

Global Trends in Equivalent Water Height with GRACE and GLDAS

Christopher McCullough

GEO 386G

December 5, 2013

1 Introduction and Background

The time variable nature of the Earth system is of utmost importance to the study of the climate and evolution of our planet. One aspect of note is the variation of the water cycle; most importantly its seasonal and long term evolution. This study aims to look into these signals as derived from the Gravity Recovery and Climate Experiment (GRACE) satellites and the Global Land Data Assimilation System (GLDAS). More specifically, for a time period from July 2003 to December 2010, the study looks at how global land water storage is evolving and how each data set correlates with one another. In addition, GRACE data is used to examine sea level variations and their evolution over time.

1.1 Gravity Recovery and Climate Experiment (GRACE)

The GRACE mission is jointly implemented by NASA and the German Aerospace Center (DLR) under the NASA Earth System Science Pathfinder (ESSP) Program and was launched in March 2002 [1]. It consists of two co-orbiting satellites flying in a near polar orbit separated by an along track distance of 220 kilometers. The primary objective of the GRACE satellites is to obtain highly accurate global models of the mean and time variable portions of the Earth's gravitational field. This done by precisely measuring the distance between the two spacecraft. As the satellites move over an area of increased mass concentration, the lead satellite is accelerated more quickly than the trailing satellite, and vice versa as the satellites move away. This continuous variability in inter-satellite range is then used to estimate the gravity field. The data from GRACE used in this study is published monthly and is given in terms of spherical harmonic coefficients.

1.2 Global Land Data Assimilation System (GLDAS)

The GLDAS was developed, in tandem, by the National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (GSFC) and the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Prediction (NCEP) [2]. It gives accurate, high resolution estimates of terrestrial water and energy storage by using a combination of ground and space observation systems, as well as land surface models. The data product used here publishes monthly equivalent water height in a one degree by one degree grid.

2 Data Collection and Pre-Processing

In each case (GRACE/GLDAS), the data was downloaded into a separate ascii text file for each month of data. These ascii text files were then preprocessed using the Texas Advanced Computing Center (TACC) supercomputer (Lonestar) to create new ascii text files which

can be imported into ArcGIS. The ArcGIS format that allows for ascii to raster conversion is shown below in table 2.

Table 1: Ascii file format that ArcGIS can convert into a raster.

```

NCOLS 360
NROWS 180
XLLCORNER 0.000
YLLCORNER -90.000
CELLSIZE 1
NODATA_VALUE 32767.000
<ROW 1 OF DATA >
<ROW 2 OF DATA >
<... >
<ROW 180 OF DATA>

```

2.1 GRACE Data

The GRACE gravity field data is published in monthly spherical harmonic coefficients, out to degree/order 60, and requires a large amount of pre-processing in order to import it into ArcGIS. The data was downloaded from <http://www.csr.utexas.edu/grace/RL05.html>, courtesy of the University of Texas at Austin (UT) Center for Space Research (CSR) [3]. Due to data outages and calibration maneuvers, some months are not available; therefore, the longest continuous stretch of data was found and used for this study. It spans from July 2003 to December 2010.

The spherical harmonic coefficients provided by the GRACE mission, given a suitable value for the Earth's equipotential surface at mean sea level, are used to calculate the geoid height [4]; however, we would like a suitable measure of the time variable water storage. This can be done using a procedure outlined by Wahr [5]. Here, assuming all background processes have been modeled (atmosphere, ocean, solid Earth, etc) and assuming all time variable signal is concentrated near the surface of the Earth, the spherical harmonic coefficients are converted to surface density by equation 1.

$$\Delta\sigma(\phi, \lambda) = a_e \rho_w \sum_{l=0}^{\infty} \sum_{m=0}^l \bar{P}_{lm}(\cos\phi) (\Delta\hat{C}_{lm} \cos(m\lambda) + \Delta\hat{S}_{lm} \sin(m\lambda)) \quad (1)$$

In this expression, $\Delta\sigma$ is the surface density, a_e is the mean Earth radius, ρ_w is the density of water, \bar{P}_{lm} are the normalized associated Legendre polynomials, ϕ is the geocentric latitude, λ is the longitude, l is the degree, m is the order and $\Delta\hat{C}_{lm}/\Delta\hat{S}_{lm}$ are variations from the mean of rescaled spherical harmonic coefficients given by equation 2.

$$\begin{aligned}\Delta\hat{C}_{lm} &= \frac{\rho_{ave}}{3\rho_w} \frac{2l+1}{1+k_l} \Delta C_{lm} \\ \Delta\hat{S}_{lm} &= \frac{\rho_{ave}}{3\rho_w} \frac{2l+1}{1+k_l} \Delta S_{lm}\end{aligned}\tag{2}$$

In this expression, ρ_{ave} is the average density of the Earth (5517 kg/m³), ρ_w is the density of water, l is the degree, k_l is the Love number of degree l used to account for the change in shape of the elastic Earth, and $\Delta C_{lm}/\Delta S_{lm}$ are variations from the mean of the normalized spherical harmonic coefficients given by the CSR data product. The value used throughout this study is equivalent water height, and is given by $\frac{\Delta\sigma}{\rho_w}$ (from equation 1).

In addition to the process outlined above to convert the spherical harmonic coefficients into equivalent water height, the GRACE data contains north-south “stripes” due to high frequency errors in the harmonic coefficients. In order to remove these errors, a Gaussian smoothing process is used. In the cases shown in this study, a Gaussian smoothing kernel with a radius of 400 kilometers was used in conjunction with the process outlined by Wahr [5] to create an ascii text file in the proper ArcGIS format, shown in table 2.

2.2 GLDAS Data

The GLDAS data was acquired in monthly gridded, 1°x1°, data sets in the form of text files. These were downloaded from <http://grace.jpl.nasa.gov/data/gldas/> which used the Goddard Earth Sciences Data and Information Center. The data was already given in a format identical to the pre-processed CSR (GRACE) data, in equivalent water height. This data was then preprocessed in order to convert the text files from their downloaded format into the ascii format readable by ArcGIS.

