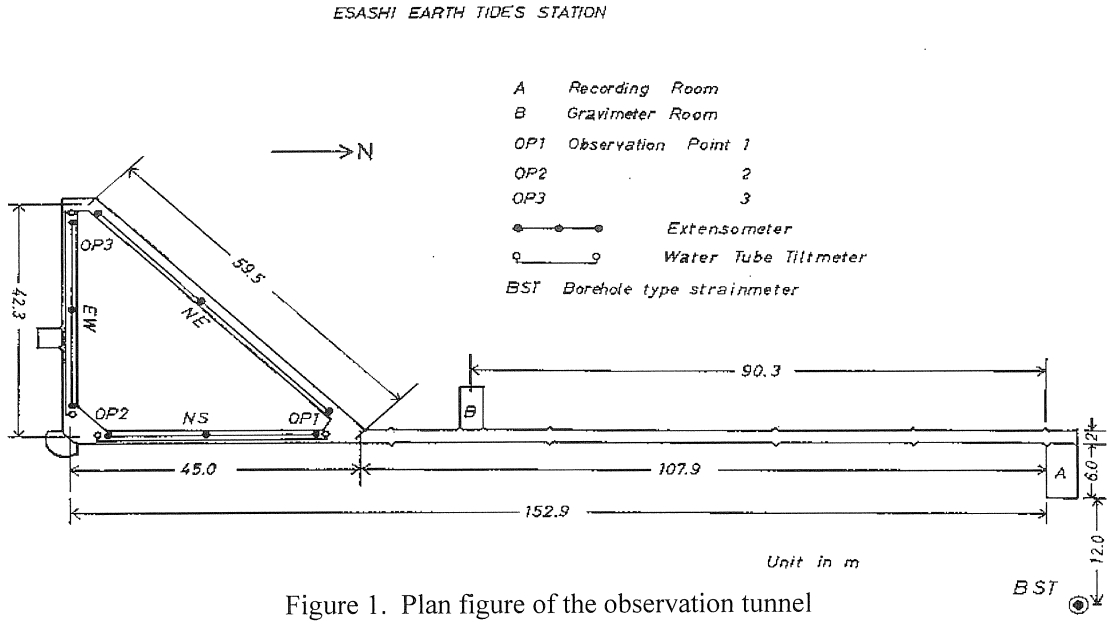


Figure 1. Plan figure of the observation tunnel

Seventeen years (i.e. the period from January 25, 1980 to January 25, 1997) strain data obtained from the 3 extensometers at both the free ends and the middle points are used in this study. Due to the limitation of computer memory, but in order to get high time resolution for estimating the admittance of diurnal and semi-diurnal tidal constituents, the raw 1-minute data were decimated at the rate of every 30 minutes by using moving average method to make a standard database for tidal analysis, although usually 1-hour data are used for the tidal analysis. The original data include irregular parts due to the malfunction of the instrument, the works in the observation tunnel, the power failure and so on. Commonly, the magnitude of step like changes due to these irregular origins are carefully estimated and corrected using the data based on the trend component that was estimated by using the tidal analysis program 'BAYTAP-G' (Tamura et al., 1991) without estimating the response to the atmospheric pressure changes, in the empirical way.

In our data analysis, the input data for BAYTAP-G are 30 minutes sampling standard database with its corresponding scale factors, the barometer data obtained at the same period with the same sampling rate as strain data and its scale factors. The barometer data is applied to estimate the strain response to the variation of atmospheric pressure in the tunnel. When running BAYTAP-G, relatively short span and time lag are set at first step. The strain steps, the strain drift and the atmospheric response are estimated together with the amplitudes and phase lags for various tidal constituents in this procedure.

Using the data with the sampling period of 30 minutes can improve the time resolution and precision of the final analysis results for the tidal admittances. By adjusting the step like changes together with solving atmospheric response, it can reduce the systematical biases between the estimated trend and original data to less than 1.5×10^{-8} strain units. If not introducing atmospheric response here, the biases will be 1 order larger or more for the data of last several years. The results from data pre-processing are given in Figure 2.

At east-west direction, the interseismic shortening rate has been investigated by many researchers based on continuous GPS observation in northeast Japan. Two groups of Sheng-Tu & Holts (1995) and Mazzotti (et al., 2000) obtained an average strain rate of about -3×10^{-8} per year in the EW direction. Heki (2004) pointed out that even in the same area, the shortening rate is not exactly the same by different place, and it is about 1×10^{-7} per year around Esashi station in recent years. By linear fitting the drifts of EW free end strain component in Figure 2, its has been obtained as $+1.8 \times 10^{-7}$, -4×10^{-8} and -2.0×10^{-7} for the year periods of 1979.4-1983.1, 1983.1-1994.6, and 1994.6-1997.0, respectively. The average over the whole observation period is estimated to be at -3×10^{-8} per year, which agrees with the mean strain rate from GPS data analysis (Sheng-Tu & Holts, 1995; Mazzotti, et al., 2000).

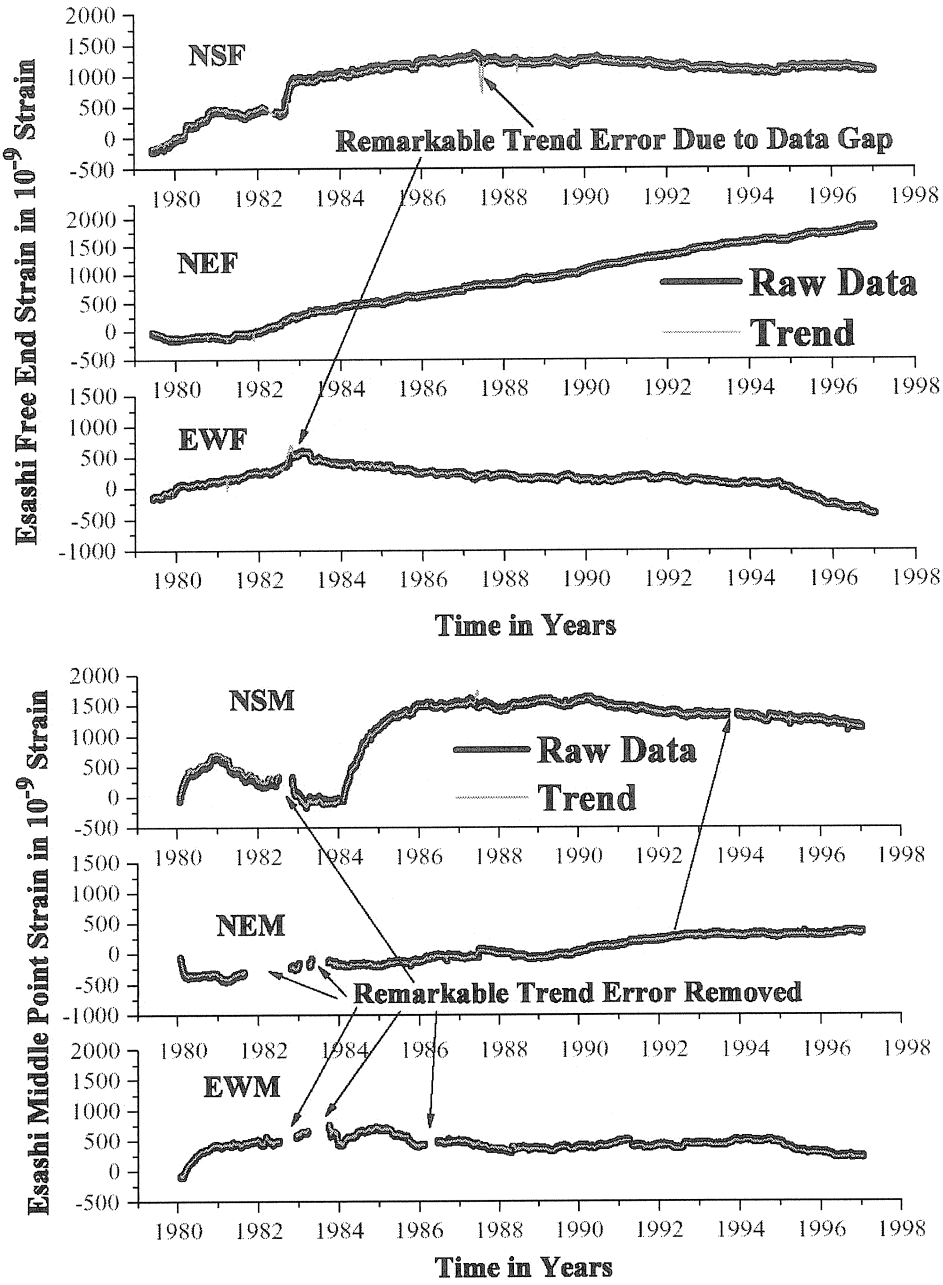


Figure 2, 17 years strain data obtained by Esashi extensometers, which were calibrated by introducing scale factor time series, and were corrected for the irregular steps. 3 free end and 3 middle point components. Thick line and thin line show the observed strain and the strain trend estimated by BAYTAP-G, respectively. Due to the large data gaps in strain observable, BAYTAP-G cannot handle the database perfectly, and introduces remarkable error into trend estimation. However, these errors can be simply removed or be replaced by more reasonable interpolating data.

3. Free core resonance from calibrated strain data

The long term tidal observation will give an independent way to estimate the parameters of fluid core resonance (FCR). Due to the slight misalignment between the rotation axes of mantle and outer core and due to the ellipticity of the core-mantle boundary (CMB) of the Earth, the Free Core Nutation (FCN) will be excited by the pressure torque acting through the CMB, due to the core-mantle coupling. When we observe the FCN on the Earth's surface, it shows a period

close to a sidereal day, therefore, the FCN is also called the nearly diurnal free wobble (NDFW). It is responsible of the resonant behavior of the Earth tides in the diurnal band. Precise observation of the FCR parameters gives an opportunity to retrieve the information on CMB related to the core-mantle coupling.

After the drifts and steps of strain components were being estimated, the steps were corrected to the original data. Finally the tidal admittances were estimated from the corrected strain data in one step using BAYTAP-G software. The precision of the result is improved by a factor of 5 or better, compared with the formal errors obtained by Sato(1989) only using 3 years data. The diurnal tidal constituents of O1, P1, K1, Psi1 and Phi1 are used to estimate the FCR parameters in this study. As an example, the observed tidal strain admittance for the NS middle point is given in Table 1.

Table 1. Tidal Admittance of NS Middle-Point Strain. Period: 1980.01.25.0 – 1997.01.21.0

SYMBOL	FACTOR (RMS)	PHASE (RMS)
O1	.27010 (.00029)	0.581° (.062°)
P1	.25669 (.00060)	3.432° (.134°)
K1	.18496 (.00019)	6.967° (.060°)
Psi1	.89915 (.02479)	13.830° (1.579°)
Phi1	.40144 (.01398)	3.286° (1.995°)

Table2.1 FCR parameters estimated from observed and corrected tidal admittances

Input Data	Correction	PFCR(s.d.)	PFCR_err	1/Q	1/Q_err
Free_end	none	429.387	4.66	3.139×10^{-4}	6.22×10^{-5}
Free_end	local	417.318	1.31	1.053×10^{-4}	1.35×10^{-5}
Free_end	local&ocean	421.732	1.6	9.499×10^{-5}	2.12×10^{-5}
Mid_Point	none	418.273	14.4	1.725×10^{-4}	2.86×10^{-4}
Mid_Point	local	404.544	17.5	2.835×10^{-4}	1.43×10^{-4}
Mid_Point	local&ocean	410.628	1.13	1.456×10^{-4}	1.77×10^{-5}

Table2.2 FCR parameters estimated from different combinations of global tidal strain with local and ocean corrections.

Combination Case	PFCR(s.d.)	PFCR_err	Q	Q_err
1.Free_end_only	421.732	1.6	10500	2350
2.Mid_Point_only	410.628	1.13	6870	830
3.Free_end+Mid_Point	419.867	1.31	6670	770

Two corrections were applied to the observed strain admittances. One is the local correction, which include the following four effects: namely, the topography, the cavity, the regional and local geologies. Another is the ocean loading correction. The transform matrices estimated by Sato (1989) were adopted to get homogeneous strain from the observed strain, however, the possible error of the matrices was not taken into account in this work. The ocean loading effects were estimated based on NAO'99b ocean tide model, which had been obtained from

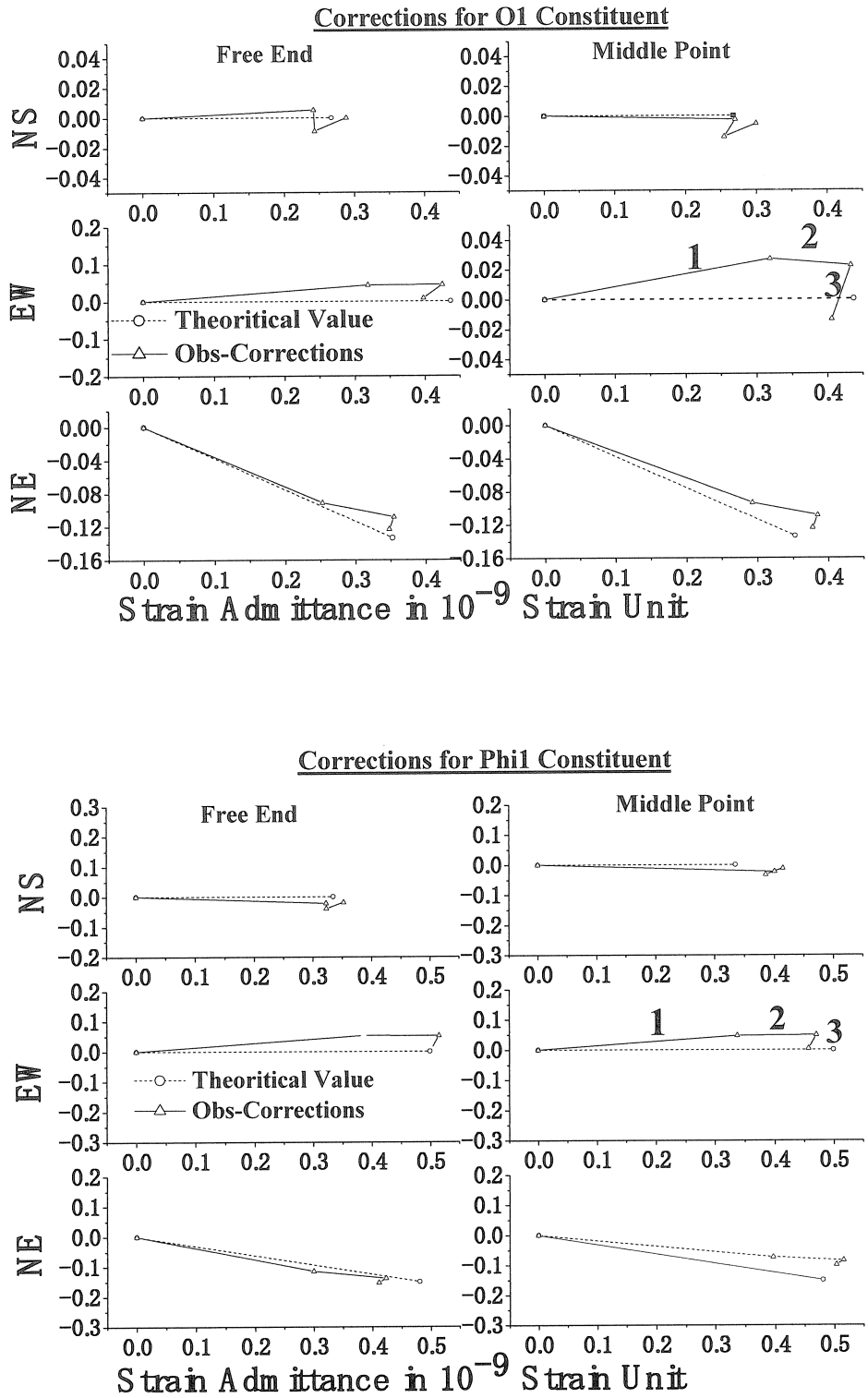


Figure 3. Phasor plots of the observation and theoretical body tides of O1 and Phi1 constituents. The dash line is the theoretical body tide. The solid lines 1,2,3 indicate the observed strains, local effects and ocean loadings respectively.

TOPEX/POSEIDON altimeter data (Matsumoto *et al.*, 2000) and Green's function for 1066A Earth model by using a modified GOTIC2 software (Matsumoto 2001). The strain corrected for these effects is consistent with that corresponding to the strain due to the theoretical body tidal strain. The corrections for O1 and Phi1 constituent of all 6 components are shown in Figure 3 in the form of the phasor plot. The discrepancy is remarkably reduced between the observed tide and the theoretical body tide by correcting for the local effects and the oceanic loading effect. The theoretical admittances of the body tide are calculated by using the Love numbers for the 1066A Earth mode (Wahr, 1981) with a FCR period obtained from VLBI method (Defraigne, *et al.,* 1994&1995).

Based on the Eqs. 6 & 7 in Sato (*et al.*,1994), the admittance of **O1** is used to normalize tidal admittances for other waves. From these normalized body tidal admittances, the SALS (Nakawaga, 1984) code with so called "Marquardt method" for nonlinear least-squares fitting is used to estimate the complex resonance strength parameters **Bs**, FCR period (**PFCR**) and quality factor **Q**. Here PFCR and Q are assumed as the common or global parameters for all strain components adopted in the fitting. After finding a set of closest initial value for the parameters to be solved by using a relatively complicated procedure with artificial constrains, all of the parameters were estimated without any constraint conditions for all parameters. The results for FCR parameters estimated from the observed tidal admittances and from corrected tidal admittances are listed in Table 2.1.

Because the observation is obtained at the same station, local effects may introduce large biases into the data analysis. This can be noticed by the large estimation errors for both of **PFCR** and **1/Q** in the table in the case of using observed admittances without any correction. Also, due to the ocean loading effects, **PFCR** will be shortened, and **Q** will be reduced for the global FCR parameters. These phenomena can also be seen in Table 2.1. For the final results, corrected for local effects and ocean loading, FCR parameters are estimated from 3 kinds of combinations of input data sets, calibrated free end data only, calibrated middle point data only, and the combination of them. For the first two cases, the global FCR parameters are set independently and separately. For the third case, the FCR parameters are set the same for both free end and middle point data. The results are given in Table2.2. A shorter FCR period and a relatively larger quality factor are obtained. Among them, result in case 3 is more reasonable than in cases 1 & 2, because the complex resonance strength parameters **Bs** are usually different for each strain component, where the global parameter FCR period **PFCR** and quality factor **Q** are homogeneous ones, and will be the same for all strain components.

4. Discussions

Long period tidal data, i.e. strain, tilt and gravity, obtained by Esashi Earth Tide Station may contribute to geophysical research in many directions. In this paper, we calibrated 17 years strain data of the three horizontal components for both free end and middle point transducers of

the extensometers, estimated the strain rate at EW direction for Esashi station, and then estimated the FCR parameters as a first report of the research work on long period tidal measurements at Esashi station.

FCR parameters have been obtained from different kinds of observations. Using super conducting gravity data of GGP, Sato (2004) got a new result of 429.66+/-1.43s.d. & 9,350-10,835; from VLBI observation, Defraigne, (*et al.,* 1994&1995) got a result of 433.9 +/-0.5 & 40,000 for PFCR and Q, respectively. Till now, all of the estimated eigenperiods are significantly shorter than the theoretical value, i.e. 460s.d. given by Wahr (1981). It can be explained by a departure from the hydrostatic flattening at the core-mantle boundary. Our result is the shortest among them. This may be partly explained by the possible error in the local transform matrices, and leave this problem as an open topic for the future work. And/or there is a possibility that the small number of the tidal waves used in the fitting is responsible to the obtained short FCR period. This will be checked, and the data analysis will be renewed by introduce more diurnal tidal constituents in number into the fitting procedure. However, as an independent average result from 17 years stable and sensitive strain data, it is precise enough to compare with others. The shorter period of FCR may indicate a larger bias or a departure of the figure of CMB that is expected from hydrostatic theory.

References

Defraigne, P., V. Dehant and Jhinderer; 'Stacking gravity tide measurements and nutation observations in order to determine the complex eigenfrequecy of the nearly diurnal free wobble'*J. Geopys. Res.*, 99, No.B5, 9203-9213, 1994.

Defraigne, P., V. Dehant and J. Hinderer; 'Correction to "Stacking gravity tide measurements and nutation observations inorder to determine the complex eigenfrequency of the nearly diurnal free wobble" by P. Defraigen, V. Dehant and J. Hinderer' *J. Geopys. Res.*, 100, No.B2, 2041-2042, 1995.

Heki, K., Space geodetic observation of deep basal subduction erosion in the Northeastern Japan, *Earth Planet. Sci. Lett.*, in press, 2004.

Matsumoto, K., T. Takanezawa and M. Ooe; 'Ocean Tide Models Developed by Assimilating TOPEX/POSEIDON Altimeter Data into Hydrodaynamical Model: A Global Model and a Regional Model around Japan', *J. Oceanography,* 56, 567-581, 2000.

Matsumoto, K., and Sato, T.; 'GOTICe: A Program for Computation of Ocean Tidal Loading Effect', *J. Geod. Soc. Japan*, 48, No.1, 243-248, 2001.

Mazzotti, S., X. Le Pichon, P. Henry and S. Miyazaki; 'Full interseismic locking of the Nankai and Japan- west Kurile subduction zones: An analysis of uniform elastic strain accumulation in Japan constrained by permanent GPS', *J. Geopys. Res.*, 105, 13159-13177, 2000.

Nakawaga, T. and Y. Oyanagi; '*Analysis of experimental data by means of least squares method, Program SALS*', in Japanese, Tokyo Univ. Press. 1984

Sato, T., M. Ooe and N. Sato; 'Observations of the Tidal Strain at the Esashi Earth Tides Station', *J. Geod. Soc. Japan*, 26, 35-49, 1980.

Sato, T.; 'Fluid core resonance measured by quartz tube extensometers at the Esashi Earth Tides station', in *Proc. 11th Int. Sympos. On Earth Tieds*, Helsinki, 573-582 1989.

Sato, T., Y. Tamura, T. Higashi, S. Takemoto, I. Nakagawa, N. Morimoto, Y. Fukuda, J. Segawa, and N. Seama; 'Resonance Parameters of the Free Core Nutation Measured from Three Superconducting Gravimeters in Japan', *J. Geomag. Geoelectr.*, 46, 571-586, 1994.

Sato, T., Y. Tamura, K. Matsumoto, Y. Imanishi, and H. McQueen, 'Parameters of the fluid core resonance estimated from superconducting gravimeter data', *BIM* 138, in publish, 2004.

Sheng-Tu, B. & Holts, W. E. 'Interseismic horizontal deformation in northern Honshu and its relationship with the subduction of the Pacific plate in the Hapan trench', *Geophys, Res. Lett.* 22, 3541-3544, 1995.

Tamura, Y., T.Sato, M. Ooe, and M. Ishiguro; 'A procedure for tidal analysis with a Bayesian information criterion', *Geophys. J.*

Int., 104, 507-516,1991.

Tsubokawa, T. and K. Asari; 'Designs of Displacement Meter Used Differential Transformer', *Proceedings of the International Latitude Observatory of Mizusawa*, 18, 94-123, 1979.

Wahr, J.M.; 'Body tides of an elliptical, rotating, elastic and oceanless Earth', *Geophys. J.R. ast. Soc.*, 64, 677-703, 1981.

Application of the Folding-Averaging Algorithm for the Determination of the Periods of the Earth's Free Oscillation Using Superconducting Gravimeter Data

J. Y. Guo[a,b], H. Greiner-Mai[b], O. Dierks[b], L. Ballani[b], J. Neumeyer[b], C. K. Shum[c]

[a]The Key Laboratory of Geospace Environment and Geodesy, Ministry of Education, School of Geodesy and Geomatics, Wuhan University, 129 Luoyu Road, 430079 Wuhan, China. Email: junyiguo@public.wh.hb.cn

[b]GeoForschungsZentrum Potsdam, Section 1.3, Telegrafenberg A17, D-14473 Potsdam, Germany

[c]Laboratory for Space Geodesy and Remote Sensing, The Ohio State University, 2070 Neil Avenue, Columbus, Ohio 43210-1275, USA

Abstract

A stacking method, which is referred as folding-averaging algorithm (FAA) in this paper, was frequently used for evaluating discrete Fourier transform (DFT) before the fast Fourier transform (FFT) technique was conceived. In this paper, we reformulate the FAA to precisely determine periods of signals which may be present in a time series. The basic principle of the FAA is to rebuild for every test period a new short time series by cutting the original time series to shorter ones of which the length is equal to the test period (at the end of the time series, a small fraction shorter than the test period may be discarded), and then stacking the short time series by averaging. In this stacking process of averaging, the amplitude of the possible signal with a period equal to the test period remains the same, but signals of different periods are averaged out and the random error is reduced. Amplitude and phase of the possible signal with a period equal to the test period can then be estimated using the averaged short time series. By searching for the maximum extremes of the amplitude by varying the test period, the periods of the signals which may be present in the time series can be very precisely determined. The FAA is distinct from DFT as follows: in FAA, periods of possible real physical signals in the time series are sought; but in DFT, sinusoidal functions with prescribed periods which are submultiples of the length of the time series are used to represent the time series exactly. The usefulness of the FAA is illustrated by applying it to determine the periods of the Earth's free oscillations using superconducting gravimeter (SG) data after the Peruvian Earthquake of magnitude 8.4 in 2001.

keywords: Time series; Folding-averaging algorithm; Period determination; Earth's free oscillation; superconducting gravimeter observation

1 Introduction

Retrieving periodical signals buried in a time series has been a topics heavily investigated in many branches of science, notably geophysics. Undoubtedly, the most well known method used nowadays in time series analysis is the DFT, or called FFT as a fast evaluation version. Having an inverse transform, the DFT gives exact representations of a time series using sinusoidal functions of which the periods are submultiples of the length of the time series. But in real physical problems, the period of a signal depends on its physical cause other than the length of observation, and hence, is not necessarily equal to a certain submultiple of the length of the time series obtained from observation. From this point of view, the DFT is not really relevant in determining periods of real physical signals in time series. A remedy to this weakness is to fit the Fourier spectrum around a peak with a resonance function (Bolt and Brillinger, 1979; Dahlen, 1982; Masters and Gilbert, 1983).

Unlike DFT, as can be seen from the process of constructing the averaged short time series stated in the abstract, the FAA is designed for searching for periods of real physical signals in time series other than giving exact representations of time series using prescribed periodical functions. For example, assume that a sinusoidal signal with a period of 0.985 hour is present in a time series of 100 hours. When using DFT to analyze this time series, the periods of sinusoidal waves we obtain nearest to 0.985 are $100/102 \approx 0.98$ and $100/101 \approx 0.99$. But when using the FAA for searching for periods of signals, a test period may be chosen as close to 0.985 as possible by increasing the density of test periods in the search. The precision of the period finally found depends on the sampling interval as well as the error of observation. Problems of precision will be discussed in the next section together with the FAA it self. Here we only point out that, if there is only one sinusoidal signal in a time series and there is no observation error, the precision of the period obtained for the signal is about $2(T/L_{\mathrm{TS}})\Delta t$, where T is the period of the sinusoidal wave, L_{TS} is the length of the time series, and Δt is the sampling interval. For the above numerical example, if the sampling

interval is 0.1 hour, the precision of the period obtained according to this criteria should be about 0.002 hour that is more accurate than that given by the DFT. See the next section for more discussion on the comparison between the FAA and the DFT and FFT.

The FAA was frequently used for evaluating DFT before the invention of FFT (e.g. Bartels, 1935). It was also used for studying periodical phenomena of which the periods are not submultiples of the time series, such as for example, variation of geomagnetic filed (Bartels, 1935; Pollak, 1930) and tides (Darwin's method of tidal analysis, see Melchior (1978) for example). As the FFT become so popular soon after its invention, the FAA is no longer much emphasized in modern literatures since then. Recently, a slight variant of it was also used to detect nonharmonic periodicities in biology (the linear stacking method of Hoenen el al. (2001)). In this work we propose to use FAA for precisely determining periods of harmonic signals by intensive search, since periods are required to be determined as precisely as possible in various problems, such as for example, the free oscillation of the Earth. Illustrative example of numerical computation is made for determining some periods of the Earth's free oscillation using a time series of gravity observed by the SG of GeoForschungsZentrum Potsdam (GFZ) installed in Sutherland, South Africa after the 2001 Peruvian Earthquake. A review on the study of the Earth's free oscillation using the worldwide network of SGs in the frame of Global Geodynamics Project (GGP) was given recently by Widmer-Schnidrig (2003).

In fact, the intensive search of periods based on the FAA requires a lot of computation. However, this is not a problem nowadays due to the advance in digital computers.

Another method which is closely related to the FAA is the phasor-walkout method, also known as graphical Fourier transform, summation dial, complex demodulation (Bartels, 1935; Bolt and Brillinger, 1979; Zürn and Rydelek, 1994). This method is particularly suitable for testing if a certain periodicity exists in a time series (Bolt and Brillinger, 1979; Zürn and Rydelek, 1994), though it can be used to search for periods precisely as well. The FAA can be considered as a variant of the phasor-walkout method for searching for periodicities with some approximation in amplitude and phase in favor of quick evaluation. More explanation on their relation will be given in next section.

Based on our analysis and numerical test, we recommend the FAA, which has a fairly simple algorithm, as an alternative method, among other methods being used, for example, the autoregressive method (e.g. Chao and Gilbert, 1980) and its variant, the Sompi method (e.g. Hori et al., 1989), the method of fitting a resonance function to the Fourier spectrum (Bolt and Brillinger, 1979; Dahlen, 1982; Masters and Gilbert, 1983), the interpolated FFT (IFFT) method, the iterative phase average (IWPA) method and the ESPRIT method (e.g. Santamaría et al., 2000) etc, for retrieving periodical signals from time series, and determining their periods with high accuracy. It is expected that it would find additional applications in geophysics.

2 The folding-averaging algorithm

For identifying possible sinusoidal signals in a time series, an amplitude spectrum is to be built. The basic principle is to estimate, for every one of an array of preassumed test periods, the amplitude and phase using the FAA. This section is divided into two subsections. The first one explains how to estimate the amplitude and phase of a signal with known period. The second one explains how to build the amplitude spectrum and accurately estimate the periods of the possible signals.

We make two assumptions on the time series: (1) the length of the times series is at least as long as tens or hundreds of the periods of the signals to be studied, (2) the sampling interval is at least as short as a tenth of the periods of the signals to be studied. These assumptions are fulfilled in numerous cases in contemporary geophysical researches.

2.1 Estimation of amplitude and phase of a signal with known period

Assume that a signal with known period T is present in the time series being analyzed and we are estimating the amplitude and phase of the signal.

First of all we analyze an ideal case that the sampling interval Δt can divide the period T exactly, i.e., the number of observation data in every period, $(T/\Delta t)$, is integral. We denote with $V_0, V_1, V_2, \cdots$ the data in the original time series sampled at equally spaced time $t = 0, \Delta t, 2\Delta t, \cdots$. Denote with M the number of short time series, each of them having $N_s = (T/\Delta t)$ data (the length of the test period), that can be cut from the original time series, discarding a small fraction shorter than the test period at the end of the original time series if exists. Arrange the short time series row by row as shown in Table 1. The averaged short time series $\bar{R}_0$, $\bar{R}_1$, $\cdots$ $\bar{R}_{N_s-1}$ is then obtained by averaging every column of the table,

$$\bar{R}_j = \frac{1}{M} \sum_{k=0}^{M-1} V_{k(T/\Delta t)+j} \,, \tag{1}$$

as the values of the signal in all elements of the same column are the same.

Table 1: Short time series row by row

V_0	V_1	$\cdots$	V_{N_s-1}
$V_{(T/\Delta t)}$	$V_{(T/\Delta t)+1}$	$\cdots$	$V_{(T/\Delta t)+N_s-1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$V_{(M-1)(T/\Delta t)}$	$V_{(M-1)(T/\Delta t)+1}$	$\cdots$	$V_{(M-1)(T/\Delta t)+N_s-1}$

We remark that (1) may be modified to take into account missing data or gaps in the time series by averaging only the data present (excluding the missing data in the sum, and replacing M by M minus the number of missing data).

In Table 1, we have not written the subscript of the first data in the short times series, $k(T/\Delta t)$, $k = 0, 1, \cdots, M-1$, as kN_s, because they will be different in the case when $T/\Delta t$ is not an integral number discussed later.

As the length of the times series is assumed to be much longer than the period of the signal, M is very large. Thus, in the averaging process, signals with different periods other than T is practically averaged out. (In fact, signals of which the periods are submultiples of T remain in the averaged short time series. This problem will be discussed later. Now we simply assume such case does not appear.)

The random error in the averaged short time series is much less than that in the original time series. Denote the root mean square error of the observations in the original time series with σ. The root mean square error of the average values in the short time series is then

$$m = \sigma/\sqrt{M-1} \tag{2}$$

By assumption, M is very large. If we have $M = 100$, the signal to noise ratio of the averaged short time series is theoretically about 10 times of the original time series. If we have $M = 10000$, this value raises to 100.

As signals with different periods are averaged out, and the signal to noise ratio is drastically enhanced, the averaged short time series should be in fact almost a sinusoidal curve, as indicated by numerical examples in the next section.

