Least squares spectral analysis of gravity data from the Canadian superconducting gravimeter: an ongoing project report

Mensur Omerbashich[1]
Petr Vanícek[2]
Department of Geodesy and Geomatics Engineering
University of New Brunswick, P.O. Box 4400
Fredericton, N.B. Canada E3B 5A3

Abstract

Fourier analysis is not a suitable tool for analysing data series with gaps. Filling the gaps with invented data becomes unacceptable when large gaps (several years in our case) are present. The Least Squares Spectral Analysis (LSSA) has been shown to be more suitable for analysing such data series in a variety of applications in astronomy, geophysics and geodesy.

In the past half a year, we have been processing such a gappy series, covering the period between 1989-1999 with a one-second step, of gravity data from the Canadian superconducting gravimeter. We started with the highest frequencies and are heading towards looking for lower and lower frequencies. In our preliminary tests, we have detected two frequency bands, around the periods of 7 and 20 seconds, with significant power. A noise of an unknown origin (microseismic, cultural noise, instrumental?) appears to excite the gravimeter characteristic frequencies in these two bands. For the purpose of our analyses, we can suppress this noise by suitably averaging the data.

We also had the first look at periods of the order of minutes to a few hours and have not seen any spectacular signal on these periods. In the next step we will focus on time intervals immediately following large earthquakes to see if the earth characteristic frequencies can be gleaned from the data. After deciding on suitable averaging/filtering of the data (45 minute averages, 1.5 hour averages?), analysis of the tidal signal, and its removal from the series will follow. Given the length of the series, the tidal analysis should yield exceptionally accurate (home generated) results, which alone are of interest. The really interesting part of the research will be then done on the residual series.

[1] Omerbasic@unb.ca , http://einstein.gge.unb.ca
[2] Vanicek@unb.ca , http://einstein.gge.unb.ca

1. A few remarks on the method

Given a time-series $\boldsymbol{f}(t) \in \mathrm{H}$, $t_i = 1,2\ldots m$, and its $\boldsymbol{C}_f$, the LSSA detects the periodicities (periodic or systematic signals) in $\boldsymbol{f}$ when $\boldsymbol{f}$ contains either random or systematic noises, or both. The *Least Squares Spectrum* is defined as (Vanícek, 1971)

$$\boldsymbol{s}(w_i) = \frac{\boldsymbol{f}^T \cdot \boldsymbol{C}_f^{-1} \hat{\boldsymbol{g}}(w_i)}{\boldsymbol{f}^T \cdot \boldsymbol{C}_f^{-1} \cdot \boldsymbol{f}}, \quad i = 1, 2, \ldots k\,; \quad k \in \mathrm{R} \tag{1}$$

for trigonometric base functions Φ, where $\boldsymbol{g} = \mathrm{F} \cdot \boldsymbol{x}$ models $\boldsymbol{f}$ by Vandermonde matrix $\mathrm{F} = [\mathrm{F}_S, \mathrm{F}_N]$ and a vector of unknown parameters $\boldsymbol{x}^T = [\boldsymbol{x}_S \mid \boldsymbol{x}_N]^T$ that includes signal (sub. S) and noise (sub. N) components together. Practically, the signal-noise separation is achieved by

$$\hat{\boldsymbol{r}} = \boldsymbol{f} - \hat{\boldsymbol{g}} = \boldsymbol{f} - \mathbf{F} \cdot (\mathbf{F}^T \boldsymbol{C}_f^{-1} \mathbf{F})^{-1} \cdot \mathbf{F}^T \boldsymbol{C}_f^{-1} \boldsymbol{f} \tag{2}$$

where decomposition of $\boldsymbol{f}$ into $\boldsymbol{g}$ and $\hat{\boldsymbol{r}}$ follows from the *projection theorem* $\hat{\boldsymbol{r}} \perp \hat{\boldsymbol{g}}$. For the time-series' that are long enough (over 150 data points) and with high significance levels, it does not matter whether the series $\boldsymbol{f}$ does or does not have its (known) covariance matrix $\boldsymbol{C}_f$.