3 ArcGIS Processing and Results

The general ArcGIS processing consisted of importing each monthly file for each data set. These individual raster files were then compiled to create 2 multiband rasters, one for the CSR data and one for the GLDAS data. These multiband rasters each contain 90 different bands (1 for each month), creating 2 four dimensional data sets. These data sets were then manipulated using various tools in ArcGIS to compute global averages, trends for each grid point, and correlation between each data set.

3.1 4D Data Creation

First, the four dimensional data sets had to be created for CSR and for GLDAS. These were created using a variety of steps.

1. Each ascii file was imported using the function *ASCII to Raster*. This allowed for the creation of individual rasters for each month of data by importing each ascii file, formatted as shown in table 2. To expedite the process, the function *ASCII to Raster* was used in batch mode by right clicking on the function in the ArcToolbox. A screen shot shows the beginnings of this process in figure 1.

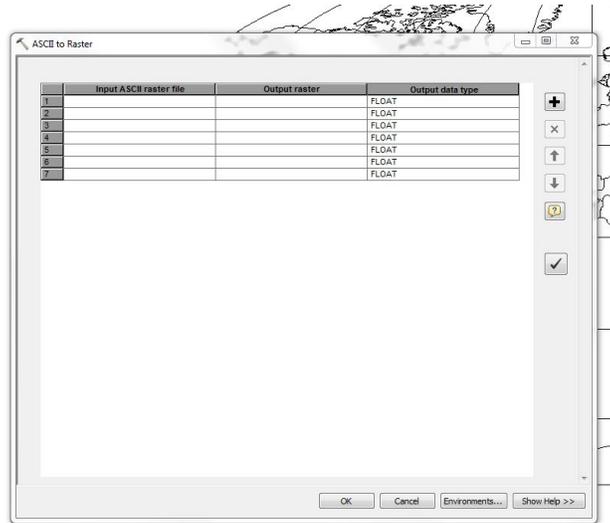


Figure 1: Screen shot of *ASCII to Raster* in batch mode.

2. Next, each individual raster was compiled into one multiband raster. This was done so that one raster file could be used to display all time steps. In other words, each data set (CSR and GLDAS) was compiled into a 90 band raster file, with each band corresponding to a different month. This compilation was performed using the *Composite Bands* tool and is shown in figure 2.

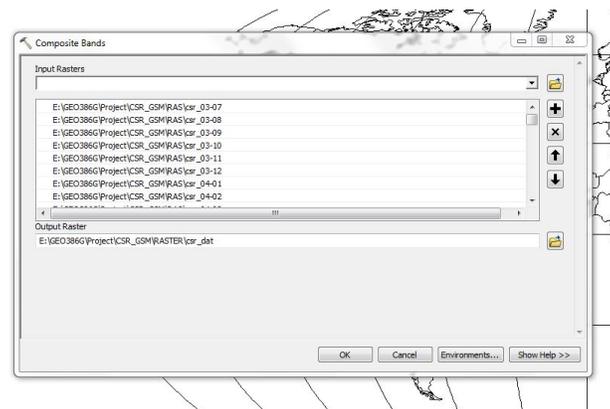


Figure 2: Screen shot of *Composite Bands* to combine individual rasters into one multiband raster

3. Next, each multiband raster had to be given a spatial reference frame in order for

ArcMap to project it properly. The spatial reference given to these data sets was WGS 1984.

4. The final step in the creation of the four dimensional data sets was specific to the CSR data. The GLDAS data set is only valid over land, so this data only contained values over these regions (the rest were null values); however, the CSR data set is global. In order to make comparisons, the CSR data had to be masked to cover the same land area as the GLDAS data. This was done using the *Extract by Mask* tool. Here the GLDAS data was used as the mask so the null values in the GLDAS data would become null values in the CSR data, as shown in figure 3.

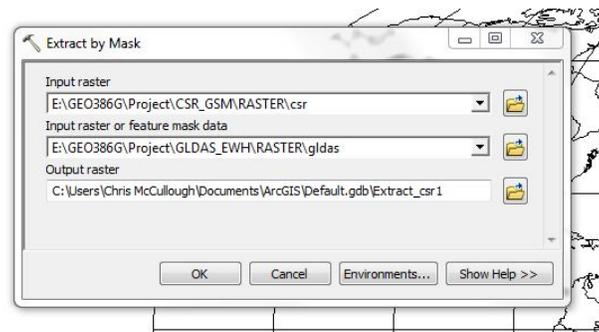
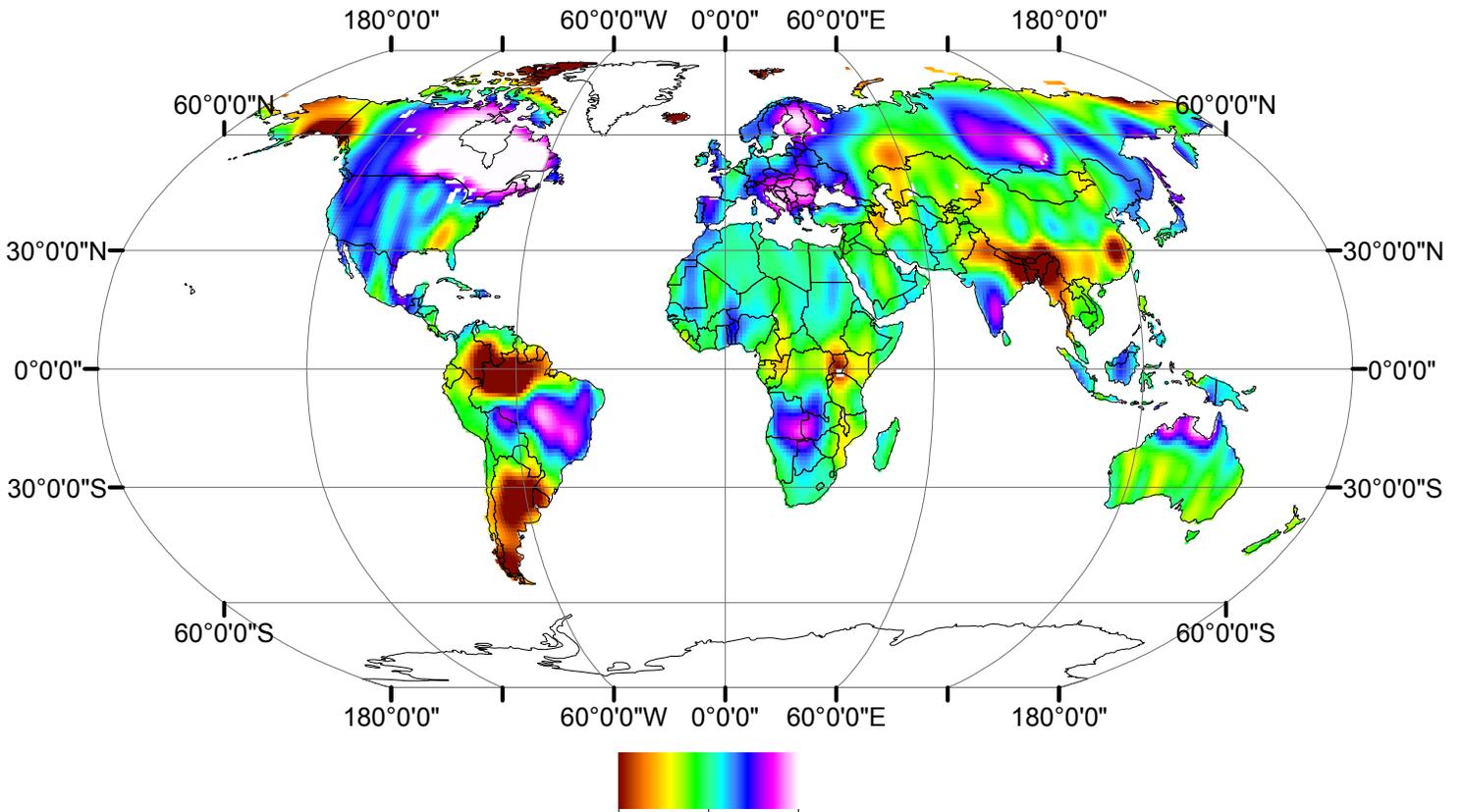