The amplitude and phase can be estimated using the averaged short time series, of which the signal to noise ratio is assumed to be raised to reasonable level. Express the signal as

$$s(t) = a\sin[(2\pi/T)t + \phi] \tag{3}$$

which is the same form for the original as well as the averaged short time series. The basic relations for determining amplitude and phase are

$$\int_0^T s(t) \left\{ \begin{array}{c} \sin \\ \cos \end{array} \right\} [(2\pi/T)t]dt = \frac{T}{2} a \left\{ \begin{array}{c} \cos \\ \sin \end{array} \right\} \phi\,. \tag{4}$$

As the sampling interval is assumed much shorter than the period of the signal, we can evaluate the integral in the above equation numerically using the left Riemann sum (which is identical to the Trapezoid sums due to the periodicity), replacing the discrete values of s, $s(k\Delta t)$, with $\bar{R}_k$, to obtain estimates for $a\cos\phi$ and $a\sin\phi$:

$$a \left\{ \begin{array}{c} \cos \\ \sin \end{array} \right\} \phi = \frac{2}{T}\Delta t \sum_{k=0}^{N_s-1} \bar{R}_k \left\{ \begin{array}{c} \sin \\ \cos \end{array} \right\} [(2\pi/T)(k\Delta t)]\,. \tag{5}$$

Estimates of amplitude and phase can then be computed according to

$$a = \sqrt{(a\sin\phi)^2 + (a\cos\phi)^2}\,, \quad \phi = \mathrm{atan2}(a\sin\phi, a\cos\phi) \tag{6}$$

where the function $\mathrm{atan2}(x, y)$ is provided in practically all programming languages.

In the above formulation, we assumed that the sampling interval Δt divides the period T exactly, i.e., the number of observation data in every period, $(T/\Delta t)$, is integral. But in practice, this rarely happens. As a result, we abandon this assumption. In the more general situation, the expressions $N_s = T/\Delta t$ and $k(T/\Delta t)$, $k = 0, \cdots, M-1$, in the subscripts of data in the short time series listed in Table 1 and equation (1) are no longer integral. Like Darwin's method of tidal analysis (e.g. Melchior, 1978), we replace $N_s = T/\Delta t$ and $k(T/\Delta t)$ with the integers closest to them, and still build the averaged short time series according to equation (1). We see that the values of the signal in the elements of the same column in Table 1 are no longer the same, but shifted in phase. Fortunately, the phase shifts between the corresponding elements of any two rows (each row is a short time series) are the same. Hence, in the averaged short time series, this phase shift does not influence the period, but only influences the estimates of amplitude and phase. As we assumed that the sampling interval is much smaller than the period of the signal, the phase shift is small, and thus its influences on the estimates of amplitude and phase are also small. The level of this influence will be formulated in the next subsection.

Finally, we discuss the relation between the FAA and the phasor walkout method. In the phasor walkout method, the expression $\sum_{k=0}^{N-1} V_k \exp\{-i(2\pi/T)k\Delta t\}$ is evaluated graphically for any test period T, where N is the total

number of data in the original time series, and $i = \sqrt{-1}$ (e.g. Zürn and Rydelek, 1994). It can be readily seen that the same expression is evaluated in the FAA outlined above (apart from a common multiplier and the approximation due to the small phase shifts in the FAA), but expressed in terms of amplitude and phase (e.g. Bartels, 1935).

2.2 Amplitude spectrum and accurate determination of periods

Now we assume that sinusoidal signals exist in a time series, but the periods are not known, and the study is intended to identify the periods.

We begin by analyzing the signature of a sinusoidal signal, of which the exact value of period T_E is not know, in the averaged short time series built with a test period T . We denote with ΔT the difference between T and T_E, so that

$$T = T_E + \Delta T\,. \tag{7}$$

The formula of the signal is the same as (3), but with T replaced by T_E. For clarity, we rewrite it out:

$$s(t) = a\sin[(2\pi/T_E)t + \phi]\,. \tag{8}$$

Evidently, the signature of this signal in the averaged value $\bar{R}_j$ given by equation (1) is

$$\bar{s}_j = \frac{1}{M}\sum_{k=0}^{M-1} a\sin\{(2\pi/T_E)[k(T/\Delta t) + j]\Delta t + \phi\} \tag{9}$$

which can be further written as

$$\bar{s}_j = \frac{1}{M}\sum_{k=0}^{M-1} a\sin[(2\pi/T_E)(j\Delta t + k\Delta T) + \phi]\,. \tag{10}$$

We see that $\bar{s}_j$ is the average of the values of the sinusoidal signal at M nodes equally spaced by ΔT: $j\Delta t, j\Delta t + \Delta T, \cdots, j\Delta t + (M-1)\Delta T$. When ΔT is very small as compared to T_E, it can also be understood as the average value of the signal in the interval between the epochs $j\Delta t$ and $j\Delta t + M\Delta T$. There are two possibilities for the result of (10).

1. According to (10), the values of the signal at all of the M nodes are equal to its value at $t = j\Delta t$ when $\Delta T = 0, T_E, 2T_E, \cdots$. Hence, the value of $\bar{s}_j$ is also equal to the value of the signal at $t = j\Delta t$. So, if a signal with a period T_E is present in the time series, and if we have built the average short time series for test periods $T = T_E, 2T_E, \cdots$, all the averaged short time series contains the signal. Reversely, this means that, in the averaged short time series built with a test period T, there may be a signals with periods $T, T/2, T/3, \cdots$. But when we use (5) and (6) to estimate amplitude and phase, the results obtained are just of those of the signal with period T, the signals with periods $T/2, T/3, \cdots$ have no contribution to the results, a direct consequence of the orthogonality of the trigonometric functions.

2. Otherwise, $\bar{s}_j \to 0$ when $M \to \infty$. This result is a consequence of the fact that the average of a sinusoidal signal in any time interval equal to its period is equal to zero. In practice, M is always a finite number. and hence $\bar{s}_j$ can not really vanish. According to what was stated below (10), we see that $\bar{s}_j$ does vanish when $M\Delta T = T_E, 2T_E, \cdots$. So in the spectrum built, small sidelobes appear around the highest central peak representing the signal, the larger is M, the smaller the sidelobes.

Based on the above properties, we can design a process for searching for periods of possible signals: estimating the amplitude a according to (5) and (6) for an array of test period T: $T_0, T_1 = T0 + dT_0, T_2 = T_1 + dT_1, \cdots$, the periods of possible signals are the one which have maximum extreme values of a. But we have to answer another question: what are the favorite values of $dT_0, dT_1, \cdots$ in the computation? If the values of $dT_0, dT_1, \cdots$ are too large, we may not see the maximum extreme of the estimate of a at all. The choice of $dT_0, dT_1, \cdots$ depends on the accuracy we expect. Take reference to (10) again. If T_E is in a interval $[T_k, T_{k+1}]$, for the estimate of a at T_k or T_{k+1} to be significantly distinguishable as nearly maximum extreme, $\Delta T = T_{k+1} - T_E$ or $\Delta T = T_k - T_E$ must be small enough in magnitude. We see that the absolute values of either $\Delta T = T_{k+1} - T_E$ or $\Delta T = T_k - T_E$ should be smaller than $dT_k/2$. So, in the search of signals, we choose $MdT_k/2 \geq M\Delta T$ to be only a small portion of T_k (or T_{k+1}). The smaller this value is, the clearer the maximum extreme we see. We chose $MdT_k/2$ to be only one n-th of T_k, whence

$$dT_k = 2T_k/(nM)\,. \tag{11}$$

This implies that $\bar{s}_j$ given in (10) is the average of the value of the signal at M equally spaced nodes in a interval of which the length is at maximum one n-th of the test period (see the analysis below equation (10)). The practical computation for the search of signals and accurate determination of periods is done in two steps:

1. Build a less accurate amplitude spectrum by choosing n to be moderately large, for example, between 4 to 16. Peaks in the spectrum may represent signals;

2. Pick out the peaks and search intensively for the maximum extremes of the amplitude near the peaks by setting n to its maximum meaningful value to be discussed below.

For any signal with a period T_E, we can not expect that $T_E/\Delta t$ be integral. Thus the approximation by phase shifting as mentioned in the last subsection should always be assumed. So, making the interval $M\Delta T \leq MdT_k/2$ smaller than the the possible unavoidable phase shift is meaningless. The phase shifts may be as large as $\Delta t/2$ to both the left and right sides. So the smallest meaningful value for $MdT_k/2$ is Δt. This implies that the maximum theoretical resolution for period is $dT_k = 2\Delta t/M$, which is approximately $2(T_k/L_{\rm TS})\Delta t$ as mentioned in the introduction, as M is approximately $L_{\rm TS}/T_k$. The value for n is then $T_k/\Delta t$. This ideal resolution may be achieved only under the ideal situation that only one signal is present in the times series and there is no error of observation. This is certainly not the case in practice. Unfortunately, we can not give an adequate estimate of error in the periods found using this method, just the same as the phasor walkout method (Zürn and Rydelek, 1994). The issue of precision will be discussed later. The interval of period in which signals are sought is chosen based on our a-priori knowledge on the problem studied.

The estimates of phase and amplitude using the FAA are not those of the original time series, but of the averaged short time series. In this paragraph we study the differences between the phase and amplitude of the original time series and the averaged short time series. These differences represent the minimum errors of the estimates using the FAA. We know from the statement after (10) that the value of the averaged short time series at epoch $j\Delta t$, $\bar{s}_j$, can be understood as the average of the real signal between the epochs $j\Delta t$ and $j\Delta t + M\Delta T$, where ΔT is the difference between the test period T and the real period of the signal T_E defined in (7). Now we assume that $\bar{s}_j$ is the value of a sinusoidal curve $\bar{s}(t)$ at the epoch $t = j\Delta t$. According to the FAA, $\bar{s}(t)$ is the approximation of the signal $s(t)$, and $\bar{s}(t)$ can be understood as the average of $s(t)$ in the interval between the epochs t and $t + M\Delta T$. According to this relation, we can analyze the signature of the phase and amplitude of $s(t)$ in $\bar{s}(t)$. We assume $s(t)$ attains its maximum extreme a at $t = t_m$, i.e. $(2\pi/T_E)t_m + \phi = \pi/2$. It can then be seen that $\bar{s}(t)$ attains its maximum extreme at the epoch $t = t_m - M\Delta T/2$, since at this epoch $\bar{s}(t)$ is the average of $s(t)$ in the interval between the epochs $t_m - M\Delta T/2$ and $t_m + M\Delta T/2$ containing t_m as midpoint. So the phase of $\bar{s}(t)$, $\bar{\phi}$, is $(2\pi/T_E)(M\Delta T/2) = \pi M\Delta T/T_E$ in advance (as the maximum arrives earlier) compared to that of $s(t)$, i.e.

$$\bar{\phi} = \phi + \pi M\Delta T/T_E\,. \tag{12}$$

The amplitude of $\bar{s}(t)$, $\bar{a}$, is the maximum extreme of $\bar{s}(t)$ which can be obtained to be

$$\bar{a} = [(aT_E)/(\pi M\Delta T)]\sin[(2\pi/T_E)(M\Delta T/2)]\,. \tag{13}$$

These relations of phase and amplitude can be used to estimate errors of phase and amplitude of the signals obtained using the FAA.

Now we turn to analyze the errors of amplitude and phase of a real signal in the FAA spectrum. Imagine that T_E falls in between the test periods T_k and T_{k+1}. The estimate of period is then T_k or T_{k+1} depending on to which one T_E is closer, and $\Delta T = T_k - T_E$ at T_k, $\Delta T = T_{k+1} - T_E$ at T_{k+1}. The phase and amplitude of $\bar{s}(t)$ are, according to (12) and (13),

$$\bar{\phi}_k = \phi + \pi M(T_k - T_E)/T_E\,, \tag{14}$$

$$\bar{a}_k = \frac{aT_E}{\pi M(T_k - T_E)}\sin\left\{\left(\frac{2\pi}{T_E}\right)\left[\frac{M(T_k - T_E)}{2}\right]\right\} \tag{15}$$

at T_k, and

$$\bar{\phi}_{k+1} = \phi + \pi M(T_{k+1} - T_E)/T_E\,, \tag{16}$$

$$\bar{a}_{k+1} = \frac{aT_E}{\pi M(T_{k+1} - T_E)}\sin\left\{\left(\frac{2\pi}{T_E}\right)\left[\frac{M(T_{k+1} - T_E)}{2}\right]\right\} \tag{17}$$

at T_{k+1}. Noticing $T_k \leq T_E \leq T_{k+1}$, wee see that the estimate of phase is in backward at T_k, and in advance at T_{k+1} as compared to that of the signal $s(t)$. Consider the worst situation that T_E is at the midpoint between T_k and T_{k+1}, i.e. $T_k - T_E = -dT_k/2$, $T_{k+1} - T_E = dT_k/2$. Then, by using (11), we see that the phase and amplitude are

$$\bar{\phi}_k = \phi - (2\pi/T_E)[T_k/(2n)]\,, \tag{18}$$

$$\bar{a}_k = \{(aT_E)/[\pi T_k/n]\}\sin\{(2\pi/T_E)[T_k/(2n)]\} \tag{19}$$

at T_k, and

$$\bar{\phi}_{k+1} = \phi + (2\pi/T_E)[T_k/(2n)]\,, \tag{20}$$

$$\bar{a}_{k+1} = \{(aT_E)/[\pi T_k/n]\} \sin\{(2\pi/T_E)[T_k/(2n)]\} \tag{21}$$

at T_{k+1}. In (18) to (21), setting $T_k = T_E$ gives good approximations. So we have

$$\bar{\phi}_k = \phi - \pi/n\,, \tag{22}$$

$$\bar{a}_k = (na/\pi)\sin(\pi/n) \tag{23}$$

at T_k, and

$$\bar{\phi}_{k+1} = \phi + \pi/n\,, \tag{24}$$

$$\bar{a}_{k+1} = (na/\pi)\sin(\pi/n) \tag{25}$$

at T_{k+1}. These relations give the differences between a and $\bar{a}$, and ϕ and $\bar{\phi}$, which represent the highest precision in the estimates using FAA for any chosen value of n. This precision may be attained only at the extremely ideal situation: only one signal is present in the time series, and the observation is error free, just like the situation for attaining the maximum precision for period. One can easily estimate the precision when highest resolution in period is made by setting $n = T_k/\Delta t$.

In practice, the time series may be more complicated. For example, it may contain linear tendency which perturbs the estimate of a and makes it unclear as maximum extreme at the period T_E. In the averaged short time series, residuals form signals of longer periods not cancelled by averaging are also similar to linear tendency, or even like a quadratic curve. So, it is preferable that signals outside the interval of interest be filtered out before using the FAA.

For having more accurate estimates of a and ϕ, or for having error estimates for them, instead of using (5) and (6), a least square fit of the averaged short time series by a sinusoidal curve with the period found may be used. This may be done either by using the cosiner algorithm (Nelson et al., 1979) or by linearizing the problem using the estimates of a and ϕ according to (5) and (6) as initial guess. A linear, and even a quadratic curve, may be combined with the sinusoidal curve for reducing the errors of estimates. Nevertheless, the error estimates represent the misfit to the averaged short time series, but not the original time series, as already discussed.

The same as the phasor walkout method (Zürn and Rydelek, 1994), we can not give adequate estimates of errors in the periods found using this method, which is already mentioned earlier in the text. In this work we propose to use an indirect method to infer errors in the estimates of periods. First, we do the analysis for the time series. Second, according to the periods, amplitudes and phase of the signals found, we add into the time series some synthetic signals which are similar to the signals found in the first step, but with periods slightly shifted for not overlapping with the signals. And third, redo the analysis for the time series containing the synthetic signals. For the synthetic signals, as the exact values of their periods, amplitudes and phases are known, the absolute errors of their estimates may be determined. These absolute errors of the estimates of the synthetic signals can then serve as a good reference of errors for the estimates of the real signals.

2.3 Relation with the DFT and the FFT with zero padding

First we discuss the relation between the FAA and the DFT. In the DFT, the coefficients of the complex fourier series of the time series are $(1/N)\sum_{j=0}^{N-1} V_j \exp\{-i(2\pi/T_k)j\Delta t\}$, where N is the total number of data in the time series, $i = \sqrt{-1}$ and $T_k = L_{\rm TS}/k$. This is the same as the FAA if $T_k = L_{\rm TS}/k$ is chosen as the test period (notice that we used the amplitudes and phases of the sine functions in the FAA). So, if we chose $\cdots, L_{\rm TS}/3, L_{\rm TS}/2, L_{\rm TS}$, i.e. the periods of the sinusoidal functions used in DFT, as test periods for the FAA, the FAA spectrum is identical to the DFT spectrum (apart from the approximation due to the small phase shifts in the FAA). So their difference lies on the differences of nodes of periods of the spectrums built using the two methods respectively. The nodes of periods of the FAA spectrum was given by (11) which may be made as dense as one want by choosing the value of n large enough, without exceeding the limit of maximum resolution as discussed before. For the DFT spectrum, the nodes of periods are submultiples of the length of the time series. Hence, at any node $T = L_{\rm TS}/k$, the step to the next node is

$$dT = L_{\rm TS}/(k-1) - L_{\rm TS}/k = T/(k-1)\,. \tag{26}$$

Notice that the meaning of k is similar to that of M in (11), i.e. the number of periods in the times series. So, when k or M is large, and when n in (11) is set to 2 for the FAA, the distances between two nodes for the DFT spectrum is the same order of magnitude as that of the FAA spectrum at comparable nodal periods in the two spectra. However, Even when n in (11) is set to 2 for the FAA, the nodal periods of the DFT and FAA spectrums are not necessarily equal because the choice of test periods for the FAA is not unique like DFT. Thus we conclude that, when the value of n in (11) is set to 2, the FAA gives a spectrum with practically the same resolution of period as the DFT. The precision of phase and amplitude of the FAA discussed above also applies to DFT by setting $n = 2$. So, according to (22) to (25), we can also conclude that, the largest possible errors in the estimates of phase and amplitude in DFT may attain $\pi/2$

and $[(\pi-2)/\pi]a = 0.36a$ even if the observation is error free. This represent in fact the DFT precision in phase and amplitude, while $dT/2 = T/[2(k-1)]$ represents the DFT precision in period.

Since its invention, the FFT has being used extensively to evaluate the DFT. Here, for illustrative purpose, we assume without loosing generality that the FFT requires the total number of data in a time series to be power of 2 (though this is not a prerequisite of the FFT). If the criteria is not satisfied, zeros are customarily added by the end of the time series to the length required, called zero padding. Now we continue to use N to denote the total number of data in the original times series, and to use N' to denote the total number of data (that is power of 2) after zero padding, i.e. $N'-N$ zeros are added to end of the original time series. In the FFT, the coefficients of the complex fourier series of the zero-padded time series are $(1/N')\sum_{j=0}^{N'-1} V_j \exp\{-i(2\pi/T_k)j\Delta t\}$, where $T_k = L'_{\rm TS}/k$, $L'_{\rm TS}$ being the length of the zero-padded time series. Remembering that V_j is equal to zero when $N \le j < N'$, we can readily see that, if we chose $\cdots, L'_{\rm TS}/3, L'_{\rm TS}/2, L'_{\rm TS}$, i.e. the periods of the sine functions used in the FFT, as test periods for the FAA, the FAA spectrum is identical to the FFT spectrum multiplied with the factor N'/N (apart from the approximation due to the small phase shifts in the FAA). So, the same as the FAA, the FFT can also be used to build spectrums with high resolution of period of signals by padding the time series with a lot of zeros (Santamaría et al., 2000). But the FAA is superior for accurate determination of the periods of the signals by extensive search near the peaks of the spectrums built (the second step of accurate determination of periods stated in the last subsection), since achieving the highest resolution that the FAA can attain using the FFT requires a huge amount of computation (though possible).

Due to the above relation of the FAA to the DFT and FFT, we can apply tapers to the FAA in the same way as to the DFT and FFT.

3 Application for the determination of the periods of the Earth's free oscillation using SG data

In this section we apply the FAA to estimate some periods of the Earth's free oscillation using SG data. The objective is to demonstrate the applicability of the FAA. The detail of the subject itself on the study of the Earth's free oscillations using the worldwide network of SGs in the frame of GGP is referred to the recent review by Widmer-Schnidrig (2003).

A property of the Earth's free oscillations, the decay, was not considered in the last section in building the FAA. A decaying sinusoidal signal can be written as (Zürn and Rydelek, 1994)

$$s(t) = a\exp\{-[\pi/(QT)]t\}\sin[(2\pi/T)t+\phi] \tag{27}$$

where Q is the quality factor. It is straightforward to redo the formulation in the last section while replacing (3) with this signal. The main characteristics of the conclusion may be inferred by inspection. As the decaying factors of the signal of period T in all the short time series in Table 1 are the same, they can be safely averaged, obtaining as result an averaged short time series which contains mainly this signal. But the amplitude of the signal in the averaged short time series is the average all over the original time series. We have also attempted to recover the quality factor by fitting the averaged short time series with the decaying signal. This seams impossible because the difference in amplitude in one cycle is too small as the quality factor Q is large in general. There are spectral analysis methods specifically conceived for retrieving decaying signals like the Earth's free oscillation (e.g. Bolt and Brillinger, 1979; Chao and Gilbert, 1980; Dahlen, 1982; Masters and Gilbert, 1983; Lindberg and Park, 1987; Park et al., 1987; Hori et al., 1989). The FAA is understood as an independent method for period determination. For determining periods of decaying signals, we cannot use too long time series, since, when the amplitude of the signal become too small, using more data implies adding more noise. Dahlen (1982) recommended that the length of the time series be Q cycles of the signal to be recovered.

The main Peru Earthquake of magnitude 8.4 occurred on 20:33:14 UTC, 23 June, 2001. Aftershocks of magnitudes 6.7, 6.6 and 7.6 occurred on 04:18:31 UTC, 26 June, 13:53:48 UTC, 5 July and 09:38:43 UTC, 7 July, respectively. The gravity time series is sampled at 5 seconds interval, which is first filtered using the least squares band pass filter of TSoft provided by the Royal Observatory of Belgium (http://www.observatoire.be/SEISMO/TSOFT/tsoft.html), to retain signals only between 0.15 to 2 mHz. Then the air pressure effect is corrected using a frequency dependent method (e.g. Neumeyer, 1995), as this effect may be important for the low frequency or long period free oscillation band (Van Camp, 1999; Zürn and Widmer, 1995). Therefore the 5 sec air pressure data is filtered the same way, and a transfer function between the air pressure and gravity data is calculated using blocks of length 12 hours. This transfer function is multiplied to the Fourier spectrum of the air pressure data which is then subtracted from the gravity data Fourier spectrum. The corrected gravity data are finally received using the inverse Fourier transformation. We do see significant improvement of signal to noise ratio under 1.3 mHz. The time series finally used in this study after filtering and air pressure correction is from 08:00:00 UTC on June 24 to 02:00:00 UTC on July 4, its length being 234 hours. Based on a visual inspection of the spectra built using the data of various time spans, the noise level has the lowest noise level.

Table 2: Synthetic signals added into the time series. The will serve as a tool to estimate the precision of the estimates of periods by comparing the assigned and estimated values of them.

Mode	Period s	Initial amplitude nm s^{-2}	Quality factor	Initial phase
${}_0W_0$	1303.000	0.0047	5700	1.0
${}_1W_0$	596.000	0.0026	1800	2.0
${}_0W_2^{-2}$	2942.000	0.0024	500	2.0
${}_0W_2^{-1}$	2902.000	0.0013	500	1.0
${}_0W_2^{0}$	2864.000	0.0002	500	0.0
${}_0W_2^{1}$	2826.000	0.0030	500	-1.0
${}_0W_2^{2}$	2792.000	0.0018	500	-2.0

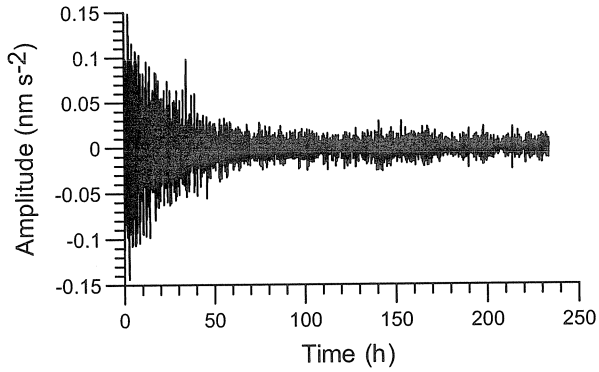


Figure 1: The Sutherland SG gravity time series after the 2001 peruvian earthquake after bandpass filtering and atmospheric pressure correction used in the study. The initial time corresponds to 08:00:00 UTC on June 24, and the end time corresponds to 02:00:00 UTC on July 4. The length is 234 hours. Based on a visual inspection of the spectra built using the data of various time spans, the noise level has the lowest noise level. The synthetic signals listed in Table 2 have been added to the gravity time series.

We use ${}_0S_0$, ${}_1S_0$ and ${}_0S_2$ to test the method. The modes ${}_0S_0$ and ${}_1S_0$ are the simplest because each of them has only one singlet in the spectrum. The mode ${}_0S_2$ is split into five singlets ${}_0S_2^m$, $m \in \{-2,-1,0,1,2\}$.

As proposed in the last section, the precision of estimates of periods will be inferred by comparing the exact and estimated values of periods of similar synthetic signals added into the time series. The synthetic signal corresponding to a real signal is chosen as similar as possible to the real signal in property, except a small difference in periods, necessary for distinguishing them in the spectrum. For conformance with the damping nature of free oscillation, the quality factor Q is also considered in each of these synthetic signals, though we don't estimate Q using FAA. The synthetic signals corresponding to ${}_0S_0$, ${}_1S_0$ and ${}_0S_2$ are ${}_0W_0$, ${}_1W_0$ and ${}_0W_2$, where ${}_0W_2$ is also split into five singlets like ${}_0S_2$, ${}_0W_2^m$, $m \in \{-2,-1,0,1,2\}$. We chose the periods and quality factors of ${}_0W_0$ and ${}_1W_0$ respectively to be close to those of ${}_0S_0$ and ${}_1S_0$ given by Riedesel el al. (1980). The choice of ${}_0W_2^m$, $m \in \{-2,-1,0,1,2\}$ is more sophisticated, as the property of splitting of ${}_0S_2$ should be retained. We chose their periods by shifting the frequencies of ${}_0S_2^m$, $m \in \{-2,-1,0,1,2\}$ given by Rosat et al. (2003a) by about the same amount, and their quality factors to be close to that of ${}_0S_2$ given by Dziewonski and Anderson (1981). For all the synthetic signals, the initial amplitudes are chosen to be close to those of the corresponding real signals determined by analyzing the time series without the synthetic signals, and the phases, arbitrarily. The parameters chosen for the synthetic signals are listed in Table 2. The graph of the time series with synthetic signals added is shown in Figure 1.

As mentioned in the last section, we search for periods in two steps. Firstly, we build an amplitude spectrum between 0.2 and 2 mHz by setting $n = 16$. Secondly, we pick out the peaks from the spectrum and search more accurately for the periods around these peaks by setting $n = T_k/\Delta t$, representing maximum possible resolution. Here we add a third step: estimate the amplitudes and phases using least squares fits to the averaged short time series built using the periods determined in the second step. The amplitude spectrum built at first step is shown in Figure 2. The splitting of ${}_0S_2$ and ${}_0W_2$ is shown in Figure 3, which is the amplification of a portion of Figure 2. As illustration, we also present the graphs of the averaged short time series of ${}_0S_0$ and ${}_1S_0$ in Figure 4, which are indeed quite similar to sinusoidal curves. The results from the second and third steps are listed in Table 3. As the quality factor is not estimated in our approach, the amplitudes we obtain represent only the average, thus not comparable with the initial values as given in Table 2. Nevertheless, we still listed them in the table. More digits are given for comparison between

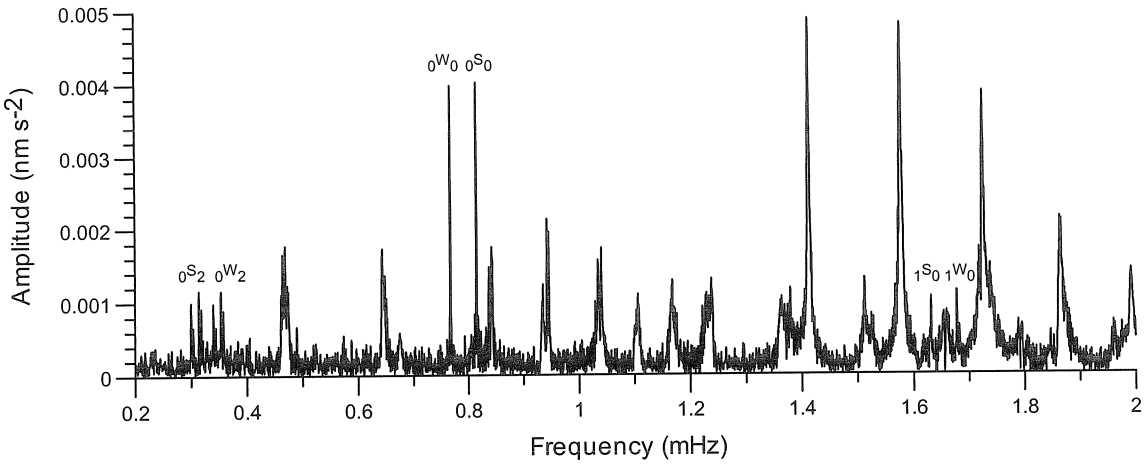


Figure 2: The amplitude spectrum of the gravity time series shown in Figure 1 built using the FAA. The peaks ${}_0S_0$, ${}_1S_0$ and ${}_0S_2$ are the free oscillation modes studied as examples in this work. The peaks ${}_0W_0$, ${}_1W_0$ and ${}_0W_2$ are the corresponding synthetic signals added to the time series for assessing precision of the estimates of periods by comparing the assigned and estimated values of them.

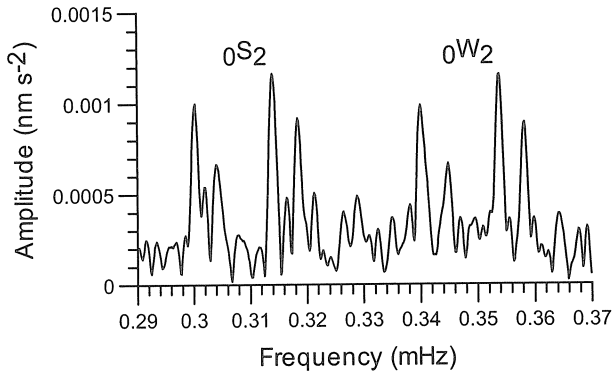


Figure 3: The splitting of ${}_0S_2$ (left) and ${}_0W_2$ (right). This is a portion of the Figure 2 enlarged for showing the detail in this band.

the estimates of the parameters and errors. The results in the table will be discussed in more detail in the following two paragraphs.

First, we discuss the results for ${}_0S_0$ and ${}_1S_0$. Levels of error in the estimates of their periods are inferred from comparing the assigned and estimated periods of the synthetic signals ${}_0W_0$ and ${}_1W_0$ listed in Tables 2 and 3, which are lower than 50 ppm for both of them. As the maximum resolution of periods of the FAA as discussed in the last section is far more accurate, this level of error is the result of the overall influence of all other signals and random errors in the time series. For the Earth's free oscillation, the estimates of the period of a mode obtained using different data should be the same. Here we compare our results with those of Riedesel el al. (1980) who stacked 9 IDA records, and estimated the period of ${}_0S_0$ to be 1227.500 ± 0.005 (or ± 4 ppm) seconds using a time series of 2000 hours, and the period of ${}_1S_0$ to be 612.929 ± 0.018 (or ± 30 ppm) seconds using a time series of 300 hours. We see that our results are in close agreement with those of Riedesel el al. (1980).

In this paragraph, we discuss the results of the 5 singlets of the mode ${}_0S_2$. For both ${}_0S_2$ and ${}_0W_2$, the singlets corresponding to the values of the azimuthal order number $m = -2, -1, 0, 1, 2$ are from left to right in Figure 2. We see that the $m = 0$ singlets of them cannot be clearly seen. Rosat et al. (2003a) gave in their Figure 5 the same graph for ${}_0S_2$ obtained using the data of the Strasbourg SG after the same Earthquake, where the singlet of $m = 0$ can be clearly seen. But when we compare the spectrums from the Sutherland and Strasbourg SG data in their Figure 4, we see that the singlet of $m = 0$ for the Sutherland instrument can neither be clearly seen. In fact, the amplitude of the singlets of ${}_0S_2$ with azimuthal number $m = -2, -1, 0, 1, 2$ depend on the amplitude of $P_2^m(\cos\theta)$ (θ is the colatitude) that is respectively 0.18, 0.23, -0.07, -1.36, 2.14 at Sutherland, and 0.11, -0.25, 0.34, 1.49, 1.31 at Strasbourg. We see that the magnitude of $P_2^0(\cos\theta)$ at Sutherland is only one fifth of that at Strasbourg. So we attribute this low

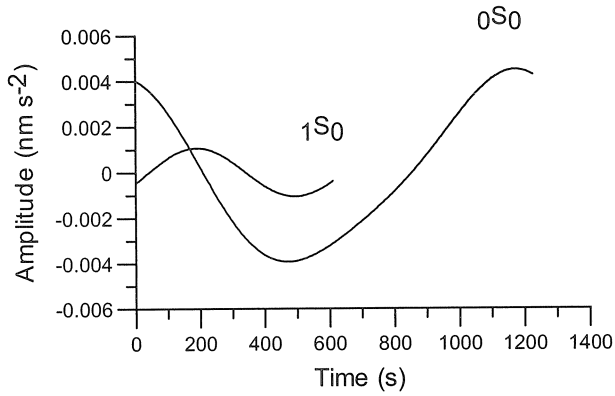


Figure 4: The averaged short time series of ${}_0S_0$ and ${}_1S_0$ corresponding to the maximum amplitude extremes. They are used for estimating amplitudes and phases. As the modes studied decay, only their average amplitudes can be estimated.

Table 3: Estimates of parameters of the synthetic and real signals. The differences between the amplitudes and phases of the synthetic signals ${}_0W_0$, ${}_1W_0$ and ${}_0W_2$ in this table and their corresponding exact values listed in Table 2 serve as error estimates for the free oscillation modes ${}_0S_0$, ${}_1S_0$ and ${}_0S_2$. The amplitudes are not comparable because those listed in Table 2 are the initial values, and those in this table is the average ones. The singlets ${}_0W_2^0$ and ${}_0S_2^0$ are too small in amplitudes, and are not estimated.