Unlike the *Fast Fourier Transform* (FFT), by using the LSSA we do not remove periodicities – rather we model these. The advantage is in avoiding what is unavoidable with the FFT – the bias that comes from a lack of general knowledge about the periods in nature. On the other hand, introducing or enforcing a systematic function (model) that can be, e.g., periodic, does not affect a signal significantly. An important property of the LSSA lays then in its invariance with respect to the enforced systematic noise. Besides, the functional (model-) representation of the naturally periodic phenomena enables one to use the presently available knowledge on these phenomena (tides, seasonal weather, etc.), and regard it as sufficient. Periods can indeed be regarded as unknown and theoretically treated as such thereof, but in practical work there is no a viable justification for this.

Since interested in physics behind the data, we intend to apply an approach that represents the above phenomena mathematically: from the astronomical facts – luni-solar tides, from the measurement records – pressure, temperature and the instrument behavior with error assessment, and from the meteorological data – the weather formations. The latter phenomenon is at the same time the most challenging one, and we hope for it to attain a clearer form once our task of analyzing the entire gravity data altogether with the year-gaps, by least squares, is brought to an end. Then, we ought to be able to see such known constituencies like, e.g., Chandler period (~435 days), lunar perigee (~8½ years), crustal ringing due to post-quake impulse oscillations, and others, but also to find more about the not-so-well-known ones: El-Niño and La-Niña (3½-7 years?), etc.

More on the Least Squares Spectral Analysis and on its statistical aspects can be found in, e.g., Pagiatakis (1999).

2. Data set

The data that we had at our disposal consisted of 7,399 files covering the period between 1989-1999. Each file represented a single day in the superconducting gravimeter's operation, opened at 00:01 hrs and closed at 24:00 hrs. The instrument was employed at the *f* = 1Hz level, with a satisfying level of maintenance, revealing on average less than a few hundred false measurements or measurements not taken at the correct time (each second) per file (day). Some years contained no gravity records (1989, 1990) but were fully to nearly-fully populated with meteo-data (pressure). The largest gap in gravity data coverage was between years 1992-1996, whereas year 1993 contained no data of any kind. The total size of the dataset was 2.22 GB. The calibration factor used in our analysis was 78.48 μGal/Volt, as per personal communication with Spiros Pagiatakis.

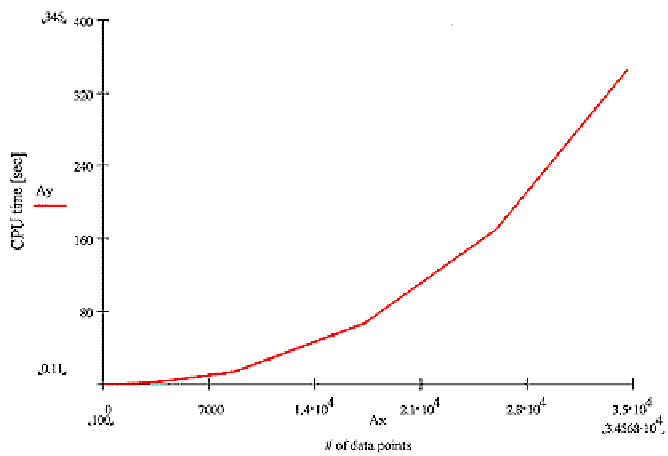


Figure 1. Data size *vs*. the processing speed.

In order to be able to acquire a certain data density, we ran several tests using ten two-thirds-of-a-day fully populated data sets at 1Hz, chosen at random among the years 1991 and 1992 data. These were from all four seasons, so as to enable propagation of the possible microseismic impact on the instrument's behavior, most strongly pronounced in winter. Resolution used was between 1000-2000 points. This quite a detail does not come as a surprise, since the irregularity in peak-generation by graphing software could result in artificially smoothed peaks. This could cause unreliable judgements, and thus seriously endanger the importance carried by the shape of the peaks. On the other hand, whenever possible, to avoid the above (*Fig*.1) processing speed problems the resolution was kept down on the 1000 pts level.