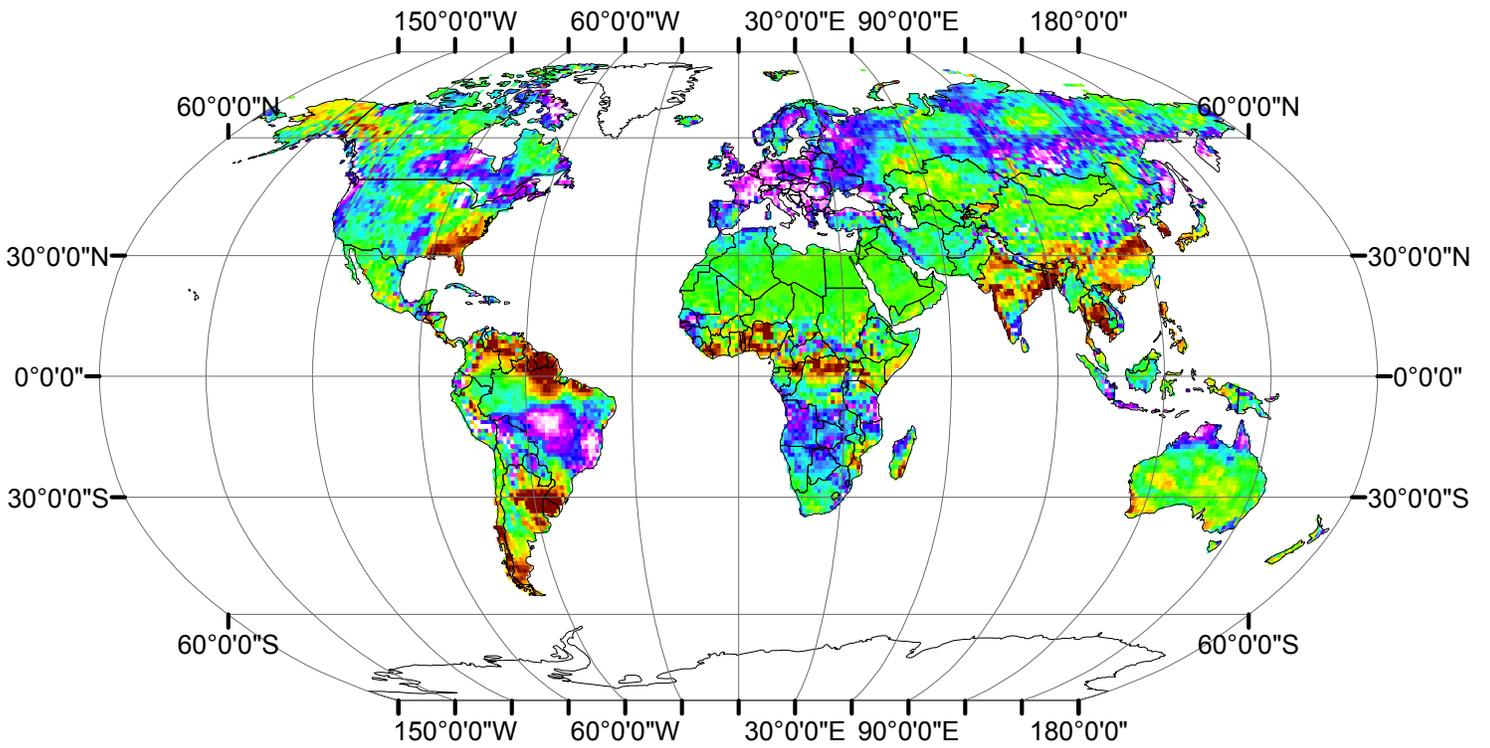
Figure 3: Screen shot of *Extract by Mask* to allow CSR data to cover same land areas as GLDAS data.

The outcome of this process, when performed for the CSR and GLDAS data sets, is two 90 band raster data sets. The following map showcases the band corresponding to July 2003 of each of these data sets. Even though each data set has 1° grids, it is apparent how the truncation at spherical harmonic degree 60 and the 400 kilometer smoothing performed on the CSR data set effects its appearance, when compared to the GLDAS data. In addition, the smoothing also creates a zero bias, as evidenced by the high and low values shown in the CSR data compared to the GLDAS data. However, this is not an issue when looking at the time variability, as opposed to the absolute monthly values.

