Proven Methods for filtering SG Data

Herbert Weidner[A]

Abstract: Subject are different methods for noise-free data reduction of low-frequency geophysical data and identification of interference sources. The most important method is the selective compensation of strong spectral lines.

Introduction

Superconducting gravimeters provide the most accurate long-term data of gravitational force and are therefore ideally suited to search for extremely weak signals. Different influences – also in the course of data processing – reduce the quality, some are avoidable.

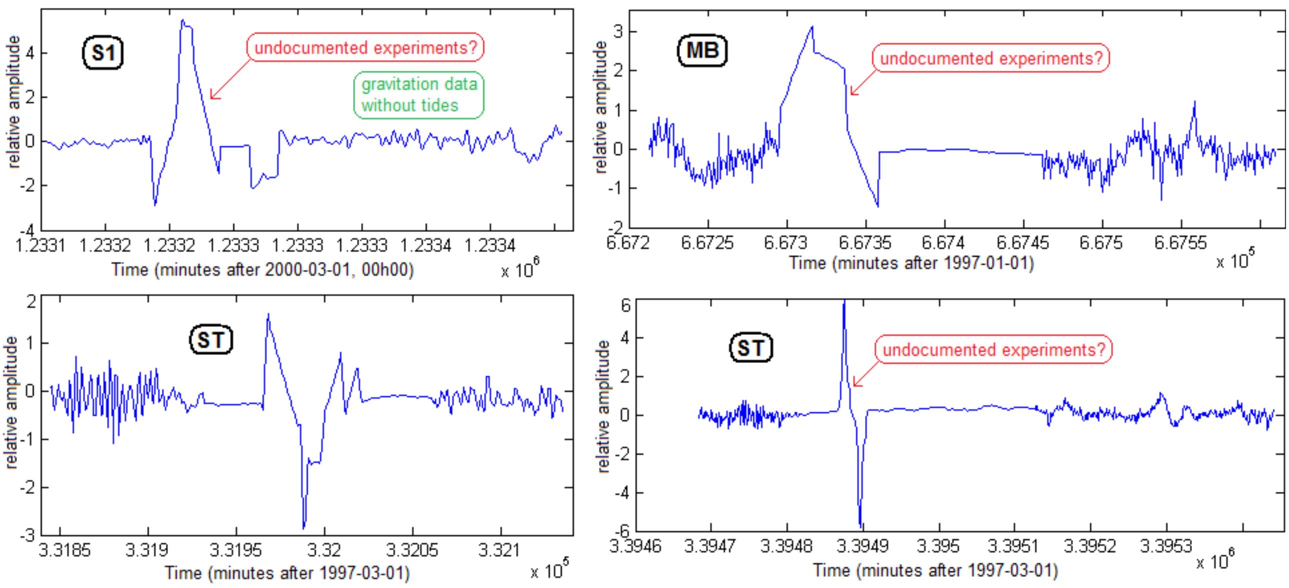
In this study, the advantages and disadvantages of different filter methods are discussed, which can support the search for extremely weak oscillations in the range below 200 µHz. Thorough research has shown that there exist very few process steps that do *not* affect SNR. Particularly advantageous are those which even reduce the noise in the frequency range of interest.

Errors in CORMIN Data

SG usually measure the gravitational force in a second cycle. In this way, huge amounts of data are generated, which must be compressed to a sensible level, because vibrations are sought, which only change after hours or days. It is neccessary to convert the data into minute data, whereby sometimes first errors creep in. The reason are data gaps, which range from a few minutes to several days.

These data gaps can be bridged in different ways: either by many zeros, by linear interpolation or by theoretical values which are supposed to replace the tides. Below, it is shown that the latter method usually generates additional noise, which must be laboriously removed. Obviously, during data pre-processing, care is taken to ensure that there are no gaps. The actual content must be ignored if it has no geophysical cause and acts exclusively as a noise generator.

