

---

# Correlation between GRACE data and Evapotranspiration data for the High Plains Aquifer

---

Corinne Vassallo

## 1 Introduction and Background

The understanding of Earth's climate over time is essential to society. The Gravity Recovery and Climate Experiment (GRACE) satellites aim to contribute to the study of Earth's climate by modeling Earth's gravity fields, which give us information about global land water storage over time. Scientists use data from GRACE in conjunction with other data sets to understand the water cycle. This particular study aims to see how GRACE data and Evapotranspiration data from the U.S. Geological Survey correlate with one another from January 2005 through December 2009.

### 1.1 Gravity Recovery and Climate Experiment (GRACE)

The GRACE mission was launched in 2002 by NASA and the German Aerospace Center (DLR). It consists of two identical spacecraft co-orbiting in a polar orbit about 500 km above the surface of Earth [1]. As the satellites pass over areas of large mass concentrations, the lead satellite accelerates more quickly than the trailing satellite, and vice versa as they move away from these mass concentrations. GPS and microwave ranging are used to make measurements of the distance between the two satellites, enabling highly accurate global models of Earth's gravitational field.

## 2 Data Collection

### 2.1 GRACE Data

GRACE data is published monthly as spherical harmonic coefficients to degree/order 60 and require significant pre-processing before it is usable in ArcGIS. At the University of Texas' Center for Space Research (CSR), spherical harmonic coefficients are used to calculate the geoid height and then, the time variable water storage. This can be done using the procedure outlined in Wahr [2].

Monthly global total water storage data (in centimeters) from mid-2003 through September 2015 is available from CSR in the form of a netCDF file (multi-dimensional gridded data). Due to a combination of GRACE data outages and the availability of Evapotranspiration data, the chosen timespan is January 2005 through December 2009.

### 2.2 Evapotranspiration Data

Evapotranspiration data was obtained from the U.S. Geological Survey (USGS) [3]. Evapotranspiration is the sum of evaporation and transpiration. More specifically, evapotranspiration is defined as the water lost to the atmosphere from the ground surface, evaporation from the capillary fringe of the groundwater table, and the transpiration of groundwater by plants whose roots tap the capillary fringe of the groundwater table [4]. Data obtained from the USGS provided monthly evapotranspiration data for the high plains aquifer region in millimeters for the chosen timespan. I will use the abbreviation HPA (High Plains Aquifer) to refer to this dataset.

### 3 ArcGIS Processing

In order to see how the two data sets relate, the Pearson correlation coefficient was computed at every cell between the two data sets. The Pearson correlation coefficient,  $r$ , is given by equation 1.  $N$  represents the total number of months (60 for this experiment),  $X$  and  $Y$  each correspond to the GRACE and the HPA datasets respectively with the index  $i$  representing an individual month, and the mean values of the time series given by  $\bar{X}$  and  $\bar{Y}$ . The closer  $r$  is to 1, the more correlated the two datasets are. The procedure for making a raster of correlation coefficients given the two datasets is outlined below:

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

1. First, I created monthly individual GRACE raster layers from the gridded netCDF file using a model built in ArcGIS in batch that made use of the *Make Raster Layer* tool (Fig. 1).