Global Equivalent Water Height for July 2003 (cm) for CSR (Top) and GLDAS (Bottom)



High : 28.5 Low : -22.4



High : 201.828 Low : -68.107

Robinson Projection
1:200,000,000
WGS 1984

3.2 Global Averages

The next step performed in the sequence of ArcGIS processing steps was the collection of global averages by band. As stated in previous section, each data set contains 90 bands corresponding to different months of data. If the average value of each band is collected, a global average of water storage over land can be examined. This was computed using the *Band Collection Statistics* tool, as shown in figure 4.

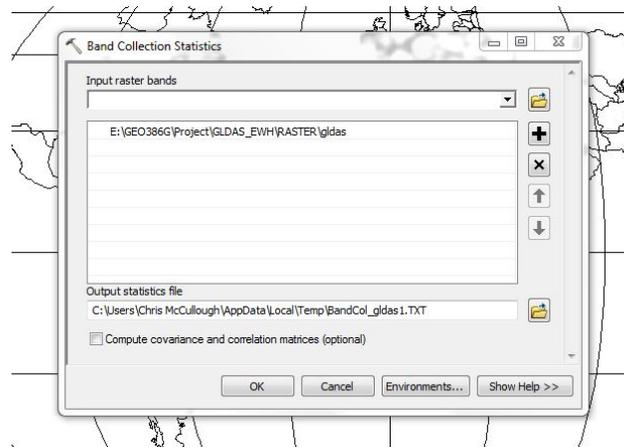


Figure 4: Screen shot of *Band Collection Statistics* tool to calculate global mean water storage per band.

The output of this tool is a text file containing a list of the minimum value, maximum value, mean value, and standard deviation for each separate band. To get global average water storage, the mean value is used. A plot of the time series of global water storage, for CSR and GLDAS, is shown in figures 5 and 6 respectively.

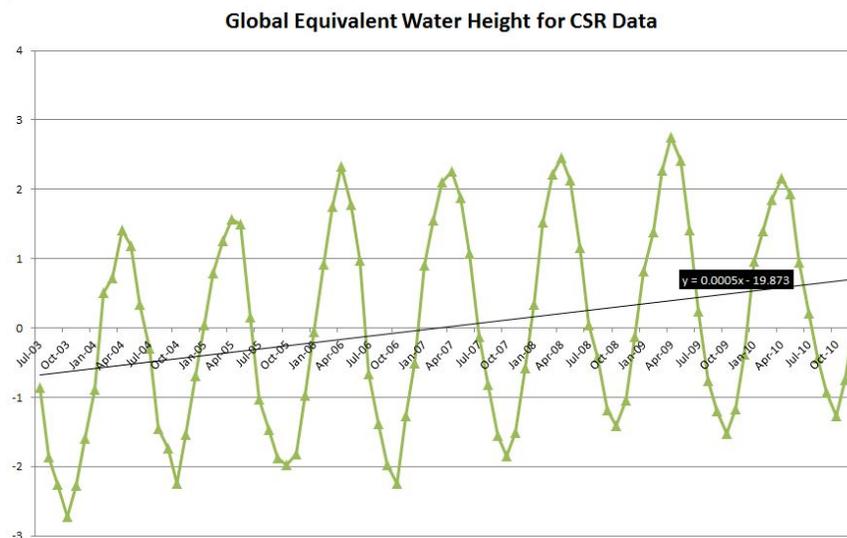


Figure 5: Global average water storage (cm) over land for CSR data set (trend = 1.8 mm/yr).

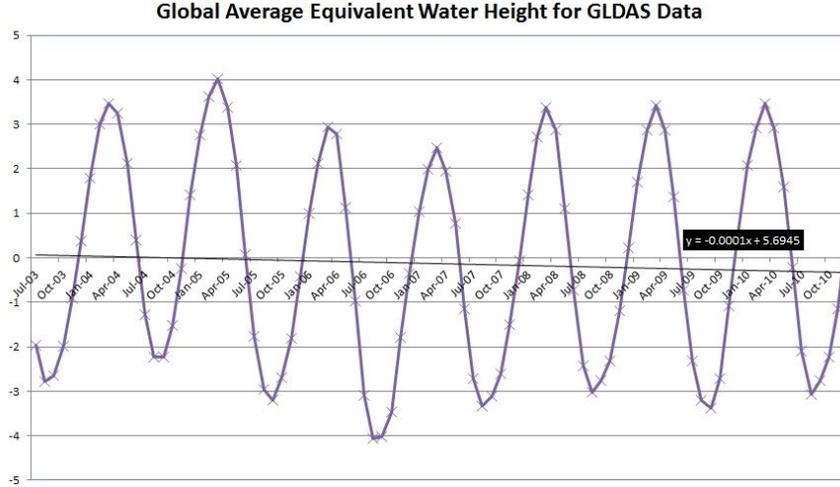


Figure 6: Global average water storage (cm) over land for GLDAS data set (trend = -0.37 mm/yr).

The figures show a strong seasonal signal along with some slight trends; however, the trends are very small and do not agree with one another. One possible explanation for their disagreement is the smoothing process performed on the CSR data set. The smoothing process, while removing correlated errors, also causes signal “leakage” between spatial areas. Therefore, it is possible that some signal from the oceans has been captured on the land area due to the smoothing. In addition, each of the trends observed are very small and may just be a product of noise in each data set. Therefore, these trends are inconclusive.

3.3 Gridded Water Storage Trends

Expanding upon the ideas of the previous section, a trend can be fit at each grid point. In other words, for each data set and at each grid point, a time series of 90 months is available at each 1°x1° grid point. A slope can be fit to each one of these time series and displayed on a map. Since there is no explicit tool in ArcGIS to do this, a variety of map algebra operations had to be used.