Mode	Period s	Amplitude nm s^{-2}	RMS error nm s^{-2}	Phase	RMS error
${}_0W_0$	1303.063	0.003996	0.000027	1.0788	0.0068
${}_1W_0$	596.018	0.001144	0.000003	2.0012	0.0027
${}_0S_0$	1227.509	0.004042	0.000025	2.0167	0.0063
${}_1S_0$	613.010	0.001055	0.000003	-0.3793	0.0029
${}_0W_2^{-2}$	2941.289	0.000984	0.000042	1.9352	0.0429
${}_0W_2^{-1}$	2899.545	0.000632	0.000173	0.6680	0.2744
${}_0W_2^0$	-	-	-	-	-
${}_0W_2^1$	2826.855	0.001134	0.000099	-0.7947	0.0873
${}_0W_2^2$	2791.475	0.000904	0.000026	-2.1804	0.0285
${}_0S_2^{-2}$	3330.855	0.001020	0.000018	-1.7162	0.0181
${}_0S_2^{-1}$	3289.067	0.000638	0.000026	2.0646	0.0404
${}_0S_2^0$	-	-	-	-	-
${}_0S_2^1$	3184.754	0.001121	0.000056	0.4112	0.0499
${}_0S_2^2$	3140.636	0.000922	0.000025	-2.5517	0.0272

Table 4: Comparison between the estimates of the periods of the ${}_0S_2$ singlets of Rosat et al. (2003b) obtained by fitting a resonance function to each Fourier spectral peak and those obtained using the FAA using the Strasbourg SG data (Unit: second). The first and second lines are respectively the estimates and their errors of Rosat et al. (2003b). The third line is our estimates. And the last line is the disagreement between the estimates of Rosat et al. (2003b) and the ours.

	${}_0S_2^{-2}$	${}_0S_2^{-1}$	${}_0S_2^0$	${}_0S_2^1$	${}_0S_2^2$
Rosat et al.	3334.91	3284.05	3235.72	3186.09	3143.57
	$\pm$0.59 (177 ppm)	$\pm$0.73 (222 ppm)	$\pm$0.60 (188 ppm)	$\pm$0.52 (163 ppm)	$\pm$0.63 (200 ppm)
Our results	3336.61	3284.65	3235.19	3185.28	3144.70
Disagreement	$\pm$1.70(509ppm)	$\pm$0.60(182ppm)	$\pm$0.53(164ppm)	$\pm$0.81(254ppm)	$\pm$1.13(360ppm)

amplitude of the $m = 0$ singlet of ${}_0S_2$ to the data of the Sutherland instrument, and is not studied. As ${}_0W_2$ was made as closer to ${}_0S_2$ as possible, it is natural that its $m = 0$ singlet is neither clear in the spectrum. Level of errors in the estimates of the periods of the ${}_0S_2$ singlets are estimated by comparing the assigned and estimated periods of the ${}_0W_2$ singlets listed in Tables 2 and 3. We see that the result of ${}_0W_2^{-1}$ has the largest error, which is 845 ppm. This is conceivable as this singlet has the smallest amplitude. The error level of other singlets are below 300 ppm. For comparison, we list in Table 4 the estimates of for the periods of the ${}_0S_2$ singlets obtained by Rosat et al. (2003b) using the Strasbourg SG data. Their disagreement from our estimate using the Sutherland SG data listed in Table 3 are 4.06 (1216ppm), 5.02 (1525ppm), 1.34 (419ppm), 2.93 (934ppm) for $m = -2, -1, 1, 2$ respectively. We see that these differences are large that seems difficult to accept. If we compare the ${}_0S_2$ spectrum of the Strasbourg SG data given in Figure 2 of Rosat et al. (2003b) with our Figure 2, we see that the Strasbourg spectrum is very clean, but the Sutherland spectrum is polluted by some lower peaks which may be local background noises, or hums. Perhaps, the periods of the Sutherland spectrum are biased by these hums with extremely close periods. For example, for the lowest peak, ${}_0S_2^{-1}$, which may be more affected by the hums, our result agrees with that of Rosat et al. (2003a) at 1525 ppm, but for the highest peak, ${}_0S_2^1$, which may be less affected by the hums, the agreement is as good as 419 ppm, quite close to our error estimate using synthetic signal.

As supplement of comparison, we have also applied in the same way the FAA to a 228 h-long record of the Strasbourg SG (from 12 O'clock of Jun 25, 2001 to 0 O'clock of July 5, 2001, this time span is chosen such that the noise is lowest in the ${}_0S_2$ band in the spectrum according to the visual examination). The results are also listed in Table 4. The last row of the table is the disagreements between our results and those of Rosat et al. (2003b), which should be acceptable as compared to the error estimates of Rosat et al. (2003b) since the data sets used are in fact not exactly the same.

As demonstration of the use of taper, we have applied a Hanning window to the same data set used by Rosat et al. (2003b) after least square band pass filtering to keep signals in the range 0.2-0.4 mHz (the band of the mode ${}_0S_2$), and then determined the periods using the FAA. The periods for the 5 singlets of the mode ${}_0S_2$ for $m = -2, -1, 0, 1, 2$ are respectively 3334.99, 3283.17, 3235.65, 3186.03, 3142.86 seconds, which are in general closer to the results of Rosat et al. (2003b) than those obtained in the last paragraph (listed in Table 1).

4 Concluding remarks

In this work, the FAA is applied to seek periodical signals in time series and to determine their periods accurately. Various aspects of the FAA are discussed, including the signal-to-noise ratio improvement in the averaged short time series, the highest possible accuracy of the estimates of the periods (for an ideal case), the signature of the real signal in the short time series, the errors in the estimates of amplitudes and phases caused by errors in the estimates of periods. The relations of the FAA with the DFT and the FFT with with zero padding are also investigated, showing that tapers can be used for the FAA in the same way as for the DFT and FFT.

A weakness of the FAA is that it does not provide with estimates of errors for the estimates of periods. However, we used an indirect method for assessing the errors of the periods which consists of adding into the time series synthetic signals that are quite similar to the real signals found, and then using the differences between the exact and estimated values of periods of these synthetic signals as a measure of errors for the estimates of periods for real signals.

To test the technique, we have applied it to numerous synthetic time series that consist of sinusoidal signals and random noises of different level. The results demonstrate the feasibility of the method.

For geophysical application, we used it to determine the periods of the Earth's free oscillations using a time series of gravity observed by the GFZ SG installed in Sutherland, South Africa after the 2001 Peru Earthquake. Comparison of our results with previous works are made using the modes ${}_0S_0$, ${}_1S_0$ and ${}_0S_2$. For all the synthetic signals corresponding the these modes added for accessing precision, our estimates of periods show very close agreements with the exact values assigned. Our estimates of periods for ${}_0S_0$ and ${}_1S_0$ are also in very good agreements with the very elaborated estimates of Riedesel el al. (1980). But for the mode ${}_0S_2$, the the agreement between our estimates of period and the recent estimates of Rosat et al. (2003b) we have chosen for comparison is less good, which may be due to the high noises very close to the frequency of ${}_0S_2$ in the Sutherland data. We have also applied the FAA to the Strasbourg SG data that Rosat et al. (2003b) used, and the results show very good agreement.

Based on our experiments, we conclude that the FAA as a valuable method, among other methods being used, for retrieving periodical signals from time series and determining their periods with high accuracy. It is expected to find other applications in geophysics.

Acknowledgment We gratefully acknowledge B. Ducarme, M. Van Camp and S. Rosat for helpful discussions. S. Rosat has also provided the preprocessed Strasbourg SG data. J.Y. Guo is supported by the Natural Science Foundation of China under grant 40274028.

References

Bartels, J., 1935. Random fluctuations, persistence and quasipersistence goephysical and cosmical periodicities. Terr. Magn. Atmos. Electricity, 40, 1-60.

Bolt, B. A., and Brinllinger, D. R., 1979. Estimation of uncertainties in eigenspectral estimates from decaying geophysical time series. Geophys. J. R. Astr. Soc., 59, 593-603.

Chao, B. F., and Gilbert, F., 1980. Autoregressive estimation of complex eigenfrequencies in low-frequency seismic spectra. Geophys. J. R. Astr. Soc., 63, 641-657.

Dahlen, F. A., 1982. The effect of data windows on the estimation of free oscillation parameters. Geophys. J. R. Astr. Soc., 69, 537-549.

Dziewonski, A. M., and Anderson, D. L., 1981. Preliminary reference Earth model, Phys. Earth planet. Inter., 25, 297-356.

Hoenen, S., Schimmel, M., and Marques, M. D., 2001. Rescuing rhythms from noise: A new metrhod of analysis. Biological Rhythm Research, 32, 271-284.

Hori, S., Fukao, Y., Kumazawa, M., Furomoto, M., and Yamamoto, A., 1989. A new mothod of spectral analysis and its application to the Earth's free oscillations: The Sompi method. J. geophys. Res., 94(B6), 7535-7553.

Masters, G., and Gilbert, F., 1983. Attenuation in the Earth at low frequencies. Phil. Trans. Roy. Soc. Land., Ser. A, 308, 479-522.

Melchior, P., 1978. Tides of the planet Earth. Pergamon Press.

Nelson, W., Tong, Y. L., Lee, J. K., and Halberg, F., 1979. Methods for cosinor-rhythmometry. Chronobiologia 6(4), 305-324.

Neumeyer, J., 1995. Frequency dependent atmospheric pressure correction on gravity variations by means of cross spectral analysis. Bull. Inf. Marrées Terr., 122, 9212-9220.

Park, J., Lindberg, C. R., and Thomson, D. J., 1987. Multiple-taper spectral analysis of terestrial free oscillations: part 1. Geophys. J. R. Astr. Soc., 91, 755-794.

Lindberg, C. R., and Park, J., 1987. Multiple-taper spectral analysis of terestrial free oscillations: part 2. Geophys. J. R. Astr. Soc., 91, 795-836.

Pollak, L. W., 1930. Prager Geophysikalische Studien. Heft 3 (Čechoslovak. Statistik, Reihe 12, Heft 13), Prague.

Riedesel, M. A., Agnew, D. C., Berger, C., and Gilbert, F., 1980. Stacking for the frequencies and Q's of ${}_0S_0$ and ${}_1S_0$. Geophys. J. R. Astr. Soc., 62, 457-471,

Rosat, S., Hinderer, J., Crossley, D., and Rivera, L., 2003a. The search for the Slichter mode: comparison of noise levels of superconducting gravimeters and investigation of a stacking method. Phys. Earth planet. Inter., 140, 183-202.

Rosat, S., Hinderer, J., and Rivera, L., 2003b. First observation of ${}_2S_1$ and study of the splitting of the football mode ${}_0S_2$ after the June 2001 Peru earthquake of magnitude 8.4. Geophys. Res. Lett., 30(21), 2111, doi:10.1029/2003GL018304.

Santamaría, I., Pantaleón, C., and Ibañez, J., 2000. A comparative study of high-accuracy frequency estimation methods. Mechanical Systems and Signal Processing, 14(5), 819-834.

Van Camp, M., 1999. Measuring seismic normal modes with the GWR C021 superconducting gravimeter. Phys. Earth planet. Inter., 116, 81-92.

Widmer-Schnidrig, R., 2003. What can superconducting gravimeters contributte to normal-mode seismology? Bull. seism. Soc. Am., 93, 1370-1380.

Zürn, W., and Rydelek, P. A., 1994. Revisiting the phasor-walkout method for detailed investigation of harmonic signals in time series. Survey in Geophys. (15), 409-431.

Zürn, W., and Widmer, R., 1995. On noise reduction in vertical seismic records below 2 mHz using local barometric pressure. Geophys. Res. Lett., 22(24), 3537-3540.

INSTITUTO DE ASTRONOMÍA Y GEODESIA (C.S.I.C. – U.C.M.)

INTERNATIONAL SEMINARY ON THE APPLICATIONS OF THE COMPUTER PROGRAM VAV – VERSION 2003 FOR TIDAL DATA PROCESSING MADRID, 21 – 24 OCTOBER 2003.

PROF. ANGEL VENEDIKOV

Geophysical Institute, Bulgarian Academy of Sciences

and

PROF. RICARDO VIEIRA

Instituto de Astronomía y Geodesia (C.S.I.C. – U.C.M.), ESPAÑA

GUIDEBOOK FOR THE PRACTICAL USE OF THE COMPUTER PROGRAM VAV – VERSION 2003

The major part of this guidebook has been presented and tested at the INTERNATIONAL SEMINARY ON THE APPLICATIONS OF THE COMPUTER PROGRAM VAV – VERSION 2003 FOR TIDAL DATA PROCESSING, held in the Institute of Astronomy and Geodesy (CSIC – UCM) in Madrid, 21 – 24 October 2003. Afterwards we have introduced some improvements of the text, as well as a few new examples, e.g. application of VAV on ocean tidal data for the detection of the shallow water tides.

For practical use of this program the users can download from the ICET WEB site http://www.astro.oma.be/ICET:

* The program VAV, together with all materials, necessary for the execution of the examples of the guidebook, as well as for other applications of the program.

* A version of the guidebook in PDF format.

A CD-ROM with the same material can be sent on request. Please contact

Prof. B.ducarme, ICET Director, Ducarme@oma.be

We hope that the experience in the tidal data processing, incorporated in the program VAV and this guidebook will be helpful to the tidal community for investigation and practical work in both solid Earth and ocean tidal domains.

Brussels, December 15, 2004

A. GENERAL REMARKS ABOUT THE GUIDEBOOK.

The guide has 7 initial sections, including this one, numerated by A through G. Section B provides a bit of the theory of the VAV program. The remaining sections C through G deal with general practical elements.

The idea of the guide is to explain the use of VAV by showing examples. In this sense, 21 sections with examples follow the general sections. They are numerated by **Example 1** through **Example 21**. The examples demonstrate the use of the options, which may be of most common interest. A list of the options used in the examples can be found in an appendix to the guide.

This guidebook does not deal with the theory of the methods and the algorithms, used by VAV. We have confined by a very brief consideration of the models of the tidal signal and the drift in Section B. The users, interested in a deeper understanding, we may use the papers, listed in the attached Bibliography, in particular the recent publications (Venedikov, Arnoso, Vieira, 2001, 2003), devoted to VAV. Many of the papers in the Bibliography are not referred in this text but they are implicitly used. The authors of VAV are, of course, at disposition for personal consultations.

The tables and figures in the text are numerated within the corresponding sections. E.g. Table D.2 is the 2nd table in Section D and Figure 17.3 is the 3rd figure in Example 17.

A variant to rapidly start using VAV is (i) learning from Section C how to install the necessary data, then (ii) go directly to Example 1.

B. GENERAL INFORMATION ABOUT THE VAV PROGRAM.

The main algorithm of VAV is tidal analysis according to the Method of the Least Squares (MLS). The application of MLS is based on **a model of the tidal signal**, which consists in the following.

Every tidal phenomena has a corresponding theoretical tidal signal whose expression $s(t)$ at time t is

$$s(t) = \sum_{\omega} h_{\omega} \cos\left(\varphi_{\omega}(t)\right) = \sum_{\omega} h_{\omega} \cos\left(\varphi_{\omega}(0) + \omega t\right) \qquad (1)$$

Here ω is frequency of a tide or tidal wave, taking m known discrete values $\omega = \omega_1, \omega_2, \ldots \omega_m$. In this expression ω is expressed in radians per unit of time. In practice, we deal with frequencies ω in deg/hr (degrees of arc per hour), in cpd (cycles per day) and in cph (cycles per hour).

VAV uses $s(t)$ provided by the development of Tamura (1987) of the tide-generating potential with $m = 1198$ tides with $\omega > 0$ and 2 constant terms with $\omega = 0$. A list of the tides is stored in the file **tamura.inp** (see Section C).

The quantities h_{ω} and $\varphi_{\omega}(t)$ in (1) are called theoretical amplitude and phase at time t of the tide with frequency ω. All parameters of (1), i.e. ω, h_{ω} and $\varphi_{\omega}(t)$ or $\varphi_{\omega}(0)$ are precisely known quantities.

In the observed data of the tidal phenomena we have an observed tidal signal whose expression $S(t)$ at time t is

$$S(t) = \sum_{\omega} H_{\omega} \cos \Phi_{\omega}(t) = \sum_{\omega} H_{\omega} \cos\left(\Phi_{\omega}(0) + \omega t\right) \qquad (2)$$

where ω takes the same known discrete values as in (1).

The quantities H_{ω} and $\Phi_{\omega}(t)$ are called observed amplitude and observed phase at time t. Unlike the theoretical h_{ω} and $\varphi_{\omega}(t)$, the observed H_{ω} and $\Phi_{\omega}(t)$ are unknown quantities, which are subject to estimation by the tidal analysis.

The relation between (2) and (1) can be described by using the so-called amplitude factor $\delta_{\omega} = H_{\omega} / h_{\omega}$ and phase shift or phase lag $k_{\omega} = \Phi_{\omega}(t) - \varphi_{\omega}(t) = \Phi_{\omega}(0) - \varphi_{\omega}(0)$ (denoted also by the Greek α). Namely, we have

$$S(t) = \mathrm{Re} \sum_{\omega} x(\omega) h_{\omega} \mathrm{Exp}\left(\mathrm{i}\varphi_{\omega}(t)\right) \text{ where } x(\omega) = \delta_{\omega} \mathrm{Exp}\left(\mathrm{i}\kappa_{\omega}\right) \qquad (3)$$

In the classical methods of analysis in the Earth tide domain the directly estimated unknowns used to be a set of the observed H_{ω} and $\Phi_{\omega}(0)$ (actually $H_{\omega} \cos \Phi_{\omega}(0)$ and $H_{\omega} \sin \Phi_{\omega}(0)$). This kind of unknowns does not allow a correct application of MLS, because it is impossible to take into account all tides. The obstacle is that the tides are concentration in frequency bands or tidal groups with very close ω, which cannot be separated. I.e., if we include all tides in the observational equations of MLS, they become linearly dependent. If we include a restricted number of main tides, the equations are not correct and we shall get biased estimates of H_{ω} and $\Phi_{\omega}(0)$ and incorrect estimates of the precision.

VAV uses the unknowns proposed and applied by Venedikov (1961, 1966). They are a set of values of the $x(\omega)$, more concretely a set of the quantities $\xi_{\omega} = \delta_{\omega} \cos \kappa_{\omega}$ and $\eta_{\omega} = -\delta_{\omega} \sin \kappa_{\omega}$. Under the assumption that $x(\omega)$ is a constant for groups of tides with very close ω we can create correct linearly independent equations, including the model (3) with all tides and a moderate number of unknown coefficients $x(\omega)$. Through the estimates of the unknowns, respectively of δ_{ω} and κ_{ω} we can get the estimates of all observed H_{ω} and $\Phi_{\omega}(0)$, not only of a small number of biased amplitudes and phases of some main tides.

The use of this kind of unknowns for the Earth tide data is based on the theory of the Earth deformation, as well as on an abundant experience.

It is easy to show that (3) is equivalent to the model of the ocean tidal signal, proposed by Munk and Cartwright (1966). The admittance function $x(\omega)$ in their model is a continuous function of ω, theoretically defined for all ω.

In our model $x(\omega)$ is a stepwise function, remaining constant for some short intervals of ω, covering a tidal group, which is not defined for intervals of ω, empty of tidal energy. We do not see a contradiction with the model of Munk and Cartwright which makes us believe that VAV can also be efficient for the ocean tidal data with a careful selection of the tidal groups.

VAV uses **an approximation of the drift** by independent polynomials in short time intervals of length ΔT hours. The coefficients of the polynomials, which are different in different intervals, are treated as unknowns by MLS. In our examples we shall use $\Delta T = 48$ and $\Delta T = 24$ but other values of the same order are also available.

The application of MLS uses a separation of the drift unknowns. This operation appeared to be equivalent to a filtration of the intervals, which separates the main tidal

frequencies and eliminates the drift. Through the filtration we get the data transformed from the time domain in the time/frequency domain, i.e. in a set of filtered numbers. In this sense we can consider ΔT as a time window, through which pass the data. Further MLS is applied namely in the time/frequency domain. This means that we create and solve the observational equations about the filtered numbers.

The use of the transformed or filtered data allows getting frequency dependent residuals and thus - frequency dependent estimates of the precision. In such a way VAV takes into account the colored character of the noise (red noise).

VAV allows the data to have gaps, jumps and perturbations of any kind between the intervals, as well as small gaps within the intervals. The principle to deal with the gaps is the most natural one: we create observational equations for the existing data and we do not create them for not existing data, as well as for perturbed data, which we want to ignore. This allows avoiding the interpolations and the reparations of doubtful data, operations, introducing anomalies and noise with very unpleasant properties.

An important task, followed by VAV is to study carefully the data. The purpose is, on the one side, to clean the data from anomalies and thus to get better analysis results. On the other side, we take into account, in particular for the Earth tide data, that every anomaly may be a geophysical signal.

In principle, the Earth tide observations, at the moment, may reveal with the highest possible sensibility and precision the slightest motions at the Earth surface. Hence, if any kind of deformation can serve as an earthquake and volcano precursor, it can be caught by our observations. One of the deplorable consequences of the interpolation/reparation of the data is that we cannot distinguish when an anomaly is a signal and when it is artificially introduced.

In the creation of VAV we have tried to include in one and the same program many different options. Nevertheless, the practical use of VAV is very simple, but the explanation of how to work is not so simple. Due to this in the following text we have chosen to give the explanations by using examples of the most commonly used applications of VAV. We hope that it is enough to start with a restricted volume of options, after which, through consultations with the authors of VAV, we can go further.

It may be encouraging for the users of VAV to know that this program is the product of the work of experienced specialists during many years. This work has started by the creation of the first method and computer program for tidal analysis (Venedikov, 1966), which has used successfully MLS by taking into account all tides, providing a frequency dependent estimation of the precision and dealing with data having gaps and arbitrary length.

C. INSTALLATION OF THE PROGRAM VAV AND DATA ORGANIZATION.

The CD attached to this guide contains 3 main folders:

\aaavav_03 - called executable folder
\a_data - folder with a data bank, used for the examples in this guide.
\aaafor_03 (connected to **\aaavav_03**) – the codes of all routines of VAV.

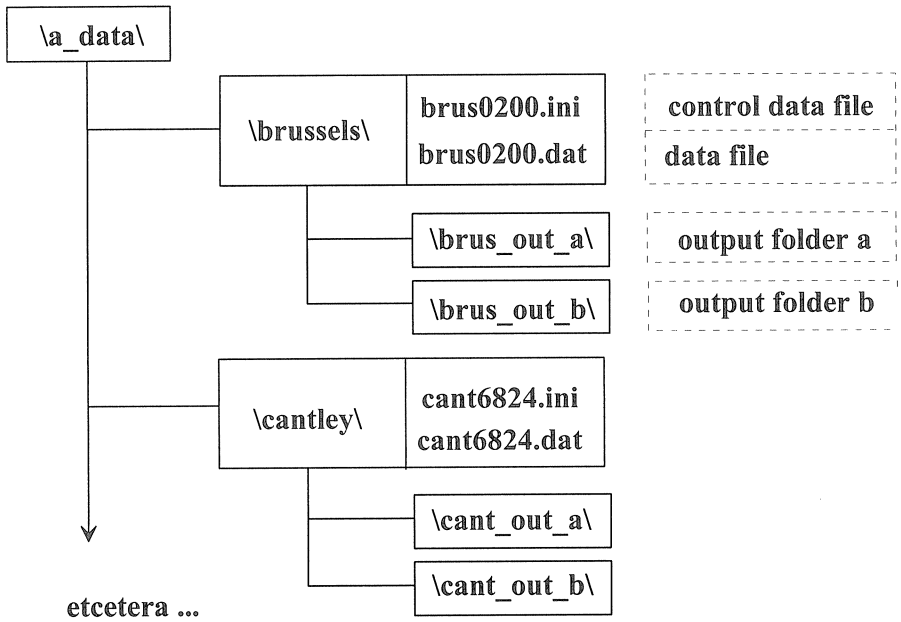
The folder **\aaavav_03** can be installed directly on the disk (C:) or in any folder in (C:). It is possible to change the name **\aaavav_03**.

The initial content of **\aaavav_03** is listed in Table C.1.

Table C.1. Content of the executable folder **\aaavav_03**

File name	Short comment
vav_03.exe	Executable (exec) file
a_list_files.inp	Description of the input data and output folders
all_ini_inp	Control data file, common for all series of data
groups.inp	Variants of tidal groups
tamura.inp	Development of tide generating potential of Tamura
cwords.inp	Input of the available names of options
out_txt.inp	Input of the descriptions of various output tables
Shallow.inp	Input, defining a set of shallow water tides

The folder **\a_data** should be copied directly in (C:). Its structure is shown by the following scheme.



Every series of data has its own folder. Within it are stored the data file and a corresponding control data file. To the folder are connected two output folders.

The user can change this scheme, e.g. several data files can be stored in one and the same folder. We use different pairs of output folders for every series, but it is possible to use one and the same pair of output folders for all series of data. We use one and the same names for the data and the control data files with different extensions: **.dat** and **.ini.** This is convenient, but not necessary.

Obligatory are only two things: (i) we must create at least two different output folders and (ii) the organization chosen and all names of folders and files should be correctly described in the file **a_list_files.inp**.

In Table C.2 we give the **a_list_files.inp,** corresponding to the scheme given above and the data, included in **\a_data**. We have four series of super conducting tidal gravity data (SG data) from the Global Geodynamics Project (Crossley, 2000, Ducarme and Vandercoilden, 2000) and a series of ocean tidal data from San Juan, kindly

provided to us by Dr. Harald Schmitz-Hübsch from Deutsches Geodätisches Forschungsinstitut in Munich. The data ere recorded by the University Hawaii Sea Level Center.

Table C.2. Content of **a_list_files.inp** with information about the names and the disposition of series of data, as well as of the output folders.

Tidal station	Content of **a_list_files.inp**	Comments
SG data Brussels, Belgium	**\a_data\brussels\brus0200.ini**	Control Data file
	\a_data\brussels\brus0200.dat	Data file
	\a_data\brussels\brus_out_a	Two output folders
	\a_data\brussels\brus_out_b	
SG data Cantley, Canada	**\a_data\cantley\cant6824.ini**	Control Data file
	\a_data\cantley\cant6824.dat	Data file
	\a_data\cantley\cant_out_a	Two output folders
	\a_data\cantley\cant_out_b	
SG data Strasbourg, France	**\a_data\strasbourg\stra0306.ini**	Control Data file
	\a_data\strasbourg\stra0306.dat	Data file
	\a_data\strasbourg\stra_out_a	Two output folders
	\a_data\strasbourg\stra_out_b	
SG data Vienna, Austria	**\a_data\vienna\vien0698.ini**	Control Data file
	\a_data\vienna\vien0698.dat	Data file
	\a_data\vienna\vien_out_a	Two output folders
	\a_data\vienna\vien_out_b	
Ocean data San Juan, Puerto Rico	**\a_data\sanjuan\sanjuan.ini**	Control Data file
	\a_data\sanjuan\sanjuan.dat	Data file
	\a_data\sanjuan\san_out_a	Two output folders
	\a_data\sanjuan\san_out_b	
	end	Terminates this list

D. CONTROL DATA FILES.

Every data file is accompanied by a control data file with some data, called here CDATA. The CDATA consist of what we call options. We shall explain the format of an option by the following example:

>F-T-window: 24 *comment: we choose a time window of 24 hours
comment is also every line, which has NOT ">" in the 1st column

Here **>F-T-window:** is a control word or a name of an option, "24" is an input and the text after the starlet * is an arbitrary comment. Comments are also all lines without ">" in 1st column, like the second line above, as well as any blank line.

This control word is a kind of an acronym of: "Filters, time window to be used".

Through this option we choose the length of the time window to be 24 hours. The time window is actually the length of our filters, which eliminate the drift.

We can arbitrarily use letters in upper and lower case, with or without the ":" at the end, e.g. **>f-T-wInDOw 24** . Many of the names of the option can be abbreviated, e.g. **>f-T-wInDOw 24** can be replaced by **>f-T-W: 24** or by **>f-T-W 24**

If **>F-T-window: 24** is not used, VAV will accept the default value of the time window, which is 48 hours.

The statement has not an effect if it is replaced by one of the following forms

>F-T-window:	Option, without any input
>F-T-window: *24	Input, preceded by a starlet *
***>F-T-window: 24**	Starlet *, preceding the option

A part of the CDATA for given series, e.g. the options defining the coordinates, is more conservative and it may be kept in all or almost all variants of the processing. We shall call it **"Basic CDATA"** (see Example 1, Table 1.1). The control data files in our **\a_data** are prepared with an initial content, which is namely the **Basic CDATA**.

For every particular processing we may add some "**New CDATA**" to the **Basic CDATA**. In the examples, which will be demonstrated, the **New CDATA**, prepared for a next example, should replace the **New CDATA** already used, while the **Basic CDATA** are generally kept. So that generally, the CDATA are used according to the following scheme:

All CDATA	**New CDATA**	To be replaced in a next variant of processing
	Basic CDATA	To be kept in almost all variants of processing

E. COMMON CONTROL DATA FILE "all_ini.inp"

The file **all_ini.inp** is situated in the executable folder **\aaavav_03**. It contains CDATA like the other control data files. The difference is that the effect of the CDATA is applied to all series we shall process. Hence, **all_ini.inp** is comfortable to be used when we have to deal with several series of one and the same type, on which one and the same options are applied. It is possible to leave **all_ini.inp** empty, without any information in it. We shall initially use the **all_ini.inp**, shown in Table E.1.

Table E.1. Content of **all_ini.inp**

Options	Comments
>D–origin-of-time: 1982 01 01 00	The origin of time $t = 0$, used in various output files and for the definition of the phases (see Section B) is chosen at the epoch 0^h,01.01.1982.
>OUT-compare-tide-name: Q1	In the files like **a_delta.dat** with results from various analysis of one and the same series and **alldelta.dat** with results from different series, the results about the tides Q1, K1, N2, M2 will be listed. It is possible to use till 26 tidal names, but a small number is more practical.
>OUT–compare-tide-name: K1	
>OUT–compare-tide-name: N2	
>OUT–compare-tide-name: M2	

F. DATA FORMAT.

The data format can be chosen by the option **>D-format:** The default case is the international format, accepted by ICET (International Centre for Earth Tides). Our SG data are prepared in this format.

The data are disposed in columns:

Date1	Date2	Channel 01	Channel 02	Channel 03	....

Here Date1 = year, month, day and Date2 = hour, minute, second of the data in the corresponding line, and channel 01, 02, ... are columns with data, e.g. tidal, air-pressure, temperature, etc. The dates should be in UT (universal time).

An inconvenience is that the data in this format cannot be directly used to apply standard plotting systems like GRAPHER. In order to avoid this inconvenience, we have an automatic output of the data in file **vav_out.dat** where we have 2 columns with time in days and hours, which allows using common ways to draw a graph of the data.

In our Basic CDATA we have included the option **>D-format: unf**. During the first application of VAV on given series of data this option should be excluded, e.g. by setting ***>D-format: unf**.

After the 1st run of VAV on given series, we get an unformatted copy of the data. E.g. for **cant6824.dat** we shall get the unformatted **cant6824.unf**. It is recommendable in all next runs to include again **>D-format: unf**, which accelerates considerably the data input.

An example of a different format is the data of San Juan. Another format, which can be used by VAV, is described in Example 20.

G. OUTPUT FILES AND FOLDERS.

After every run of VAV on given series we get the main analysis results in the **output folder a**, file **analysisNN.dat**. Here NN is a 2 digit number, starting by 99 for the first run, i.e. from the 1st run we get **analysis99.dat**. After the 2nd run we get **analysis98.dat**, by keeping **analysis99.dat**. Then, after the 3rd run we get **analysis97.dat**, etc. E.g., by applying 5 runs, we shall get the content of the **output folder a**, shown in Table G.1.

Table G.1. Content of **output folder a** after 5 runs.

Run 5	**Analysis95.dat**
Run 4	**Analysis96.dat**
Run 3	**Analysis97.dat**
Run 2	**Analysis98.dat**
Run 1	**Analysis99.dat**

The process can continue till we get **analysis40.dat**. Then VAV interrupts the experiments, by asking all files **analysisNN.dat** to be deleted.

The main analysis results are at about the bottom of **analysisNN.dat** and they look like those in Table G.2 (without the number of the columns which are added here).

