December 5th | 2014

GLDAS and EartH2Observe Compared

Written by: Frank Schalla

CE 394K Term Project Final Report

Table of Contents

Introduction

Global Land Data Assimilation System (GLDAS) processes satellite and ground-based observational data products, using advanced land surface modeling and data assimilation techniques, in order to generate fields of land surface states and fluxes (Rodell et al., 2004). GLDAS produces high-quality raster surfaces to support weather and climate predictions, water resources applications, and water cycle investigations.

EartH2Observe (E2O) is a similar European project funded by the European Centre for Medium-Range Weather Forecasting (ECMWF). E2O has similar objectives, "to contribute to the assessment of global water resources through the use of new Earth Observation datasets and techniques (earth2observe.eu)." The system also integrates observation-based data and models to construct a large temporal dataset of at least 30 years.

These two systems have global application abilities. Similar systems such as North American Land Data Assimilation System (NLDAS) and European Centre for Medium-Range Weather Forecasts (ECMWF) have already been successfully implemented in the USA and European Member States and Co-operating States, respectively.

One motivating factor and potential application of satellite systems is in rainwater collection system design. The design of a rainwater collection system requires at least 10 years of continuous monthly or daily (if available) rainfall data near the system's location. Often in sparsely populated rural regions in developing countries access to long-term rain-gauge data is not available or financially feasible. The application of GLDAS or E2O data may be justified as a design tool if the data is similar to rain-gauge data.

This report will compare the accuracies of both satellite systems over the countries of Guatemala (Central America) and Ghana (West Africa). The author has worked and lived in Guatemala and Ghana, has knowledge of the countries and rain-gauge data sources and has design experience with rainwater catchment systems.

Guatemala Compared

Guatemala is a geographically diverse country in the north of Central America. The coastal pacific region is located in the south of the country. Within around 40 miles the coastal lowlands transition to mountainous highlands. These mountainous highlands have elevations around 4,000 ft and the highest peak in Central America, at nearly 14,000ft. Moving north past the highland mountains are the northern lowlands.

There are two seasons in Guatemala, the wet and dry season. The wet season ranges from April to November, while the dry season is from December to April. In addition to being geographically diverse Guatemala has spatially diverse rainfall as well.

Monthly rainfall data was collected through a Guatemala governmental agency (Instituto Nacional de Sismología, Vulcanología, Meteorología y Hidrología, http://www.insivumeh.gob.gt/) for 44 meteorological sites across the country. The year of 2009 was used to compare with monthly satellite data since it contain the most complete monthly rainfall data. The data collection process proved tedious because the web site (which was since revised) had no easy downloading abilities. Rain-gauge locations were only provided on a large country map with an associated city and did not contain latitude or longitude coordinates. The rain gauge sites required an estimation of the latitude and longitude based on judgment. Even with the estimation of rain-gauge location, the data was compared with ½° to ¼° satellite grids, hence accuracy was not decreased.

Below is a map showing the location of the meteorological sites and the surrounding Guatemalan terrain.