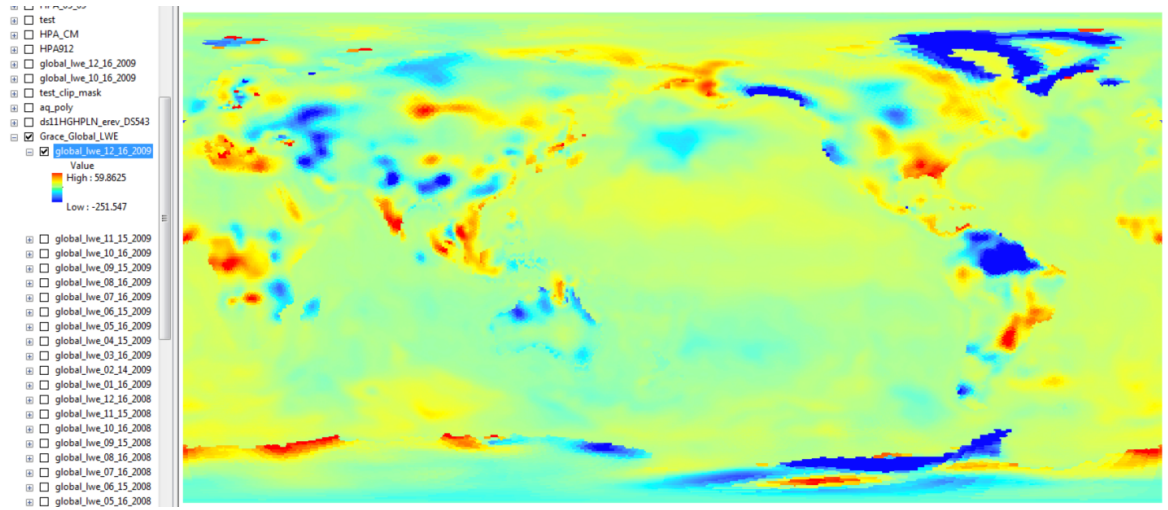


Figure 1: This image shows a sample global raster layer of a single month (December 2009) created from the GRACE data gridded netCDF file. The explanation of the colors is shown to the left of the image. The units are in centimeters. For reference, the United States is on the right side of the image.

2. Then, I made sure both datasets had the same spatial reference. I changed that of the HPA data set to match that of the GRACE data using the *Project* tool in batch.
3. Next, I created a model to more efficiently convert the HPA dataset to centimeters using the raster calculator and running the model in batch (Fig. 2 and 3).
4. Now that both datasets shared the same spatial reference and units, I had to make sure they had the same resolution and covered the same land area for quicker processing and to calculate  $\bar{X}$  and  $\bar{Y}$ .
  - (a) First, I created a multi-band raster out of the 60 monthly HPA rasters using the *Composite Bands* tool. Then, I resampled the HPA data using the *Resample* tool to change the resolution to  $0.5^\circ \times 0.5^\circ$  to match that of the GRACE data (Fig. 4).
  - (b) Next, I created a multi-band raster out of the 60 monthly GRACE rasters also using the *Composite Bands* tool (Fig. 5). I then clipped the global GRACE data to the resampled HPA data using the *Clip* tool.

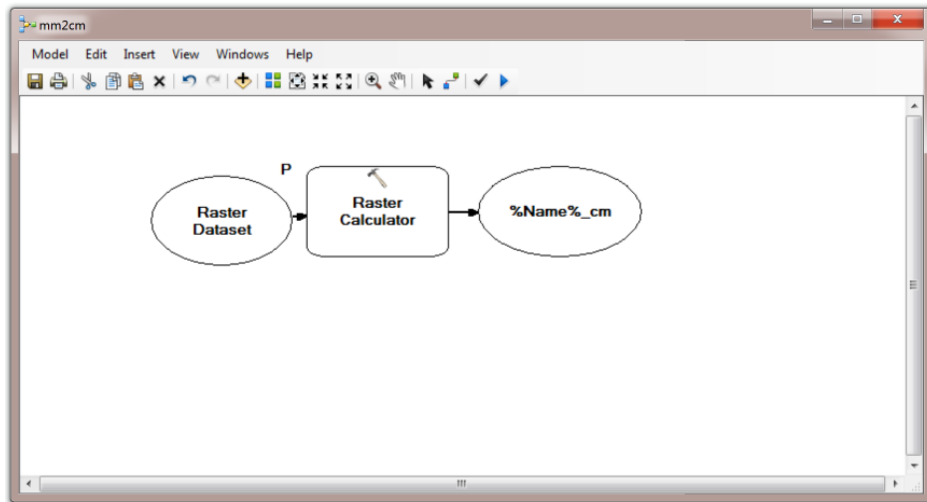


Figure 2: This shows the model built to more efficiently convert the HPA dataset to centimeters. The oval on the left represents the input raster which is then fed to the raster calculator. The rightmost oval represents the output raster. Note, the % signs that surround “Name” in the oval representing the output allow the user to use the input raster name as part of the output raster name, allowing for quicker file naming.