A slope can be fit to a time series using least squares methods. Usually this involves operations with matrices; however, we are only in need of the slope so the operations are simplified into cell by cell operations of multiplication. The general least square process can be broken down into equations as shown in equation 3.

$$\frac{1}{t_1^2 + t_2^2 + \dots + t_n^2} (t_1 y_1 + t_2 y_2 + \dots + t_n y_n) \quad (3)$$

Here, t is a time value (in this case each month), y is the data value (the value of equivalent water height in each cell), and the subscript shows the associated raster band. These

operations were performed cell by cell to determine the slope at each grid value using the following steps.

1. First, the value of the first part of equation 3 can be easily computed and loaded into a constant raster. Assuming that the time values are as given as fractions of a year, as shown in table 1, the constant raster is created using the *Create Constant Raster* tool.

Table 2: Values used for time in least squares calculation of slopes.

t_1	$\frac{1}{24}$
t_2	$\frac{3}{24}$
t_3	$\frac{5}{24}$
\vdots	\vdots
t_{90}	$\frac{179}{24}$

Computing the term on the right hand side of equation 3 yields a constant value of approximately 0.0006. In other words,

$$\frac{1}{t_1^2 + t_2^2 + \dots + t_n^2} = 0.000593610883$$

This value is then used to create the constant raster.

2. Next, constant rasters are created corresponding to each t_i value. This is done using the *Create Constant Raster* tool again, except using the batch mode. This is shown in figure 7.

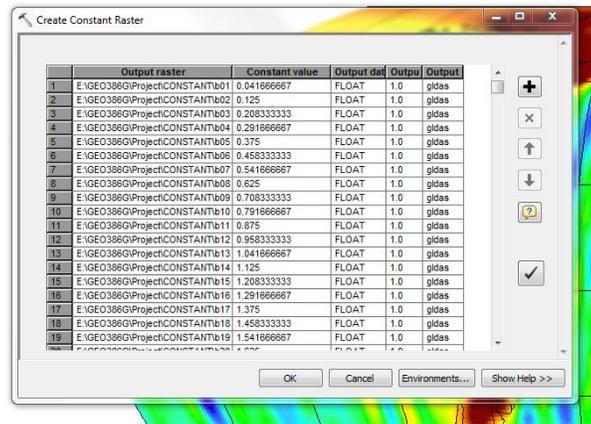


Figure 7: Screen shot of the *Create Constant Raster* tool.

- Once the constant rasters are created, the next step is to multiply each constant raster by its corresponding band of data. This is done using the raster math tool *Times* in batch mode. A screen shot is shown in figure 8.

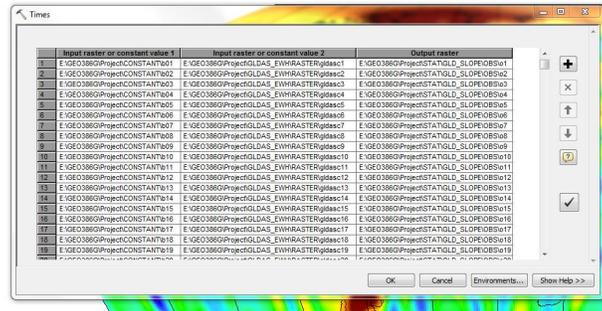
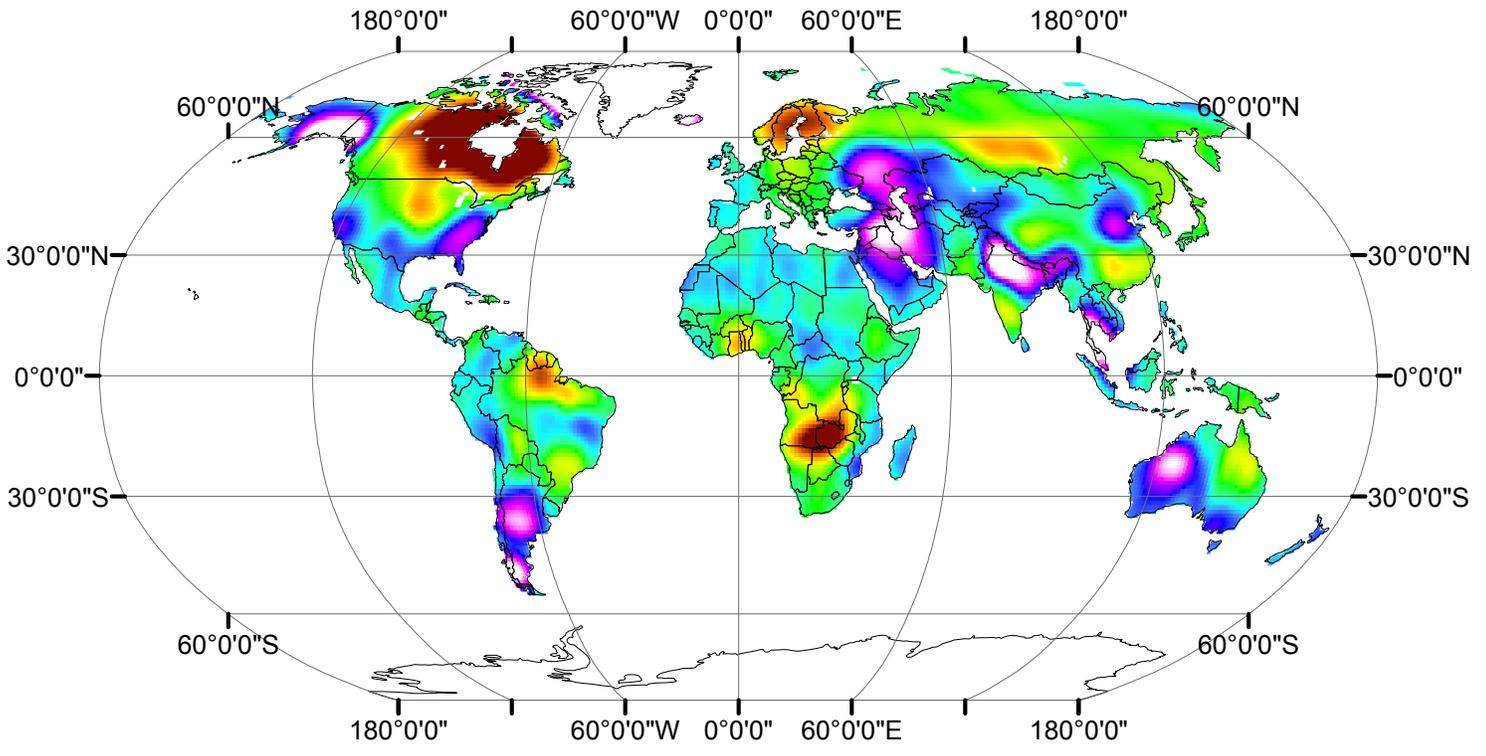


Figure 8: Screen shot of the *Times* tool.