3. Characteristic frequencies of the gravimeter

A zone of peaks at ~20sec and ~7sec periods in eight out of ten test results was detected, most remarkably represented by Fig 02, and it could not, using the available literature (Bengert, 1996), be related to a known instrumental behavior. It could perhaps be ascribed to the cultural or other causes, as its microseismic origin would not appear to us in a clear form (test datasets spanned over all four seasons). The other 2 tested datasets had nearly flat LS spectrum.

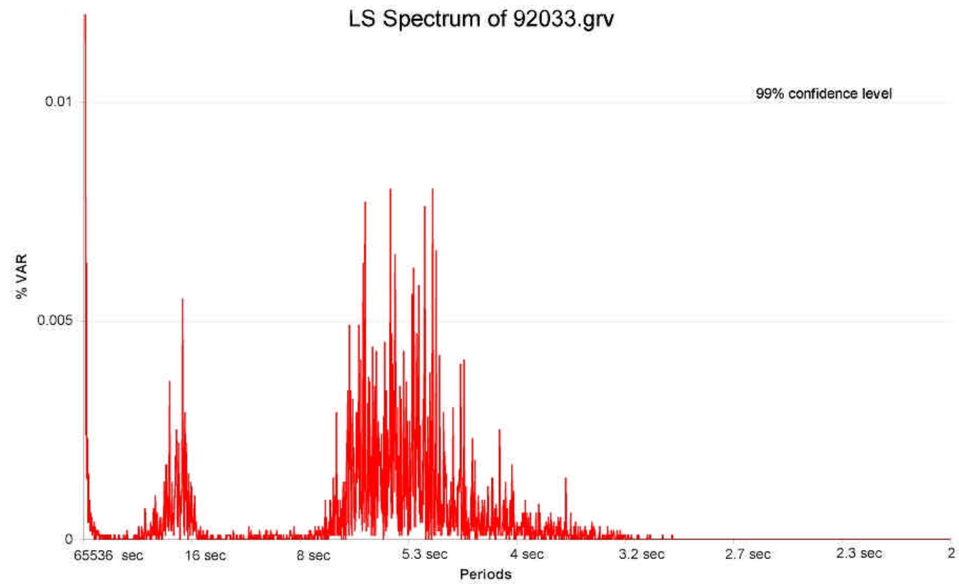


Figure 2. The most significant LSS of ten examined randomly chosen days.

The same data (Fig 2) was then, as containing the most significant high-frequency noise characteristics, picked for determining the proper decimation technique: if a constructed filter satisfied for this one, it will as well satisfy for all other data.

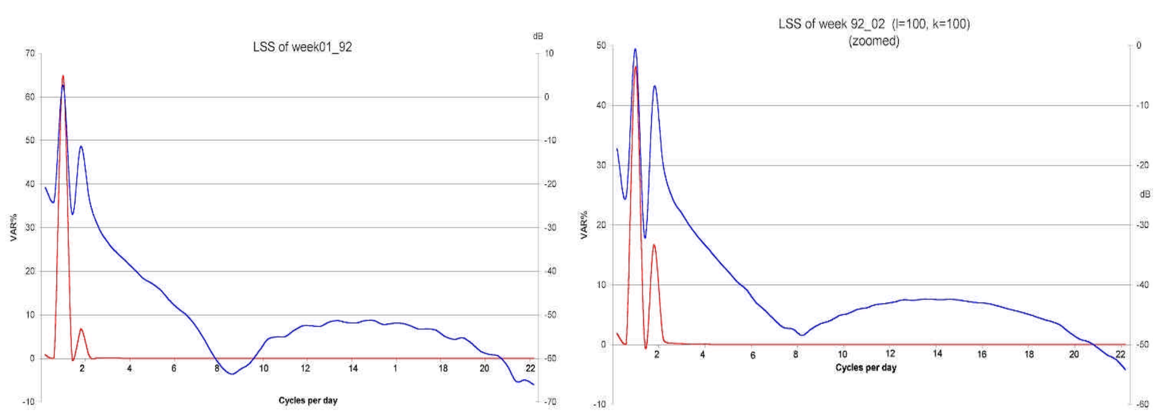


Figure 3. A pair of one-week data sets; red is the LS spectrum,
blue is power spectrum (in dB).