Table G.2. Sample of analysis results in a file **analysisNN.dat**.

tidal group frequency interval (cycles/day)		numb of tides	name of group	analysis results main observed amplit H	MSD of H	amplit. factor delta	MSD of delta	phase lag kappa	MSD of kappa
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
results at filter frequency 1.000 cpd with MSD of unit weight = 9.789									
0.00015 :	0.24995	279	MF	61.7920	3.0175	1.20399	0.05880	-3.307	2.857
0.72150 :	0.90631	143	Q1	67.0935	0.3020	1.15120	0.00518	-0.378	0.258
0.92194 :	0.94049	58	O1	351.0383	0.4174	1.15321	0.00137	0.094	0.068
0.95809 :	0.97419	48	NO1	27.9444	0.3465	1.16741	0.01448	-0.020	0.714
0.98905 :	1.01369	56	K1	489.2783	0.3015	1.14302	0.00070	0.078	0.035
1.02855 :	1.04480	40	J1	28.2464	0.3303	1.18006	0.01380	-0.767	0.667
1.06484 :	1.21640	105	OO1	15.2541	0.2860	1.16490	0.02184	-0.123	1.071
results at filter frequency 2.000 cpd with MSD of unit weight = 4.464									
1.71938 :	1.87214	99	2N2	8.8120	0.1022	1.15856	0.01343	5.266	0.672
1.88839 :	1.90646	50	N2	67.3205	0.1787	1.17121	0.00311	3.742	0.153
1.92377 :	1.94275	56	M2	355.7213	0.1882	1.18489	0.00063	2.866	0.030
1.95823 :	1.97693	39	L2	10.0183	0.0817	1.18059	0.00962	1.905	0.468
1.99179 :	2.18284	133	S2	167.3661	0.1643	1.19824	0.00118	0.992	0.057
results at filter frequency 3.000 cpd with MSD of unit weight = 3.075									
2.75324 :	3.08125	81	M3	3.9825	0.1152	1.06735	0.03088	-1.339	1.659
3.00000 :	3.00000	1	S3	0.3358	0.1161	76.3731	26.4121	62.514	19.823
results at filter frequency 4.000 cpd with MSD of unit weight = 2.715									
3.79196 :	3.93790	10	M4	0.1422	0.0968	3.31073	2.25433	249.553	38.756
4.00000 :	4.00000	1	S4	0.1257	0.1022				
results at filter frequency 5.000 cpd with MSD of unit weight = 2.539									
4.83068 :	4.83068	1	M5	0.2758	0.0955				
5.00000 :	5.00000	1	S5	0.1744	0.0955				
results at filter frequency 6.000 cpd with MSD of unit weight = 2.218									
5.79682 :	5.79682	1	M6	0.0885	0.0836				
6.00000 :	6.00000	1	S6	0.0826	0.0834				

Columns 1 through 4 in Table G.2 describe the used tidal groups. E.g. in the first line with numerical data we have a group named MF which includes 279 tides from the development of Tamura, whose frequencies ω are $0.00015 \text{ cpd} \le \omega \le 0.24995 \text{ cpd}$.

In many cases the name of a group coincides with the name of the main tide (highest amplitude) in the group as it is given in **tamura.inp**. This is the case in Table G.2. When variants with a higher number of groups are used, the name of the group may be conventional because it does not exist in our **tamura.inp**.

Columns 5 through 10 contain results from the analysis of the data about the quantities, defined in Section B.

Column 5 provides the estimated observed amplitude H of the main tide in the corresponding group, i.e. the tide with the highest amplitude.

Columns 7 & 9 provide the estimates of the parameters δ and κ, which are accepted to be one and the same for all tides in the corresponding group.

MSD in columns 6, 8 & 10, as well as further in this guide means: "mean square deviation" or "standard deviation" or "estimated standard", as well as "mean square

error". The MSD of a quantity x will be often denoted by $\sigma(x)$. By using this, we may say that columns 6, 8 & 10 contain $\sigma(H)$, $\sigma(\delta)$ and $\sigma(\kappa)$ respectively.

VAV uses a set of complex filters, amplifying selected basic frequencies Ω. In the example above we have used the default values $\Omega = 1, 2, 3, 4, 5$ & 6 cpd. The precision is estimated by the residuals of the filtered numbers and thus we get frequency dependent MSD of unit weight, i.e. MSD $\sigma(\Omega)$ (MSD at frequency Ω). It also computes a hypothetical MSD σ(White Noise) under the unrealistic assumption for white noise. These estimates can be found in **analysisNN.dat** in a table under the title

frequency dependent MSD (mean square deviations)

At the same place we can also find the Akaike Information Criterion AIC (Sakamoto et al., 1986) determined at every frequency (frequency dependent AIC) and a global value, corresponding to σ(White Noise).

Figure G.1 shows the MSD taken from this output. The graphics show that using the assumption for white noise will overestimate the precision at the lower frequencies and underestimate it at the higher frequencies.

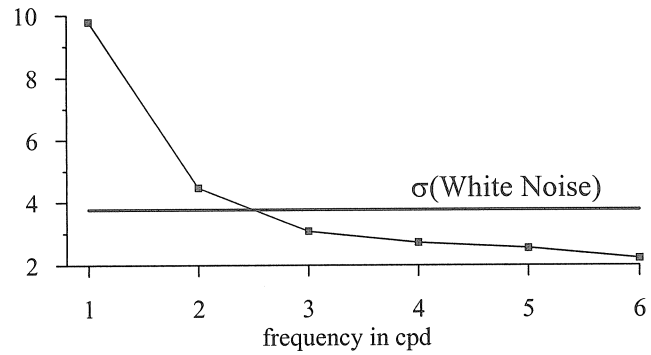


Figure G.1. Frequency dependent MSD $\sigma(\Omega)$ determined by VAV and MSD σ(Wite Noise) obtained under the false assumption for white noise.

Through the rows **results at filter frequency ...** the output in Table G.2 shows the relation between results and the filters used. E.g. the row

```
-------------------------------------------------------------------------------
 results at filter frequency  2.000 cpd with MSD of unit weight = 4.464
-------------------------------------------------------------------------------
```

indicates that the following results for 2N2 through S2 are mainly related with frequency $\Omega = 2$ cpd and the precision is estimated by using $\sigma(2\text{ cpd}) = 4.464$.

In parallel to the output **analysisNN.dat** VAV provides analysis results in file **all_tides.dat** in the **output folder b**. We have the estimated amplitudes and phases for all tides, e.g. for the 1198 tides of the development of Tamura. This output is a complement to the output in **analysisNN.dat**, which may be more interesting for the ocean tidal data. Table G.3 is an example of this output. Explanations about the content of this output are given in the file **all_tides.dat** itself.

The unit of the amplitudes and their MSD in all output files is equal to the unit of the input data. The phases and the phase lags are always in degrees of arc. The amplitude factor δ is a measureless quantity, when the unit of the data is the same as

the unit of the theoretical tides. More about the unit of the data can be found in Example 21.

In **output folder b** are stored most of the other output results, associated with the analysis results, e.g. some residuals, samples of the analysis results, etc.

Table G.3. Sample of the data in an output file **all_tides.dat**.

Nr	Doodson	Darwin	freq_cpd	freq_cph	amplitude	msd_ampl	phase
555	165-465		1.002575549	0.041773981	0.055855	0.000034	36.8468
556	165-545		1.002590816	0.041774617	9.689590	0.005971	70.1791
557	165-545		1.002590816	0.041774617	0.009408	0.000006	250.1791
558	165-555	K1	1.002737909	0.041780746	489.278322	0.301484	195.0274
559	165-555		1.002737909	0.041780746	0.088547	0.000055	15.0274
560	165-557		1.002738171	0.041780757	0.017545	0.000011	220.3863
561	165-565		1.002885003	0.041786875	66.384081	0.040905	139.8756
562	165-565		1.002885003	0.041786875	0.010377	0.000006	319.8756
563	165-575		1.003032097	0.041793004	1.424856	0.000878	264.7239
564	165-655		1.003047364	0.041793640	0.159813	0.000098	298.0563
565	165-665		1.003194458	0.041799769	0.066009	0.000041	242.9045
566	165-765		1.003503912	0.041812663	0.012928	0.000008	345.9334
567	165-775		1.003651006	0.041818792	0.011081	0.000007	290.7816
568	166-544		1.005328594	0.041888691	0.012005	0.000007	248.0961
569	166-554	PSI1	1.005475688	0.041894820	3.826102	0.002358	192.9444
570	166-556		1.005475949	0.041894831	0.058176	0.000036	218.3033
571	166-564		1.005622782	0.041900949	0.069257	0.000043	137.7926
572	167-355		1.007594819	0.041983117	0.242863	0.000150	10.1625
573	167-365		1.007741913	0.041989246	0.067411	0.000042	315.0107
574	167-455		1.007904274	0.041996011	0.028234	0.000017	113.1914
575	167-465		1.008051367	0.042002140	0.010594	0.000007	58.0396
576	167-553		1.008213466	0.042008894	0.096037	0.000059	190.8613
577	167-555	PHI1	1.008213728	0.042008905	6.966006	0.004292	36.2202
578	167-565		1.008360822	0.042015034	0.266872	0.000164	161.0685

Example 1. Start of VAV and simple tidal analysis, data Cantley.

The Basic CDATA for Cantley are shown in Table 1.1.

Table 1.1. Basic CDATA in **cant6824.ini**.

Options or statements	Short comments
>D-format: unf	Unformatted data, after the 1st application of VAV.
>MCH-model: 1	Multi-channel (MCH) analysis, model 1.
>MCH-channel: 2	MCH analysis with air-pressure in channel 02
>D-nm-input-channels: 4	Number all data channels in **cant6824.dat**
>D-max-nm-data: 77105	Number = or > of the data in **cant6824.dat**
>D-tidal-channel: 1	Indicates that the tidal data are in channel 01
>ST-long-e: 284.1929	Geographic longitude, positive to the East, degrees of arc
>ST-latit-n: 45.585	Geographic latitude, positive to the North, degrees of arc
>ST-altit-meters: 269.0	Altitude in meters
>ST-name: Cantley ...	Till 20 lines with name of station, data, instrumentation, etc., till 80 columns in a line
>ST-name: SG GRAV...	

We shall apply on the data Cantley a very simple tidal analysis, which means that we shall use default cases and values, without the MCH analysis. For that purpose we shall exclude the first 3 options in Table 1.1, by setting

***>D-format: unf**	VAV will read formatted data
***>MCH-model: 1**	Simple analysis without multi-channel (MCH) analysis
***>MCH-channel: 2**	

After this preparation, VAV should be started in one of the following ways:

1. In DOS we should go to the folder **\aaavav_03**, write there the command "**vav_03**" and press "return" key.
2. By using the "windows explorer" we should click on folder **\aaavav_03** then click on file **vav_03**.
3. Through a FORTRAN project, including all routines in **\aaafor_03**.

After VAV is started we get displayed the list of the data files, taken from **a_list_files.inp**, followed by a message of VAV, as shown in Table 1.2.

Table 1.2. Initial message of VAV with information, taken from **a_list_files.inp**

```
      data at disposition:
\a_data\brussels\          brus0200.dat ------- file nr: 01
\a_data\cantley\           cant6824.dat ------- file nr: 02
\a_data\strasbourg\        stra0306.dat ------- file nr: 03
\a_data\vienna\            vien0698.dat ------- file nr: 04
\a_data\sanjuan\           sanjuan.dat -------- file nr: 05
```

followed by a message of VAV:	Our answer:
`Enter Nr of a file, "return" is accepted as 05:`	2

The number "2", written as answer of the message, chooses the data Cantley for processing.

After pressing the "RETURN" key, VAV is executed. We get the 1st results in

\cant_out_a\ analysis99.dat.

A sample of results about the tides, selected in **all_ini.inp,** is also displayed in **\cant_out_b**, files: **a_delta.dat, a_kappa.dat** and **a_hamplit.dat**.

We shall use only **a_delta.dat**. Table 1.3 shows some selected data, taken from **a_delta.dat** in this example.

Table 1.3. Cantley, sample of results by simple analysis, amplitude δ factors and MSD $\sigma(\delta)$, taken from **a_delta.dat**.

RUN_NR	δ(Q1)	δ(K1)	δ(N2)	δ(M2)	AIC
RUN_01	1.16284	1.14830	1.20964	1.20360	134965
MSD σ(δ)	±.00159	±.00021	±.00052	±.00010	

An output, very useful to study various anomalies, is stored in **residfn_00.dat**. Namely, the columns "1.0_cpd" till "6.0_cpd" provide the modulus of the residuals of the filtered numbers at the corresponding frequencies Ω = 1, 2, ... 6 cpd. Figure 1.1 represents these residuals for frequencies Ω = 1, 2 & 3 cpd, i.e. for the D, SD and TD tidal species.

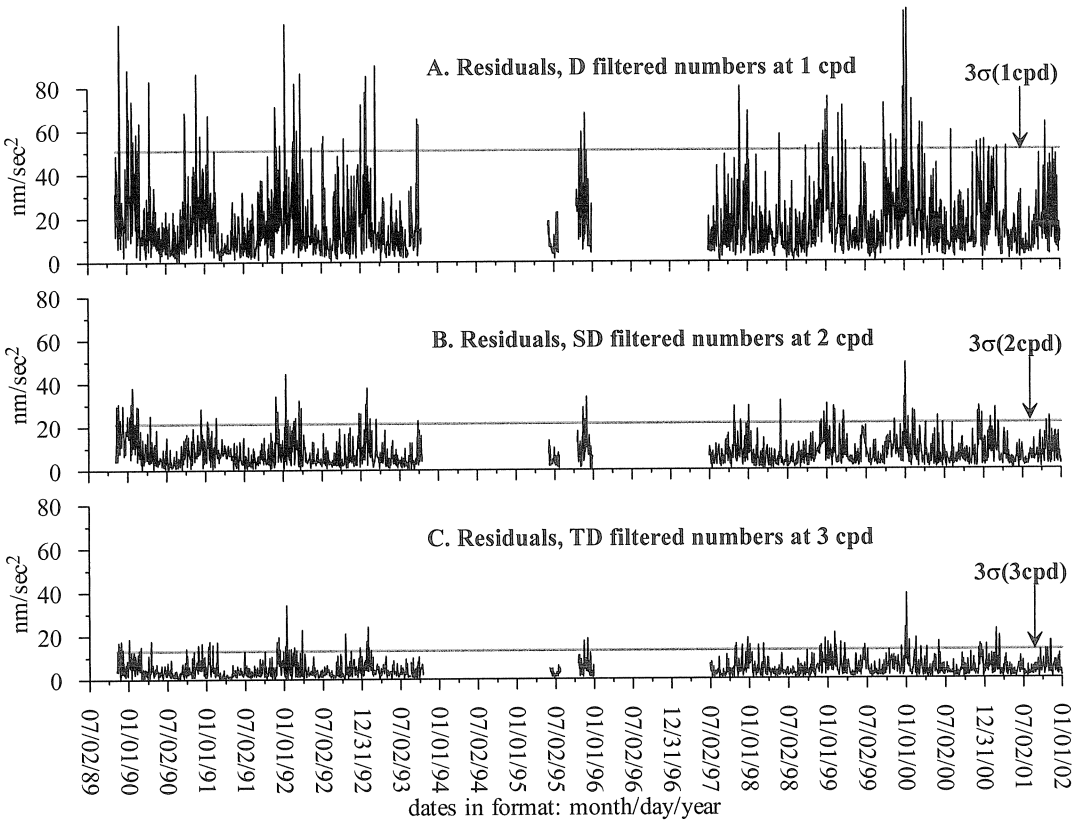


Figure 1.1. Cantley, modulus of the residuals of the filtered numbers and 3σ threshold levels, taken from **\cant_out_b\residfn_00.dat**.

The thick straight line is a threshold level. Its ordinate is $3\sigma(\Omega)$ where $\sigma(\Omega)$ is the MSD, estimated by using the residuals at the frequency Ω. The coefficient "3" is the t-coefficient of Student, corresponding to very high confidential probability (the rule 3 sigma). As shown in Figure 1.1 we have too many residuals over the threshold level, i.e. we have something systematically wrong in the analysis. Due to this we have to look for a more sophisticated processing.

By the way, the threshold level $3\sigma(\Omega)$ can be replaced e.g. by $1.96\sigma(\Omega)$ or $3.5\sigma(\Omega)$ etc., by using the options

>E-Student-t: 1.96 or **>E-Student-t: 3.5** etc.

Actually, we know beforehand that the SG data are subject to strong effect of the air-pressure. Due to this, in the next Example 2, we shall apply a multi-channel analysis, involving the air-pressure.

Example 2. Multi-channel (MCH) analysis with air-pressure, data Cantley.

We shall skip the starlets * in the first 3 statements in **cant6824.ini,** i.e. we shall restore the Basic CDATA in Table 1.1. Then we shall add to the Basic CDATA

New CDATA	With effect
>Out-hr-residuals: yes	Output of hourly residuals in **residall.dat**

and run VAV in the same way as in Example 1, i.e. by choosing file Nr 2.

We shall get in **\cant_out_a** the new results in file **analysis98.dat** by keeping the earlier results in **analysis99.dat,** arranged according to Table G.1.

Now we shall get 2 new rows in **a_delta.dat**, so that the results, shown in Table 1.3 become as those in Table 2.1.

Table 2.1. Cantley, sample of results by MCH analysis with air-pressure, amplitude δ factors and MSD σ of δ taken from **a_delta.dat**.

Variant	RUN_NR	δ(Q1)	δ(K1)	δ(N2)	δ(M2)	AIC
Simple analysis	RUN_01	1.16284	1.14830	1.20964	1.20360	134965
	MSD σ(δ)	±.00159	±.00021	±.00052	±.00010	
Multi-channel analysis	RUN_02	1.16556	1.14789	1.20934	1.20349	91047
	MSD σ(δ)	±.00045	±.00006	±.00052	±.00005	

We get an extremely strong improvement of the precision: nearly 4 times lower MSD for the D tides and 2 times for the SD tides. I.e., our MCH analysis is O.K.

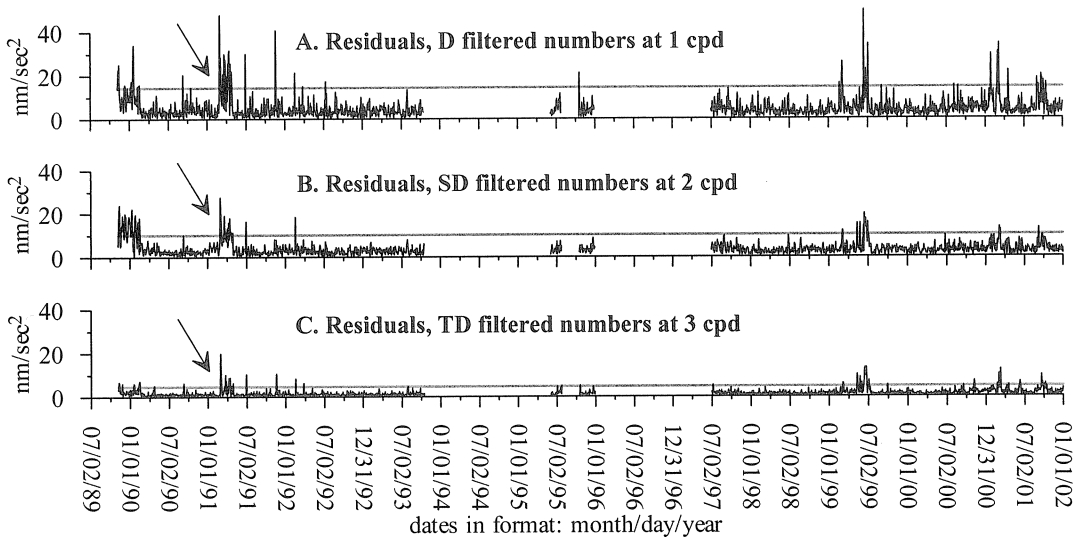


Figure 2.1. New residuals (modulus) of the filtered numbers and threshold levels from file **residfn_00.dat** after run 2 of VAV with MCH analysis including air-pressure data.

The improvement of the precision is also clearly demonstrated by Figure 2.1, where the new residuals are plotted. The comparison with Figure 1.1 shows that we have a considerably lower level of the residuals.

It is very important that the residuals, exceeding the threshold levels, are considerably fewer. Nevertheless, we have a curious accumulation of anomalous residuals at several time intervals. We shall try to study the particular "swarm" of residuals, indicated by arrows, somewhere between January and July 1991.

In the next Figure 2.2A the residuals of the D-filtered numbers around and in the anomalous area are plotted. Now we use time in hours, in order to get exactly the place of the anomaly. Figure 2.2B, obtained due to **>Out-hr-residuals: yes,** represents the residuals of the hourly data, which clearly confirm the anomaly. By using these figures we have got the start and the end of the anomaly, as they are indicated in Figure 2.2B.

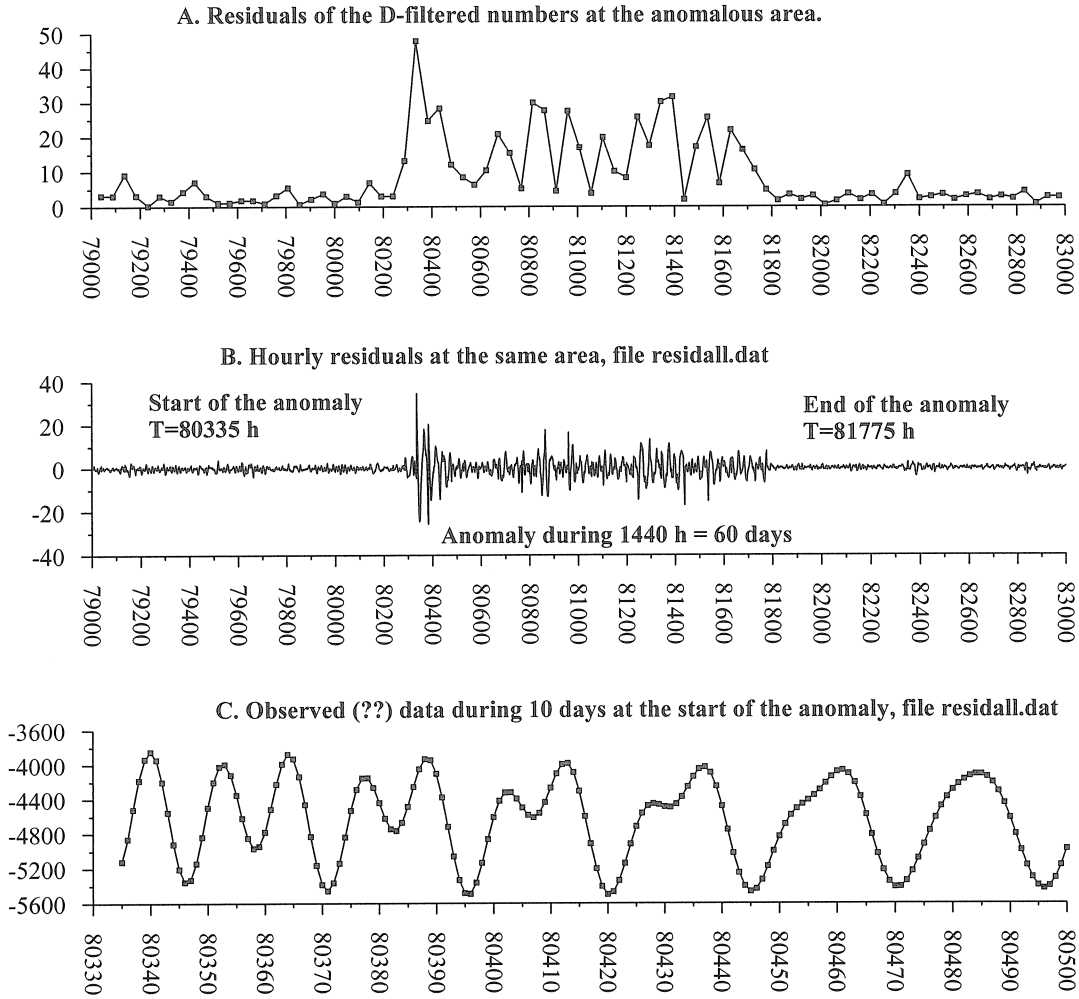


Figure 2.2. Some data at the anomalous area, indicated in Figure 2.1 by arrows.

Example 3. Elimination of data intervals, data Cantley.

Figure 2.2C shows that there are not visual anomalies in the observed data. We can let this mystery for further investigations and now continue the analysis without this anomaly, i.e. without the data from time 80335^h till 81775^h.

There is another anomaly at the start of the data, clearly seen in Figure 2.1B. We have roughly estimated that it is better to abandon the data from the initial time 68816^h till time 72240^h.

The elimination of these data is done by adding to the Basic CDATA:

New CDATA	Comments
>E-T-interval: 68816 72240	Eliminates the interval from the start till 72240^h
>E-T-interval: 80335 81775	Eliminates the interval from 80335^h till 81775^h

For our experiments, it is better the New CDATA to replace the New CDATA used before, i.e. the option **>OUT-HR-residuals: yes**, whose execution takes time.

Table 3.1 gives an extraction of the results in **a_delta.dat** after this elimination of the data. Now we do not show the results from RUN 1, since they are no more interesting.

Table 3.1. Cantley, elimination of 2 time intervals, sample of amplitude δ factors and MSD σ of δ taken from **a_delta.dat**.

RUN	δ(Q1)	δ(K1)	δ(N2)	δ(M2)	Number of data used
RUN_02	1.16556	1.14789	1.20934	1.20349	77232
MSD σ(δ)	**±.00045**	**±.00006**	**±.00025**	**±.00005**	
RUN_03	1.16515	1.14781	1.20939	1.20340	72384
MSD σ(δ)	**±.00040**	**±.00005**	**±.00020**	**±.00004**	

We have a reduction of the MSD by some 10% at the D tides and some 20% at the SD tides. We would get still better results if we have more carefully studied the whole data set and eliminate more data.

Notice that we use the term "eliminate" but actually the data are not eliminated and they remain on their place. We merely exclude them from the processing.

Example 4. Automatic elimination of anomalies, data Cantley.

An easier way to look for, find and eliminate anomalous data is to replace the New CDATA in Example 3 above by:

New CDATA	Effect
>E-nm-iteratons: 5	Will find too big residuals and eliminate them in 5 iterations

More explicitly, the effect of this option is the following. In **iteration 0** we get the residuals in **residfn_00.dat**, as shown in Figure 2.1. The residuals over the threshold level indicate that the corresponding filtered intervals are **"bad intervals"**, i.e. they comprise some anomalies.

In the next **iteration 1** the analysis is applied on the data, from which the bad intervals from **residfn_00.dat**, i.e. those, located by **iteration 0**, are excluded. Thus we get new analysis results and new residuals, stored in **residfn_01.dat**. The next **iteration 2** will eliminate the new bad intervals from **residfn_01.dat** and this will continue according to the number of iterations chosen in **>E-nm-iteratons: 5.**

After the last iteration, we get in **badint_out.dat** a list of all bad intervals. For every one of them we have an indication at which iteration it has been eliminated.

Figure 4.1A is the same as Figure 2.1A but a different vertical scale is used. Figure 4.1B shows the effect of **iteration 1**. We have namely eliminated considerable number of the anomalies, detected in **iteration 0**. As a result, according to the threshold level, the level of the noise is reduced some 1.5 times.

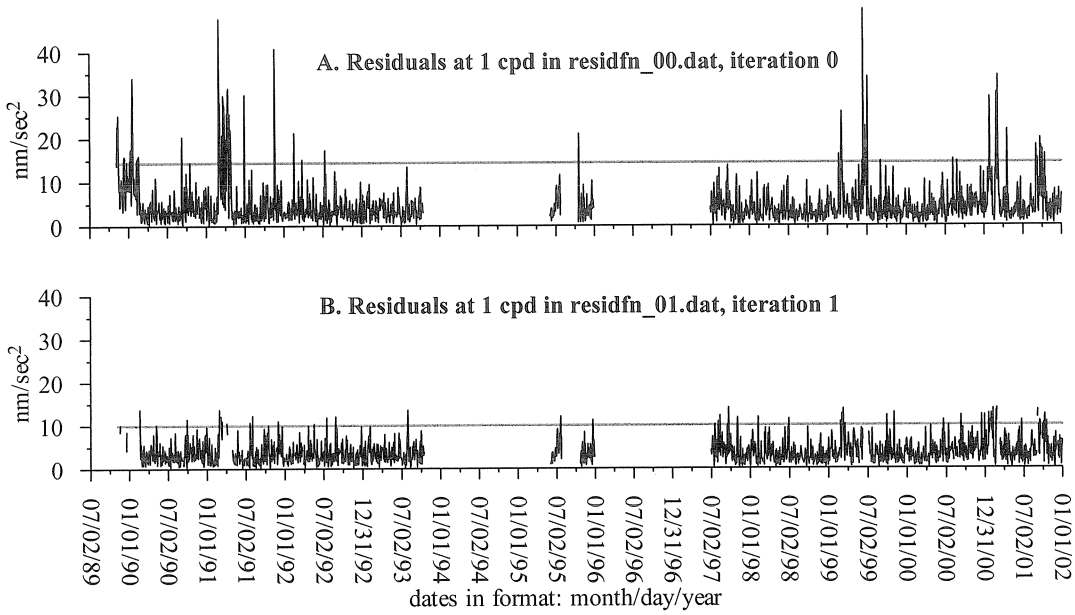


Figure 4.1. Residuals (modulus) and threshold level 3σ at 1 cpd, obtained by **iteration 0** and **iteration 1**.

This reduction of the noise is not yet a guarantee for better results. It is important to check the precision of the analysis results, which may be not improved. It is even possible the precision to be badly worsened.

The analysis results are demonstrated in Table 4.1.

Iteration 2 has eliminated 12.1% of the data. For normal data, the reduction of their quantity by 12.1% should increase the MSD σ(δ) at least $\sqrt{100/87.9} = 1.07$ times. Here we have a decrease of σ(δ) for Q1 from 0.00045 to 0.00029. Such an effect would be obtained if we disposed by $(45/29)^2 = 2.4$ times longer series of data. It is more than obvious that this is a very useful procedure.

Table 4.1. Cantley, elimination in 5 iterations, amplitude δ factors and MSD σ of δ taken from **a_delta.dat** (the earlier results in **a_delta.dat** are not shown).

RUN/ iteration	δ(Q1)	δ(K1)	δ(N2)	δ(M2)	Number Data	Elimi-nated %
RUN_04/00	1.16556	1.14789	1.20934	1.20349	77232	0.0%
MSD σ(δ)	**0.00045**	**0.00006**	**0.00025**	**0.00005**		
RUN_04/01	1.16531	1.14790	1.20930	1.20344	72144	6.6%
MSD σ(δ)	**0.00033**	**0.00004**	**0.00018**	**0.00003**		
RUN_04/02	1.16554	1.14787	1.20925	1.20347	67920	12.1%
MSD σ(δ)	**0.00029**	**0.00004**	**0.00017**	**0.00003**		
RUN_04/03	1.16547	1.14785	1.20918	1.20349	65568	15.1%
MSD σ(δ)	**0.00028**	**0.00004**	**0.00016**	**0.00003**		
RUN_04/04	1.16536	1.14784	1.20915	1.20350	64272	16.8%
MSD σ(δ)	**0.00027**	**0.00004**	**0.00016**	**0.00003**		
RUN_04/05	1.16526	1.14783	1.20914	1.20350	63600	17.7%
MSD σ(δ)	**0.00027**	**0.00004**	**0.00016**	**0.00003**		

Example 5. Elimination of one and the same data, data Cantley.

It is reasonable, among all iterations in Example 4, to choose **iteration 2** because (i) we have less important reduction of the MSD after **iteration 2** and (ii) in principle, it is better not to go very far with the quantity of the eliminated data. In order to use in further analyses one and the same eliminated data, without the execution of the iterations, i.e. without **>E-nm-iterations:**, we have first to store

The output file **badint_out.dat**	as a file named **badint.dat**

Then we have to use, instead of **>E-nm-iterations: 5**

New CDATA	Effect
>E-badint.dat: 2	Will provide the same results as at iteration 2

Table 5.1. Cantley, elimination by using **badint.dat**, amplitude δ factors and MSD σ of δ taken from the last two rows of **a_delta.dat**.

RUN	δ(Q1)	δ(K1)	δ(N2)	δ(M2)	Number data	Elimi-nated %
RUN_05	1.16554	1.14787	1.20925	1.20347	67920	12.1%
MSD σ(δ)	**0.00029**	**0.00004**	**0.00017**	**0.00003**		

The effect of **>E-badint.dat: 2** is that VAV will take those bad intervals from the new file **badint.dat**, which have been eliminated when **iteration 2** has been executed. Then a new run of VAV will produce the results in Table 5.1, which are identical to RUN_04/02 in Table 4.1 above.

The idea is that thus we have prepared "cleaned" data, ready for further investigation. Notice, that in all these procedures "elimination" means only "not use at the moment for the processing while the original data remain intact".

The "elimination" introduces considerable quantity of gaps. Nevertheless, despite the superstitious fears of many scientists from the gaps, just the gaps introduced here provided a considerable improvement of the precision.

Example 6. Another analysis in 5 iterations, data Strasbourg

The Basic CDATA for Strasbourg are:

***>D-FORMAT: unf**
>D-NM-INPUT-CHANNELS: 4
>D-MAX-NM-DATA: 115800
>D-TIDAL-CHANNEL: 1
>MCH-MODEL: 1
>MCH-CHANNEL: 2
>ST-LONG-E: 7.684
>ST-LATIT-N: 48.6223
>ST-ALTIT-METERS: 180.0
>ST-GRAV-GALS: 981.
>ST-NAME: STATION 0306 STRASBOURG ...
>ST-NAME: Both series of data ...

The experiment with the MCH analysis in Cantley is so convincing, that we need not here to make more experiments, i.e. we can directly use the MCH options in the basic CDATA. We shall also directly apply the automatic elimination of the data, i.e. we shall use

New CDATA
>E-nm-iteratons: 5

Thus we get the first analysis results in **\stra_out_a\analysis99.dat.** A sample of these results is shown in Table 6.1.

Table 6.1. Strasbourg, MCH analysis with elimination in 5 iterations: amplitude δ factors and MSD σ of δ taken from **\stra_out_b\a_delta.dat.**

RUN_NR	δ(Q1)	δ(K1)	δ(N2)	δ(M2)	Number data	Elim. Data %
RUN_01/00	1.14608	1.13576	1.17200	1.18552	115824	0.0%
MSD σ(δ)	±.00055	±.00007	±.00030	±.00006		
RUN_01/01	1.14637	1.13592	1.17236	1.18571	107424	7.3%
MSD σ(δ)	±.00032	±.00004	±.00020	±.00004		
RUN_01/02	1.14632	1.13605	1.17248	1.18580	99552	14.0%
MSD σ(δ)	±.00026	±.00004	±.00017	±.00003		
RUN_01/03	1.14639	1.13608	1.17254	1.18581	94128	18.7%
MSD σ(δ)	±.00025	±.00003	±.00017	±.00003		
RUN_01/04	1.14650	1.13611	1.17257	1.18583	91392	21.1%
MSD σ(δ)	±.00024	±.00003	±.00016	±.00003		
RUN_01/05	1.14652	1.13610	1.17264	1.18583	90336	22.0%
MSD σ(δ)	±.00024	±.00003	±.00016	±.00003		

Iteration 2 reduces σ(δ) for Q1 from 0.00055 to 0.00026. Such an improvement of the precision can be obtained, theoretically, through an increase of the number of the data by $(55/26)^2 = 4.5$ times, i.e. if the observations have been obtained during more than 50 years! We have got this improvement in the opposite way – through the decreasing the number of the data by 14.0% and introducing a number of gaps.