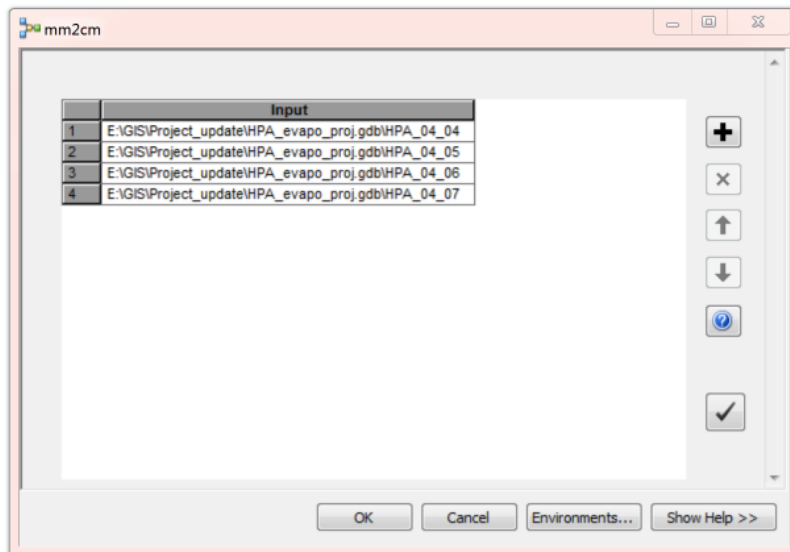


Figure 3: This shows the mm2cm model used to convert the HPA dataset to centimeters being run in batch.

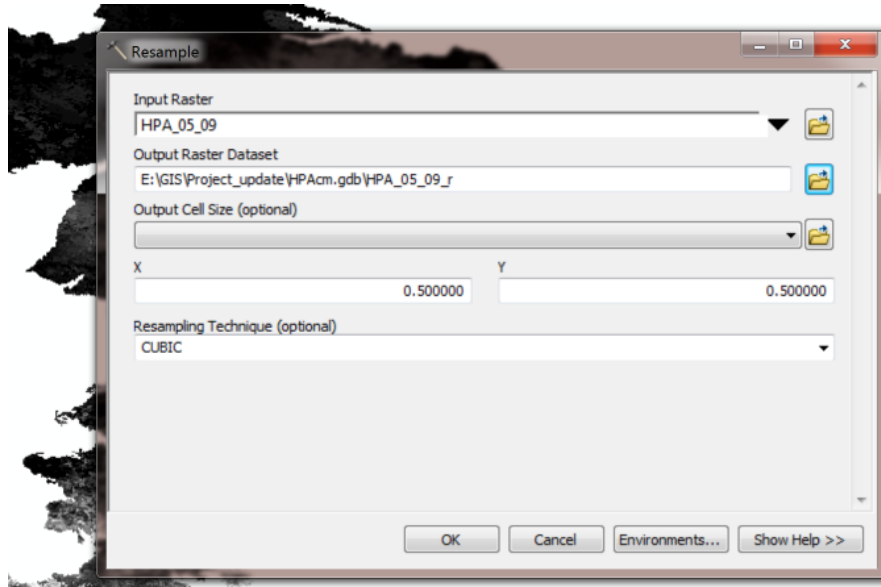


Figure 4: This shows the resampling of the multi-band HPA raster to the the resolution to  $0.5^{\circ} \times 0.5^{\circ}$ .