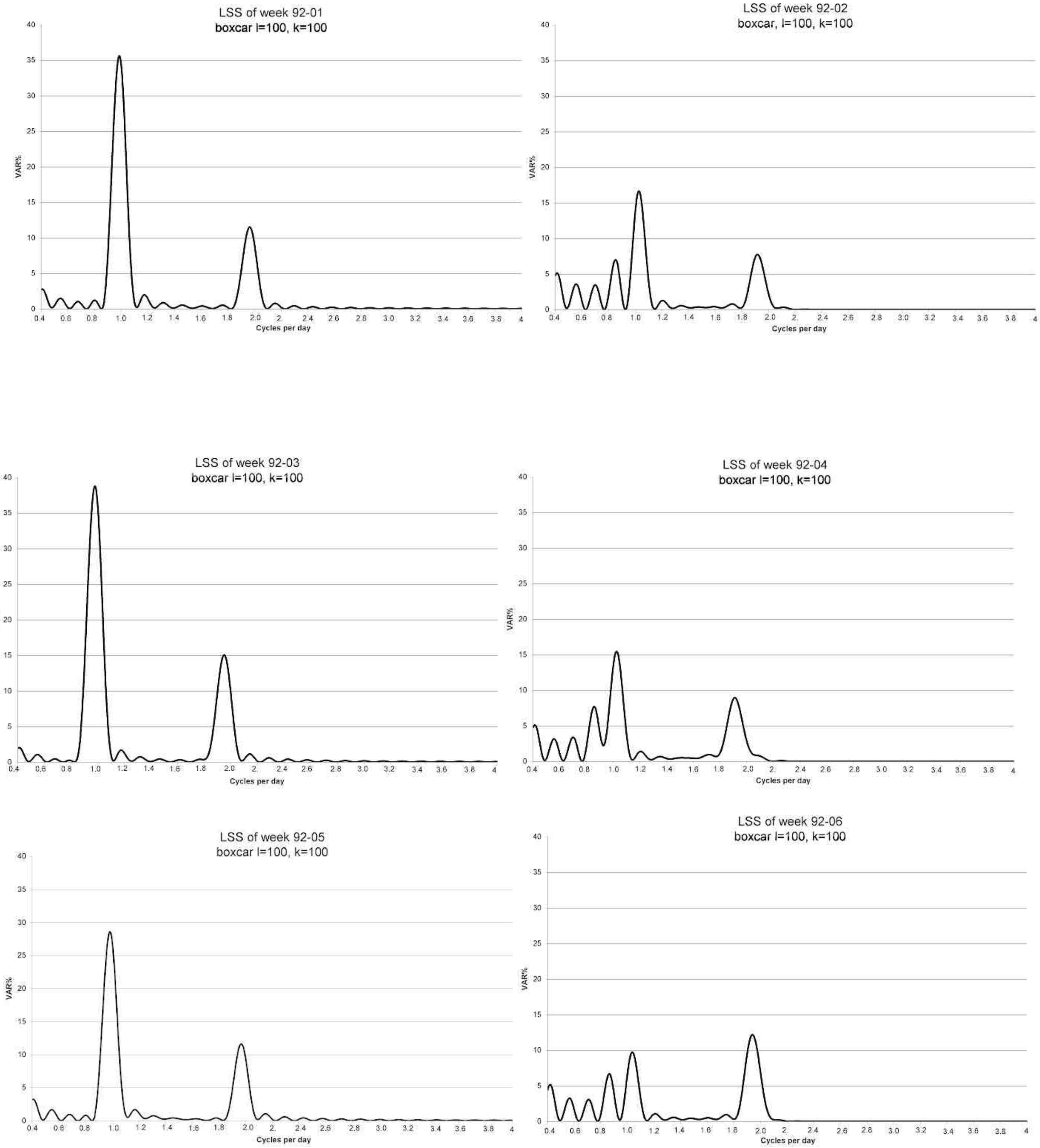


Figure 4. The Least Squares Spectra of the weeks 05 and 06 of year 1992, individually. On the left hand-side are odd, and on the right-hand side even weeks. Data decimated using the boxcar filter with *l* =100, *k*=100. Bandwidth of interest: tidal (periods between ¼ of a day to 2½ days). Note the disturbance in the even-week LS spectra, due to the fortnightly tidal periods.