The classical spectral analysis needs continuous data without gaps. When the data have gaps, they generate spikes. VAV uses MLS which is a more general and more flexible method, able to take into account the existence of the gaps.

We may choose as optimum the variant **RUN_01/02**, i.e. elimination at iteration 2. At this stage already we get a considerable improvement of the precision, while in the next iterations the improvement is not very important.

Then we have to transform

File **badint_out.dat** in file **badint.dat**

If we accept permanently iteration 2, we have to replace **>E-nm-iterations: 5** by

>E-badint.dat: 2

The next run of VAV will provide results, identical to **RUN_01/02** in Table 6.1.

In such a way we have prepared cleaned data Strasbourg for other investigations.

Example 7. Analysis of variances or factorial analysis, data Strasbourg.

The data Strasbourg are composed by two series, obtained by two different instruments. These parts are separated by a big gap around the time point $T = 5400$ days (Figure 7.1). It is certainly interesting to check whether there are some systematic differences between the two series. In another slang the task can be formulated as: "to check whether the instrument used is a factor". Such a problem can be solved through the method analysis of variances, based on the famous ratio F of Fisher.

This example does will not offer surprising results with particular meaning. It is only instructive, by showing that some conclusions need making statistical inferences and that such inferences need MLS estimation of the precision, based on the sum or squares of residuals.

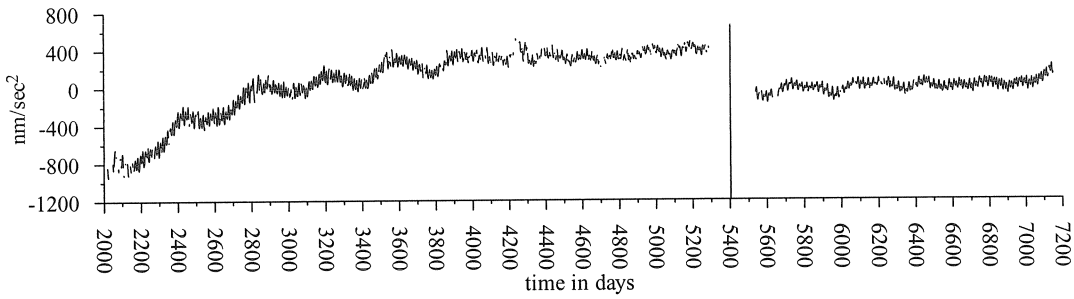


Figure 7.1. Strasbourg, drift, taken from file **drift.dat**.

Before starting a new investigation by VAV it is good

to delete **a_delta.dat** and all files **analysisNN.dat**

Then we shall analyse (i) the 1st series, before $T = 5400$ days, (ii) the 2nd series, after $T = 5400$ days and (iii) both series together, i.e. all data.

This is done in 3 runs of VAV, by using consecutively:

Run_01, New CDATA for the analysis of the 1st series	Effect
>E-badint.dat: 2	Eliminates the data, determined at iteration 2
>E-t-interval 5400 7300 days	Eliminates the whole 2nd series, remains the 1st series

Run_02, New CDATA for the analysis of the 2nd series	Effect
>E-badint.dat: 2	Eliminates the data, determined at iteration 2
>E-t-interval 2000 5400 days	Eliminates the whole 1st series, remains the 2nd series

Run_03, New CDATA for the analysis of all data	Effect
>E-badint.dat: 2	Eliminates the data, determined at iteration 2

A sample of the results is given in Table 7.1.

Table 7.1. Strasbourg, sample of results from **a_delta.dat** for separate analyses of two series and analysis of both series together.

Series of data	δ(Q1)	δ(K1)	δ(N2)	δ(M2)	Number data	Elim. data %
Run 1: 1st series	1.14578	1.13558	1.17184	1.18537	63504	19.2%
MSD σ(δ)	±.00032	±.00004	±.00023	±.00004		
RUN 2: 2nd series	1.14742	1.13710	1.17344	1.18651	36048	3.2%
MSD σ(δ)	±.00035	±.00005	±.00021	±.00004		
RUN 3: all data	1.14632	1.13605	1.17248	1.18580	99552	14.0%
MSD σ(δ)	±.00026	±.00004	±.00017	±.00003		

In every **analysisNN.dat** we get, after the title:

frequency dependent sum of squares of residuals:

the sums of squares of the residuals (column SSQ) at the basic frequencies and the corresponding degrees of freedom (column DFR).

We have recollected in Table 7.2 the SSQ from the 3 cases.

Table 7.2. Sum of squares of residuals $S_j(\Omega)$ and degrees of freedom $d_j(\Omega)$ for Run $j = 1,2,3$ of VAV and the ratio $F(\Omega)$ of Fisher at frequency $\Omega = 1,2,...6$ cpd.

Freq.	First seies		Second series		All data		$F(\Omega)$
Ω cpd	$d_1(\Omega)$	$S_1(\Omega)$	$d_2(\Omega)$	$S_2(\Omega)$	$d_3(\Omega)$	$S_3(\Omega)$	
1.0	2605	27322.8	1461	7757.1	4107	42068.0	19.76
2.0	2625	15939.6	1481	4635.5	4127	23088.6	23.89
3.0	2641	4507.0	1497	785.7	4143	5313.2	3.21
4.0	2641	2569.8	1497	332.0	4143	2911.5	2.78
5.0	2641	1808.0	1497	197.4	4143	2016.2	4.46
6.0	2641	1273.8	1497	142.4	4143	1421.7	3.18

By using these data we compute, for every row, i.e. for every frequency Ω the ratio (criterion) of Fisher, whose expression in this case is

$$F(\Omega) = \left(\frac{S_3 - S_1 - S_2}{d_3 - d_1 - d_2}\right) \Big/ \left(\frac{S_1 + S_2}{d_1 + d_2}\right) \text{ where } S_j = S_j(\Omega) \text{ and } d_j = d_j(\Omega)$$

The values obtained are given in the last column of Table 7.2. All of them are too big and we cane state that there are significant differences between the two series.

The SSQ and the DFR here used are also used to get a frequency dependent estimation of the precision. Other programs for tidal analysis, e.g. ETERNA, which do not use SSQ for frequency dependent estimation of the precision, are unable to use the basic statistical criteria, like *F* of Fisher and *t* of Student.

Example 8. Determination of the LP (long period) tides, data Vienna

VAV does not use neither drift, obtained by filtration, e.g. by the filter of Pertsev, nor residual drift. The drift is approximated by polynomials of low power, independently for every time window, i.e. every filtered interval. Then the coefficients of the polynomials are estimated through the application of MLS. Thus VAV provides an estimated drift. The following Figures 8.1 & 8.2 give the estimated drift at the central points of the filtered intervals (of every time window).

In Figure 8.1, since the LP tides are ignored, they remain in the drift. We can clearly distinguish the main LP tides, namely: the lunar fortnightly MF, with period close to 14 days, as well as the lunar monthly MM, with period close to 28 days.

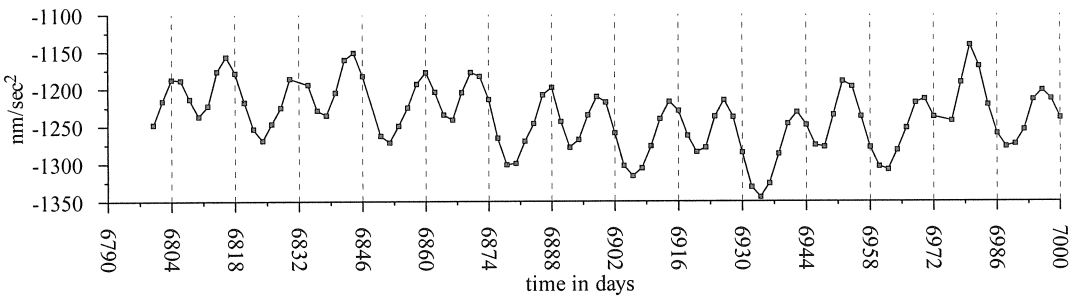


Figure 8.1. Vienna, drift estimated for every time window of 48^h by using approximation through polynomials of power $k = 2$, WITHOUT determination of LP.

Figure 8.2 shows the drift in the case when the LP tides have been taken into account. Due to this the LP tides, at least those, manifested in Figure 8.1, have disappeared. More about the determination of the LP tides can be found in (Ducarme et al., 2003).

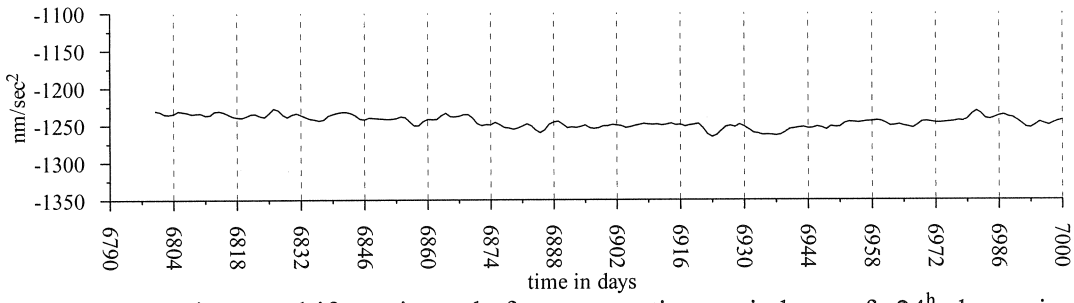


Figure 8.2. Vienna, drift estimated for every time window of 24^h by using approximation through polynomials of power $k = 0$, WITH the determination of LP.

Now we shall show how we can estimate the LP tides in Vienna, i.e. what does it mean "the LP tides have been taken into account".

We shall use **all_ini.inp**, with CDATA shown in Table 8.1. There are included the names of 6 LP tidal groups (see the input file **groups.inp** in **\aaavav_03**), because we shall deal with some or all of them.

Table 8.1. New content of **all_ini.inp**, useful for the determination of the LP tides.

>D-origin-of-time: 1982 01 01 00
>out-compare-tide-name: SSA
>out-compare-tide-name: MSM
>out-compare-tide-name: MM
>out-compare-tide-name: MF
>out-compare-tide-name: MSTM
>out-compare-tide-name: MSQM

We shall use the Basic CDATA shown in Table 8.2.

Table 8.2. Basic CDATA for Vienna in **vien0698.ini**.

***>D-format: unf**
>MCH-model: 1
>MCH-channel: 2
>D-nm-input-channels 4
>D-tidal-channel 1
>D-max-nm-data 35810
>ST-long-e: 16.3579
>ST-latit-n: 48.2493
>ST-altit-meters: 192.440
>ST-grav-gals: 981.
>St-name: STATION 0698 VIENNE ...
>St-name: 48.2493 N ...
>St-name: SUP-GWR C 025 ...

Here, as always, **>D-format: unf** is excluded by the starlet for the first run of VAV.

Before everything, we have to clean the data and prepare the file **badint.dat** as we have done with Cantley and Strasbourg. For this purpose we shall add

New Cdata
>E-NM-iterations: 5

Then we start VAV and choose file Nr 4, i.e. the data Vienna.

We shall see through the output files **\vien_out_a\analysis99.dat** and **\vien_out_b\a_delta.dat** that the elimination procedure has not very strong effect. Nevertheless, it seems reasonable to accept that iteration 2 has brought an improvement.

It is better now to skip the starlet in ***>D-format: unf.**

Our next step is to transform in **\vien_out_b**

badint_out.dat ⇒ badint.dat

Afterwards we delete **a_delta.dat,** as well as all **analysisNN.dat.**

Now we start the determination of the LP tides. We shall first use

New CDATA	Some comments
>E-badint.dat: 2	Eliminates the bad intervals, obtained after iteration 2.
>F-t-window: 24	Makes the time window = 24^h
>F-k-drift-polyn: 0	Drift, approximated by a stepwise function, remaining constant during the selected time window, i.e. 24 h.
>GR-LP: 6	Chooses LP variant 6 in **groups.dat**, i.e. 6 LP groups

Then we run again VAV and we get in **a_delta.dat** the results, given in row **>GR-LP 6** in Table 8.3

Further we replace **>GR-LP: 6** successively, one by one, by

>GR-LP: 5
>GR-LP: 4
>GR-LP: 3
>GR-LP: 2
>GR-LP: 1

and run every time VAV, by choosing, of course, the data Vienna.

Thus we get the results in the remaining rows of Table 8.3 with 5, 4, 3, 2 & 1 LP groups respectively. The arrows show how the groups in a given variant are shaped by the tides in the groups from preceding variants. In particular, variant **>GR-LP: 1**, i.e. one group, named MF, unifies all 6 groups from the previous variants, i.e. all LP tides.

Table 8.3. Vienna, determination of LP tides, data taken from **a_delta.dat.**

	δ(SSA)	δ(MSM)	δ(MM)	δ(MF)	δ(MSTM)	δ(MSQM)	AIC
>GR-LP 6 MSD σ(δ)	1.522 ±.199	1.2658 ±.0934	1.1428 ±.0152	1.1370 ±.0054	1.1286 ±.0188	1.0720 ±.0695	61347
>GR-LP 5 MSD σ(δ)	→	1.3123 ±.0847	1.1429 ±.0152	1.1369 ±.0054	1.1288 ±.0188	1.0719 ±.0695	61348
>GR-LP 4 MSD σ(δ)	→	→	1.1482 ±.0150	1.1369 ±.0054	1.1288 ±.0188	1.0728 ±.0695	61359
>GR-LP 3 MSD σ(δ)	→	→	1.1481 ±.0150	1.1370 ±.0054	1.1252 ±.0182	←	61359
>GR-LP 2 MSD σ(δ)	→	→	1.1481 ±.0150	1.1360 ±.0052	←	←	61368
>GR-LP 1 MSD σ(δ)	→	→	→	1.1372 ±.0049	←	←	61365

In this kind of analyses with variants, in which one and the same set of data is used, very helpful can be the AIC criterion of Akaike – a minimum value of AIC indicates the most promising variant. The minimum here is at the first variant, with 6 LP groups. In the same time we have: (i) the minimum is not much lower than the other values and (ii) the MSD of the group MF in the last variant with one LP group is lower by some 10% than in the case of 6 LP. In these controversial circumstances we may rely

on the principle of parsimony, recommending the variants with lower number of unknowns, in this case – with lower number of LP groups. Due to this, we would recommend the last variant, unifying all LP tides in one group.

Example 9. Variation of the drift models, ocean tidal data San Juan.

We shall use the series of ocean tidal data from the station San Juan, covering an interval of 16 years: 1.01.1985 – 31.12.2000.

We shall use an **all_ini.inp**, shown in Table 9.1.

Table 9.1. Content of **all_ini.inp**.

>D-origin-of-time: 1985 01 01 00
>out-compare-tide-name: Q1
>out-compare-tide-name: K1
>out-compare-tide-name: N2
>out-compare-tide-name: M2
>out-compare-tide-name: MF

Now the epoch, chosen by **>D-origin-of-time**, coincides with the first date of the data. Actually, this is the default case of VAV, so that this option is written here only for information.

The Basic CDATA we have to use are given in Table 9.2.

Table 9.2. Basic CDATA in the control data file **sanjuan.ini**.

>D-format: Sanjuan
***>D-format: unf**
>D-max-nm-data: 140300
>ST-long-e: -66.1167
>ST-latit-n: 18.4600
>ST-altit-meters: 0
>ST-grav-gals: 0
>ST-component: ocean
>ST-name: San Juan ...

Here are the following particularities.

(i) The data are in a very particular format for which VAV has an appropriate subroutine. It is called by using **>D-format: Sanjuan.** After the 1st run, this statement may be replaced by **>D-format: unf**.

(ii) We have only one channel with tidal data. Therefore we cannot use an MCH analysis and we do not need to use **>D-nm-input-channels** and **>D-tidal-channel** .

(iii) We have to use **>ST-component: ocean**, which indicates the type of the data. For this kind of data the theoretical amplitudes and phases are computed for the equilibrium ocean tides. In the earlier examples with gravity data we have not used **>ST-component:**, because the default option is prepared namely for gravity data.

Now we shall try to find an optimum or at least a reliable model of the drift. For this purpose we shall run 5 times VAV, by choosing the San Juan data, with the 5 variants of the New CDATA, given in Table 9.3.

Table 9.3. Five variants of New CDATA, where ΔT is the time window, k is power of the drift polynomials and the presence of MF means that one LP tidal group MF (see example 8) is included in the analysis.

Variants	New CDATA	Parameters chosen
Variant 1	**>GR-D-SD: 50** **>F-T-window: 48** **>F-k-drift-polyn: 0**	$\Delta T = 48^h,\quad k = 0$
Variant 2	**>GR-D-SD: 50** **>F-T-window: 48** **>F-k-drift-polyn: 1**	$\Delta T = 48^h,\quad k = 1$
Variant 3	**>GR-D-SD: 50** **>F-T-window: 48** **>F-k-drift-polyn: 2**	$\Delta T = 48^h,\quad k = 2$
Variant 4	**>GR-D-SD: 50** **>F-T-window: 48** **>F-k-drift-polyn: 0** **>GR-LP: 1**	$\Delta T = 48^h,\quad k = 0$, MF
Variant 5	**>GR-D-SD: 50** **>F-T-window: 24** **>F-k-drift-polyn: 0** **>GR-LP: 1**	$\Delta T = 24^h,\quad k = 0$, MF

A sample of the results from the use of these variants is given in Table 9.4.

Table 9.4. Sample of five variants of analysis results, extracted from the **a_delta.dat.**

Variants	Q1	K1	N2	M2	MF	AIC
Variant 1	1.19485 **±.01743**	1.01972 **±.00223**	0.86822 **±.00453**	0.70380 **±.00083**		**179213**
Variant 2	1.19307 **±.01699**	1.02046 **±.00217**	0.86940 **±.00447**	0.70406 **±.00082**		**176022**
Variant 3	1.19246 **±.01660**	1.02024 **±.00221**	0.86928 **±.00448**	0.70408 **±.00083**		**176229**
Variant 4	1.19458 **±.01737**	1.01971 **±.00222**	0.86805 **±.00453**	0.70366 **±.00083**	0.76055 **±.09314**	**179042**
Variant 5	1.19310 **±.01428**	1.01971 **±.00194**	0.86905 **±.00338**	0.70384 **±.00064**	0.78523 **±.07882**	

In all variants **>GR-D-SD: 50** makes use of 50 D and SD tidal groups (variant 50 in **groups.inp**). In variants 1 through 4 we use a time window of 48 hours, i.e. the drift is approximated by some polynomials, separately in intervals of 48 hours. In variants 1, 2 & 3 we raise the power of the polynomials from 0 to 2. In variant 4 the power is reduced to 0, i.e. the drift is approximated by a stepwise function, remaining a constant during 48 hours. Instead, we include an LP tidal group. In the last variant 5 we

have a shorter time window of 24 hours and one LP group as in variant 4. This is just the variant, used in the previous Example 8.

The variants 1 through 4 can be compared through the AIC value in Table 9.4. Its lowest value is at variant 2, i.e. $\Delta T = 48$ and $k = 1$ should be considered as most reliable. These four variants cannot be compared through AIC with the last variant 5 due to the use of different time windows, which means different number of filtered quantities. In this case we can deal with the MSD. Since they are clearly lower for variant 5 we can decide that this is the best variant for the analysis of these ocean data.

Example 10. Variation of the grouping, data San Juan.

By taking into account the results in Example 9 we shall use

New CDATA
>F-T-window: 24
>F-k-drift-polyn: 0
>GR-LP: 1

This is the definition of the drift model in the last variant 5 in Example 9.

Further we shall add to the New CDATA consecutively, one by one, **>GR-D-SD: 63, >GR-D-SD: 50, >GR-D-SD: 32, >GR-D-SD: 30, >GR-D-SD: 14, >GR-D-SD: 13, >GR-D-SD: 11,** every time running VAV.

Here 63, 50 etc. are conventional numbers of a variant of grouping in **groups.inp**, but in these cases these numbers are also equal to the number of the groups in the corresponding variant.

Table 10.1 shows a sample of the results of this series of analyses.

Table 10.1. San Juan, sample of analysis results, taken from **a_delta.dat** and obtained by using different variants of grouping.

Variant used	δ(Q1)	δ(K1)	δ(N2)	δ(M2)	δ(MF)	AIC
>GR-D-SD: 63	1.19290 **±.01477**	1.01745 **±.00203**	0.87211 **±.00342**	0.70571 **±.00065**	0.78433 **±.07971**	**335749**
>GR-D-SD: 50	1.19310 **±.01428**	1.01971 **±.00194**	0.86905 **±.00338**	0.70384 **±.00064**	0.78523 **±.07882**	**335053**
>GR-D-SD: 32	1.19616 **±.01443**	1.01974 **±.00196**	0.86516 **±.00360**	0.70395 **±.00069**	0.77046 **±.07714**	**339302**
>GR-D-SD: 30	1.19616 **±.01443**	1.01974 **±.00196**	0.86516 **±.00361**	0.70395 **±.00069**	0.75448 **±.07693**	**339341**
>GR-D-SD: 14	1.19833 **±.01398**	1.01973 **±.00198**	0.86743 **±.00358**	0.70396 **±.00070**	0.74759 **±.07742**	**340298**
>GR-D-SD: 13	1.19831 **±.01429**	1.01955 **±.00202**	0.86743 **±.00358**	0.70396 **±.00070**	0.74771 **±.07912**	**341663**
>GR-D-SD: 11	1.19832 **±.01435**	1.01358 **±.00193**	0.86743 **±.00359**	0.70395 **±.00070**	0.74491 **±.07944**	**342002**

There are not essential differences in the MSD. However, the lowest value of AIC shows as a most reliable the variant 50 with 50 D-SD tidal groups. Very closely to it is situated the variant with 63 tidal groups. These are variants with a very detailed separation, which are seldom used in the Earth tide domain.

Example 11. Prediction of the tidal signal through analysis, data San Juan.

Here we shall analyze a short interval of 2 months: from 0^h,1.01.1986 till 23^h,28.02.1986. By the same run of VAV we shall make a prediction of the tidal signal for a larger remote interval, namely for the interval of 1 year from 0^h,1.01.2000 till 23^h,31.12.2000.

For this purpose we shall use

New CDATA	Comments
>F-T-window: 24	Chooses time window 24 hours
>F-k-drift-polyn: 0	Drift as stepwise function
>GR-LP 1	Determines the LP tides in 1 group
>E-T-interval: 1985 1 1 0 1985 12 31 23	After the elimination for analysis remain the data 0^h,1.01.1986÷23^h,28.02.1986
>E-T-interval : 1986 3 1 0 2001 01 01 00	
>OUT-predict: 2000 1 1 0 2000 12 31 23	Prediction to be made in the interval 0^h.1.01.2000 ÷ 23^h.31.12.2000

In Example 3 we have used **>E-T-interval:** T_1 T_2 where T_1 & T_2 are in hours and in Example 7 **>E-T-interval:** T_1 T_2 **days**, where due to the presence of the word "**days**" T_1 & T_2 are in days.

A third format of this option is used here. Namely, when **>E-T-interval:** is followed by 8 numbers they are perceived as two dates, each date represented by year, month, day & hour. Thus the effect of

>E-T-interval 1985 1 1 0 1985 12 31 23

is the elimination of the interval from 0^h,1.01.1985 till 23^h,31.12.1985, i.e. from the beginning till the second date.

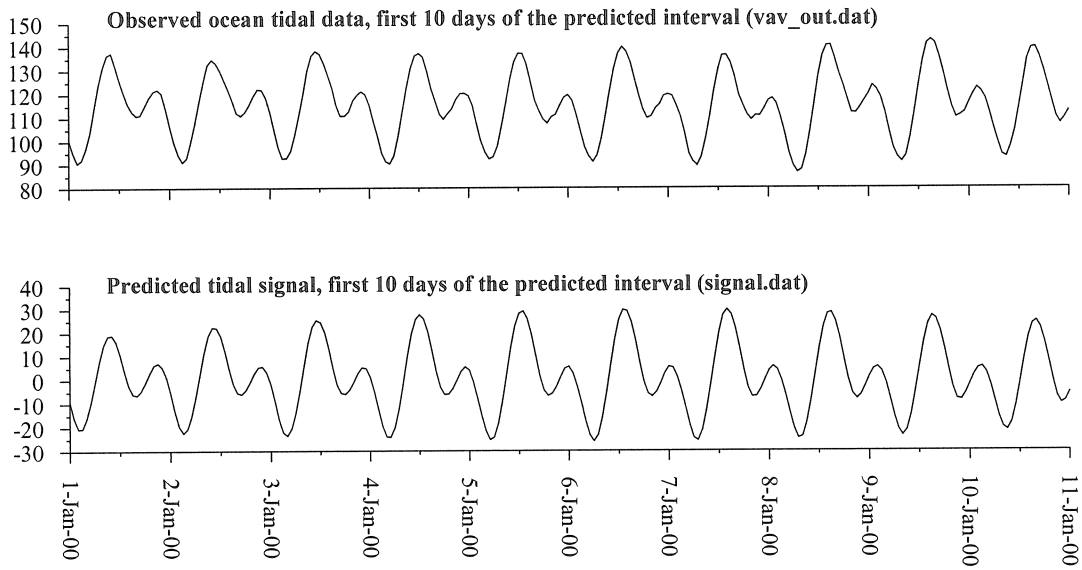


Figure 11.1. San Juan, ocean data; sample of observed data and predicted tidal signal (column "pred_sig" in **signal.dat**) during the first 10 days of the predicted interval.

Now we cannot use a very detailed separation of the tidal groups, because we analyse a very short part of the data. Therefore we do not use here **>GR-D-SD: 50** as

well as any other variant of grouping. In such a case VAV chooses a default variant of grouping, corresponding to the data length.

After running VAV on the data San Juan we get the predicted signal for every hour in the chosen interval, displayed in **signal.dat**. Figures 11.1 and 11.2 show a comparison between the predicted signal and the corresponding observed data. The observed data are plotted by using **vav_out.dat**.

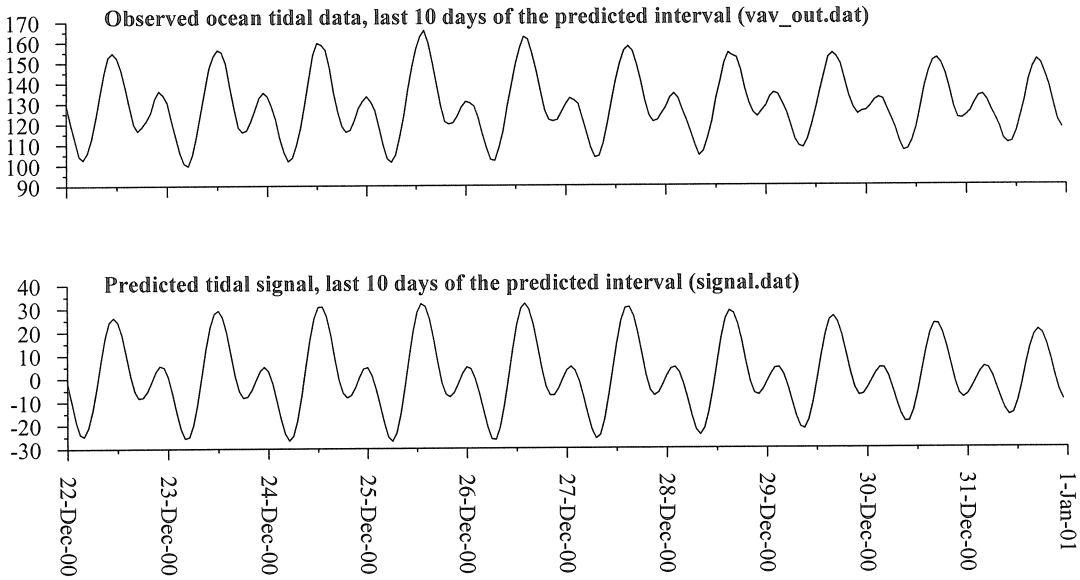


Figure 11.2. San Juan, ocean data; sample of observed data and predicted tidal signal (column "pred_sig" in **signal.dat**) during the last 10 days of the predicted interval.

Notice, that we have a prediction of the tidal signal, without the drift. Due to this we have a difference in the general level of the observed and predicted curves.

Example 12. Study and prediction of the zero-line (mean sea level), data San Juan.

Figure 12.1 shows the estimated drift of the data San Juan, with one point per day, over the whole data interval of 16 years. It is a rather complicated curve, but one thing is obvious – we have an annual component, most likely of meteorological origin.

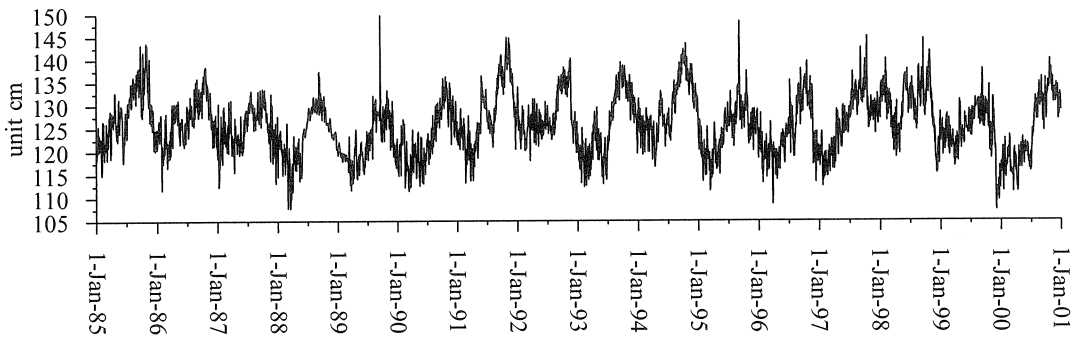


Figure 12.1. Estimated drift curve of the data San Juan (column "drift_obs" in **drift.dat,** produced by Example 10, variant **>GR-D-SD: 50**).

The Z-options of VAV allow to investigate this curve and approximate it, so that we can get a smoothed zero line of the data. If this line involves a constant term, the latter can be accepted as the mean sea level of the data.

The parameters of the zero line, including the constant term, representing the mean sea level, are included as unknowns of the global analysis, in parallel with the tidal unknowns. Thus we get these parameters correctly estimated by MLS, free of the effect of the tidal signal, with a correct estimation of the precision.

Now we shall use

New CDATA	Comments
>F-T-window: 24	Chooses time window of 24 hours
>F-k-drift-polyn: 0	Drift as stepwise function
>GR-LP : 1	Determines the LP tides in one tidal group
>GR-D-SD: 50	Variant 50 of D-SD tidal groups
>Z-k-degree-polyn: 0	Polynomial of power 0 = mean sea level
>Z-freq: 1 cpy	Annual component 1 cycle/year
>Z-freq: 2 cpy	Annual component 2 cycles/year

The effect of the New CDATA is that the equations of the analysis include a general model of the estimated drift in Figure 12.1, composed by unknown constant and two annual periodic components with frequencies 1 and 2 cpy (cycles/year). The unknown constant will represent the mean sea level, obtained by taking into account, i.e. by eliminating, the tidal components of the data, as well as the annual components.

The result from the analysis is represented on Figure 12.2, drawn by using the file **drift.dat**. The curve denoted as "constant + annual components" is approximation of the drift curve by all components (column "drift_adj" in **drift.dat**). The curve denoted "constant term" represents only the polynomial component of power 0 (column "polynom" in **drift.dat**), which is actually the estimated mean sea level. The grey line representing the drift is taken from column "drift_obs" in **drift.dat**.

The file **drift.dat** should be kept under, say, the name **drift_mem.dat** in order to draw later the drift curve in Figure 12.4.

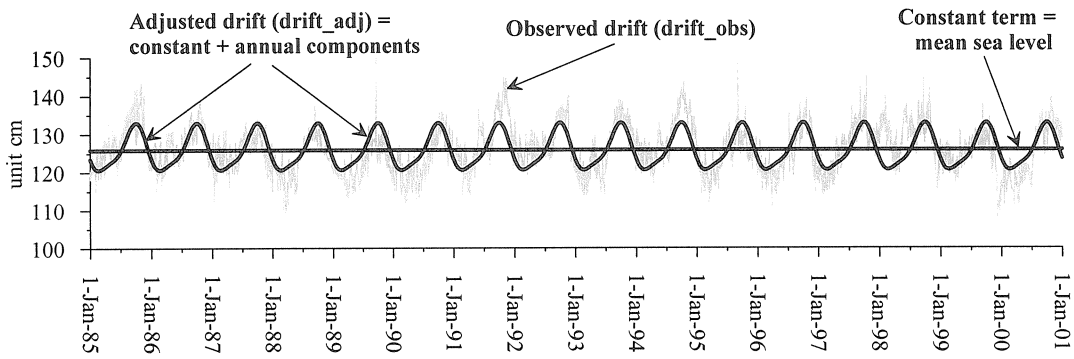


Figure 12.2. Approximation of the drift by constant and annual components.

A careful study can show us that we have some points of the curve in Figure 12.2, where the drift is changing its general behavior. Hence, a better approximation can be obtained by partition the data in several segments and experiment an approximation by polynomials, different in the different segments.

