Article

Preprocessing of Gravity Data

Jana Izvoltova 1,*, Dasa Bacova 1, Jakub Chromcak 1 and Stanislav Hodas 2

1 Department of Geodesy, University of Zilina, Univerzitna 8215/1, 010 26 Zilina, Slovakia; bacova@uniza.sk (D.B.); jakub.chromcak@uniza.sk (J.C.)
2 Department of Railway Engineering and Track Management, University of Zilina, Univerzitna 8215/1, 010 26 Zilina, Slovakia; stanislav.hodas@uniza.sk
* Correspondence: jana.izvoltova@uniza.sk; Tel.: +421-41-5135550

Abstract: The paper deals with computation techniques applied in preprocessing of gravity data, which are based on time series analysis by using mathematical and statistical smoothing techniques such as moving average, moving median, cumulative and moving average, etc. The main aim of gravity data preprocessing is to avoid abrupt errors caused by a sudden movement of the subsoil due to human or natural activities or systematic instrumental influences and so provide relevant gravity values, which are then subjected to further processing. The new approach of the described research involves the preprocessing phase in gravity data analysis to identify and avoid gross errors, which could influence the size of unknown parameters estimated by the least square method in the processing phase.

Keywords: gravity data; time series analysis; moving average; moving median

1. Introduction

Gravimetry is a science discipline used to provide important information about the earth’s geodynamical processes. Therefore, it is an integral part of almost all geosciences, including geophysics, geodesy, geology, geotechnics, etc. Gravity observations are useful for providing information on mass distribution within the earth or space, but they depend on time variation. Gravimetric measurements bring temporal gravity variations involving natural and human influences. These gravity variations can be recorded by the absolute or relative gravimeters, depending on particular technology and purpose. While absolute gravity measurements determine the gravity from the fundamental acceleration quantities and time, the modern, relative ones use a counterforce for the determination of gravity differences between different stations or, if carried out in the stationary mode, the variations of gravity with time [1]. The second mentioned observation method brings so-called stationary gravity values, organized in a time series dataset. The final gravity at the station depends on an instrument’s precision and the used processing method, which involves external and internal influences known as gravity corrections [2], instrumental drift, earth tides, and random noise. The precision of the reached gravity depends on identifying measurement errors and their subsequent reduction or elimination if possible. The raw gravity dataset, arranged in a time series, could be sometimes influenced by the sudden changes in readings brought by different causes: sudden earth movements, reading error, internal instrument effects, external influences, etc. These sudden changes in measured signals are called abrupt errors and are expressed as “jumping signals” in graphs.

The paper’s primary goal is to focus attention on the process of identification of abrupt errors of gravity data by using the mathematic-statistical technique used to be preferred mainly in financial engineering. The above-mentioned preprocessing analysis of time series data proceeds to the phase of creating a smoothed model, which is then submitted to further processing. The processing phase includes the minor gravity corrections involving external and internal influences on gravity data. It is not the main topic of the paper. Hence, it is mentioned only marginally at the end of the paper.
2. Time Series Analysis of Gravity Data

The preprocessing phase of gravity data consists of time series analysis to identify abrupt errors in the gravity dataset. Mathematics offers many tools to smooth such errors occurred in a continuous signal, from which methods of moving averages seem to be very useful. It belongs to the mathematic-statistical techniques that are useful for analyzing time series data, mainly in the trading world. In quantitative trading, time series analysis is about fitting statistical models, inferring underlying relationships between series or predicted future values, and generating trading signals. The method of time series analysis is a relatively old method, well described by many scientists like Anderson (1977), Kendall and Stuart (1979), Box, Brockwell (1991), Montgomery (1991), Jenkins and Reinsel (1994), Chatfield (1996), etc. [2–8]. It is applied by many users from various scientific disciplines, which operate with some quantity that is measured sequentially in time over some interval. In geodesy, time series analysis is often used to process long-term global navigation satellite system observations to determine earth and ocean tidal effects in a permanent station network.