4. Decimation technique

Decimation is a necessity: by approximating the *CPU time vs. data size* plot (Fig.1) by a third-order polynomial accurate within 2%, one can easily extrapolate that a six-week-dataset (such as the one used in Fig.5) LS spectrum computation (without enforcing periods), if taken at the rate at which the data was collected (i.e. 1Hz), would require over 4½ years of processing time on a PIII @ 500MHz.

The decimation technique used in the following was *boxcar*. This is a moving average filter with a uniform distribution. The algorithm is simply:

$$\frac{1}{2k+1}\sum_{i=-k}^{k} l_i \,, \tag{3}$$

where, after conducting several tests, the parameter values of $l = 100$ and $k = 100$ where selected.

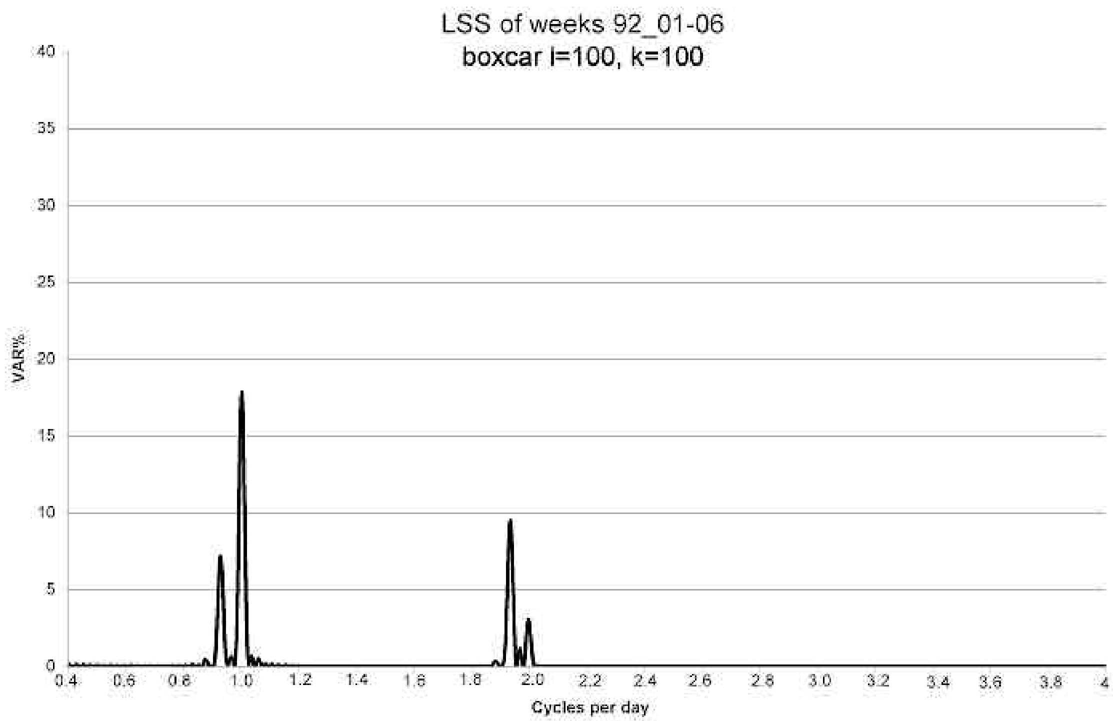
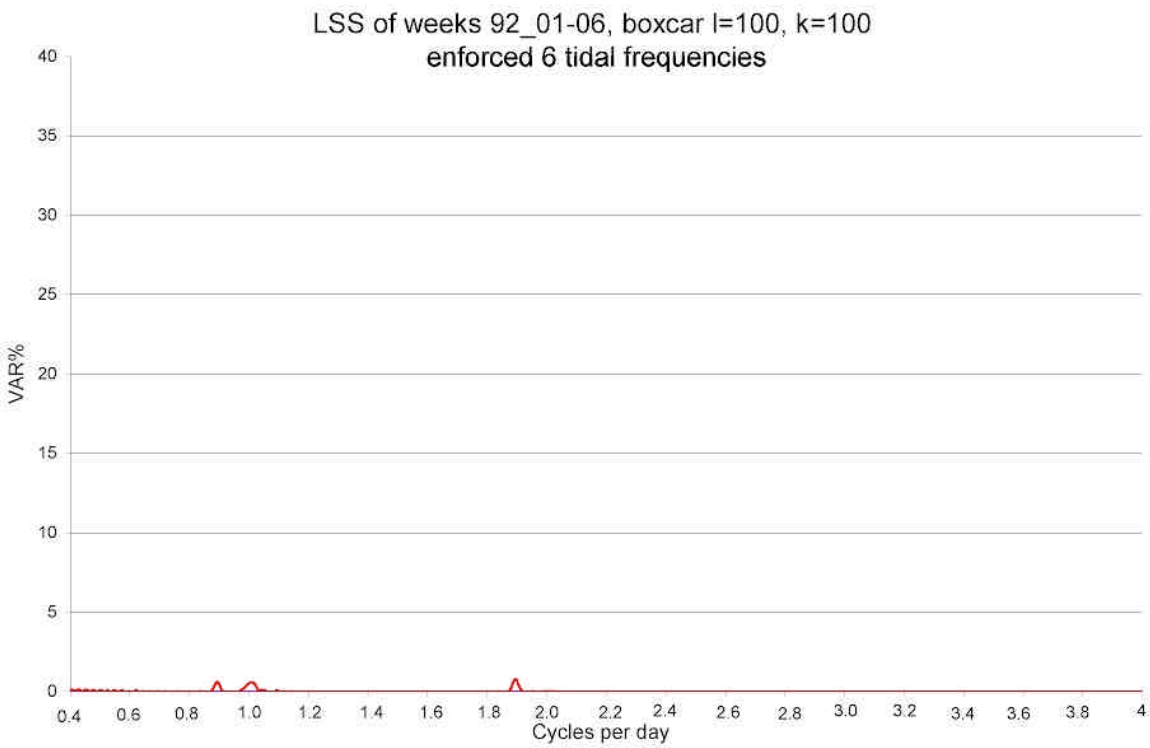