This can be done by using

New CDATA	Comments
>F-T-window: 24	Chooses time window of 24 hours
>F-k-drift-polyn: 0	Drift as stepwise function
>GR-LP 1	Determines the LP tides in one tidal group
>GR-D-SD: 50	Variant 50 of D-SD tidal groups
>Z-k-degree-polyn: 3	Polynomials of power 3 will be applied
>Z-segm-time: 1989 5 20 0	The data are partitioned in 4 segments by the selected dates and in every segment the drift is approximated by polynomial of power 3
>Z-segm-time: 1994 11 10 0	
>Z-segm-time: 1997 10 15 0	
>Z-freq: 1 cpy	The same annual components as those, used above
>Z-freq: 2 cpy	

The result is represented in Figure 12.3. We have certainly a better approximation but we loose the possibility to derive a constant mean sea level.

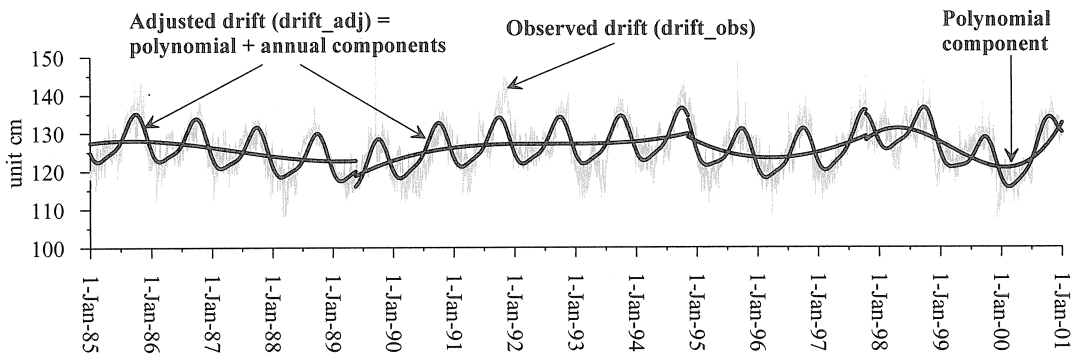


Figure 12.3. Approximation of the drift by polynomials of power 3 in 4 separate segments and annual components of 1 & 2 cpy.

The curve "polynomial + annual components" is taken from column "drift_adj" and "polynomial component" from column "polynom" in **drift.dat**.

It is of course interesting to predict the zero line. For this purpose it is not convenient to use the sophisticated approximation in Figure 12.3, because we have not the guarantee that the changes will go in the same direction. It is more reasonable to remain at the simple case in Figure 12.2, with the hope that the mean constant level and the annual components will keep their behavior.

Now we have to return to the New CDATA as those, used for Figure 12.1, with some additional options, namely we have to use

New CDATA	Comments
>F-T-window: 24	Chooses time window of 24 hours
>F-k-drift-polyn: 0	Drift as stepwise function
>GR-LP 1	Determines the LP tides, 1 tidal group
>Z-k-degree-polyn: 0	Polynomial power 0 = mean sea level
>Z-freq: 1 cpy	Annual component 1 cycle/year
>Z-freq: 2 cpy	Annual component 2 cycles/year
>E-t-int 1995 1 1 0 2002 1 1 0	Analysis of the data till the end of 1994
>OUT-predict: 1995 1 1 0 2000 12 31 23	Prediction till the end of 2000

The result of the prediction of the zero line is shown by Figure 12.4.

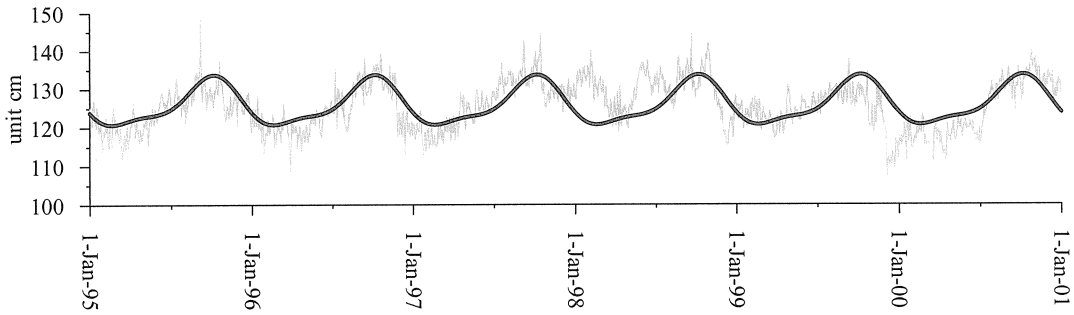


Figure 12.4. San Juan, observed drift (grey line), and predicted zero-line (thick line), obtained by using 10 years data from 1.01.1985 till 31.12.1994.

The observed drift is in column "drift" in **drift_mem.dat**, kept earlier, and the predicted zero-line – in new column "pre_zero_line" in **signal.dat**.

Through this application of VAV we get now in the column "pred_sig" of **signal.dat** predicted tidal data = predicted tidal signal + predicted zero line, i.e. predicted tidal signal + predicted drift. A sample of the result is shown by Figure 12.5, drawn by using the file **vav_out.dat** for the existing data and predicted data in column "pred_sig" in **signal.dat**.

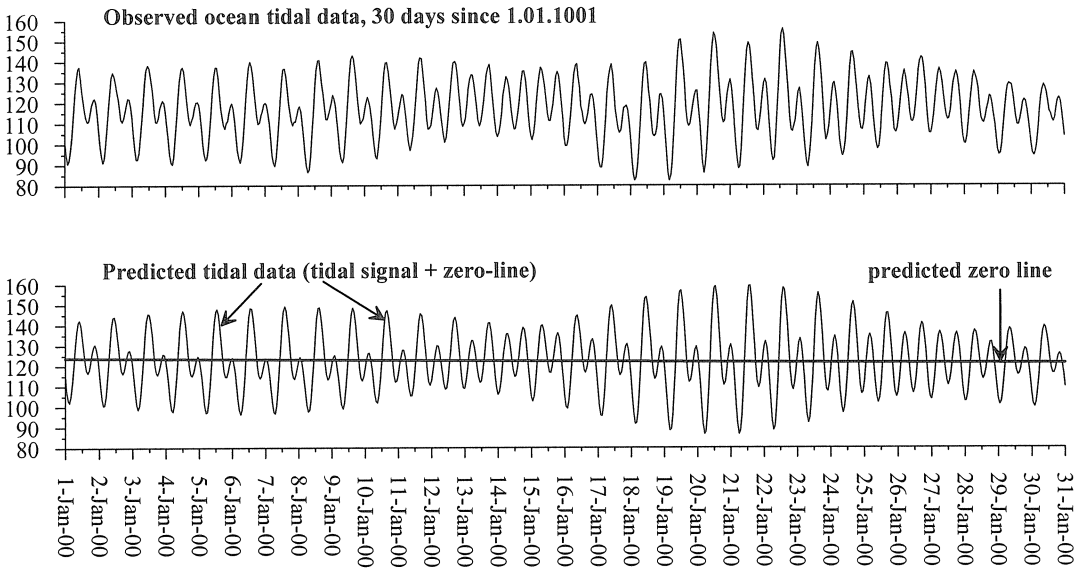


Figure 12.5. San Juan, predicted tidal data = predicted signal + predicted zero-line.

Notice, that the inclusion of the predicted drift suppress the systematic difference in the general level between observed and predicted data in Figures 11.1 & 11.2.

Example 13. Determination of high frequency tides, data San Juan

Under high frequency (HF) tides we shall understand tides or whatever waves, with frequencies higher than the frequencies of the SD tides. Theoretically, most important HF are the TD (third-diurnal) tides with frequencies near 3 cpd and the QD (quarter-diurnal) tides with frequencies near 4 cpd.

The development of Tamura provides 1198 tides, including a set of TD and QD tides with frequencies till 4 cpd. All of them have theoretical amplitudes and phases and we can define and look for their parameters δ and κ.

Our table of this development in **tamura.inp** includes additional HF tides S4, S5, ... S11 with frequencies 4, 5, ... 11 cpd respectively. They are harmonics of the main meteorological tide S1 whose period is exactly 24 hours. This meteorological S1 does not exactly coincide with the theoretical S1 in **tamura.inp** .

The file **tamura.inp** also includes additional tides M5, M6, ... M11 with frequencies 5, 6, ... 11 cpld (cycles/lunar day). They are harmonics of the D lunar tide with period 1 lunar day, denoted as M1X in **tamura.inp**.

Traditionally, we use the acronyms D, SD, TD and QD for sets of tides, which are around 1, 2, 3 & 4 cpd or cpld. The use of HF, which may go till 11 cpd (even much further for minutes data), needs new acronyms. It is convenient a set of tides, which are around K cpld and/or K cpd, to be denoted by DK. Thus D, SD, TD & QD may be replaced by D1, D2, D3 & D4 respectively. Practically more important is that in such a way the sets M5, S5, ... M11, S11 can be considered as 7 sets of tides denoted by D5, ... D11 respectively.

The default case of filter's frequencies in VAV is $\Omega = 1, 2, ... 6$ cpd. This set allows the determination of D1 through D6, but not the higher frequencies. If we want also to determine the D7 through D11 waves, we need to raise the filter frequencies till 11 cpd.

The frequencies and the phases of the additional tides S4, M5, S5, ... M11, S11 are defined like the other tides through the argument coefficients of Doodson. However, they have not theoretical amplitudes and we cannot define the amplitude factor δ. Hence we cannot deal with the unknowns ξ and η (see Section B). Instead, for tides of such kind, VAV uses as unknowns the observed amplitudes and phases. Everyone of the tides is considered as a group, composed by a single tide and at the output we get an amplitude and a phase, but not the usual δ and κ. This is possible because S4, M5, S5, ... M11, S11 have considerably different frequencies, allowing their separation.

We shall determine all HF tides, including S4, M5, S5, ... M11, S11 in the ocean data of San Juan by using

New CDATA	Comments
>GR-D-SD: 50	Options, which remain the same as before.
>F-T-window: 24	
>F-k-drift-polyn: 0	
>GR-LP: 1	
>GR-TD: 7	Determines TD separated in 7 tidal groups
>GR-QD: 7	Determines QD separated in 7 tidal groups, including S4
>F-Highest-freq: 11 cpd	VAV will use the filter frequencies 1, 2, ... 11 cpd which makes possible the determination of all HF waves, till D11
>GR-HF: 2	Includes in the analysis the tides M5, S5, ... M11, S11

The effect of **>GR-TD: 7** is that the TD tides are separated in 7 groups. One of them is the group S3, composed by a single, very small theoretical tide S3. The result we shall obtain about S3 is actually for a meteorological tide S3 with period 8 hours.

The effect of **>GR-QD: 7** is that the QD tides are separated in 7 groups. One of them is the group S4, composed by the added tide S4, which is actually a meteorological tide with period 6 hours.

The results in **analysisNN.dat** about TD and QD are, as usually, δ and κ about the selected tidal groups and amplitudes of the main tide in every group. An exception is S4, having only amplitude.

In Table 13.1 the results about the TD tides are given, but only the amplitudes, which are more interesting in the case of ocean data.

The names of groups MO3, MK3 and SK3 are the names of the shallow water waves whose frequencies enter in the corresponding frequency interval of the group. The main tides in **tamura.inp** with these names do not coincide exactly with the actual MO3, MK3 and SK3.

Table 13.1. San Juan, TD (third-diurnal) tides determined when the option **>GR-TD: 7** is used.

Frequency interval of the tidal group in cpd	Number of tides in the group	Name of the group	Amplitude of the main tide in cm	MSD of the amplitude in cm
2.75324 - 2.86971	36	MO3T	0.0177	±0.0067
2.89264 - 2.89826	6	M3к	0.0219	±0.0070
2.89841 - 2.90389	7	M3	0.0296	±0.0071
2.92711 - 2.94033	14	MK3T	0.0499	±0.0061
2.96599 - 2.97191	5	S3X	0.0009	±0.0064
3.00000 - 3.00000	1	S3	0.0780	±0.0070
3.00305 - 3.08125	13	SK3T	0.0007	±0.0008

The results (amplitudes only) about the QD tides are given in Table 13.2. All groups have the names of shallow water waves, used in the same way as in the case of the TD groups.

Table 13.2. San Juan, QD (quarter-diurnal) tides determined when **>GR-QD: 7** is used.

Frequency interval of the tidal group in cpd	Number of tides in the group	Name of the group	Amplitude of the main tide in cm	MSD of the amplitude in cm
3.79196 - 3.79196	1	N4	0.0203	0.0059
3.79682 - 3.79682	1	3MS4	0.0096	0.0059
3.82826 - 3.83311	2	MN4	0.0726	0.0058
3.86440 - 3.86455	2	M4	0.1609	0.0058
3.90084 - 3.90146	2	KN4	0.0049	0.0026
3.93775 - 3.93790	2	MK4	0.0203	0.0054
4.00000 - 4.00000	1	S4	0.0622	0.0058

Further, due to the options **>F-Highest-freq: 11 cpd** and **>GR-HF: 2** we get the amplitudes of the D5, D6, … D11 tides. The results are shown in Table 13.3.

One of the general results is that the meteorological S5, S6, S7 & S9 have significant amplitudes. Another one is that the even lunar harmonics M6 and M8 are clearly significant, as well as M4 in Table 13.2. It is interesting that the amplitude of M4 is higher than the amplitude of M3, while the theoretical amplitude of M4 is considerably lower than M3. Also, if M6 and M8 were derived from the tidal potential, they would have considerably lower theoretical amplitudes than M3.

All odd lunar harmonics, with the exception of M3 in Table 13.1 and a doubtful exception of M5, are not significant. Not significant are also the highest frequencies 9, 10 and 11 cpd, may be because they are very close to the Nyquist frequency of 12 cpd.

Table 13.3. San Juan, high frequency tides in the frequency domains D5, D6,…D11, determined when the options **>F-Highest-freq: 11** and **>GR-HF: 2** are used.

Frequency in cpd	Tidal Name	Amplitude In cm	MSD of Amplitude	Significant or not sign.
4.83068	M5	0.0085	0.0043	Yes ?
5.00000	S5	0.0811	0.0043	Yes
5.79682	M6	0.1264	0.0042	Yes
6.00000	S6	0.0410	0.0042	Yes
6.76296	M7	0.0023	0.0035	No
7.00000	S7	0.0342	0.0035	Yes
7.72909	M8	0.0474	0.0035	Yes
8.00000	S8	0.0048	0.0035	No
8.69523	M9	0.0030	0.0033	No
9.00000	S9	0.0308	0.0033	Yes
9.66137	M10	0.0025	0.0032	No
10.00000	S10	0.0029	0.0032	No
10.62750	M11	0.0044	0.0032	No
11.00000	S11	0.0047	0.0031	No

Example 14. Shallow water tides, data San Juan.

Shallow water (ShW) tides can be defined by their frequencies but they have not theoretical amplitudes. Due to this VAV can deal with them in a way, similar to S4, M5, S5, … in Example 13. I.e., each ShW tide shape a group of one tide and VAV estimates its amplitude and phase.

The user can define a set of shallow water (ShW) tides by the input file **shallow.inp** in **\aaavav_03**. A sample of **shallow.inp**, prepared for this example, is given in Table 14.1. After a run of VAV using **shallow.inp**, we get a list of the ShW tides in the file **shallow_out.inp**, also in **\aaavav_03**.

Table 14.1. Sample of the content of **shallow.inp**; the complete file contains 107 ShW tides.

SD or D2 shallow water tides
OQ2= O1 + Q1
MNS2= M2+N2-S2
OP2= O1 + P1
MKS2= M2+K2-S2
LDA2= 2 1 -2 1 0 0
KJ2= K1 + J1
2SM2= 2*S2 - M2
TD or D3 shallow water tides
NO3= N2 + O1
MO3= M2 + O1
M3= 3*M1X
NK3= N2 + K1
SO3= S2 + O1
MK3= M2 + K1
SP3= S2 + P1
S3= 3*S1
SK3= S2 + K1

. . . . . .

D6 shallow water tides
2NM6= 2*N2 + M2
ST12= 0.2387380574 cph
2MN6= 2*M2 + N2
ST13= 0.2402502093 cph
ST41= 0.2413060429 cph
M6= 3*M2
MSN6= M2 + S2 + N2
MKN6= M2 + K2 + N2
ST42= 0.2441279756 cph
2MS6= 2*M2 + S2
2MK6= 2*M2 + K2
NSK6= N2 + S2 + K2
2SM6= 2*S2 + M2
MSK6= M2 + S2 + K2
S6= 3*S2

. . . . . .

end

In every line we have at the 1st place the name of the ShW tide (can be arbitrary name of 1 to 4 letters), followed without a blank by "=". Then we have the definition of the corresponding tide in one of the following 3 ways.

(i) As a linear combination of some D and SD tides.

Such are most of the cases in Table 14.1, as well as in **shallow.inp**. It is possible to use combinations of one to 5 tides, whose names exist in our **tamura.inp**. The star in these expressions denotes multiplication; the blanks after "=" are ignored. When given ShW is defined in this way, VAV takes the argument coefficients of the indicated D and SD tides, computes the corresponding combinations and thus the argument coefficients of the new ShW tide are obtained.

One particular moment is the D tide M1X. This is a small tide with Doodson argument number 155.555, i.e. this is an exactly lunar D tide. Thus the tides defined as K*M1X are exactly lunar harmonics. Our **shallow.inp** includes all M tides, used in Example 13, defined as 3*M1X, 4*M1X, ... 11*M1X.

Another particularity is the use of S1. In the development of Tamura S1 is the tide 164.556 which is slightly different from the meteorological wave with frequency just 1 cpd. In the expressions above S1 is considered as 164.555 whose frequency is just 1 cpd, so that the waves defined as K*S1 are exactly meteorological harmonics. Our **shallow.inp** includes all S tides, used in Example 13, defined as 3*S1, 4*S1, ... 11*S1

(ii) By the argument coefficients.

E.g. **LDA2= 2 1 -2 1 0 0** defines a wave with number of Doodson 263.655.

(iii) Directly by the frequency of the ShW tide.

E.g. **ST12= 0.2387380574 cph** defines **ST12** as a wave with frequency 0.2387380574 cycles/hour. It is possible to replace "cph" by "cpd" or "deg/hr" when the frequency is given in cycles per day or degrees/hour respectively.

In order to introduce in the analysis the ShW tides from **shallow.inp** we shall use

New CDATA	Comments
>GR-D-SD: 50	Options, which remain the same as before.
>F-T-window: 24	
>F-k-drift-polyn: 0	
>GR-LP: 1	
>GR-TD: 0	The zeros exclude from the analysis these tidal groups, used in Example 13, because we shall use ShW tides in the same frequency domains.
>GR-QD: 0	
>GR-HF: 0	
>F-Highest-freq: 11 cpd	Same as before
>GR-Shallow: 3 11	Will include in the analysis the ShW tides in the frequency domains D3, D4, ... D11, i.e. all of them, without D2=SD.

The option **>GR-Shallow: 3 11** includes in the analysis 87 of the tides, defined by our file **shallow.inp**. This increases the number of unknowns by 174, namely 87 amplitudes and 87 phases.

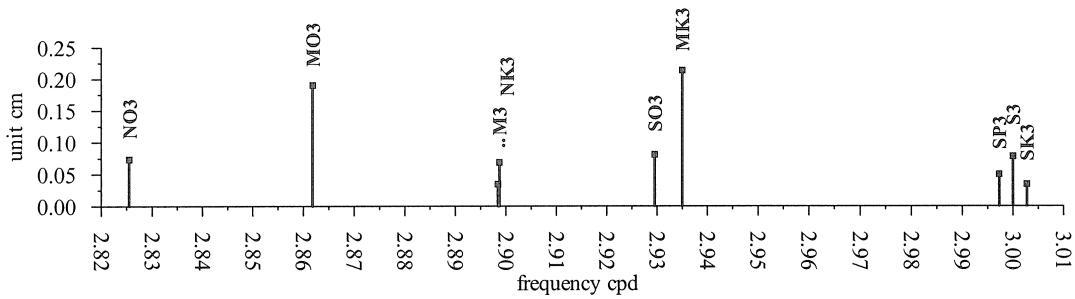


Figure 14.1. Shallow water tides in D3 frequency domain; estimated amplitudes from file **all_tides.dat**, column "amplitude".

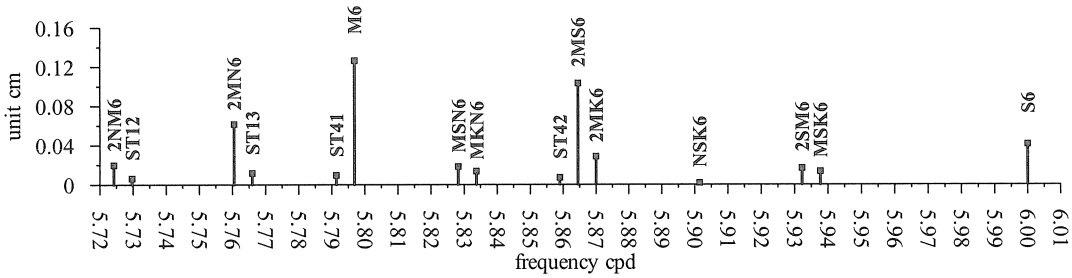


Figure 14.2. Shallow water tides in D6 frequency domain; estimated amplitudes from file **all_tides.dat**, column "amplitude".

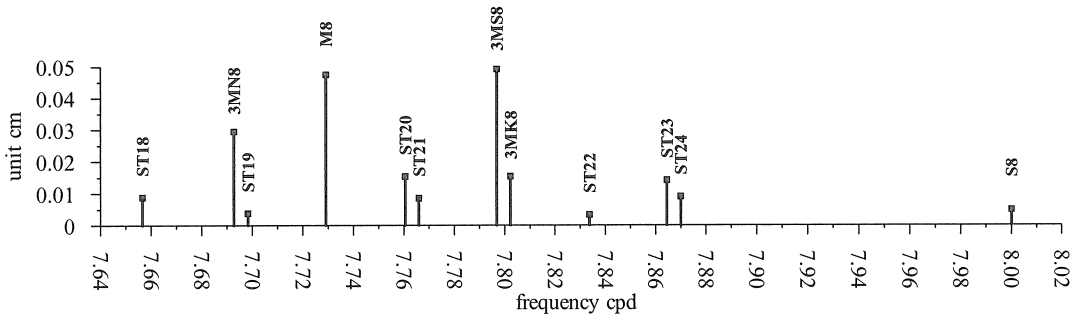


Figure 14.3. Shallow water tides in D8 frequency domain; estimated amplitudes from file **all_tides.dat**, column "amplitude".

After running the program with this input and selecting the data San Juan, we get the results in the corresponding **analysisNN.dat**, as well as in **all_tides.dat**. The ShW tides in **analysisNN.dat** are represented only by amplitudes.

The detailed output in **all_tides.dat** provides the observed amplitudes and phases. The phases are relative to the meridian of the observation point and the time origin used. It is given in the head of the output data. We are ready to introduce other definition of the phases under requests of the users.

A sample of the output in **all_tides.dat** is represented by figures 14.1, 14.2 & 14.3.

In a next run of VAV we may get synthesized or predicted signal of the ShW tides by using **all_tides.dat**. E.g. the D3 signal in January 1985 will be obtained through:

New CDATA	Comments
>OUT-predict 1985 01 01 00 1985 01 31 00	The signal will be created for January, 1985
>W-all-tides: 3 3 cpd	Input of the D3 amplitude and phases from **all_tides.dat** and creating the D3 signal.

The results are now displayed in **analysisNN.dat** in a table under the head

Predicted signal by using file all_tides.dat

Figure 14.4 represents this output.

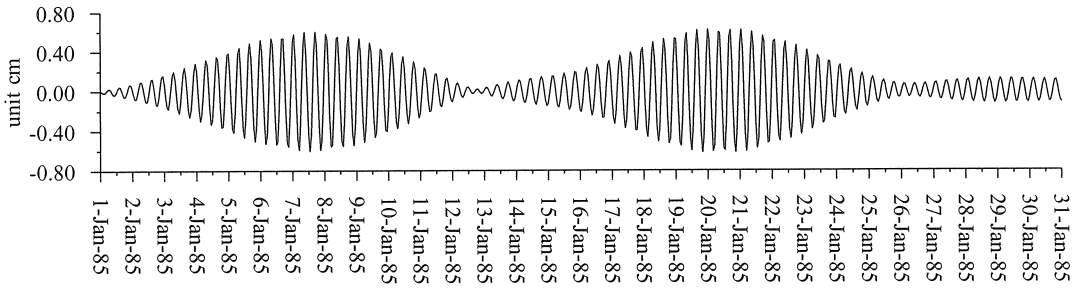


Figure 14.4. Predicted ShW signal in D3 domain; option **>W-all-tides: 3 3 cpd**.

The next figures 14.5, 14.6 & 14.7 are obtained through the consecutive replacement of **>W-all-tides: 3 3 cpd** by the options, shown in the figures.

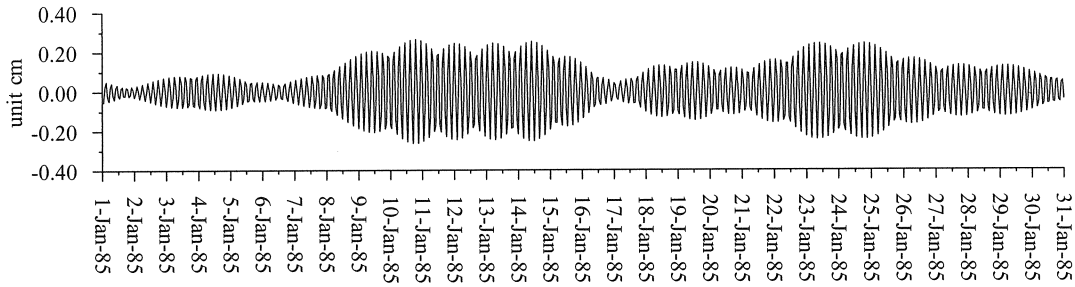


Figure 14.5. Predicted ShW signal in D6 domain; option **>W-all-tides: 6 6 cpd**.

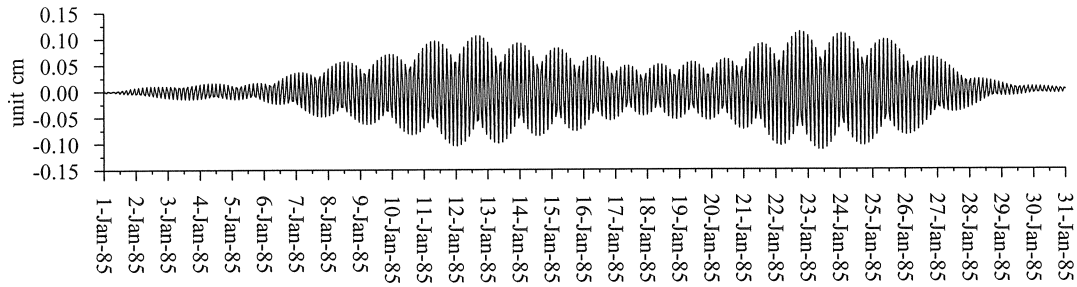


Figure 14.6. Predicted ShW signal in D8 domain; option **>W-all-tides: 8 8 cpd**.

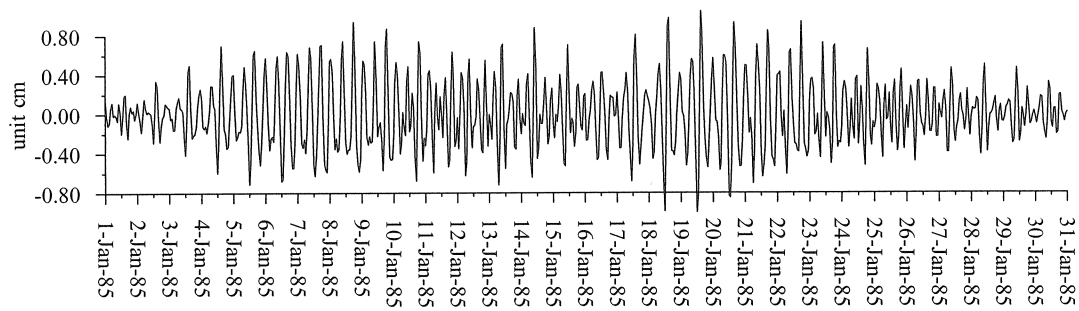


Figure 14.7. Predicted ShW signal in all HF domains D3, D4, … D11; option **>W-all-tides: 3 11 cpd**.

By the way, if we have used **>W-all-tides: 0 11 cpd**, we would get the predicted curve of the total tidal signal.

Example 15. Looking for new waves, data San Juan.

Under new waves or new frequencies we understand waves with frequencies, which do not exist in the development of the tide-generating potential. In this sense they can also be called non-tidal waves, respectively non-tidal frequencies. According to this definition, non-tidal are all waves, which have been determined by Example 13 and the ShW in Example 14. The advantage of the option used here with respect to the techniques, used in the previous example are: (i) we may look for ShW in some frequency intervals, without preliminary fixation of their frequencies and (ii) we may look for any frequency, which may be a signal, nothing to do with the tidal signals.

The general form of the option to be added is

>W-non-tid-freq-interval: ω_a ω_b $\Delta\omega$ cpd

If "cpd" is added, as above, then ω_a ω_b and $\Delta\omega$ are frequencies in cpd; if "cpd" is not added - ω_a ω_b and $\Delta\omega$ are frequencies in degrees/hour, i.e. angular frequencies.

The effect is that we get a variable frequency

$$\omega = \omega_a + k\Delta\omega \quad \text{where } k = 0, 1, 2, \ldots$$

which vary in the interval $(\omega_a \ \omega_b)$ by a step $\Delta\omega$.

The values of ω are included one by one, independently in the analysis. For each ω VAV creates a group, composed by a single wave with frequency ω and unknown amplitude and phase. Thus the equations, which are composed for the usual tidal groups, now are raised by two new equations. The new system is solved and we get estimates of the elements of the new wave.

Then ω is replaced by a new frequency, increased by $\Delta\omega$. The process continues until all values in the interval $(\omega_a \ \omega_b)$ with step $\Delta\omega$ are exhausted.

This option seems to be similar to the spectral analysis but actually it is not the same. The spectral analysis will estimate the amplitude at given frequency ω without taking into account the existence of strong signals at the tidal frequencies. Our procedure takes into account the main tidal signal, because the main tidal constituents remain in the system of equations.

Now we shall again analyze the San Juan data by using

New CDATA
>F-T-window: 24
>F-k-drift-polyn: 0
>GR-LP: 1
>F-Highest-freq: 11 cpd
>GR-HF: 2
>E-T-interval: 1985 1 1 0 1999 12 31 23
>W-non-tid-freq-interval: 5.7 6.1 0.001 cpd

Since we need the creation and the solution of many systems of equations, this operation is slow. In particular for such large series as San Juan, we have to use very small frequency step $\Delta\omega$, i.e. a very great number of frequencies and systems of equations.

Therefore through **>E-T-interval: 1985 1 1 0 1999 12 31 23** we exclude all data before 1.01.2000 and thus we shall deal only with 1 year data: 1.01.2000 – 31.12.2000.

Through **>W-non-tid-freq-interval: 5.7 6.1 0.001 cpd** we shall check the frequency interval from 5.7 cpd till 6.1 cpd with a step 0.001 cpd. This is an interval in which we may expect to find some ShW tides.

This interval covers the tides M6 and S6, which are included, as a default case, in the equations. Since the frequency interval covers M6 and S6, they are automatically excluded from the usual analysis. Otherwise they would interfere with some of the new frequencies and we would get linearly dependent equations.

The results from this processing are stored in the output file **nontid_freq.dat**. For every frequency VAV computes the amplitude and its MSD σ. Thus we get in **nontid_freq.dat** for every frequency the amplitude (column "amplit") and two threshold levels: 1.96σ and 3σ in columns "conf95" and "conf3s" respectively.

Figure 15.1 is built up by using **nontid_freq.dat** when the options above are used. We have got a pick at the frequency of M6, confirming the result in Table 13.3. However, we have got several new picks, i.e. new non-tidal waves, the most important being at frequency 5.8650 cpd. It remains now to compare with the ShW determined earlier.

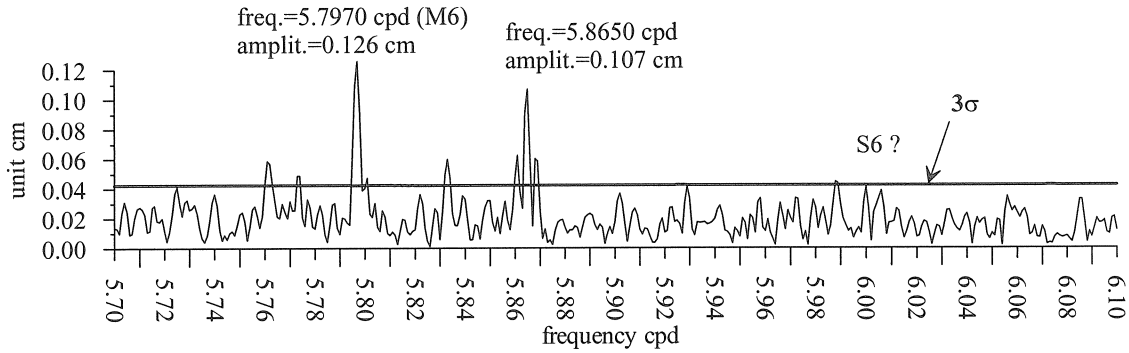


Figure 15.1. San Juan, looking for new waves between 5.7 & 6.1 cpd.

Now, in order to study the spectrum at D8 we shall replace, in the New CDATA above **>W-non-tid-freq-interval: 5.7 6.1 0.001 cpd** by

>W-non-tid-freq-interval: 7.5 8.1 0.001 cpd

The result is presented in Figure 15.2. Now the analysis confirms the result about M8 in Table 13.3 but we have again a new pick, i.e. a new non-tidal wave at frequency 7.797 cpd. It is again interesting to compare with the earlier results about the ShW tides.