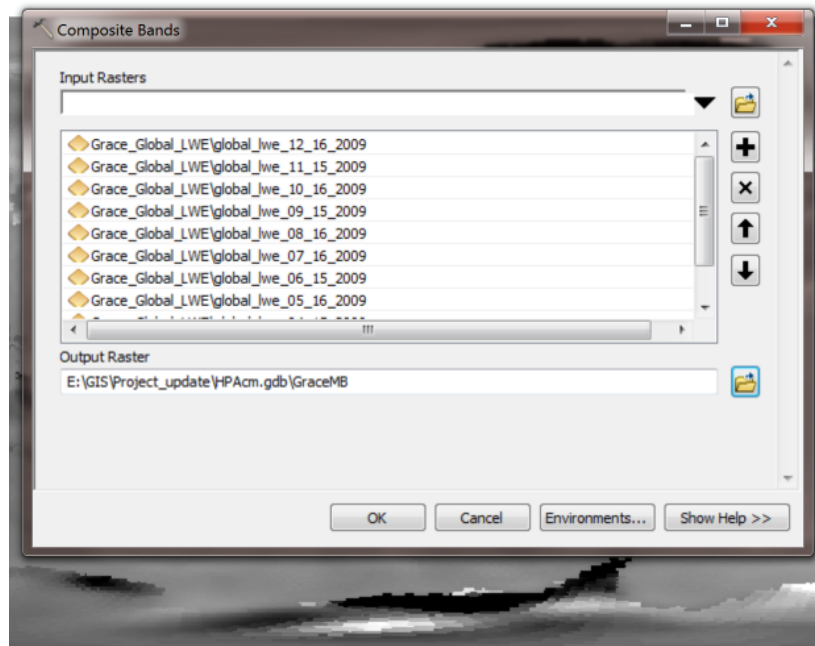


Figure 5: This shows the creation of a multi-band raster out of the 60 monthly GRACE rasters.

- (c) Lastly, I used the *Cell Statistics* tool on the resulting GRACE and HPA rasters that now shared the same resolution and land area to calculate the mean values at each grid point of each dataset, resulting in  $\bar{X}$  and  $\bar{Y}$ .
5. Now that I had calculated  $\bar{X}$  and  $\bar{Y}$ , I needed access to the monthly rasters for each of the two datasets after they were resampled and clipped. I did this by creating a model to extract the bands from the multi-band rasters that were created in steps 4.a and 4.b. This model made use of the *Make Raster Layer* tool in batch. Once complete, there were 120 total rasters - 60 rasters from the monthly GRACE data representing the  $X_i$ 's and 60 rasters from the monthly HPA data representing the  $Y_i$ 's.
6. Then, the numerator was calculated:
- Next, the *Raster Calculator* was used in batch to create the time series values minus the mean ( $X_i - \bar{X}$ ) and ( $Y_i - \bar{Y}$ ). This created 120 new rasters.
  - After each ( $X_i - \bar{X}$ ) and ( $Y_i - \bar{Y}$ ) were created, the *Raster Calculator* was again used within a model and in batch to multiply each ( $X_i - \bar{X}$ ) with its corresponding ( $Y_i - \bar{Y}$ ), creating 60 new rasters (Fig. 6 and Fig. 7).

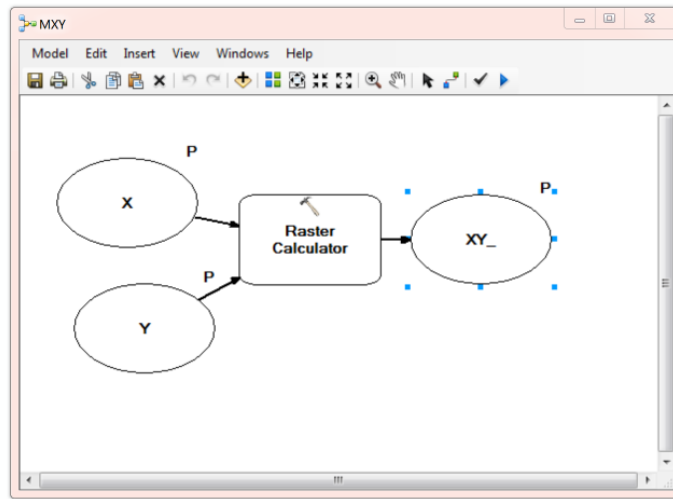


Figure 6: This shows the model built to more efficiently multiply each ( $X_i - \bar{X}$ ) with its corresponding ( $Y_i - \bar{Y}$ ). Note that the X and Y ovals on the left are modeled parameters, meaning that the user selects them. They are then fed to the raster calculator to be multiplied together. The output layer name is also modeled as "XY\_" allowing the user to input a number after the underscore.