It would be desirable to mark all data segments that were altered other than by low-pass filters or that were influenced by experiments in the immediate vicinity of the measuring instrument. The following images show segments that were most likely not generated by geophysical causes, but all warnings are missing. In any case, they raise the noise level.



(A) 17. Mar. 2019, email: herbertweidner@gmx.de

Experiments are necessary and do not bother if the corresponding periods are clearly marked by a chain of zeros. The search for extremely weak signals with oscillation durations of a few hours (Slichter triplet) becomes illusory if the records contain data from experiments imitating strong signals with approximately this rhythm. These signals act like broadband jamming transmitters. They hugely hinder the search for hidden signals if they are not detected and removed.

How to improve the search for extremely weak signals?

All the following considerations apply to noisy geophysical records and to the search in the frequency range 35 µHz to 200 µHz. Just below this range, the gravimeters measure extremely strong signals here on the earth generated by the moon and planets. Since previous attempts with high pass and notch filters disappointed, the cause was investigated. The result was surprising: all these filters increase the mean noise level, because they inevitably produce unwanted mixed products. The mathematical background is obvious: A series of measurements (a record) is a long chain of numerical values measured at regular intervals. The record can be understood as a superposition of many oscillations of different frequencies. If you want to emphasize or attenuate individual frequencies, this is usually done with FIR or IIR filters.

Each digital filter is based on the idea of adding *modified* adjacent readings to each individual reading. It is not about how big the factors are and how they are determined, it is only important that the readings are *changed* by multiplication. Technically, this is a modulation and thereby new frequencies are generated, which act as additional noise.

If you do not want to degrade the SNR, no standard filters should be used, especially no IIR filter. Unfortunately, there are very few straightforward procedures that will be discussed now.

Step 1: The comb filter

When analyzing the SG data, the extremely strong tides interfere the most. Their amplitude is about the factor 1,000,000 higher than the amplitudes of the sought signals, which may even be below the noise level of the earthquakes. As you can not avoid earthquakes, you have to do everything possible to prevent the noise level from rising.

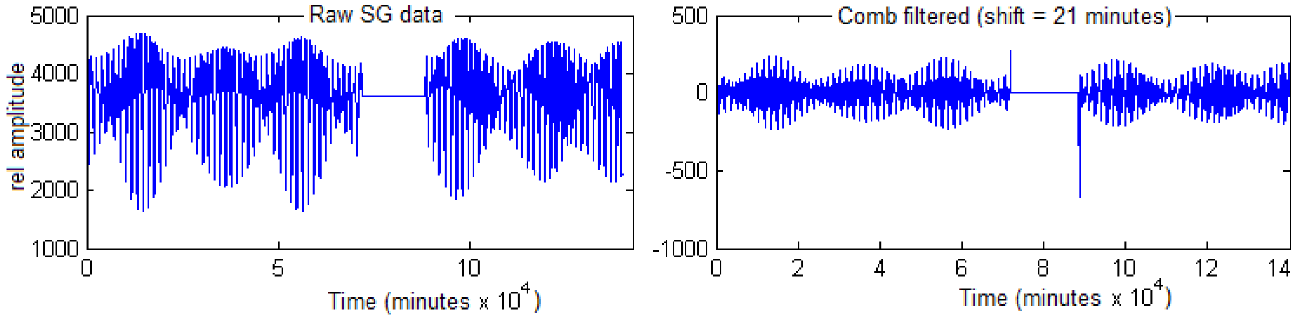
An SG measures the relative movement between the housing and the measuring ball. This can be caused by two different physical processes: changes in the gravitational force or vibration of the housing (structure-borne sound microphone). It would be very instructive to hang a SG on elastic bands so that no rapid ground movements can be transmitted to the housing. This wishful dream will probably never be fulfilled.

Every earthquake generates non-periodic changes in the measured values, sometimes only individual spikes. The frequency spectrum of irregularly occurring peaks is determined by two properties: the interval between adjacent spikes determines the low-frequency component (this interferes with our examination) and the duration of the peaks determines the high-frequency component (irrelevant in the present study).