Figure 5. The Least Squares Spectra of a six-week data set (first six weeks of year 1992). Data decimated using the boxcar filter with l =100, k=100. Bandwidth of interest: tidal (periods between ¼ of a day to 2½ days).

Enforced periods (cf. Hou, 1991):

13.4098257	15.0424341	28.9840259
13.9426854	15.5742444	29.5159428
14.5057965	28.4395041	29.9977268 °/hr



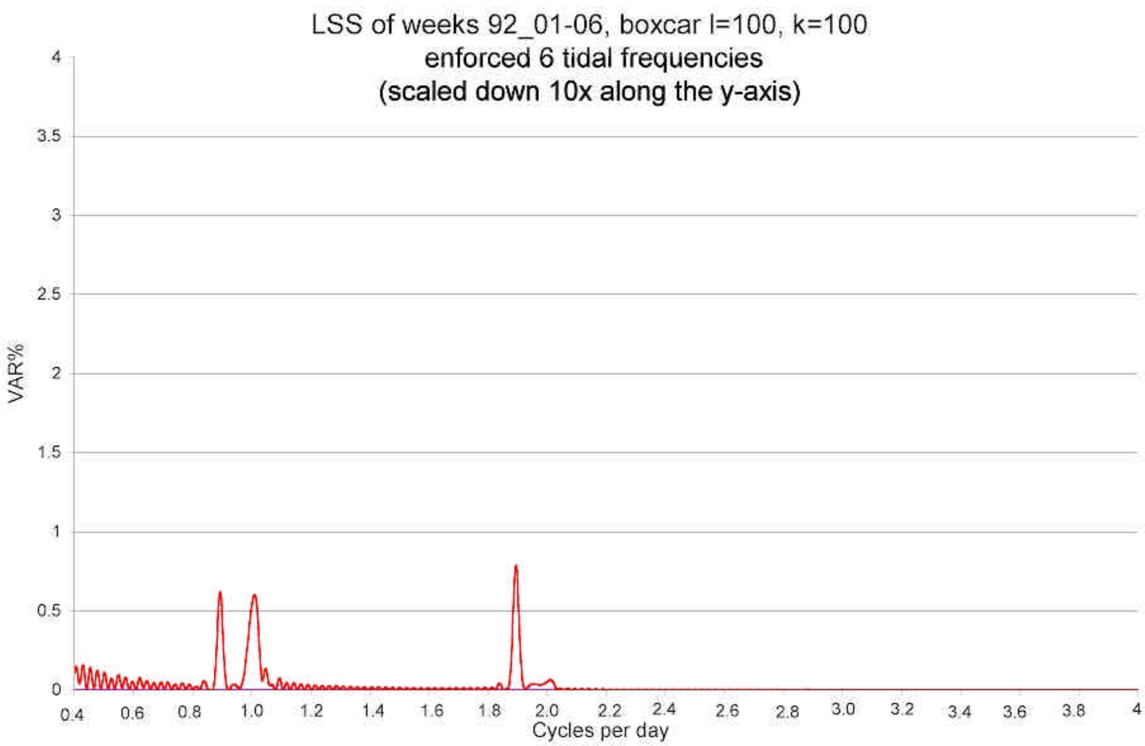
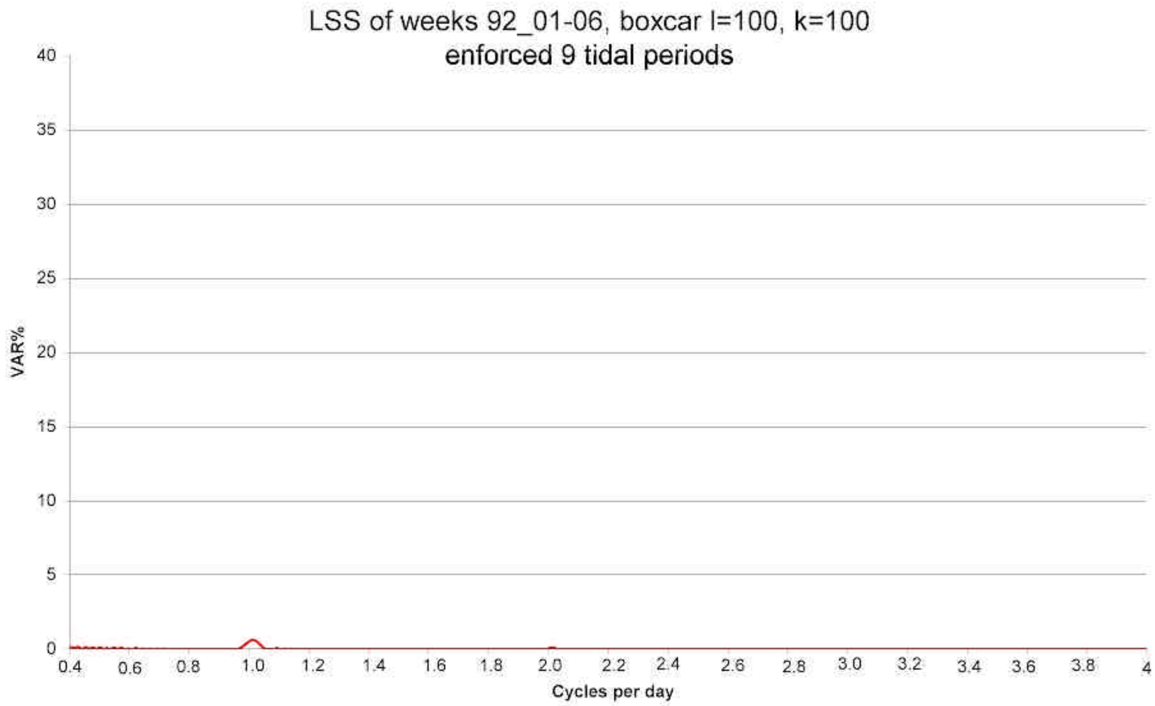


Figure 6. The Least Squares Spectra of a six-week data set (first six weeks of year 1992). Data decimated using the boxcar filter with *l* =100, *k*=100. Bandwidth of interest: tidal (periods between ¼ of a day to 2½ days). Shown is the LS spectrum (top) after enforcing 6 tidal periods from the Table of Lumped Constituents (Hou, 1991). The bottom image shows the same, but scaled down 10x along the y-axis.