Figure 2 - Guatemala and surrounding countries

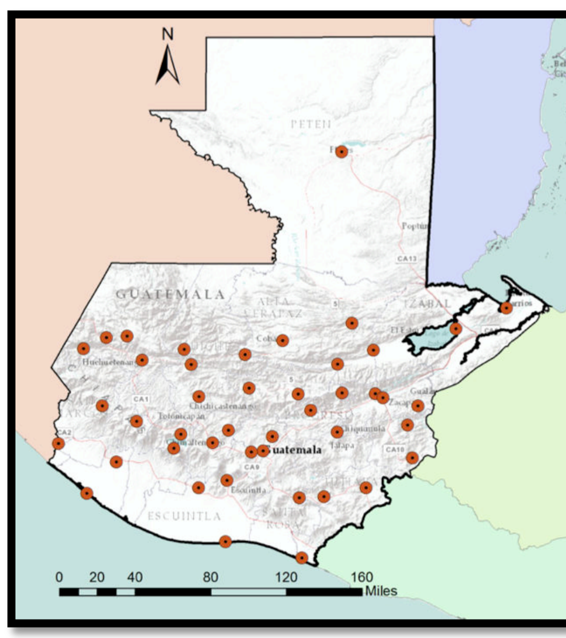

Figure 2 - Guatemala terrain and rain-gauge locations

The vast majority of rain-gauge locations are in the central highlands and southern coastal region of the country. Due to the lack of rain-gauge locations in the far north of the country the interpolated rasters for that region are not assumed to be as reliable as the central and southern portions.

A variety of interpolation tools were tested to accurately represent the rainfall data spatially. Spline with Barriers proved to provide the most visually appealing and realistic representation of the spatial variation of rainfall across Guatemala.

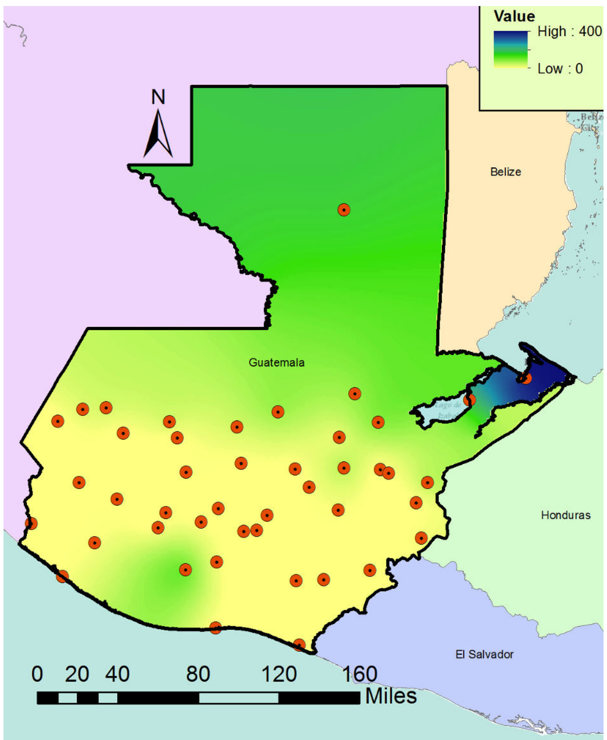

Figure 4 - January 2009 Rainfall (mm)

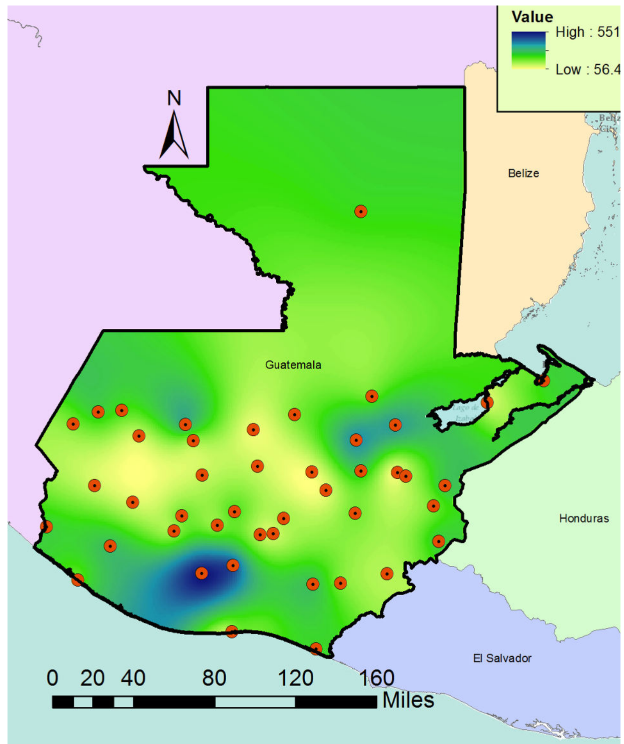

Figure 4 - September 2009 Rainfall (mm)

The month of January is in the middle of the dry season. The southern coastal lowlands and central highlands have fairly spatially consistent rainfall, while large rainfall (400mm) was recorded for the small coastal region in the east of the country.