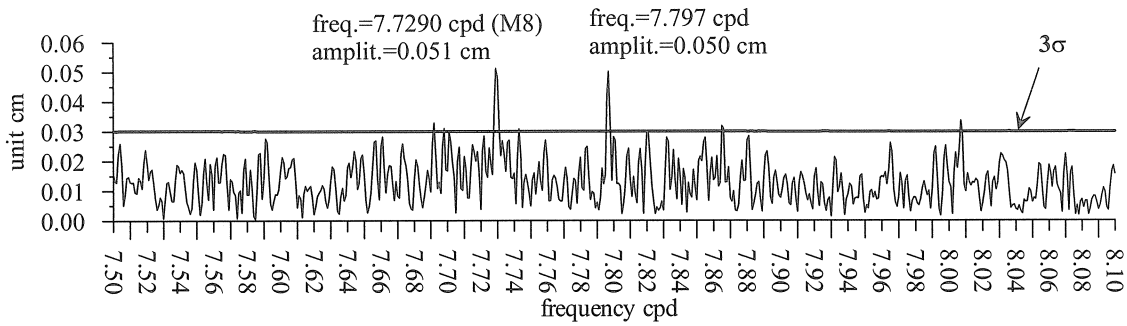


Figure 15.2. San Juan, looking for new waves in the interval 7.5, 8.1 cpd.

Example 16. Time variations of non-tidal waves, data San Juan

VAV can study the time variations of the tidal parameters by partition the data in short segments and getting the results from each segment. As we shall see, the segments can be with or without overlapping.

Now we shall use the data San Juan with

New CDATA
>F-T-window: 24
>F-k-drift-polyn: 0
>GR-LP: 1
>F-Highest-freq: 11 cpd
>GR-HF: 2
>W-non-tid-freq: 5.7 6.1 0.001 cpd
>TV-length-segment-days: 366
>TV-shift-segments-days: 183

Compared to the case in Example 15, we have skipped the **>E-T-interval...** because here we deal with the whole series of data and we have added the last two options **>TV-** related with "Time Variations". Their effect is the following.

The whole series of data is partitioned in segments of length 366 days. The segments are moved by 183 days, i.e. every 2 neighboring segments have 50% overlapping. The data of every segment is analyzed separately and we get for every segment, i.e. for its central epoch the amplitudes in the selected frequency range. Thus we get the amplitudes H as a function of both frequency and time, i.e. $H = H(\omega,T)$ in a time frequency domain. The amplitudes $H(\omega,T)$ are computed together with their MSD $\sigma = \sigma(\omega,T)$. This allows building up confidential intervals and finding points of ω and T where $H(\omega,T)$ is significantly different from zero. I.e. segments at epoch T where the frequency ω is really manifested.

The file **tvar_nontid_2d.dat** represents $H(\omega,T)$ and $3\sigma(\omega,T)$ (very high confidential threshold level) in a table. It can be used to draw 2-D spectra, similar to Figure 15.1, but at a series of epochs, namely at the epochs of all segments.

The file **tvar_nontid_3d.dat** provides $H(\omega,T)$ in column "amplit", $1.96\,\sigma(\omega,T)$ in column "conf_95%" and $3\,\sigma(\omega,T)$ in column "conf_3sg". In columns "damp95%" and "damp3sg" we have the quantities

$$\text{damp95\%} = \begin{cases} \text{amplit} - \text{conf\_95\%} & \text{when amplit} > \text{conf\_95\%} \\ 0 & \text{when amplit} \le \text{conf\_95\%} \end{cases}$$

$$\text{damp3sg} = \begin{cases} \text{amplit} - \text{conf\_3sg} & \text{when amplit} > \text{conf\_3sg} \\ 0 & \text{when amplit} \le \text{conf\_3sg} \end{cases}$$

When we have $\text{damp95\%} > 0$ we may pretend that the amplitude is significant with a risk for error 5%. When we have $\text{damp3sg} > 0$ we may pretend that the amplitude is significant with a very small risk for error.

Figure 16.1 represents the damp3sg as a function of the time/frequency domain, obtained by the application of the New CDATA shown above.

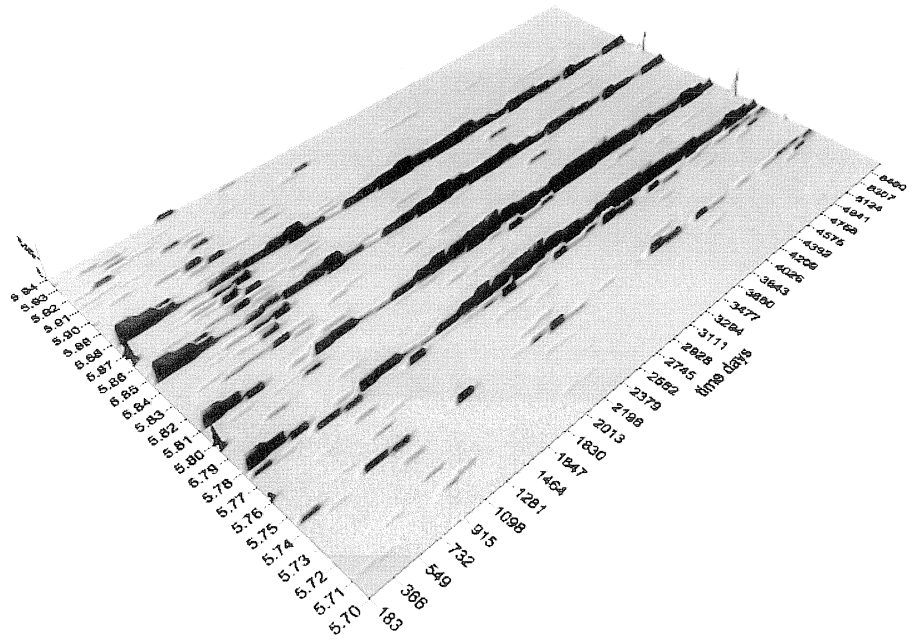


Figure 16.1. San Juan, 3-D picture in a time frequency domain of damp3sg from file **tvar_nontid_3d.dat**, showing where we have significant amplitudes.

Here we have clearly significant amplitudes, rather stable in the time, at the frequencies: 5.78, 5.81, 5.85 & 5.88 cpd and very small amplitudes, which appear and disappear at frequency 5.745 cpd. If we compare with Figure 15.1, we see that here we have a small shift towards lower frequencies; in particular the most important 5.797 cpd of M6 seems shifted to 5.78 cpd.

In order to check we have drawn Figure 16.2, by using again damp3sg from **tvar_nontid_3d.dat**, but by neglecting the time.

A careful study of Figure 16.2 shows that there are not contradictions with Figure 15.1. E.g. the detail, shown on Figure 16.3, allows us to estimate that the pick shown is at or very closed to the frequency 5.797 cpd of M6, and certainly not at the frequency 5.78 cpd of Figure 16.1.

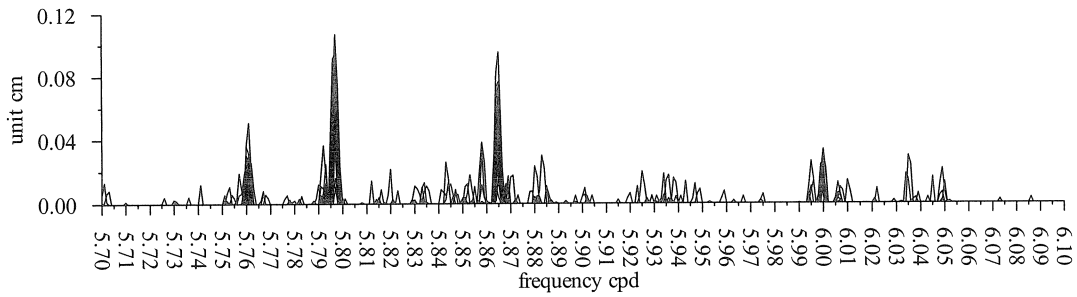


Figure 16.2. San Juan, significant amplitudes, represented by damp3sg>0, taken from **tvar_nontid_3d.dat,** all epochs represented by one curve, .

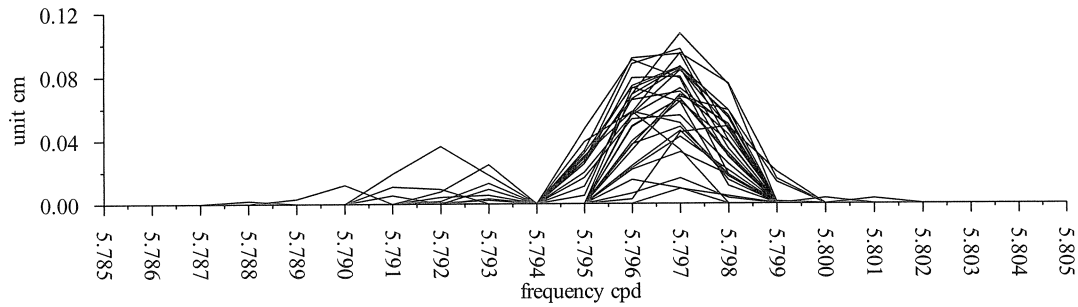


Figure 16.3. Detail of Figure 16.2, confirming the existence of the frequency of M6.

Example 17. Prediction of the tidal signal by using input tidal parameters and corrections to the absolute gravity observations.

In Example 11 we have predicted the tidal signal at a tidal station through the analysis of data, obtained in the same station. At the end of Example 14 we have shown how to predict, by using the output **all_tides.dat**, the tidal signal at given frequency bands, as well as the total signal.

The task of this example is similar, closer to Example 14. The difference is that here we shall use for the prediction an input set of tidal parameters, which may be chosen in an arbitrary way. One way is to use the analysis of a station, which does not obligatory coincides with the point of prediction. Another way is to choose personally the parameters.

One of the main purposes of this computation is to provide tidal corrections to the absolute gravity observations. It is clear that this purpose will be accomplished in the best way when the tidal parameters are obtained by the analysis of tidal gravity data recorded at the point of the absolute observations.

We shall now define and use a new folder in **\a_data**, namely:

$$\backslash a\_data \backslash predict \backslash \begin{cases} \textbf{predict.ini} \text{ (control data file)} \\ \backslash pred\_out\_a \backslash \quad \text{(output folder a)} \\ \backslash pred\_out\_b \backslash \quad \text{(output folder b)} \end{cases}$$

Unlike the other folders in **\a_data** here we have only a CDATA file and we have not a data file.

We shall add to the **a_list_files.inp** in **\aaavav_03** (Table C.2) the information about this folder, i.e. we have to write, wherever between the stations,

\a_data\predict\predict.ini
\a_data\predict\predict.dat
\a_data\predict\pred_out_a
\a_data\predict\pred_out_b

Here **predict.dat** is a dummy name about data that do not exist.

The file **predict.ini**, which will be used, is represented by Table 17.1. It is aimed to predict the tidal signal at the station Strasbourg during one month.

Table 17.1. Control data file **predict.ini** for prediction of tidal data in point Strasbourg; the numbers in parentheses are added here and they do not take part in the input.

Coordinates of Strasbourg:
>ST-long-e: 7.684
>ST-latit-n: 48.6223
>ST-altit-meters: 180.0
>ST-grav-gals: 981
Definition of the interval of prediction 0^h 25.11. – 0^h,25.12.1999:
>OUT-predict: 1999 11 25 00 1999 12 25 00
>D-step-mean-(minutes): 30 *Predicted data every 30 minutes
>W-tidal-param: yes

(1)		(2)	(3)		(4)	(5)	(6)	(7)	(8)	(9)
0.00000	:	0.00000	1.000		0.000					
0.00015	:	0.24995	279	MF	50.384	0.136	1.14315	0.00308	0.041	0.155
0.72150	:	0.83311	49	SGQ1	2.624	0.018	1.15266	0.00775	-0.702	0.386
0.85118	:	0.85969	23	2Q1	8.985	0.019	1.15066	0.00247	-0.771	0.123
0.86090	:	0.87002	20	SGM1	10.827	0.020	1.14892	0.00207	-0.619	0.103
0.88733	:	0.89613	33	Q1	67.689	0.019	1.14707	0.00033	-0.236	0.016
0.89781	:	0.90631	18	RO1	12.857	0.019	1.14707	0.00174	-0.173	0.087
0.92194	:	0.93045	32	O1	353.740	0.019	1.14773	0.00006	0.072	0.003
0.93196	:	0.94049	26	TAU1	4.669	0.018	1.16145	0.00446	0.283	0.220
0.95809	:	0.96676	30	M1	27.886	0.018	1.15043	0.00074	0.150	0.037
0.96857	:	0.97419	18	CHI1	5.343	0.019	1.15251	0.00413	0.594	0.205
0.98905	:	0.99514	7	PI1	9.625	0.019	1.14787	0.00228	0.002	0.114
0.99697	:	0.99803	10	P1	164.684	0.019	1.14836	0.00013	0.262	0.007
0.99985	:	1.00015	4	S1	3.961	0.027	1.16820	0.00811	0.700	0.399
1.00182	:	1.00365	18	K1	492.439	0.019	1.13606	0.00004	0.279	0.002
1.00533	:	1.00562	4	PSI1	4.258	0.019	1.25569	0.00559	-0.471	0.255
.....										
3.79196	:	3.93790	10	M4	0.028	0.004	0.54901	0.08720	219.130	9.100

end *end is necessary to end the input.

The prediction will use the tidal parameters, included in **predict.ini** after the statement **>W-tidal-param: yes.**

From this input data VAV reads only the rows, which have either 9 or 4 numerical data, separated by one or more blanks. The data in columns (1) and (2) define a frequency interval in cpd. The 2 zeros in the first input row include the 2 zero or constant terms in **tamura.inp**. Then VAV takes from every row the amplitude factor δ and the phase lag κ, whish have to be applied on the tides in the frequency interval

In the case of 9 numbers, δ is in column (6) and κ - in column (8). The rows of this kind in Table 17.1 are taken directly from an analysis of the data Strasbourg.

In the case of 4 numerical data, δ is in column (3) and κ - in column (4). Such format of the input is convenient, when we prepare manually the data.

In this set of data we cannot use the non-tidal waves, which have not δ factors. I.e., we have to restrict to the frequencies 4 cpd, but without S4.

When VAV is started, we have to select the data file with the dummy name **predict.dat**, but actually VAV will use only **predict.ini**

The result form the processing is stored in **\predict_out_b**, file **theosig.dat.** The output in **\predict_out_a** has not a particular interest.

Due to the option **>D-step-mean-(minutes): 30**, we get the values of the predicted signal every 2 minutes. We shall get the predicted signal every hour if we use **>D-step-mean-(minutes): 60.** It is possible to get data with a step of some seconds. E.g., **>D-step-mean-(minutes): 2 sec** , where "sec" is changing the default unit, we shall get the data every 2 seconds.

When we dispose by observed data, as in this case, it is interesting to compare them with the predicted signal. This is namely done in Figure 17.1. Generally, there should be a systematic difference, because the predicted signal does not include the drift.

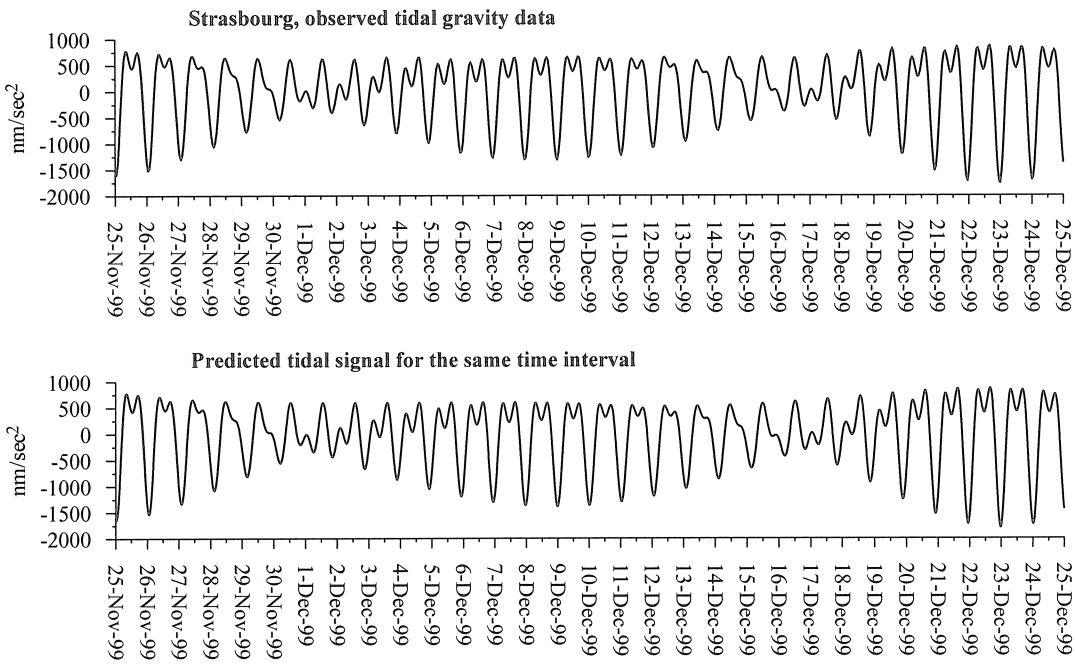


Figure 17.1. Strasbourg, observed data and predicted tidal signal.

Now we shall suppress in **all_ini.inp** and **predict.ini** the options:

>D-origin-of-time: 1982 01 01 00
>OUT-predict: 1999 11 25 00 1999 12 25 00
>D-step-mean-(minutes): 30

Instead we shall use in **predict.ini** the option **>D-abs-grav: 3700**, where 3700 means that this option is followed by no more than 3700 lines with absolute data. This makes active the data in Table 17.2, already included in **predict.ini:**

Table 17.2. Sample of absolute gravity data from Strasbourg, included in **predict.ini.**

```
>D-abs-grav:  3700
 nr    yyyy dd   hr mi se   abs_gravity     sig        nn
  1    1998 265   8  3  0   9.808774601 0.48616E-06  25.
  2    1998 265   8 18  0   9.808774579 0.33712E-06  25.
  3    1998 265   8 33  0   9.808774540 0.22667E-06  25.
  4    1998 265   8 48  0   9.808774408 0.28054E-06  25.
  5    1998 265   9  3  0   9.808774373 0.28307E-06  25.
  6    1998 265   9 18  0   9.808774301 0.43441E-06  25.
  7    1998 265   9 33  0   9.808774249 0.26813E-06  25.
  8    1998 265   9 48  0   9.808774201 0.22865E-06  25.
  9    1998 265  10  3  0   9.808774120 0.13544E-06  25.
 10    1998 265  10 18  0   9.808774057 0.24142E-06  25.
 11    1998 265  10 33  0   9.808774015 0.23549E-06  25.
 12    1998 265  10 48  0   9.808773964 0.17128E-06  25.
 13    1998 265  11  3  0   9.808773968 0.40049E-06  25.
 14    1998 265  11 18  0   9.808773889 0.15855E-06  25.
.....................................................................
439    1999 358   8 30  0   9.808774993 0.20960E-06  49.
440    1999 358   9  0  0   9.808775040 0.21614E-06  49.
end
```

These are absolute gravity observations in Strasbourg (Amalvict et al., 2001), kindly provided to us by Dr M.Amalvict. The data are before applying the tidal corrections. We have some 3700 individual data, obtained in separated series. The column "nr" is consecutive number within every series. It is followed by 5 columns with the date of the observations, including the time in minutes and seconds. Then we have the observed quantity in m/sec^2, i.e. in unit = 100 gal. The remaining data are not used here.

After **>D-abs-grav: 3700** is thus included, we run VAV by choosing **predict.dat.** Thus we get a new output in file **abs_grav.dat**. A sample of this output is shown in Table 17.3.

Table 17.3. Sample of the content in the output file **abs_grav.dat**

```
yyyymmdd hrmise  signal_par    thours      tdays  grav_obs  grav_corrected

input file:
\a_data\predict\predict.dat 30.10.2003, 12h 44m 25sec
In this case the quated input file is a dummy name, not used

yyyymmdd hrmise:date=year, month, day, hour, minute & second
signal_par: predicted tidal signal by using input tidal parameters
thours: relative time in hours
tdays: relative time in days
grav_obs: observed absolute gravity in unit=100 gals.
grav_corrected: corrected absolute gravity=
                              =grav_obs-signal_par, nm/sec**2

yyyymmdd hrmise  signal_par    thours      tdays  grav_obs  grav_corrected
19980922 080300   346.480      0.0000      0.00000 9808774601. 9808774255.
19980922 081800   291.526      0.2500      0.01042 9808774579. 9808774287.
19980922 083300   232.845      0.5000      0.02083 9808774540. 9808774307.
19980922 084800   171.473      0.7500      0.03125 9808774408. 9808774237.
19980922 090300   108.494      1.0000      0.04167 9808774373. 9808774265.
19980922 091800    45.016      1.2500      0.05208 9808774301. 9808774256.
19980922 093300   -17.846      1.5000      0.06250 9808774249. 9808774267.
19980922 094800   -78.989      1.7500      0.07292 9808774201. 9808774280.
.....................................................................................
```

Figure 17.2 represents the results about the first 96 hours of the data in **abs_grav.dat.**

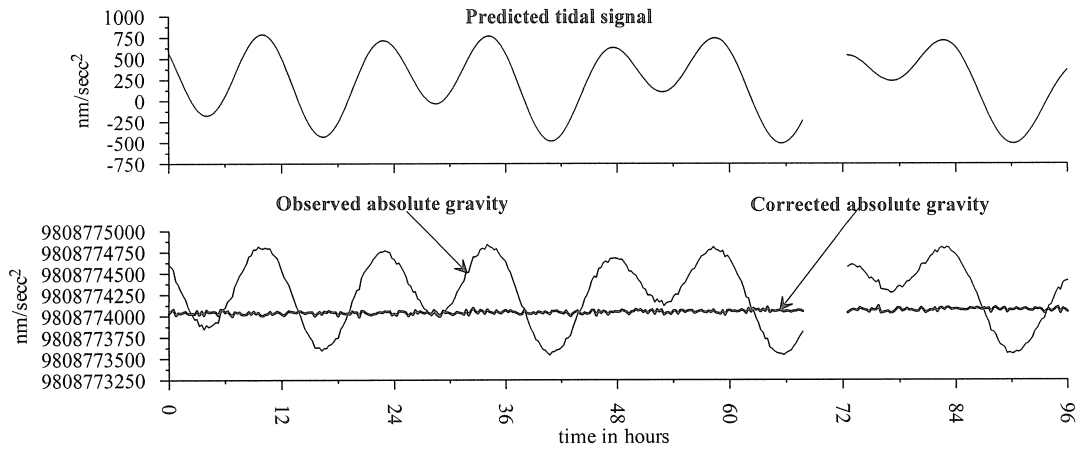


Figure 17.2. Strasbourg, sample of predicted, observed and corrected absolute gravity data.

Figure 17.3 is a demonstration of the effect of the tidal corrections over the whole set of observations, used here.

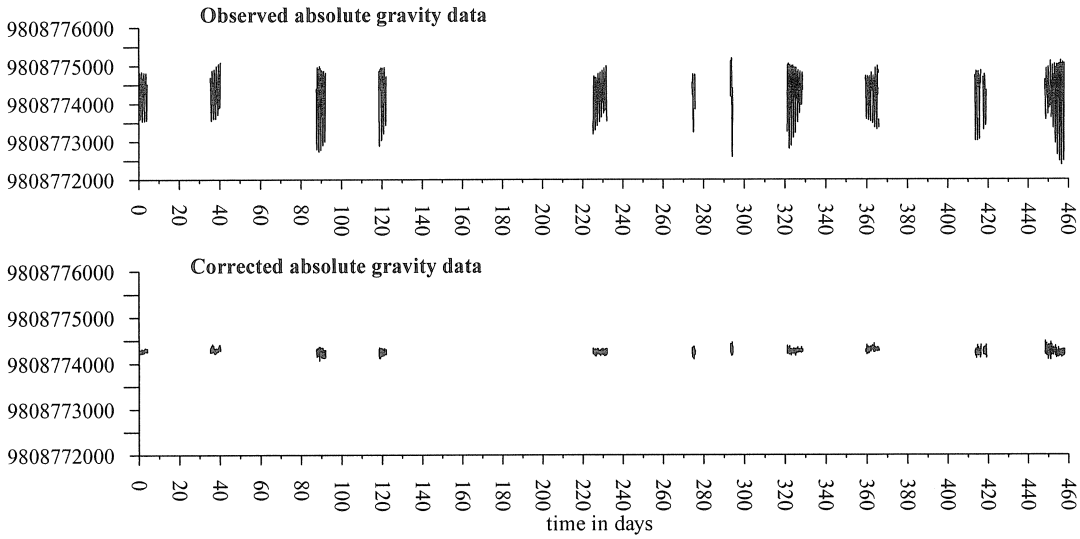


Figure 17.3. Strasbourg, all observed and corrected absolute gravity data.

Example 18. The problem of "aliasing", data Vienna.

There is a misunderstanding (Schüller, 1978, Wenzel, 1997) that our transformation from the time domain in the time frequency domain is a decimation of the data with a time step 48^h (when the time window is $\Delta T = 48$ hours). If it was so indeed, the Nyquist frequency in deg/hr would be $180^o/48 = 3.75^o/h$ or 0.25 cpd, much lower than the Nyquist frequency of $180^o/h$. Then for every tidal frequency ω all frequencies $\omega' = \omega \pm j3.75^o/h$ would be aliases to ω for any integer j. This means that it would be impossible to distinguish between them. I.e., if some aliases are included in the equations, they will become linearly dependent and thus it will be impossible to solve them.

Schüller (1978) has shown that when one presumed alias of O1 is added to the data with amplitude of the order of the amplitude of O1, it affects the results for O1. This is of course not a proof that we have really an alias. It simply shows that such systematic perturbation affects the results. Actually, it will certainly affect the results, obtained by whatever method of analysis. It is curious, that such an experiment has never been applied to other methods, e.g. to ETERNA.

Actually, our algorithm is very close to the direct processing of the hourly data. with Nyquist frequency 180^o per hour and this frequency remains intact after the filtration.

The filters we use are obtained by the transformation of the functions

$$C(\Omega,t) = \mathrm{Exp}(i\Omega t) \qquad (*)$$

where t is time, measured within the filtered intervals or what we call the time window ΔT. For the default case of hourly data and time window $\Delta T = 48^h$, $t = -23.5, -22.5, \ldots +22.5, +23.5$ hours.

In (*) Ω is frequency, which takes a series of discrete values. For the default case of VAV Ω, expressed in cpd, takes the values

$$\Omega = 1,2,3,4,5 \;\&\; 6 \text{ cpd}, \qquad (**)$$

related with the frequency domains, in which we have a tidal signal.

For the particular case of power of the drift polynomials $k = 0$, the filters coefficients are directly proportional to the functions (*), i.e. we have

$$F(\Omega,t) = c.C(\Omega,t)$$

where c is a normalizing coefficient.

For $k > 0$ $F(\Omega,t)$ are obtained through the orthogonalization of $C(\Omega,t)$ with respect to the drift polynomials, as well as between themselves. I.e., in any case, the filters are orthogonal.

This procedure is obtained by (i) creating observational equations for the original hourly data, including the model of the drift. Then (ii) we apply an MLS algorithm for the determination of the signal unknowns, which separates them from the drift unknowns. Thus we get (iii) that the application of MLS on the original unknowns can be replaced by two stages. The first stage consists in the application of orthogonal filters, which transforms the data from the time domain in a time frequency domain and

eliminate the drift. The second stage consists in creation of new observational equations about the filtered numbers and their solution by MLS.

If the sum of number $k+1$ of the polynomial terms and the number of sine and cosine terms in all $C(\Omega,t)$ is equal to the number of data in the filtered intervals, then our analysis is exactly equivalent to the direct processing of the original data. E.g., for hourly data and $\Delta T = 48^h$ the cases

$$k = 0 \text{ with } \Omega = 0.5, 1, 1.5, 2, \ldots 11.5, 12 \text{ cpd}$$
$$k = 2 \text{ with } \Omega = 1, 1.5, 2, 2.5, \ldots 11.5, 12 \text{ cpd}$$

make our analysis equivalent to a direct processing of hourly data (taking into account that for $\Omega = 12$ cpd $C(\Omega,t)$ has not a cosine term).

However, since we have not signals at all of these frequencies, we can abandon some of them. The case (**), designed for tidal data, dominated by the tidal signal is a certain deviation from the direct processing. Nevertheless, its application is not at all decimation with a time step $\Delta T = 48^h$ and the statement "it violates the sampling theorem" (Wenzel, 1997) is, softly said, not correct. By the way, an algorithm, which violates whichever mathematical theorem, is not able to provide any reasonable result.

Now we shall apply VAV on the high quality data of Vienna in 3 variants.

Variant A. We shall analyze the data such as they are.

Variant B. We shall go much further than the experiment of Schüller, by adding at once 4 supposed aliases of O1, called WAL (see Table 18.1) and apply blindly VAV on the deformed data, i.e. without taking into account the WAL.

Table 18.1. Waves, supposed to be aliases of O1, to be added to the data Vienna.

Name	Angular frequency	Period hours	Amplitude
WAL1	$\omega_1 = \omega(O1) - 7.5^\circ =$ 6.4430356067°/hour	55.874	380 nm/sec^2
WAL2	$\omega_2 = \omega(O1) - 3.75^\circ$ =10.1930356067°/hour	35.318	380 nm/sec^2
WAL3	$\omega_3 = \omega(O1) + 3.75^\circ$=17.6930356067°/hour	20.347	380 nm/sec^2
WAL4	$\omega_4 = \omega(O1) + 7.5^\circ$ = 21.4430356067°/hour	16.789	380 nm/sec^2

Variant C. We shall analyze the deformed data, by taking into account the presence of the WAL, i.e. by including the WAL in the observational equations. If WAL and O1 were aliases indeed, i.e. if our algorithm uses decimation, 10 column-vectors of the equations would be linearly dependent, the matrix of the normal equations would be non-invertible and the computation would be interrupted catastrophically.

We shall use the **all_ini.inp** as

>D-origin-of-time: 1982 01 01 00
>out-compare-tide-name: Q1
>out-compare-tide-name: O1
>out-compare-tide-name: K1
>out-compare-tide-name: N2
>out-compare-tide-name: M2
>out-compare-tide-name: S2

For Variant A we shall use directly the Basic CDATA, given in Example 8, Table 8.2, without any elimination of the data. It is important for the next variants to exclude **>D-format: unf**.

For variant B we shall use the options in Table 18.2.

Table 18.2. Options, through which the original data are charged by the "aliases" WAL in Table 18.1.

New CDATA
>D-add-artificial-wave: 6.4430356067 380
>D-add-artificial-wave: 10.1930356067 380
>D-add-artificial-wave: 17.6930356067 380
>D-add-artificial-wave: 21.4430356067 380

The effect of these options is adding the waves, defined in Table 18.1. A sample of the original data is shown in Figure 18.1. Figure 18.2 shows the same data, when the options above are applied. We get a curve, which is very difficult to be recognized as a tidal one.

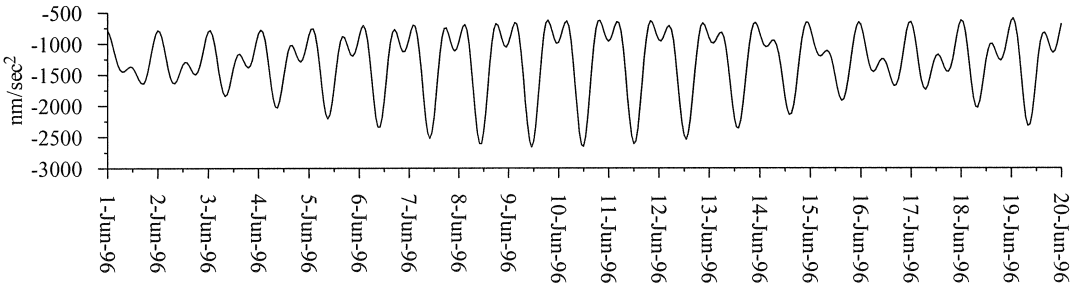


Figure 18.1. Sample of the original data Vienna.

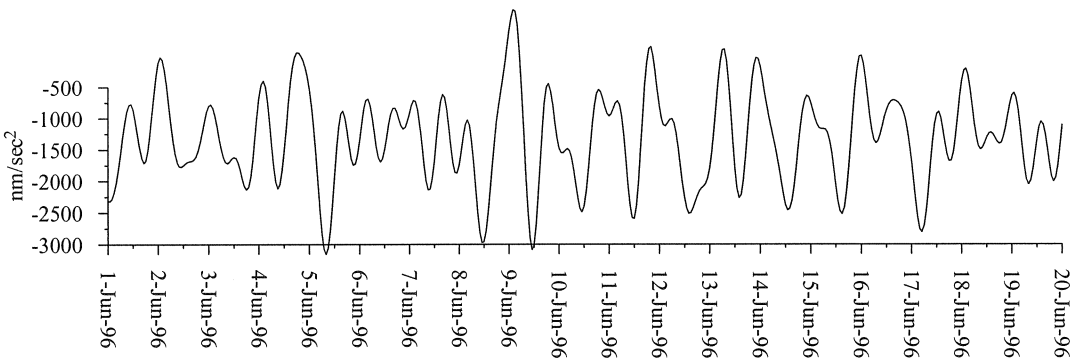


Figure 18.2. Sample of data Vienna with added four "aliases" WAL from Table 18.1.

Finally, for Variant C we shall add to the options in Table 18.2 those in Table 18.3.

Table 18.3. Options, through which we include in the equations of the analysis the waves WAL from Table 18.1.

Options, added to the New CDATA
>GR-new-wave: 6.4430356067 WAL1
>GR-new-wave: 10.1930356067 WAL2
>GR-new-wave: 17.6930356067 WAL3
>GR-new-wave: 21.4430356067 WAL4

The results of the options, used in these 3 variants are given in Table 18.4 and Table 18.5.