Some smoothing techniques have been applied to filter time series data to find the reliable function for cleaning up the raw gravity dataset (Figure 1). A moving average is an effective tool for adjusting a short-term time-series variation [9]. Several moving averages in time series analysis differ with the placement of the adjusted value. A simple “moving average” (MA) or one-sided moving average is placed at the end of values being averaged:

\[ MA_i = \frac{1}{k+1} \sum_{j=0}^{k} x_{i-j} \]  

where index \( i \) changes from \( k + 1 \) to \( n \), \( n \) is the number of data and \( k \) is the order or size of the sampling window. Two-sided MA is centered in the middle of the values being averaged:

\[ MA_i = \frac{1}{2k+1} \sum_{j=-k}^{k} x_{i+j} \]  

where index \( i \) changes from \( k + 1 \) to \( n - k \). While a two-sided moving average can estimate or see the trend, the one-sided moving average can be used as a simple forecasting method. Both types of moving averages use an odd number of periods. However, when working with time series, seasonality effects must be smoothed, which requires the period to be equal to seasonal length. “Centred moving average” (CMA) is often applied for this purpose, which distinguishes from others by using an even number of values:

\[ CMA_i = \frac{MA_{i-0.5} + MA_{i+0.5}}{2} \]

![Figure 1. Time series of gravity dataset (original figure).](image-url)
If the user wants to get the average of all data up until the current datum point, "cumulative moving average" (CuMA) is suitable. In CuMA, the data arrive in an ordered datum stream, and the user would like to get the average of all of the data up until the current datum point

$$CuMA_t = \frac{x_1 + \ldots + x_t}{t}$$

where \( t \) changes from 1 to \( n \). In the case of robust statistics, "Moving median" (MM) is the most suitable to smooth or remove the time series noise. The moving median is not so popular as the moving average, but it provides a more robust estimate of a trend compared to the moving average. It is not affected by outliers, and it removes them. The moving median in a sampling window is calculated from linear approximation defined in the interval:

$$\frac{i}{(n+1)} \leq \alpha \leq \frac{(i+1)}{(n+1)}$$

and is defined as robust statistics corresponding to 50% quantile (\( \alpha = 0.5 \)):

$$u_\alpha = (n+1) \left( \alpha - \frac{i}{n+1} \right) (u_{i+1} - u_i) + u_i$$

An overview of the values of the particular moving averages at critical points of gravity dataset is in Table 1, in which the column SMOOTHED GRAV. represents the adjusted data estimated by the least-squares method

<table>
<thead>
<tr>
<th>DAY</th>
<th>RAW GRAV.</th>
<th>MA</th>
<th>CMA</th>
<th>CuMA</th>
<th>MM</th>
<th>SMOOTHED GRAV.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7192</td>
<td>5894.2870</td>
<td>5894.2900</td>
<td>5894.2920</td>
<td>5894.2530</td>
<td>5894.2900</td>
<td>5894.2870</td>
</tr>
<tr>
<td>0.7200</td>
<td>5894.2720</td>
<td>5894.2830</td>
<td>5894.2860</td>
<td>5894.2530</td>
<td>5894.2870</td>
<td>5894.2910</td>
</tr>
<tr>
<td>0.7207</td>
<td>5894.2930</td>
<td>5894.2840</td>
<td>5894.2860</td>
<td>5894.2530</td>
<td>5894.2930</td>
<td>5894.2950</td>
</tr>
<tr>
<td>0.7215</td>
<td>5894.2950</td>
<td>5894.2870</td>
<td>5894.2880</td>
<td>5894.2530</td>
<td>5894.2940</td>
<td>5894.2940</td>
</tr>
<tr>
<td>0.7222</td>
<td>5894.3100</td>
<td>5894.3110</td>
<td>5894.3110</td>
<td>5894.2660</td>
<td>5894.3100</td>
<td>5894.3100</td>
</tr>
<tr>
<td>0.9381</td>
<td>5894.3280</td>
<td>5894.3160</td>
<td>5894.3110</td>
<td>5894.2660</td>
<td>5894.3110</td>
<td>5894.3140</td>
</tr>
<tr>
<td>0.9389</td>
<td>5894.3100</td>
<td>5894.3150</td>
<td>5894.3150</td>
<td>5894.2660</td>
<td>5894.3100</td>
<td>5894.3100</td>
</tr>
<tr>
<td>0.9396</td>
<td>5894.3110</td>
<td>5894.3160</td>
<td>5894.3160</td>
<td>5894.2660</td>
<td>5894.3110</td>
<td>5894.3110</td>
</tr>
<tr>
<td>0.9404</td>
<td>5894.3110</td>
<td>5894.3110</td>
<td>5894.3150</td>
<td>5894.2660</td>
<td>5894.3110</td>
<td>5894.3110</td>
</tr>
<tr>
<td>0.9885</td>
<td>5894.3020</td>
<td>5894.2980</td>
<td>5894.2970</td>
<td>5894.2680</td>
<td>5894.3020</td>
<td>5894.3020</td>
</tr>
<tr>
<td>0.9893</td>
<td>5894.3110</td>
<td>5894.3030</td>
<td>5894.3030</td>
<td>5894.2680</td>
<td>5894.3020</td>
<td>5894.3010</td>
</tr>
<tr>
<td>0.9900</td>
<td>5894.2920</td>
<td>5894.3020</td>
<td>5894.3010</td>
<td>5894.2680</td>
<td>5894.3020</td>
<td>5894.3010</td>
</tr>
<tr>
<td>0.9908</td>
<td>5894.2990</td>
<td>5894.2990</td>
<td>5894.2990</td>
<td>5894.2680</td>
<td>5894.2950</td>
<td>5894.3010</td>
</tr>
<tr>
<td>0.9916</td>
<td>5894.2950</td>
<td>5894.2940</td>
<td>5894.2990</td>
<td>5894.2680</td>
<td>5894.2950</td>
<td>5894.3000</td>
</tr>
<tr>
<td>Mean square error in mGal</td>
<td>0.0009</td>
<td>0.0011</td>
<td>0.0468</td>
<td>0.001</td>
<td>0.0027</td>
<td></td>
</tr>
</tbody>
</table>