A "shift + subtract" comb filter ($\alpha = -1$) with a short shift width (suggested value ≈ 20 minutes) turns every single pulse into a double pulse with opposite sign whose spectral focus is on higher frequencies. Therefore, a comb filter can attenuate the noise at low frequencies. The noise moves into another, less interesting frequency range. More importantly, not only the amplitudes of the very strong tides are greatly reduced, but a slow drift is also erased.

From the considerations of the previous chapter follows: In a comb filter, a *single* adjacent reading is multiplied by -1 and added. With a FIR filter, there are about a hundred readings that are included with different factors. An IIR filter takes even more into account. Therefore, a comb filter basically produces less noise than all other methods.

The two figures show the amplitude reduction through a comb filter. A multiple application brings no further profit.



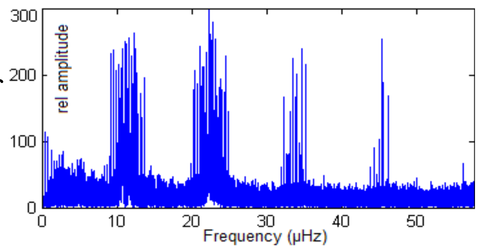
After the comb filter, there will be a small peak (twenty data points wide) at each discontinuity in successive data, which must be removed individually before the next step. In the course of this time-consuming work done by hand, data gaps should be filled by zeros. It is counterproductive to use low-pass filters at this early stage of the filtering because they significantly increase the noise level and produce numerous unwanted mixed products, which appear in the spectrum as enigmatic spectral lines. The abrupt transitions at the ends of the data gaps remain, because they do not disturb in the following step.

Step 2: Selective compensation of strong spectral lines

The signal-to-noise ratio is not changed when adding or subtracting a noiseless mathematically "simple" function. This can be a constant value (eg a straight line) or a sinusoidal function of any frequency or even a whole set of different sinusoidal functions (this is the main reason for the inverse FFT filter, which is explained below). In principle, the function could also be a polynomial, but we specialize here in geophysical data series.

The procedure now described, which in turn compensates most interfering spectral lines, is an essential and important part of the data reduction. It consists of the following steps, which are repeated several thousand times: calculate the spectrum by FFT (decimation but no windowing!) and choose the most intense spectral line. The frequency must be determined extremely accurately by interpolation so that the error remains smaller than 10^{-11} Hz. Then, amplitude and phase are determined by multi-stage iteration by tentatively subtracting a sine function with these properties from the present data. With the *correct* choice of amplitude and phase, the mean amplitude of the rest will be minimal.

As soon as the amplitudes of the strongest spectral lines have fallen below 1/6000 of their initial value, it is time to stop the iterations. After about one hour of computing time and a cup of coffee, you get an interim result of quite low amplitude, which contains hardly any annoying tides. The figure shows the typical spectrum after this second processing step.

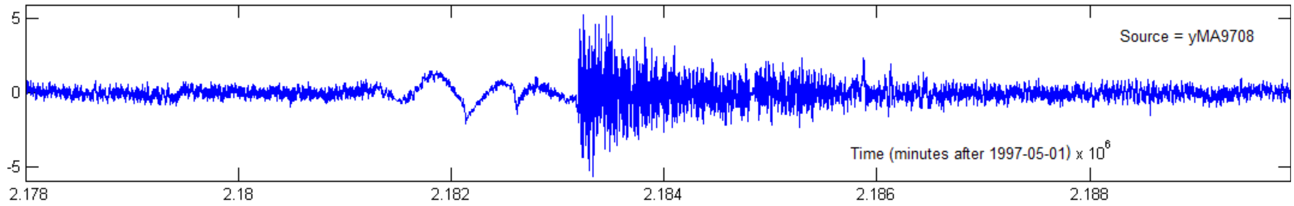


If the noise touches the zero line over and over again, there are no gross data errors. If that is not the case, there is no point in continuing. Then you should elimi-nate the obvious errors and re-run the procedure from step one.