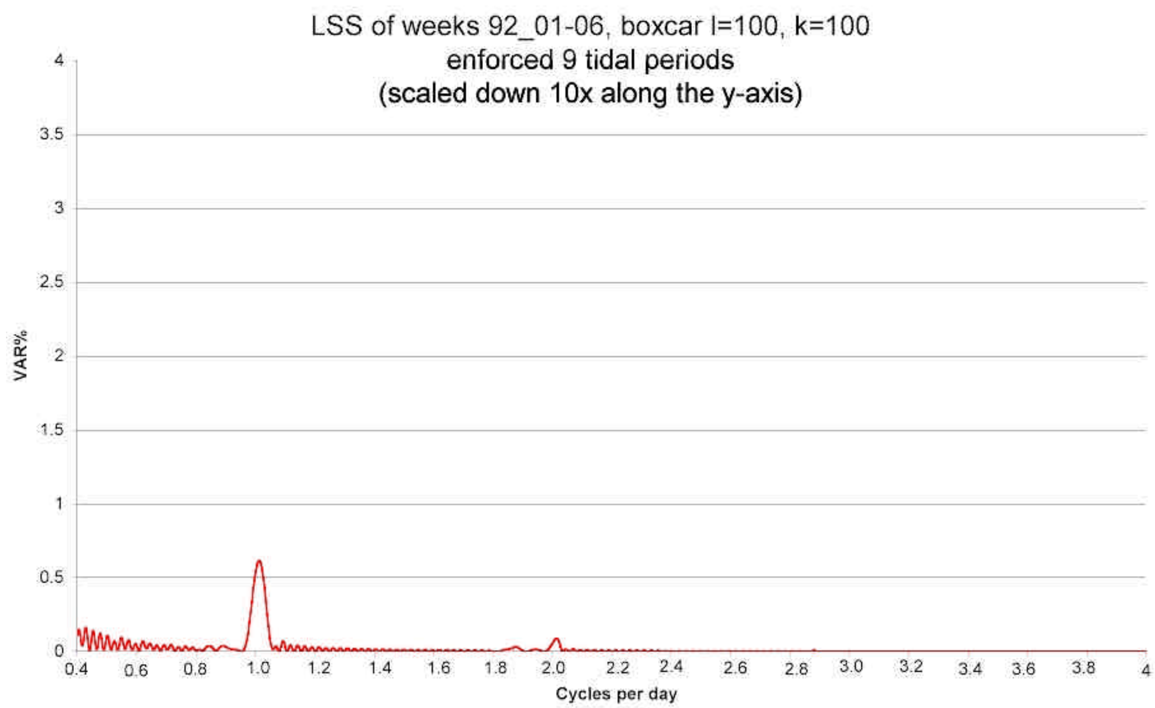


Figure 7. The Least Squares Spectra of a six-week data set (first six weeks of year 1992). Data decimated using the boxcar filter with *l* =100, *k*=100. Bandwidth of interest: tidal (periods between ¼ of a day to 2½ days). Shown is the LS spectrum (top) after enforcing 9 tidal periods from the Table of Lumped Constituents (Hou, 1991). The bottom image shows the same, but scaled down 10x along the y-axis.

References

Bengert, B. R., 1996. Spectral Analysis of Superconducting Gravimeter Data. MSc Thesis, Department of Earth and Space Science, York University, North York, Ont., Canada.

Hou, T., 1991. Sequential Tidal Analysis and Prediction. MSc Thesis, Department of Geodesy and Geomatics Engineering, University of New Brunswick, Fredericton, NB, Canada.

Pagiatakis, S. D., 1999. Stochastic significance of peaks in the least-squares spectrum. Journal of Geodesy, Springer-Verlag.

Vanícek, P., 1971. Further development and properties of the spectral analysis by least-squares fit. Astrophy Space Sci 12: 10-33.