“Weighted moving average” (WMA) is often applied in technical analysis of financial data with the specific meaning of weights that decrease arithmetical progression. In a \( p \)-day of WMA, the latest day has weight \( p \), the second latest \( p-1 \), etc., down to one:

$$WMA = \frac{p x_1 + (p - 1) x_2 + \ldots + x_i}{p(p + 1)}$$

where index \( i \) means the time and \( p \) is the weight. The weighted moving average has not been applied in the smoothing analysis of gravity data because of the unreasoning use of weights in gravity time series.

“Exponential moving average” (EMA) is also weighted toward the most recent value, but the rate of decrease between one value and its preceding value is not consistent. The difference in the decrease is exponential. The formula for EMA [10] is as follows:
\[ EMA = p(x_i + (1 - p)x_{i-1} + (1 - p)^2x_{i-2} + \ldots + (1 - p)^kx_{i-k}) + (1 - p)^{k+1}EMA_{i-(k+1)} + \ldots \] (8)

where \( k \) changes from 0 to \( n \). The weight of the general datum point is \( p(1 - p)^j \). WMA and EMA techniques seem to be more suitable for comparing more grav observed in various time intervals.

Besides the mentioned smoothing techniques, there are many other signal processing methods. Among them, wavelet denoising plays an essential role in separating clean images from noisy images or filtering airborne gravity data [11,12].

3. Abrupt Error Identifying

Scientists use many methods to identify outliers or gross errors in observation. Most of them utilize statistical hypothesis testing based on comparing the testing value with the critical one, which is represented by the quantile of a probability distribution. Abrupt error is assumed to be an outlier of the gravity dataset, which is indicated by comparing the corresponding residual value with the critical value estimated from the equation:

\[ T_{\alpha} = t \sigma_{GRAV} \] (9)

where \( \sigma_{GRAV} \) is the standard deviation of gravimeter SCINTREX Autograv CG-5 [13] defined by producer by the value 0.5 mGal and \( t \) is a confidential coefficient defined as a quantile of Student distribution dependent on specified probability. Quantitatively is \( t \) equal to 2.

Experimental measurements were obtained from three days of static observation of gravity acceleration realized by autograph with the setup read time of 60 s, which responded to 4331 observation cycles. From the Autograv optional parameters, tide and terrain corrections have been disallowed to perform the requisite functions in the processing phase.

Statistical hypothesis testing has been applied to confirm the existence of the abrupt error. It compares the testing value represented by the corresponding residual \( e_i \) (Table 2) with the critical value \( T_{\alpha} = 0.010 \) mGal by defining the probability value \( p = 0.95 \). As illustrated in Figure 1, three abrupt errors have been identified in the gravity dataset with the appropriate values of residuals and moving averages.