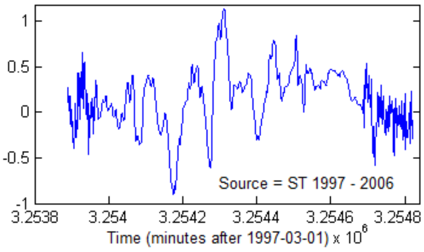
The abrupt transitions at both ends of each data gap are still present, but now they have such low amplitude that they can hardly be distinguished from the noise. No further work is required here. Do not use a low-pass or high-pass filter!

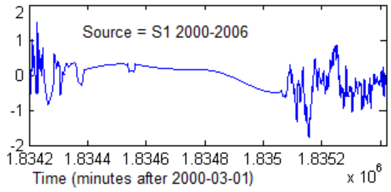
Step 3: Erase the causes of the noise.

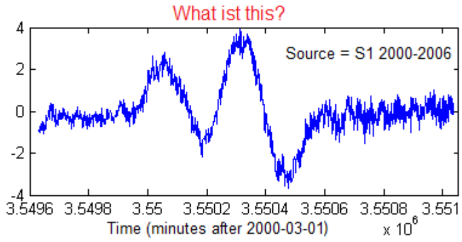
As a rule, a typical SG CORMIN file contains about a hundred segments that are *not* meaningful data but generate noise and *must* be eliminated. The following figure shows an example. Obviously, the SG was overburdened by an earthquake and did not provide meaningful data for 30 hours. The data gap was – not very well done – filled by synthetic data, which should imitate the tides. To simplify the matter, three sections were calculated and put together with visibly bad results. If one calculates the spectrum of this short segment, one realizes that particularly much noise is produced especially in the frequency range, in which the Slichter triplett is assumed.



There is only one way to eliminate such man-made noise: the segment must be replaced by zeros. The bad news: there are many more such sequences of “data” in the records of all SG stations, without exception. Here are some more data segments that are puzzling. It is difficult to understand: You are using sinfully expensive superconducting gravimeters and insert noise!







Step 4: Inverse FFT filter

All previous measures had the sole aim of reducing the amplitudes of the tides with frequencies below 40 µHz or so, *without* increasing the noise level. But the tides are not gone and a renewed application of selective compensation would be extremely time-consuming. A high pass filter should be avoided because it generates noise. But there is a simple and fast method to remove the tidesignals that are still present (= high pass) and at the same time all frequencies above 200 µHz or so (= low pass) without generating noise. This procedure is called FFT filter, you can decide on two different approaches:

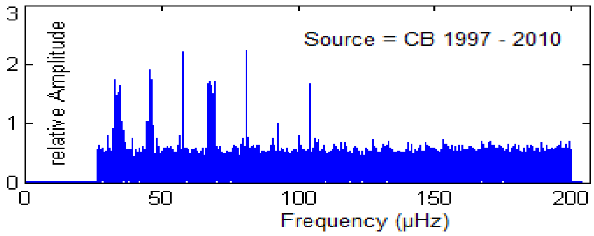
- "quick and dirty": The spectrum of the total signal is calculated with FFT, all unwanted frequency ranges are deleted and the remaining frequency components are used to reconstruct the filtered signal with iFFT. This method generates additional noise and falsifies the desired frequency range because FFT can reconstruct the signal only inaccurately. Below, the reasons are discussed in more detail.
- "clean": the spectrum of the total signal is calculated with FFT, the all amplitudes in the *desired* frequency range are deleted and the remaining frequency components are used to reconstruct an *auxiliary signal* with iFFT. Signals in unwanted frequency ranges can be largely compensated by subtracting the *auxiliary signal* from the total signal. Due to the limited frequency resolution of the FFT, the auxiliary signal is only a rough approximation to the actual signal, which is why the filter does not attenuate all frequencies outside the desired frequency range completely. But the main goal is achieved: The tides are almost extinguished and – more important – this method does *not* generate additional noise.