- Then, *Cell Statistics* tool was used to compute the sum of the rasters created in the previous step, resulting in the numerator of Eq. 1.
7. After the numerator was calculated, the denominator was calculated.
- Again, the *Raster Calculator* was used in a model in batch mode to square each ( $X_i - \bar{X}$ ) and ( $Y_i - \bar{Y}$ ), creating 120 new rasters from each  $(X_i - \bar{X})^2$  and  $(Y_i - \bar{Y})^2$ .
  - Then, the *Cell Statistics* tool was used twice to compute sums of these rasters. i.e. The two rasters,  $\sum_{i=1}^N (X_i - \bar{X})^2$  and  $\sum_{i=1}^N (Y_i - \bar{Y})^2$ , were calculated.
  - The square root of each of the two rasters created in the previous step were then calculated using the *Raster Calculator* resulting in another two rasters.
  - Lastly, the two rasters created in the previous step were multiplied together using the *Raster Calculator* resulting in the denominator of Eq. 1.
8. Finally, the numerator was divided by the denominator resulting in a single raster representing the correlation coefficient,  $r$ , of the two data sets over time (Eq. 1). This final map can be seen on the last page of this document.

	X	Y	XY
1	xlix_21	yliy_21	E:\GIS\Project_update\Corr\XY\XY_21
2	xlix_22	yliy_22	E:\GIS\Project_update\Corr\XY\XY_22
3	xlix_23	yliy_23	E:\GIS\Project_update\Corr\XY\XY_23
4	xlix_24	yliy_24	E:\GIS\Project_update\Corr\XY\XY_24
5	xlix_25	yliy_25	E:\GIS\Project_update\Corr\XY\XY_25
6	xlix_26	yliy_26	E:\GIS\Project_update\Corr\XY\XY_26
7	xlix_27	yliy_27	E:\GIS\Project_update\Corr\XY\XY_27
8	xlix_28	yliy_28	E:\GIS\Project_update\Corr\XY\XY_28
9	xlix_29	yliy_29	E:\GIS\Project_update\Corr\XY\XY_29
10	xlix_30	yliy_30	E:\GIS\Project_update\Corr\XY\XY_30
11	xlix_31	yliy_31	E:\GIS\Project_update\Corr\XY\XY_31
12	xlix_32	yliy_32	E:\GIS\Project_update\Corr\XY\XY_32
13	xlix_33	yliy_33	E:\GIS\Project_update\Corr\XY\XY_33
14	xlix_34	yliy_34	E:\GIS\Project_update\Corr\XY\XY_34
15	xlix_35	yliy_35	E:\GIS\Project_update\Corr\XY\XY_35
16	xlix_36	yliy_36	E:\GIS\Project_update\Corr\XY\XY_36

Figure 7: This shows the model built to more efficiently multiply each  $(X_i - \bar{X})$  with its corresponding  $(Y_i - \bar{Y})$  being run in batch. Note how each column corresponds to each of the modeled parameters in Fig. 6

#### 4 Conclusions

From the final map shown on the last page of this document, one can see that the two data sets are most correlated in the northwest region of the High Plains Aquifer. However, the maximum value for the correlation coefficient raster is only 0.53, indicating weak correlation between the two data sets. To further understand the meaning behind this result, future work would include analyzing a larger time span of data, as well as including precipitation data in the analysis. For future time series analysis similar to that completed here, I would also recommend using a combination of Python (or other programming languages) and ArcGIS. ArcGIS is not optimized for large time series analysis at its current version and basic raster calculations take significant time when run in batch. However, once all data is processed, ArcGIS is a wonderful tool to visualize resulting data.