<table>
<thead>
<tr>
<th>DAY</th>
<th>RAW GRAV.</th>
<th>( e_{MA} )</th>
<th>( e_{CMA} )</th>
<th>( e_{CaMA} )</th>
<th>( e_{MM} )</th>
<th>( e_{SMOOTH} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7192</td>
<td>5894.2870</td>
<td>−0.0039</td>
<td>−0.0049</td>
<td>0.0338</td>
<td>−0.003</td>
<td>−0.0039</td>
</tr>
<tr>
<td>0.7200</td>
<td>5894.2720</td>
<td>−0.0189</td>
<td>0.0188</td>
<td>−0.015</td>
<td>−0.0189</td>
<td></td>
</tr>
<tr>
<td>0.7207</td>
<td>5894.2930</td>
<td>0.0020</td>
<td>0.0075</td>
<td>0.0397</td>
<td>0.0060</td>
<td>0.0020</td>
</tr>
<tr>
<td>0.7215</td>
<td>5894.2950</td>
<td>0.0039</td>
<td>0.0417</td>
<td>0.0020</td>
<td>0.0039</td>
<td></td>
</tr>
<tr>
<td>0.7222</td>
<td>5894.2940</td>
<td>0.0029</td>
<td>0.0064</td>
<td>0.0407</td>
<td>0.0000</td>
<td>0.0029</td>
</tr>
<tr>
<td>0.9374</td>
<td>5894.3100</td>
<td>−0.0043</td>
<td>−0.0005</td>
<td>0.0443</td>
<td>0.0000</td>
<td>−0.0043</td>
</tr>
<tr>
<td>0.9381</td>
<td>5894.3260</td>
<td>0.0138</td>
<td>0.0623</td>
<td>0.0170</td>
<td>0.0138</td>
<td></td>
</tr>
<tr>
<td>0.9389</td>
<td>5894.3100</td>
<td>−0.004</td>
<td>−0.0048</td>
<td>0.0442</td>
<td>0.0000</td>
<td>−0.004</td>
</tr>
<tr>
<td>0.9396</td>
<td>5894.3110</td>
<td>−0.0029</td>
<td>0.0452</td>
<td>0.0000</td>
<td>−0.0029</td>
<td></td>
</tr>
<tr>
<td>0.9404</td>
<td>5894.3110</td>
<td>−0.0028</td>
<td>−0.0039</td>
<td>0.0452</td>
<td>0.0000</td>
<td>−0.0028</td>
</tr>
<tr>
<td>0.9885</td>
<td>5894.3020</td>
<td>0.0006</td>
<td>0.0048</td>
<td>0.0343</td>
<td>0.0060</td>
<td>0.0006</td>
</tr>
<tr>
<td>0.9893</td>
<td>5894.3110</td>
<td>0.0099</td>
<td>0.0433</td>
<td>0.0090</td>
<td>0.0099</td>
<td></td>
</tr>
<tr>
<td>0.9900</td>
<td>5894.2920</td>
<td>−0.0088</td>
<td>−0.0087</td>
<td>0.0242</td>
<td>−0.01</td>
<td>−0.0088</td>
</tr>
<tr>
<td>0.9908</td>
<td>5894.2950</td>
<td>−0.0056</td>
<td>0.0272</td>
<td>0.0000</td>
<td>−0.0056</td>
<td></td>
</tr>
<tr>
<td>0.9916</td>
<td>5894.2950</td>
<td>−0.0053</td>
<td>−0.0041</td>
<td>0.0272</td>
<td>0.0000</td>
<td>−0.0053</td>
</tr>
</tbody>
</table>

4. Regression Function as the Smoothing and Estimating Function

The smoothed method depends on the amount and character of time series data or the operator’s possibilities and preferences. However, the simplest way to avoid the abrupt error seems to be to apply the reliable smoothing and estimating technique to adjust raw
gravity data and estimate the unknown parameters. Time series gravity data assumes to use nonlinear regression function used to estimate unknown parameters. The advantage of combining both moving average and regression function in gravity dataset (Figure 1) leads to computation of the reliable earth parameters from the mathematical model:

\[
\Delta g_i = \Delta g_0 + g_1(t_i - t_0) + g_2(t_i - t_0)^2 + g_3(t_i - t_0)^3 + a\sin \frac{t_i - t_0}{T_1} + b\cos \frac{t_i - t_0}{T_1} + c\sin \frac{t_i - t_0}{T_2} + d\cos \frac{t_i - t_0}{T_2} + \Delta_i
\]

where \(\Delta g_0, g_1, g_2, g_3\) are estimated unknown parameters of a gravimeter drift \(a, b, c, d\) are unknown parameters of earth tides depending on time changes \(t_i - t_0\), which are needed to compute amplitude \(A_i\) and phase \(p_i\) of a particular tidal wave with its period \(T_i\) according to the following formulas [14]:

\[
A_i = \sqrt{a_i^2 + b_i^2}
\]

\[
p_i = \arctan \frac{b_i}{a_i}
\]