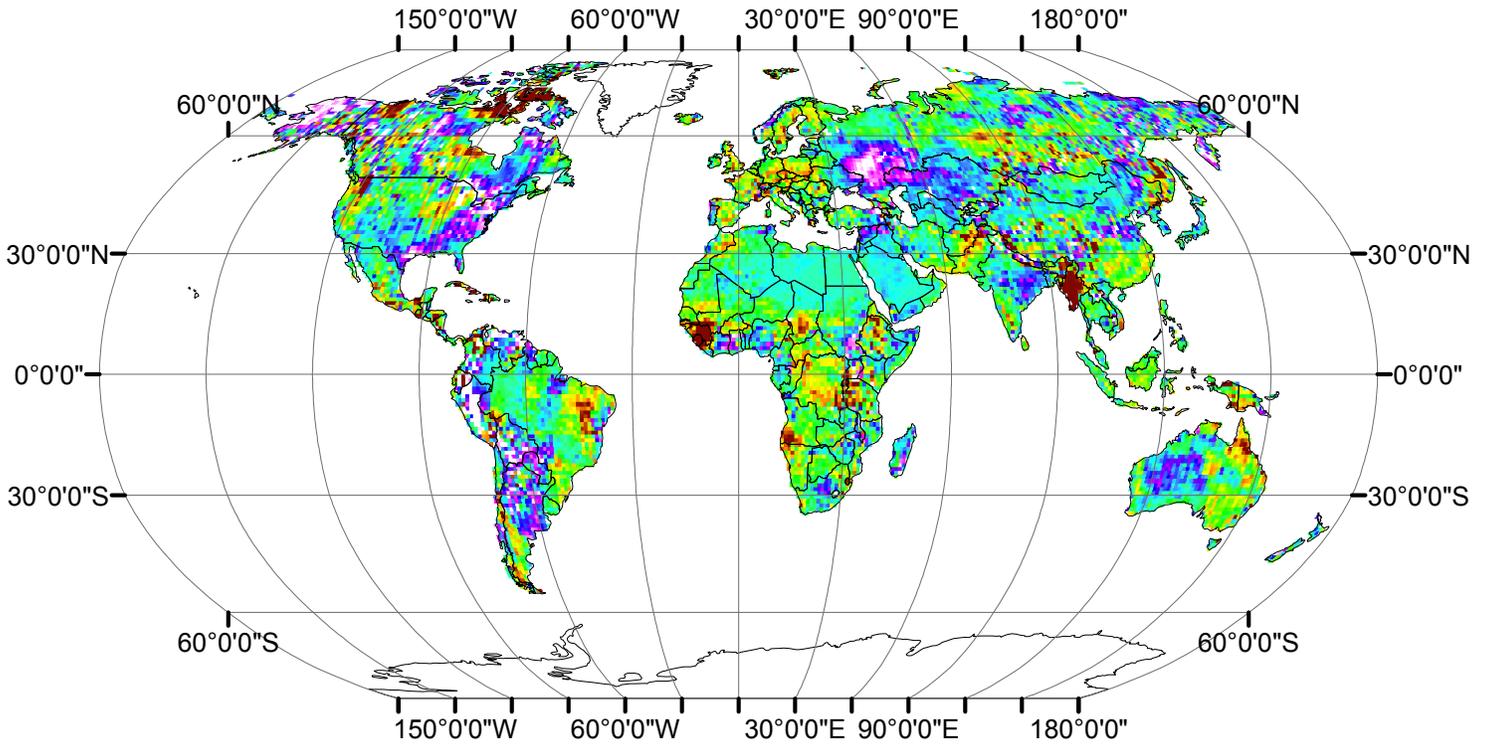
- Next, the *Cell Statistics* tool is used to compute the sum of the rasters created in the previous step. This tool takes as input multiple rasters, and then outputs a single raster containing the predetermined cell by cell statistic, in this case the sum.
- Finally, the *Times* tool is used again to multiply the raster created in previous step by the constant raster created in the first step, yielding an estimate of the slope for each grid point.

A map of these slopes is displayed on the following page. Each data set shows large trends in water storage loss in the southeastern United States, Alaska, southern South America, western Australia, central India, and Russia (near Kazakhstan) through the Middle East. There are also large gains in Northern Canada and in other localized regions.

Global Equivalent Water Height Trend for July 2003 - December 2010 (cm/yr) for CSR (Top) and GLDAS (Bottom)



High : 4.42966 Low : -5.87544



High : 12.579 Low : -11.2284

Robinson Projection
1:200,000,000
WGS 1984

3.4 Pearson Correlation of Data Sets

Another aspect of importance is the correlation between the data sets. As seen before, the trends in each data set have been established, but if they do not agree with one another, intelligent conclusions can not be made. Therefore, the Pearson correlation coefficient was computed between each time series at every grid point. The Pearson correlation between 2 time series is given by equation 4.