Rainfall varies more for the month of September, ranging from 56 mm in the mountainous highlands to 560 mm along the southern coastal lowlands. Rainfall appears to vary spatially according to the geographical terrain of the country. Low rainfall was reported in the mountainous highlands and higher rainfall in the northern and southern lowlands.

Satellite Data

GLDAS raster data was downloaded with a tool created by Gonzalo Espinoza. The tool allowed for the easy download of a specific data parameter, a range of dates and provided simple integration into an ArcGIS base map. E2O raster data was downloaded in NetCDF format from an E2O database (https://wci.earth2observe.eu/thredds/catalog.html). An ArcGIS tool was used to import the raster data into ArcGIS. The data for both systems were downloaded as monthly rainfall rate (units of kg / m^2 / s). The monthly rainfall rates were converted to monthly rainfall (mm) by using a raster calculator. This was done because the provided rainfall rates are normally very low (<0.001) and were difficult to conceptualize and compare numerically.

Below are GLDAS and E2O raster maps showing monthly rainfall for September 2009. GLDAS has a grid size of ¼° and E2O has a grid size of ½°.

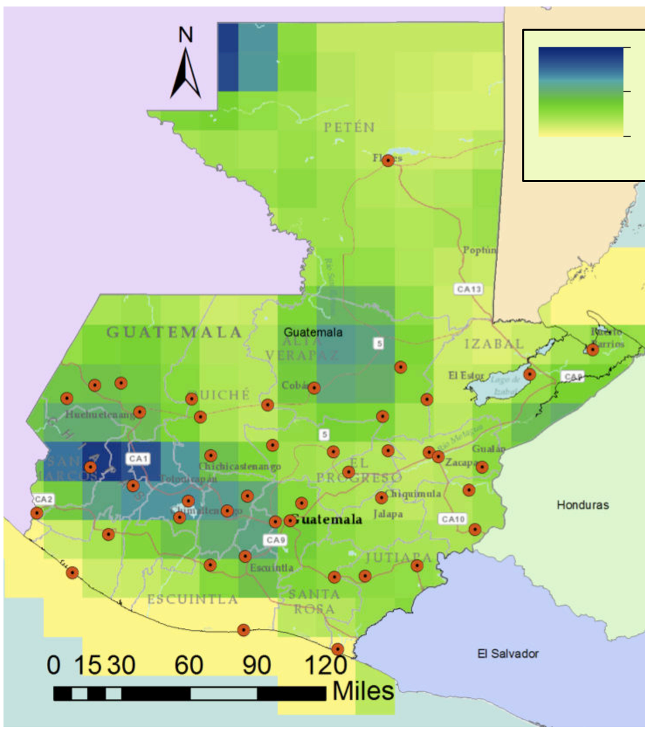

Figure 6 - GLDAS rainfall (mm) for September 2009

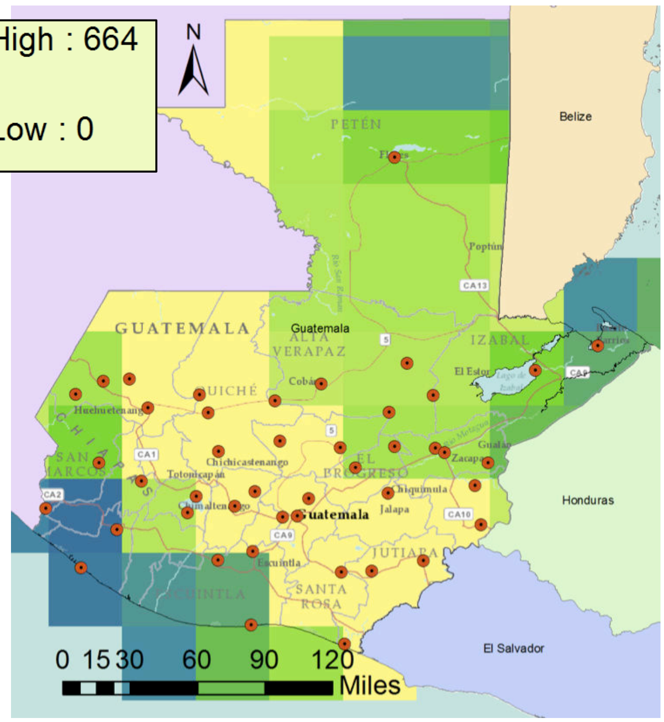

Figure 6b - E2O rainfall (mm) for September 2009