The appropriate coefficients of gravimeter drift, the earth tides parameters estimated by the periods \(T_1 = 1.075806\) and \(T_2 = 0.517525\) defined for tidal wave components O1 and M2, and calculated values of amplitude and phase are displayed in Table 3.

<table>
<thead>
<tr>
<th>(\Delta g)</th>
<th>(g_1)</th>
<th>(g_2)</th>
<th>(g_3)</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
<th>(A)</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5894.269</td>
<td>-0.000193</td>
<td>-0.000004</td>
<td>0.000000</td>
<td>-0.004785</td>
<td>-0.028121</td>
<td>-0.013337</td>
<td>0.021867</td>
<td>0.028525</td>
<td>1.402253</td>
</tr>
</tbody>
</table>

The last symbol of Equation (10) represents a random gravitational noise that can be linear or harmonic, in nature and represents the unused rest of the harmonic series. The unknown regression parameters as the drift and tidal parameters of gravity data and their relevant variance components were estimated by the least square method.

The gravity preprocessing finishes with the creation of the smoothed gravity model, displayed in Figure 2, which was prepared for the data processing to apply nominal gravity corrections.

![Smoothed gravity model (original figure)](image)

**Figure 2.** Smoothed gravity model (original figure).

### 5. Discussion

The paper’s main topic is to apply the reliable smoothing technique for the time series analysis of the gravity dataset. While the moving average, moving median, and centred moving average seems to be very useful in the process of gross error identification, the
cumulative moving average appears to be not so sensitive to smooth abrupt errors in the gravity data. Weighted and exponential smoothing techniques appear to be not suitable to use in gravity data preprocessing. Mathematics provides another smoothing technique based on data adjustment and estimation. The most suitable for gravity data seems to be the nonlinear regression with the least square restriction.

6. Conclusions

The smoothed gravity model was then subjected to the implementation of the nominal gravity corrections to find out the mathematical model of gravity acceleration at the station. The processing phase consists of involving external and internal influences represented by earth and ocean tides, hydrogeological forces, atmospheric pressure and temperature changes, seismic and terrain corrections, tilt and calibration corrections, and instrumental drift, which occurs due to the stress relaxation in the elastic quartz system of modern relative gravimeters. The most mathematical model involves random noise, representing random disturbances of gravity data.

Besides the mentioned effects, other disturbances influence the actual gravity acceleration value at the station as polar motion, instrument and station origin, earthquakes, tectonics, etc. Most of these influences are system parameters and are involved as firmware filters.

The widely used processing method of gravity data is a remove-restore technique [15,16]. This general method of gravity data processing is based on two phases. The first one, called the remove phase, is based on making nominal corrections for the largest influences, such as tides and pressure. Then, the problems are fixed in the residual signal and returned to the removed signals in the second restore phase.

Another approach to gravity data preprocessing seems to be finding the causes of abrupt error occurrences, which can have various origins consisting of external, instrumental, or human sources.

Author Contributions: Practical outputs, solution, analysis, D.B. and J.C.; Conceptualization, S.H.; Methodology, writing J.I.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Ministry of education, science, research and sporz of the Slovak Republic, grant number: VEGA 1/0643/21.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: There is only a part of data from own direct measurements in the manuscript. No conflict with their using.

Acknowledgments: This article is the result of the implementation of the project VEGA 1/0643/21, “Analysis of spatial deformations of a railway track observed by terrestrial laser scanning”, supported by the Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic and Slovak Academy of Sciences. This article is the result of implementing the project KEGA 038ŽU-4/2020, “New approaches to teaching physical geodesy in higher education”, supported by the Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic.

Conflicts of Interest: The authors declare no conflict of interest.

References