$$r = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2}} \quad (4)$$

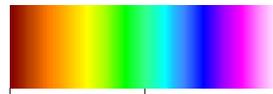
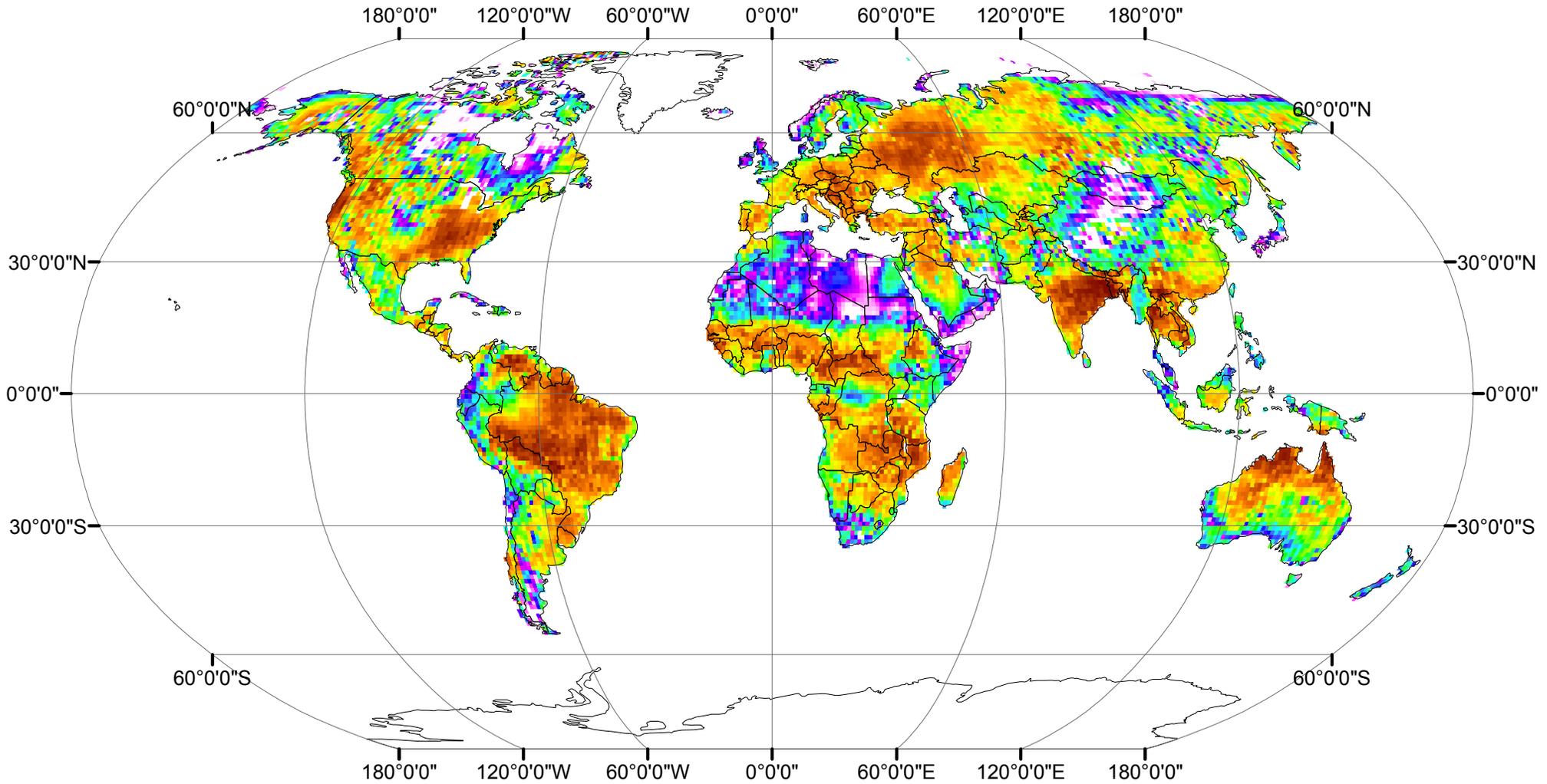
Here, r is the correlation, \bar{X} and \bar{Y} are the mean values of each times series, and X_i and Y_i are the time series values. With this statistic the closer the value is to 1, the the more correlated the series are with each other.

Since ArcGIS does not have an explicit tool to compute the correlation at each grid point, raster math must be used again. The process is very similar to that used to calculate the slopes and is outlined below.

1. First, the 90 band rasters for the GLDAS and CSR data sets can be used with the *Cell Statistics* tool to compute the mean values of each data set at each grid point. This creates a raster containing the values of \bar{X} and \bar{Y} , as referenced to equation 4.
2. Next, the raster math *Minus* tool can be used in batch mode to create rasters that contain the times series values minus the mean (i.e. $X_i - \bar{X}$ and $Y_i - \bar{Y}$). This creates 180 total rasters.
3. Next, the raster math *Times* tool is used in batch mode to multiply the corresponding rasters created in the previous step (i.e. each $X_i - \bar{X}$ by its corresponding $Y_i - \bar{Y}$).
4. Then, the *Cell Statistics* tool is used to compute the sum of the rasters created in the previous step. This yields a raster which contains the numerator in equation 4.
5. Now the denominator of equation 4 must be computed. Since the each $X_i - \bar{X}$ and $Y_i - \bar{Y}$ was already computed in step 2, these rasters can be input into the raster math *Square* tool, in batch mode, to compute $(X_i - \bar{X})^2$ and $(Y_i - \bar{Y})^2$.
6. Next, the *Cell Statistics* tool is used to compute the sum of each of the rasters computed in the previous step. This yields the 2 rasters needed to compute the denominator.
7. Finally, since 3 single band rasters have now been created, the *Rastor Calculator* can be used to compute the Pearson correlation coefficient, as given in equation 4.

Execution of these steps leads to the creation of the map displayed on the following page. The most important observation from this map is the location of areas of high correlation. They correspond directly with areas that showed large negative trends in water storage. These are, most notably: the southeastern United States, Australia, India, and parts of Russia continuing into the Middle East. Therefore, each independent data set is capturing similar time variable water storage information, lending credibility to the earlier findings.

Pearson Correlation of Global Equivalent Water Height from July 2003 - December 2003 for CSR and GLDAS



High : 0.963177 Low : -0.936006

Robinson Projection
1:140,000,000
WGS 1984

3.5 GRACE Trends over the Oceans

Finally, an attempt was made to illuminate water trends over the oceans using GRACE data. However, before this could be done a mask had to be created that removed the land from the global CSR data. The full process is outlined below.