The difference between the two methods has several reasons: The FFT procedure was invented to reconstruct a data segment of finite length by superimposing many sinusoidal and cosine functions of different frequencies. This approach has several problems:

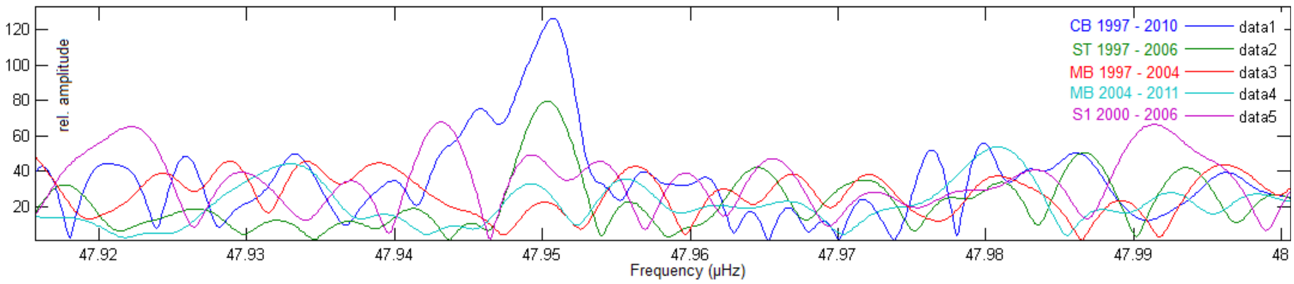
- The data segment is limited whereas sinusoids are unlimited. Therefore the reconstruction is surely faulty. These errors are expressed as increased noise especially if a window function is used.
- Before starting FFT, *you* define how many different sinusoids are allowed (256 up to 2^{22} or more). Since it is very unlikely that even one of those predefined frequencies is identical to an actually frequency occurring in the data series, the reconstruction is surely buggy.

- In most cases, the data is modified by multiplication with a bell curve (Hann window) before FFT. Each multiplication produces unwanted mixing frequencies that increase the background noise. Who needs noise?
- The Fourier transformation was *not* invented to measure frequencies. But it provides good clues as to which frequency range the true values may be found.

For all these reasons, the "clean" method was chosen to suppress the unwanted frequencies below 26 µHz and above 200 µHz.The adjacent figure shows the final spectrum, which can be searched for previously unknown signals. Of course, the usual window function (Hann or similar) is intentionally omitted, because the rectangular window (= no window) produces the least amount of additional noise. During all processing steps, strict attention was paid to the fact that noise sources are detected and deleted and no additional noise is generated.



In the above overview spectrum one recognizes some unusual spectral lines, which merit further attention. One of them can be easily detected in the records of all (previously) tested stations and has the frequency 47.9500 ± 0.0003 µHz. The error was calculated using the jackknife method. The figure shows the superimposition of several single spectra (MB was split due to excess length).



Step 5: Decimation

It is unusual that the file that has gone through the previous processing steps and contains data of a very low frequency range, has a sampling period of 60 seconds. That's too high a time resolution because we look for signals whose oscillation period is a few hours. A decimation by the factor 10 or 60 is appropriate, but requires that the data pass through a low pass filter. This may affect the signal-to-noise ratio.

Final remarks

All programs used (written in MATLAB) can be requested from the author.

It will take some time for further SG data to be prepared with the described methods in order to be able to carry out comparative investigations. A significant portion of the total time is required to detect and remove manual errors that have been added to the CORMIN files.

Acknowledge

Thanks to David Crossley for initiating the Global Geodynamics Project (GGP) and thanks the IGETS data base at GFZ Potsdam for storing the data.