#### 5 References

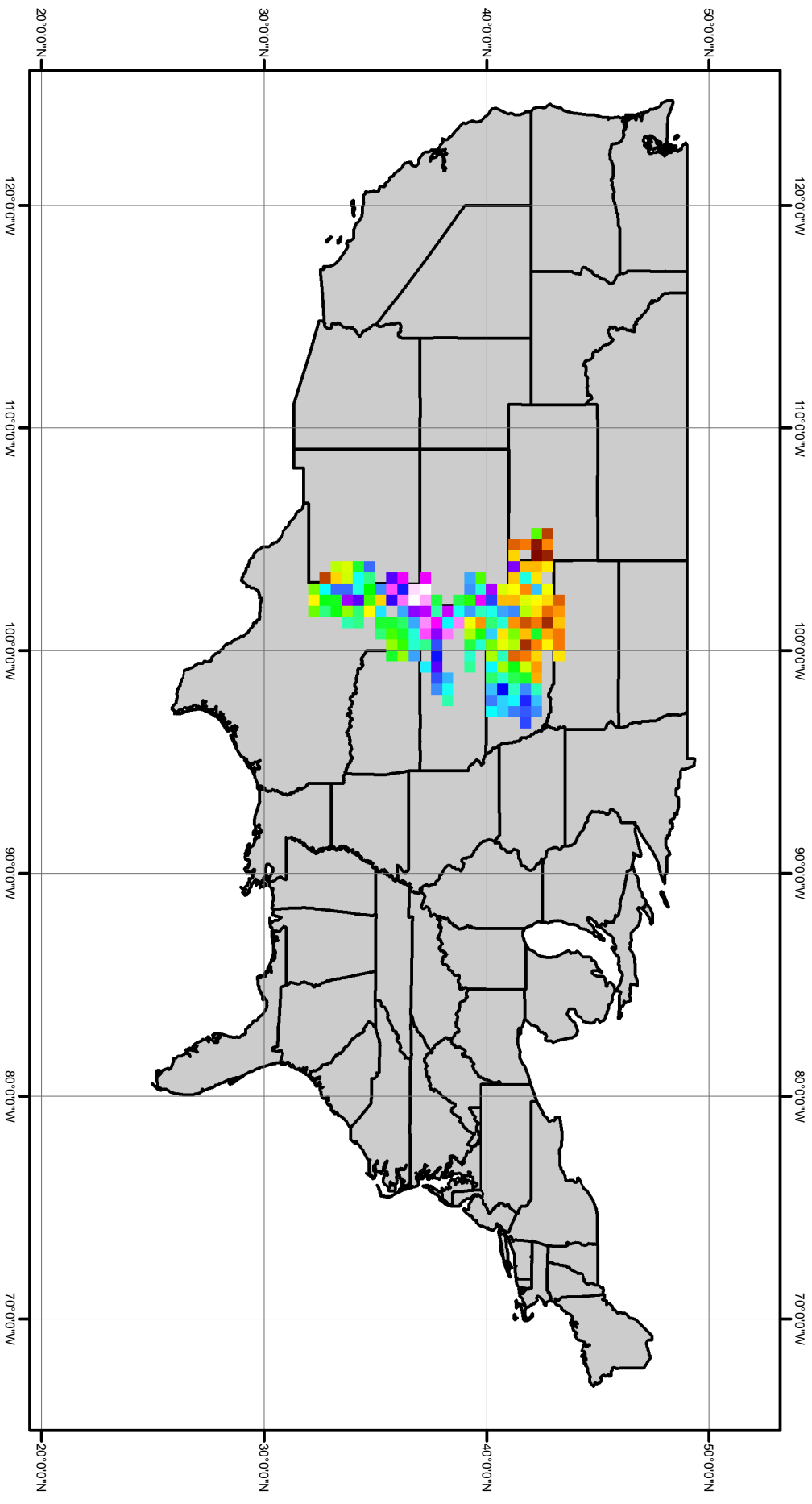
[1] Tapley B.D., S. Bettadpur, M. Watkins, and C. Reigber (2004), The gravity recovery and climate experiment: Mission overview and early results, *Geophysical Research Letters*, 31, L09607, doi:10.1029/2004GL019920.

[2] 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).

[3] U.S. Geological Survey. DS-777 Monthly Actual Evapotranspiration Rasters calculated using the Simplified-Surface-Energy-Balance (SSEB) Model from April 2000 to December 2009 for the High Plains Aquifer in Parts of Colorado, Kansas, Nebraska, New Mexico, Oklahoma, South Dakota, Texas, and Wyoming. <http://water.usgs.gov/lookup/getgislislist>

[4] U.S. Geological Survey. "Evapotranspiration- the Water Cycle." <http://water.usgs.gov/edu/watercycleevapotranspiration.html>

# Pearson Correlation of Global Total Water Storage from 2005 through 2009 for GRACE and HPA Evapotranspiration Data



**Pearson Correlation Coefficient**  
High : 0.530615  
Low : -0.216978

United States

Coordinate System: GCS WGS 1984  
Datum: WGS 1984  
Units: Degree

0 250 500 1,000 1,500 2,000 Kilometers