1. Using a polygon feature class, the same one used in Lab 2, which contains polygons outlining the countries of the world, the *Buffer* tool can be used to create a 400 kilometer buffer around the outside of the polygons. This buffer is necessary because of the smoothing operation performed on the CSR data. This smoothing can cause leakage of signal from the land into the oceans. Therefore, it is necessary to mask ocean areas near the land.
2. Next, the polygon feature created in the previous step can be converted to a raster using the *Polygon to Raster* tool. This converts the buffered polygon to a raster file which can be manipulated into a mask.
3. Then, the *IsNull* logical operator is used within the *Raster Calculator* to create a binary raster that differentiates between the land buffer and the oceans.
4. Next, the *SetNull* conditional operator is used within the *Raster Calculator* to set the land buffer areas to null values.
5. After that, the *Extract by Mask* tool is used to mask out the land buffered areas.
6. Finally, the *Band Collection Statistics* tool is used to collect a global average for each raster band in the now masked out global data set. The text file created with this tool is then used to make figure 9.

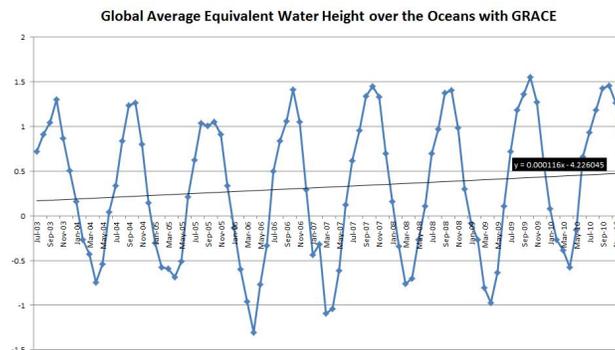


Figure 9: Global average variation of sea level (cm) using GRACE CSR data (trend is approximately 0.5 mm/yr).

This shows a small increase in sea level over the course of the 7.5 years of data; however, it is nowhere near the generally accepted value of 3 mm/yr. This shows that there is something wrong with the smoothing/filtering process used before displaying the data. While the data was smoothed to 400 kilometers, there are still high frequency errors and “stripes” that have

not been removed. This is evident when looking at the global map for one month. This is shown in figure 10.

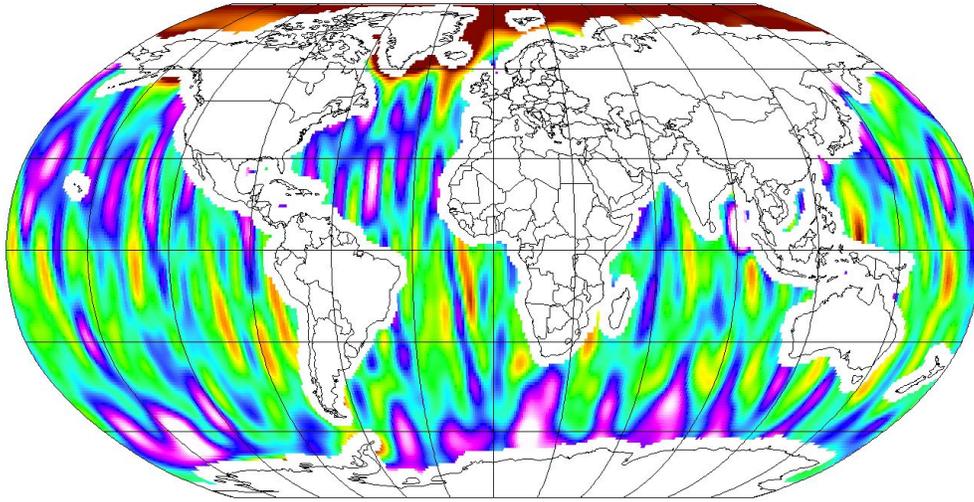


Figure 10: Equivalent water height (cm) over the oceans for July 2003 (warm colors high/cool colors low).

Here, it is apparent that there are still many correlated errors in map. This is evident in the numerous north-south “stripes.” Therefore, in order to glean more information about global sea level rise, more sophisticated filtering techniques need to be used.

4 Conclusions and Possible Future Exploration

The analysis performed showed alarming losses in water storage over specific regional areas. Some of these areas include Alaska, the southeastern United States, western Australia, India, and the Middle East. These trends are validated by the fact that each independent data set, CSR (GRACE) and GLDAS, revealed similar trends in each of these areas. However, attempts to quantify global water storage or global sea level proved inconclusive. This could be due to filtering processes performed, or some modeling deficiencies in the data analysis processes. Further exploration is necessary to draw any conclusions on a global scale. In addition, separate analysis of the ice sheets of Greenland and Antarctica could be done to reveal trends in ice melt around the world.

5 Acknowledgments

The GRACE monthly spherical harmonic coefficients were acquired from <http://www.csr.utexas.edu/grace/RL05.html> and used the GSM fields. The background model, which was

removed, is GIF48.

The GLDAS monthly land water content grids [2] were acquired from <http://grace.jpl.nasa.gov> which used the *Goddard Earth Sciences Data and Information Services Center*.

Data pre-processing was performed using the Texas Advanced Computing Center (TACC) resources located at the University of Texas at Austin.

References

- [1] Tapley B.D., S. Bettadpur, M. Watkins, and C. Reigber (2004), The gravity recovery and climate experiment: Mission overview and early results, *Geophysical Research Letters*, 31, L09607, doi:10.1029/2004GL019920.
- [2] Rodell M, P. R. Houser, U. Jambor, J. Gottschalk, K. Mitchell, C.-J. Meng, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich, J. K. Entin, J. P. Walker, D. Lohmann, and D. Toll (2004) : The Global Land Data Assimilation System. *Bulletin of the American Meteorological Society*, vol 85 (3), pp 381-394.
- [3] Bettadpur, S. UTCSR Level-2 Gravity Field Product User Handbook, May 2012.
- [4] Barthelmes, F. (2009): Definition of functionals of the geopotential and their calculation from spherical harmonic models: theory and formulas used by the calculation service of the International Centre for Global Earth Models (ICGEM), <http://icgem.gfz-potsdam.de>, (Scientific Technical Report STR, 09/02), Potsdam, Deutsches Geo-ForschungsZentrum GFZ, 32 S.: Ill., graph. Darst., 10.2312/GFZ.b103-09026.
- [5] Wahr J. and Molenaar M., Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE, *Journal of Geophysical Research*, 103, (30,205-30,229).