Outlier detection and correction for GRACE data to improve the continental water balance

M. J. Tourian1,2, J. Rieger1, B. Devaraju3 and N. Sneeuw2

1. Institute of Hydraulic Engineering, University of Stuttgart, Germany
2. Institute of Geodesy, University of Stuttgart, Germany
mohammad.tourian@iws.uni-stuttgart.de

Abstract

GRACE (gravity recovery and climate experiment) gravity measurements provide a direct measure of monthly changes in mass over the Earth’s continents. As such changes in mass mainly correspond to water storage changes, these measurements allow to close the continental water balance on large spatial scales and on a monthly time scale within the respective error bounds. When quantifying uncertainties, positive and negative peaks are detected in GRACE aggregated time series that do not correspond to hydrological or atmospheric signals. These peaks must be interpreted as outliers, which carry the danger of signal degradation. In this study an algorithm is developed to identify outliers and replace them with appropriate values. The procedure of outlier detection is verified by evaluating catchment based aggregated GRACE signals with ground truth from hydrology and atmospheric signals. The results show a significant improvement in the correlation of GRACE versus atmospheric and hydrological signals over 255 catchments. Also, the noise level is significantly reduced.

1. Problem Statement

In order to obtain meaningful error estimates for remote sensing missions an evaluation with ground truth measurements is needed. However, ground truth measurements for an evaluation of GRACE mass estimates seem to be cumbersome [5]. However on landmasses, storage changes derived from hydrological data or from atmospheric models can be used as ground truth for GRACE temporal derivatives aggregated over certain area e.g. catchment which is formulated as:

\[ \frac{dM}{dt} \rightarrow P - R - ET \]  \hspace{1cm} (1a)

\[ \frac{dM}{dt} \rightarrow MFD - R \]  \hspace{1cm} (1b)

where \( \frac{dM}{dt} \) is derivative of aggregated mass deviation over individual catchments (hereafter mass derivative), \( P \) is precipitation, \( R \) is runoff, \( MFD \) is moisture flux divergence and \( ET \) is actual evapotranspiration. In other words, hydrological \( (P - R - ET)\) and atmospheric \( (MFD - R)\) water storage changes over any spacial area are compared with mass changes rates. During comparison, positive and negative peaks are detected in GRACE aggregated time series that do not correspond to ground truth (Figure 1).

2. Data

In this study GRACE monthly solutions have been taken from different sources within the time period of February 2003 to December 2007, GFZ (GeoForschungsZentrum, Germany), JPL (Jet Propulsion laboratory, USA), CSR (Center for space research, University of Texas at Austin), ITG (University of Bonn). Subsequently, the monthly solutions have been filtered using the decorrelation filter of Swenson and Wahr (2006) in combination with a Gaussian 500 km filter [4]. The filtered monthly solutions are then aggregated over individual catchments. Precipitation, runoff and MFD data have been taken from GPCC (global precipitation climatology center), GLHCC (global runoff data center) and ERA-interim (ECMWF reanalyses products), respectively. As ET, it is considered an unknown term, for the evaluation of outliers by hydrological storage changes rates, only \( P \) and \( R \) are considered which are the most dynamic components of the water balance (equation (1)).

3. Outlier Detection Algorithm

Our outlier detection algorithm is based on the data snooping method [1, 2, 6]. Following flowchart shows the algorithm:

- By applying the algorithm over different available GRACE products (GFZ, JPL, CSR, ITG) aggregated over catchments and by cross checking with the atmospheric and hydrological signals, confidence levels of 97% (\( \alpha = 0.03 \)) for data snooping on the signal and of 98% (\( \alpha = 0.02 \)) for data snooping on the residual have been found as appropriate confidence levels.

4. Result of Outlier Identification for Different Products

By applying the algorithm over different available GRACE products (GFZ, JPL, CSR, ITG) aggregated over catchments and by cross checking with the atmospheric and hydrological signals, confidence levels of 97% (\( \alpha = 0.03 \)) for data snooping on the signal and of 98% (\( \alpha = 0.02 \)) for data snooping on the residual have been found as appropriate confidence levels.

5. Impact of Outlier Correction

The following investigations are concerned with the impact of outlier correction on correlation of hydrological signals (\( P - R \)) versus GRACE (\( MFD - R \)) and atmospheric storage changes (\( MFD - R \)) versus GRACE (\( MFD - R \)).

6. Discussion

The impact of the outlier correction emphasizes the necessity of outlier correction. Despite the effectiveness of the outlier identification and correction scheme the reasons for the outlier is not clear yet. As the different products show a similar number of outliers yet with only a small number of coincidences between the products the conclusion must be drawn, that outlier generation strongly depends on the data processing algorithms.

Table 1: Number of detected outliers from different products over 255 catchments within the time period of Feb. 2003-Dec. 2007.

<table>
<thead>
<tr>
<th>Products</th>
<th>Number of Outliers</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFZ</td>
<td>372</td>
<td>2.5%</td>
</tr>
<tr>
<td>JPL</td>
<td>316</td>
<td>2.1%</td>
</tr>
<tr>
<td>CSR</td>
<td>339</td>
<td>2.2%</td>
</tr>
<tr>
<td>ITG</td>
<td>426</td>
<td>2.8%</td>
</tr>
</tbody>
</table>

Table 2: Average of MMSNR before and after outlier correction for different products.

<table>
<thead>
<tr>
<th>Products</th>
<th>MMSNR before</th>
<th>MMSNR after</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFZ</td>
<td>1.38</td>
<td>1.63</td>
</tr>
<tr>
<td>JPL</td>
<td>1.71</td>
<td>1.81</td>
</tr>
<tr>
<td>CSR</td>
<td>1.51</td>
<td>1.66</td>
</tr>
<tr>
<td>ITG</td>
<td>1.30</td>
<td>1.62</td>
</tr>
</tbody>
</table>

Figure 1: a) Mass deviation on Okavango catchment from GFZ with a distinct isolated peak on September 2004. b) \( \frac{dM}{dt} \) (mass derivative) from GFZ, JPL, CSR and ITG over Okavango catchment with a distinct double peak derived from mass deviation signal.

Figure 2: Flowchart of outlier identification algorithm.

Figure 3: Mass deviation signal on the Okavango catchment with \( \alpha_1 \) and \( \alpha_2 \) for each year with the corrected mass deviation signal.

Figure 4: a, b, c and d of show 372, 316, 339 and 426 detected outliers from GFZ, JPL, CSR and ITG catchment based aggregated monthly solution respectively. The considered time series for each product is between Feb. 2003 and Dec. 2007 and the catchments were sorted in term of area descendingly.

Figure 5: Scatter plot of GRACE mass derivative \( (MFD - R) \) versus \( P - R \) for Columbia river catchment before (a) and after (b) outlier correction.

Figure 6: Polar diagram of mass derivative \( (MFD - R) \) versus \( P - R \) for 57 catchments. Correlation is shown in arc and ratio of standard deviation (standard deviation of mass derivative divided by standard deviation of \( \frac{dM}{dt} \) ) is depicted on radius.

Figure 7: Time-variability of the Earth’s gravity field: Hydrological and oceanic effects and their possible detection using GRACE. Journal of Geophysical Research. 103:30205-30320, 1998.

References