Table 18.4. Vienna, results from **a_delta.dat**; analyses by using default variant of the filters of VAV, i.e. time window $\Delta T = 48^h$, power of the drift polynomials $k = 2$ and filter frequencies $\Omega = 1,2,3,4,5$ & 6 cpd.

δ(Q1)	δ(O1)	δ(K1)	δ(N2)	δ(M2)	δ(S2)
Variant A: Original data without WAL, results δ_A & MSD σ_A					
1.14604 ±.00033	1.14796 ±.00006	1.13410 ±.00004	1.17538 ±.00016	1.18107 ±.00003	1.17838 ±.00008
Variant B: Data with WAL, results δ_B & MSD σ_B					
1.12584 ±.22495	1.18805 ±.04313	1.13458 ±.02904	1.17754 ±.03782	1.18175 ±.00738	1.19104 ±.01867
Variant C: Data with WAL, but WAL are included in the equations, results δ_C & σ_C					
1.14605 ±.00033	1.14793 ±.00006	1.13410 ±.00004	1.17538 ±.00016	1.18107 ±.00003	1.17839 ±.00008

Table 18.5. Estimated amplitudes of the added "aliases" WAL of O1 in Variant C, taken from the last **analysisNN.dat**.

Amplitudes in nm/sec^2		
WAL1	379.9489	±0.0811
WAL2	379.9426	±0.0250
WAL3	379.9751	±0.0207
WAL4	380.0249	±0.0829

Provided WAL were aliases, the last Variant C would be impossible. Surprisingly or not, we get results, practically identical to Variant A. Something more, as shown by Table 18.5, we get fairly good estimates of the WAL themselves.

The conclusion is that WAL are not aliases and, more generally, VAV does not use a decimation of the data.

It is not surprising that in Variant B we get strongly affected results. The effect is strongest on δ(O1), which is similar to the result of Schüller, but Variant C disproves his suppositions. Actually, our model of the drift has discontinuities every ΔT hours, which gives a great flexibility of the model. In the same time this can create a kind of a resonance between the WAL used and some tides, in this case with O1. We shall return to this moment in the next Example 19.

It is curious that actually, from a statistical point of view, the results Variant B are not good but in the same time they are correct. For example for $\delta(O1) = 1.18805 \pm 0.04313$ we have a 95% confidential interval (1.103515, 1.27258) which is largely covering the correct value $\delta(O1)=1.14796$ in Variant A. I.e., we cannot be mislead to believe that $\delta(O1)$ is significantly different from the correct $\delta(O1)=1.14796$.

Example 19. Check the aliasing problem in data Brussels without gaps and a comparison of VAV with ETERNA.

The purpose of this Example 19 is to check the aliasing problem on a larger set of data without gaps. We shall also compare the effect of the waves WAL, defined by Table 18.1 on the VAV and ETERNA analyses.

Now we shall use data without gaps (the data Vienna have a few gaps) for two reasons. First, provided the waves WAL are aliases for VAV, their effect should be stronger when there are not gaps. Second, ETERNA does not support so well the gaps as VAV and we would not like to exploit this situation. I.e., we do not intend to use VAV in a more favorable situation. In the same time, since we shall make a comparison, we shall apply VAV in its more powerful variant of the filters, which should be used when we suspect strong signals outside the tidal frequency bands.

Now we shall apply VAV on the data Brussels in just the same Variants A, B & C as in Example 18, but we shall add

New CDATA	Comments
>E-T-interval: 1982 1 1 0 1987 1 1 0	Select a part of the data without gaps.
>F-k-drift-polyn: 0	Drift, approximated by stepwise function
>GR-LP: 1	Includes the LP tides in one group
>F-lowest-freq: 0.5 cpd	The filter frequencies to be used are:
>F-delta-freq: 0.5 cpd	$\Omega = 0.5, 1, 1.5, 2, \ldots 11, 11.5$ cpd,
>F-highest-freq: 11.5 cpd	i.e. all available frequencies.

Samples of the results are given in Table 19.1.

Table 19.1. Brussels, VAV results from **a_delta.dat**; analysis by using filters with time window $\Delta T = 48^h$, power of the drift polynomials $k = 0$, one group of LP tides and filter frequencies $\Omega = 0.5, 1, 1.5, 2, \ldots .11., 11.5$ cpd.

δ(Q1)	δ(O1)	δ(K1)	δ(N2)	δ(M2)	δ(S2)
Variant A: Analysis of the original data, results δ_A & σ_A					
1.15012 ±.00043	1.15317 ±.00008	1.14001 ±.00006	1.16907 ±.00025	1.18375 ±.00005	1.19633 ±.00011
Variant B: Analysis of data charged by WAL, results δ_B & σ_B					
1.14750 ±.03727	1.16345 ±.00713	1.13988 ±.00505	1.17026 ±.03214	1.18414 ±.00619	1.19929 ±.01374
Variant C: Data charged by WAL, but WAL included in the equations, results δ_C & σ_C					
1.15012 ±.00043	1.15316 ±.00008	1.14001 ±.00006	1.16907 ±.00025	1.18375 ±.00005	1.19633 ±.00011

We can simply repeat the conclusions, concerning the results in Table 18.4. It can only be added, that now the deviations in Variant B from Variant A are less important than in Table 18.4.

Table 19.2 shows the results by the application of ETERNA on the same data in Variants A & B and Table 19.3 – a comparison between the VAV and the ETERNA results. We could not apply variant C because we do not know about the existence of such option of ETERNA.

Table 19.2. Brussels, ETERNA results.

δ(Q1)	δ(O1)	δ(K1)	δ(N2)	δ(M2)	δ(S2)
Variant A: Analysis of the original data, results δ_A & σ_A					
1.15073 ±.00037	1.15324 ±.00007	1.13987 ±.00005	1.16886 ±.00022	1.18376 ±.00004	1.19604 ±.00009
Variant B: Analysis of data charged by WAL, results δ_B & σ_B					
1.13759 ±.00147	1.14229 ±.00029	1.12915 ±.00021	1.15661 ±.00022	1.17258 ±.00004	1.18510 ±.00009

Table 19.3. Comparison of the effect of the WAL on the VAV and ETERNA results.

Q1	O1	K1	N2	M2	S2
Program VAV, differences between Variant A and Variant B $\delta_A - \delta_B$					
+0.00262	-0.01028	+0.00013	-0.00119	-0.00039	-0.00396
Program ETERNA, differences between Variant A and Variant B $\delta_A - \delta_B$					
+0.01276	+0.01095	+0.01072	+0.01225	+0.01118	+0.01094
Ratio $R_{E/V} = \lvert\delta_A - \delta_B\rvert_{ETERNA} / \lvert\delta_A - \delta_B\rvert_{VAV}$					
4.87	1.07	82.5	10.3	28.7	2.76
ETERNA, ratio $\theta_{AB} = \lvert\delta_A - \delta_B\rvert / \sigma_B$					
8.68	37.76	51.05	55.68	279.5	121.6

As shown by Table 19.3, the effect of the WAL on ETERNA is considerably stronger then their effect on VAV, in particular more than 80 times stronger for K1. The only exception is the case of O1, where we have a similar effect. This can be considered as a confirmation that there is a kind of resonance between O1 and WAL.

In the same time, since all cases in ETERNA are affected in a very similar way, we have to conclude that we have a resonance between the WAL and all tides. So to say, in VAV only O1 is saturated by WAL, while in ETERNA all waves are saturated by WAL. It is also interesting that all δ_B of ETERNA are displaced with respect to δ_A in one and the same direction, i.e. we have a kind of systematic error.

Shall we conclude that the WAL are aliases to all waves in ETERNA? Of course not! The conclusion is simply that ETERNA cannot support data with very strong perturbations, at least in some cases.

The worse in these ETERNA results is the strange behavior of the estimates σ_B of the precision in Variant B. In the case of the SD tides N2, M2 & S2 we have $\sigma_B = \sigma_A$, i.e. the addition of a tremendous noise in Variant B remained imperceptible.

For the D tides σ_B exceeds σ_A but not enough, not correspondingly to the strong effect of the noise.

In the last row of Table 19.3 we have included a ratio denoted θ_{AB}. If σ_B were good estimates, corresponding to the distortion of the results, θ_{AB} would be of the order of 1 or, better, lower than 1. We have very high values, considerably exceeding 1. This means that the ETERNA estimates of the precision failed to estimate the effect of the artificially introduced noise.

Thus, if you look only the ETERNA Variant B in Table 41 you can conclude that these are acceptable results with very high precision of the SD waves,

It is also interesting to compare the VAV and ETERNA results in Variant A.

We have almost identical values of the δ factors and very close MSD, slightly lower for ETERNA. This is almost always the case of all comparisons with ETERNA.

On this occasion we would like to give a warning.

VAV uses least squares estimates MSD, obtained through sums of squares of residuals and having given number of degrees of freedom. The MSD of ETERNA are obtained according to an intuitive algorithm, as amplitudes, without theoretically defined properties.

The case of Variant B here is a good example. Through the σ_B of ETERNA one should conclude that the ETERNA results are considerably better than the VAV results, while they are considerably worse in this case.

Example 20. Data format VAV-ICET.

The international format, which is accepted by ICET and used by our SG data, is somewhat old-fashion and not very comfortable. It uses the conventional symbols "99999999", "88888888" and "77777777" which are remainders from the time of the punched cards.

VAV may use a variant of the ICET format, called by the option

>D-format: VAV-ICET

The general format of the data is, just like the ICET format:

{Date} {channel01} {channel02} .....

{Date} is the date of the data in the channels, occupying 15 columns, namely:

year: in 4 columns
month: in 2 columns
day: 2 in columns
One (just one) obligatory blank, followed by
hour: 2 in columns, minutes: in 2 columns & seconds in 2 columns.
The blanks inside {Date} are accepted as zeros, so that

```
19970701 080102 can be replaced by
1997 7 1  8 1 2.
```

It is possible to have some blanks before {Date}, so that the following data are equivalent:

```
19970701 130000 -1495.278  -992.293   -21.316     0.264
   19970701 130000 -1495.278  -992.293   -21.316     0.264
```

{channel01}, {channel02} .... are the data in as many channels as they are defined by **>D-nm-input-channels**, but no more than 10. {Date} {channel01} {channel02} ... are separated by arbitrary number of blanks, at least one. A datum in a channel may be with or without decimal point (i.e. real or integer number) with an arbitrary number of digits. However, the length of the record should not exceed 129 columns.

Due to these rules one and the same record may be written as

```
19970701 130000 -1495.278  -992.000   -21.316     0.264
   19970701 130000   -1495.278  -992   -21.316  .264
  19970701 130000   -1495.2780000  -0992.000     -0021.316       00.264
```

Lines with a star * in column one are accepted as comments. Also comments are all non-numeric symbols inside a row, separated by at least one blank from the data. For example we may use:

```
*channel 1 gravity  (nms-2)
*        2 pressure (hPa)
*        3 pole tide(nms-2)
*        4 LOD      (nms-2)
or:
*The following row contains a doubtful air-pressure value:
19970701 080000 -2452.390 air-pr: -990.745 (???)  -21.449     0.277
```

In such a way a considerable quantity of information can be transferred from the diary of the observations to the data, by comments like:

*following 24 hours are affected by the earthquake in ....

The data may include, at the beginning or at any line some commands or options, starting by > in column 1.

The option

>Time add seconds `-17.002`

will add –17.002 seconds to all following epochs till a next appearance of a new time correction. The correction is NOT additive. I.e., if a next time correction is

>Time add seconds `19.002`

VAV will add to the following epochs just 19.002 seconds.

The option

>add to channel 2 constant -0.2567

will add to the following data in channel 2 a constant **-0.2567** till a next **>add** correction of the same channel appears, e.g. till we meet

>add to channel 2 constant 99.00

The correction is additive, e.g. in the case above, VAV will add to the following data in channel 2 **98.7433 = -0.2567 + 99.00**.

We may use at one and the same place adding corrections to different channels, e.g.

>add to channel 3 constant 0.07
>add to channel 2 constant -0.2567

The options

>multiply channel 4 by coeff -0.99871
>multiply channel..... 2 by coeff 1.00012

will multiply the following data in channels 4 and channel 2 by 0.99871 and 1.00012 respectively.

A later appearance of

>multiply channel 4 by coeff 3.0

will replace (not multiply) the multiplying coefficient of channel 4 by 3.0. The coefficient of channel 2, defined earlier, remains 1.00012.

We shall have a **gap** in the following cases.

The data are read with a test of sequence. Thus if we have missing data, e.g.

```
19970701 110000 -2181.612 -991.682 ?!? -21.369 0.269
19970701 130000 -1495.278  -992.293   -21.316     0.264
```

VAV declares a gap between these two rows automatically, without any indication. However, the gap may be explicitly indicated by writing

```
19970701 110000 -2181.612 -991.682 ?!? -21.369 0.269
>gap
19970701 130000 -1495.278  -992.293   -21.316     0.264
```

but this is not necessary.

Some authors replace the missing data by a particular number. When the option

>value accepted as gap -999999

is to used, the data

```
19970701 180000  -999999
19970701 190000  -999999
```

are understood as missing data.

When **>value accepted as gap 0** is used, we shall have a gap at the place of data, with zero value in the first channel, as well as when we have blanks in place of the numerical data, e.g. in a case like

```
19970701 180000
19970701 190000
```

Otherwise, the data in a channel cannot be replaced by blanks.

There are cases of sequential data, when we want to introduce a gap, e.g. because we have a jump or doubts about a jump. Then it is necessary to introduce the statement >gap. It is also possible to use

>gap due to a jump

One or several consecutive blank rows, are also considered as a gap in the default case of VAV. However, when

>blank lines are NOT gaps

is used, the blank rows are simply ignored.

When we have, **>Time add seconds -17.002** and, later, a different time correction **>Time add seconds 19.002** VAV introduces a gap after the second correction. However, if the second correction is equal and thus it has not any effect, a gap is not introduced.

The **>add** correction may be accompanied by a jump. Then the gap has to be explicitly declared, e.g. as

>add to channel 1 constant 0.0022
>gap due to a jump

The same is valid for the multiplying correction.

At the end of the data we do not need to put anything, but an **>end** is allowed.

In the options above, valid is only the first letter, after the ">". The remaining non-numerical data are actually comments. Table 20.1 is a summary of the options here shown and their most abbreviated format.

Table 20.1. List of the options, when format VAV-ICET is used.

Options in extended format	Options in abbreviated form
>Time add seconds 19.002	>T 19.002
>add to channel 2 constant -0.2567	>A 2 –0.2567
>multiply channel 4 by coeff -0.99871	>M 4 -0.99871
>gap	>G
>gap due to a jump	>G
>value accepted as gap -999999	>V -999999
>blank lines are NOT gaps	>B

VAV can use other data formats, under requests of the users.

Example 21. Miscellaneous: various types of tidal data, unit of the data, amplitude and phase corrections, more about the MCH (multi-channel) analysis.

We have the following options for processing of **various types of tidal data**:

>ST-component: gravity	Gravity data or vertical component (default case)
>ST-component: ocean	Ocean tidal data
>ST-component: tilt	Tiltmeters or horizontal component
>ST-component: extensometer	Horizontal extensometers
>ST-component: vertical ext.	Vertical extensometers

Processing of gravity data is a default case, so that **>ST-component: gravity** can be omitted.

Other kind of data, e.g. temperature, air-pressure and water table can be processed as gravity or ocean data. Then the δ factors are meaningless, but the amplitudes are O.K. Their unit will be the same as the unit of the data. Both gravity and

ocean tidal phenomena are theoretically in phase with the equilibrium tide. Hence, the phase lag are relative to the phases of the equilibrium tide, i.e. to the phases of the tide generating potential.

The cases of "tilt" and "extensometer" (horizontal) need to use

>ST-azim-north-clockw: a

e.g. **>ST-azim-north-clockw: 278.275** or **>ST-azim-north-clockw: -81.725**

where "a", 278.275 & -81.725 are the azimuth of the instrument, measured clockwise from the North in degrees of arc.

For all cases, except for the gravity data, it is recommendable to use

>ST-grav-gals: g

e.g. **>ST-grav-gals: 981.1234**

where "g" and 981.1234 are values of the gravity in gals, i.e. in cm/sec^2, at the point of observation.

If this option is not used or it is used with g = 0, then a normal gravity is computed by VAV.

In principle, the **unit of the data** can be arbitrary. In any case, the estimated amplitudes will be the same as the unit of the data.

However, the amplitude factor δ is a measureless quantity and it has a physical sense if the unit of the data is the same as the unit of the theoretical amplitudes (see Section B).

The unit used for the different types of data are

Gravity	1 nm/sec^2 = 1 μgal
Ocean	1 cm
Tilt	1 mas (millisecond of arc)
Extensometer	1 nstr (nanostrain = 10^{-9} strain)

When the unit used of the data is not the same, as the units above we shall get the factor δ in a deformed scale. For example, the unit of the tilt data is very often 0.1 mas (millisecond of arc). In the same time the theoretical amplitudes are computed in unit 1 mas. As a result, we shall get 10 times greater δ.

The unit of the data can be controlled by the option

>D-channel-coeff: Nrchannel coefficient

The effect is that all the data in channel number **Nrchannel** will be multiplied by the constant **coefficient**. E.g., if tilt data in 0.1 mas are in channel 3, they will be transformed in mas through

>D-channel-coeff: 3 0.1

It is possible to change the scale of several or all channels. e.g. by using

>D-channel-coeff: 01 0.1	channel 01 will be multiplied by 0.1
>D-channel-coeff: 03 3.14	Channel 03 will be multiplied by 3.14

The use of a negative **coefficient** for the tidal channel will change the phase lags by 180°

It is possible to introduce frequency dependent **amplitude and phase corrections** through a set of options (no more than 69) like

>W-amplit-phase-corr: freqa freqb cpd corr_amplit corr_phase

Here **freqa freqb** define a frequency interval in cpd. If **cpd** is not used, the frequency interval is in deg/hr. The effect of the statement on the tides in this frequency interval is: the amplitudes and the δ factors are multiplied by **corr_amplit** and the phase lags are increased by **corr_phase** in degrees of arc.

For example the effect of

>W-amplit-phase-corr: 0.5 1.5 cpd 2.0 30.0
>W-amplit-phase-corr: 1.5 2.5 cpd 1.8 60.0

is that the amplitudes and δ of all D tides are multiplied by 2 and the phase lag is increased by 30 degrees, while the amplitudes and δ of all SD tides are multiplied by 1.8 and the phase lag is increased by 60 degrees.

The **MCH analysis** has been applied by using only one MCH channel. It is possible to use till 6 channels, e.g. channels 4, 1 & 3 will be used through the options

>MCH-channel: 4
>MCH-channel: 1
>MCH-channel: 3

>MCH-model 1 uses a model of direct (without phase lag) effect of the MCH channels with one regression coefficient per channel and frequency. The variant

>MCH-model 2

accepts a possible phase lag in the effect of the MCH channels. Then we get two regression coefficients per channel and frequency and an estimated phase lag in degrees of arc. We hope that the output in **analysisNN.dat** is enough clear.

The default option of the MCH analysis provides frequency dependent regression coefficients, i.e. different regression coefficient per every one of the basic filter frequencies. It is possible to get frequency independent coefficients by using

>MCH-freq-indep: yes

Conclusions.

We hope, actually we are sure, that the tidal specialists who decide to use our program VAV will be satisfied. Nevertheless, we do not exclude the possibility of some defects and failures. All critical remarks, as well as all suggestions for further improvement of VAV, as well as of the guidebook are heartily welcome.

Good luck with the VAV program!

Acknowledgments: We wish to express our thanks to the Global Geodynamics Project (GGP) consortium, which made available the SG data we have used. Special thanks go to the persons in charge of the SG stations, namely B. Ducarme and M. Van Ruymbecke for Brussels, J. Merriam for Cantley, J. Hinderer and J.P. Boy for Strasbourg and B. Meurers for Vienna. Special thanks are due to B.Ducarme for his participation in a series of applications and improvement of VAV, as well as in the creation of this guidebook. We want also to thank Begoña Tejedor from the Institute of Oceanography in Cadiz for her help to introduce the analysis of shallow water tides.
We want also to express our gratitude to Mme M. Amalvict for the absolute gravity data from Strasbourg and H. Schmitz-Hübsch, Deutsches Geodätisches Forschungsinstitut in Munich and the University Hawaii Sea Level Centre for the data from San Juan.
The work of A. Venedikov in the Institute of Astronomy and Geodesy in Madrid on this guidebook has been supported by the State Secretariat for Education and Universities at the Spanish Ministry of Education and Culture for which we express our thankfulness to the Spanish Authorities.

BIBLIOGRAPHY

Amalvict, M., Hinderer, J., Boy, J.P., Gegout P., 2001. A three-year comparison between a superconducting gravimeter (GWR C026) and an absolute gravimeter (FG5#206) in Strasbourg (France). Journal of the Geodetic Society of Japan, 47 (1), 334-340.

Arnoso J., De Toro C., Venedikov A.P., Vieira R., 1997. On the estimation of the precision of the tidal data. *Working Group on High Precision Tidal Data Processing, Bonn, Sept. 16-19, 1996, Bulletin d'Informations des Marées Terrestres*, **127**, 9757-9767.

Arnoso J., Vieira R., Velez E.J., Van Ruymbeke M., Venedikov A.P., 2001. Studies of tides and instrumental performance of three gravimeters at Cueva de Los Verdes (Lanzarote, Spain). *J. Geodetic Society of Japan*, **47**, **1**, 70-75.

Arnoso J., Ducarme B., Venedikov A.P., Vieira R., 2002. Time variations and anomalies in the air pressure admittance of superconducting tidal gravity data. *Bulletin d'Informations des Marées Terrestres*, **136**, 10793-10808.

Chojnicki, T., 1972. Determination des parameters de marées par la compensation de observations au moyen de la méthode des moindres carrées. Publications of the Institute of Geophysics, Polish Academy of Sciences, Marees Terrestres, 55, 43-80.

Crossley, D., 2000. Report on the status of GGP. Cahiers du Centre Européen de Geodynamique et de Séismologie 17, 1-7.

De Toro C., Venedikov A.P., Vieira R., 1990. Determination of some particular waves in the Earth tide data. *Bulletin d'Informations des Marées Terrestres*, **106**, 7511-7521.

De Toro C., Venedikov A.P., Vieira R., , 1991. A particular study of the relation between Earth tide data and other time series. *Bulletin d'Informations des Marées Terrestres*, **111**, 8062-8072.

De Toro, C., Venedikov, A., Vieira, R., 1991. A particular study of the relation between Earth tide data and other time series. *БАН, Българско Геофизично Списание,* **XVII, 1**, 37-45.

De Toro C., Venedikov, A.P., Vieira, R., 1991. A particular study of the relation between Earth tide data and other time series. Bulletin d'Informations des Marées Terrestres, 111, 8062-8072.

De Toro C., Venedikov A.P., Vieira R., 1993. On the time variations of the sensitivity. *Bulletin d'Informations des Marées Terrestres*, **116**, 8546-8556.

De Toro C., Venedikov A.P., Vieira R., 1993. A new method for Earth tide data analysis. *Bulletin d'Informations des Marées Terrestres*, **116**, 8557-8586.

De Toro C., Venedikov A.P., Vieira R., 1993. tudy of the sensibility variations of the Earth tide records. *Física de la Tierra,* **5**, 61-70. Editorial Complutense.

De Toro C., Venedikov A.P., Vieira R., 1993. Study of the sensitivity variations of the Earth tide records. *БАН, Българско Геофизично Списание,* **XIX**, **2**, 49-56.

De Toro C., Vieira R., Hernandez C., Venedikov A.P., Arnoso J., 1995. Tidal models of the Mediterranean Sea. *Proceedings of the Twelfth International Symposium on Earth Tides, Beijing, Aug. 1993,* 511-528, Science Press, Beijing, New York.

Doodson, A.T., 1928. The analysis of tidal observations. Philosophical Transactions of the Royay Society of London, A227, 223-279.

Ducarme, B. and Vandercoilden, L., 2000. First results of the GGP data bank at ICET. Cahiers du Centre Européen de Geodynamique et de Séismologie, 17, 117-124.

Ducarme B., Ruymbeke M. van, Venedikov A.P., Arnoso J., Vieira R., 2003. Polar motion and non tidal signals in the superconducting gravimeter observations in Brussels. Submitted to Physics Earth and Planetary Interior.

Ducarme B., Venedikov A.P., Arnoso J., Vieira R., 2003. Determination of the long period tidal waves in the GGP superconducting gravity data. Paper, presented at the IUGG meeting in Saporo and submitted to Geodynamics.

Godin G., 1972. The analysis of tides, University of Toronto Press, Toronto (264 pp.)

Horn W., 1960. The harmonic analysis, according to the least square rule, of tide observations upon which an unknown drift is superposed. Bolletin de Geofisica Teorica e Applicata, 2 (5) 218-222.

Imbert B., 1954. L'analyse des marées par la méthode des moindres carrées. Bulletin du Comité Central Océanographique et Etude de Cotes VI, 89, 398-410.

Lecolazet R., 1958. La méthode utilisée a Strasbourg pour l'analyse harmonique de la marée gravimétrique. Bulletin d'Informations des Marées Terrestres, 10, 153-178.

Melchior, P., 1983. The tides of the planet Earth, Pergamon Press, Oxford, 641 pp.

Melchior, P., 1995. International Centre for Earth Tides (ICET). I.U.G.G. Chronicle, 228, 245-250.

Melchior, P., Venedikov, A.P., 1968. Derivation of the wave M3 ($8^h.279$) from the periodic tidal deformations of the Earth. Physics of the Earth and Planetary Interiors, 1, 362-372.

Munk, W.H., Cartwright, D.F., 1966. Tidal spectroscopy and prediction. Philosophical Transactions of the Royal Society of London A259 (1105) 533-581.

Pagiatakis, S.D., 2000. Application of the least-squares spectral analysis to superconducting gravimeter data treatment and analysis. Cahiers du Centre Européen de Geodynamique et de Séismologie 17, 103-114.

Pertsev B.P., 1958. Harmonic analysis of tides (in Russian). Izvestia Akademii Nauka SSSR, Seria Geophizicheskaya 8, 946-958.

Sakamoto, Y., Ishiguro, M., Kitagawa, G., 1986. Akaike information criterion statistics, D. Reidel Publishing Company, Tokyo (290 pp.).

Schüller, K., 1977. Tidal analysis by the hybrid least squares frequency domain convolution method. Proceedings 8th International Symposium on Earth Tides. Bonn, Germany, pp. 103-128.

Schüller, K., 1978. About the sensitivity of the **Venedikov** tidal parameter estimates to leakage effects. *Bull. Inform. Marées Terrestres,* **78**, 4635-4648.

Tamura, Y., 1987. A harmonic development of the tide-generating potential. Bulletin d'Informations des Marées Terrestres, 99, 6813-6855.

Tamura, Y., Sato, T., Ooe, M., Ishiguro, M., 1991. A procedure for tidal analysis with a Bayesian information criterion. Geophysical Journal International, 104, 507-516.

Usandivaras J.C. and Ducarme, B., 1969. Analyse des enregistrement des marées terrestres para la méthode des moindres carrées. Observatoire Royal de Belgique, Série Géophysique, 95, 560-569.

Venedikov, A.P., 1961. Application à l'analyse harmonique des observations des marées terrestres de la Méthode des moindres carrées. Comptes Rendues, Académie Bulgare des Sciences, 14 (7), 671-674.

Venedikov, A.P., 1966. Une méthode d'analyse des marées terrestres à partir d'enregistrements de longueurs arbitraires. Observatoire Royal de Belgique, Série Géophysique, 71, 463-485.

Venedikov A.P., 1989. A model for the study of the effects of the air pressure on the Earth tide data. Bulletin d'Informations des Marées Terrestres, 103, 7263-7273,

Venedikov, A.P. and Ducarme, B., 2000. Determination of the Earth tide constituents generated by the tidal potentials of degree 3 and 4 for large series of observations made by superconducting gravimeters. Bulgarian Geophysical Journal, 26, (in press).

Venedikov A.P., Vieira R., De Toro C., 1992. A new method for Earth tide analysis. *Instituto de Astronomía y Geodesia, Publicación,* **183**, 1-30.

Venedikov A.P., Vieira R., Toro C. de., 1993: Nuevas aportaciones al análisis de observaciones de mareas terrestres. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales*. **LXXXVII**, 471-478.

Venedikov, A.P., Vieira, R., Toro, C. de, Arnoso, J., 1997. A new program developed in Madrid for tidal data processing. Bulletin d'Informations des Marées Terrestres, 126, 9669-9704.

Venedikov, A.P., Arnoso, J., Vieira, R., 2001. Program VAV/2000 for tidal analysis of unevenly spaced data with irregular drift and colored noise. Journal of the Geodetic Society of Japan, 47 (1), 281-286.

Venedikov A.P., Vieira R., De Toro C., 1995. The computer program NSV used in Madrid for tidal data processing. *Meeting of the Working Group on High Precision Tidal Data Processing, Bonn, Aug. 30- Sept. 2, 1994, Bulletin d'Informations des Marées Terrestres*, **121**, 9108-9126.

Venedikov A.P., 1995. Remarks about the MV66 and ETERNA 3.1 tidal analysis methods (a discussion). *Meeting of the Working Group on High Precision Tidal Data Processing, Bonn, Aug. 30- Sept. 2, 1994, Bulletin d'Informations des Marées Terrestres*, **121**, 9127-9138.

Venedikov A.P., Vieira R., De Toro C., 1995. A new program developed in Madrid for tidal data processing. *Instituto de Astronomía y Geodesia, Publicación,* **189**, 1-47.

Venedikov A.P., Vieira R., De Toro C., 1997. On the determination of the D and SD Earth tides generated by the tidal potential of the third order. *Bulletin d'Informations des Marées Terrestres*, **126**, 9635-9637.

Venedikov A.P., Vieira R., Toro C. de, Arnoso J., 1997. A new program developed in Madrid for tidal data processing. *Bulletin d'Informations des Marées Terrestres*, **126**, 9669-9704.

Wenzel, H.-G., 1994. Earth tide analysis package ETERNA 3.0. Bulletin d'Informations des Marées Terrestres, 118, 8719-8721.

Wenzel, H.G., 1997. Analysis of Earth tide observations. Lecture Notes in Earth Sciences, 66, Springer, 59-75.

Venedikov A.P., Ducarme B., 2000. Determination of the Earth tide constituents generated by the tidal potentials of degree 3 and 4 for large series of observations made by superconducting gravimeters. *БАН, Българско Геофизично Списание*, **26**, 61-74.

Arnoso J., Vieira R., Velez E.J., Cai W.-X., Tan S.-L., Jun J., Venedikov A.P., 2001. Monitoring tidal and non-tidal tilt variations in Lanzarote Islands (Spain). *J. Geodetic Society of Japan*, **47**, **1**, 456-462.

Venedikov A.P., Arnoso J., Vieira R., 2001. Program VAV/2000 for Tidal Analysis of Unevenly Spaced Data with Irregular Drift and Colored Noise. *J. Geodetic Society of Japan*, **47**, **1**, 281-286.

Venedikov A.P., Arnoso J.,Vieira R., 2003. VAV: A program for tidal data processing. *Computer & Geosciences,* **29**.

Appendix: List of the options used in the examples of the guidebook.

Options	Examples
D- options, related with the data	
>D-abs-grav: 3700	17
>D-add-artifical-wave: 6.4430356067 380	18
>D-channel-coeff: 01 0.1	21
>D-format: Sanjuan	9
>D-format: unf	1
>D-max-nm-data: 77105	1
>D-nm-input-channels: 4	1
>D-origin-of-time: 1982 01 01 00	8
>D-step-mean-(minutes): 30	17
>D-tidal-channel: 1	1
E-options, elimination of data	
>E-badint.dat: 2	5
>E-nm-iterations: 5	4
>E-T-interval: 1985 1 1 00 1985 12 31 23	11
>E-T-interval: 5400 7300 days	7
>E-T-interval: 68816 72240	3
>E-Student-t: 1.96	1
F- options, related with the filters	
>F-delta-freq: 0.5 cpd	19
>F-highest-freq: 11.5	19
>F-Highest-freq: 11 cpd	13
>F-k-drift-polyn: 0	8
>F-lowest-freq: 0.5 cpd	19
>F-T-window: 24	8
GR- options, groups of tides, tidal unknowns	
>GR-D-SD: 50	9
>GR-HF: 2	13
>GR-LP: 6	8
>GR-new-wave: 6.4430356067 WAL1	18
>GR-QD: 7	13
>GR-Shallow: 3 11	14
>GR-TD: 7	13
MCH-options, multi-channel analysis	
>MCH-channel: 2	1
>MCH-model: 1	1
>MCH-model: 2	21
>MCH-freq-indep: yes	21
OUT-options, output, results	
>OUT-compare-tide-name: Q1	9
>OUT-hr-residuals: yes	2
>OUT-predict: 2000 1 1 0 2000 12 31 23	11

ST-options, station, instruments	
>ST-altit-meters: 269.0	1
>ST-azim-north-clockw: 278.275	21
>ST-component: extensometer	21
>ST-component: gravity	21
>ST-component: ocean	21
>ST-component: tilt	21
>ST-component: vertical ext.	21
>ST-grav-gals: 981	21
>ST-latit-n: 45.585	1
>ST-long-e: 284.1929	1
>ST-name: Station 6824 Cantley...	1
TV-options, time variation	
>TV-length-segment-days: 366	16
>TV-shift-segments-days: 183	16
W-options, waves, tides	
>W-all-tides: 3 3 cpd	14
>W-amplit-phase-corr: 0.5 1.5 cpd 2.0 30.0	21
>W-non-tid-freq-interval: 5.7 6.1 0.001 cpd	15
>W-tidal-param: yes	17
Z-options, zero line, drift, mean sea level	
>Z-freq: 1 cpy	12
>Z-k-degree-polyn: 0	12
>Z-segm-time: 1989 5 20 0	12

Remark: The names of the options can also be copied from the file **cwords.inp** in the executive folder **\aaavav_03\.**