There are general similarities that can be noted qualitatively between the two satellite systems. There are higher rainfall amounts in the south and north of the country and lower rainfall in the central mountainous highlands. For the GLDAS raster there appears to be low rainfall in the southern coastal region of the country. Those are actually no data grids and were unfortunately colored the same as no rainfall.

Ghana Compared

Ghana is a country located in the middle of West Africa. The south of the country is located on the Atlantic Ocean, more specifically the Gulf of Guinea. The country's elevation is generally consistent, at around 500 - 1,000 ft (not considering the coastal region). The highest peak in Ghana is located in the east, and has an elevation of around 3,000 ft. There are two seasons, the wet and dry season. The wet season ranges from April to November, while the dry season ranges from December to April. The dry season aligns with Harmattan, a dry desert wind that blows south from the Sahara Desert.

Monthly rainfall data was accessed through the Global Precipitation Climatology Center (GPCC, gpcc.dwd.de) a German Federal Ministry that provides raster data compiled from local rain-gauge stations worldwide. The GPCC raster grid size was ¼°. Satellite data was once again accessed with GLDAS and E2O. Monthly data was also compared for all months of 2009. Rainfall sample locations were created at the center of each GPCC raster. Raster values were extracted at those points for GPCC, GLDAS and E2O data.

Below are maps showing Ghana with rainfall sampling locations and the GPCC rainfall raster for January 2009.

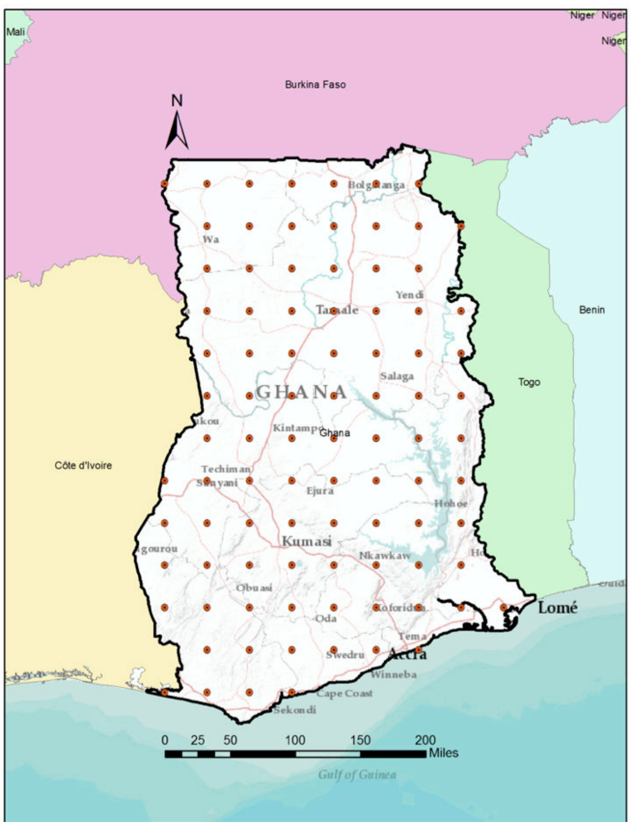

Figure 8 - Rainfall sampling locations for Ghana

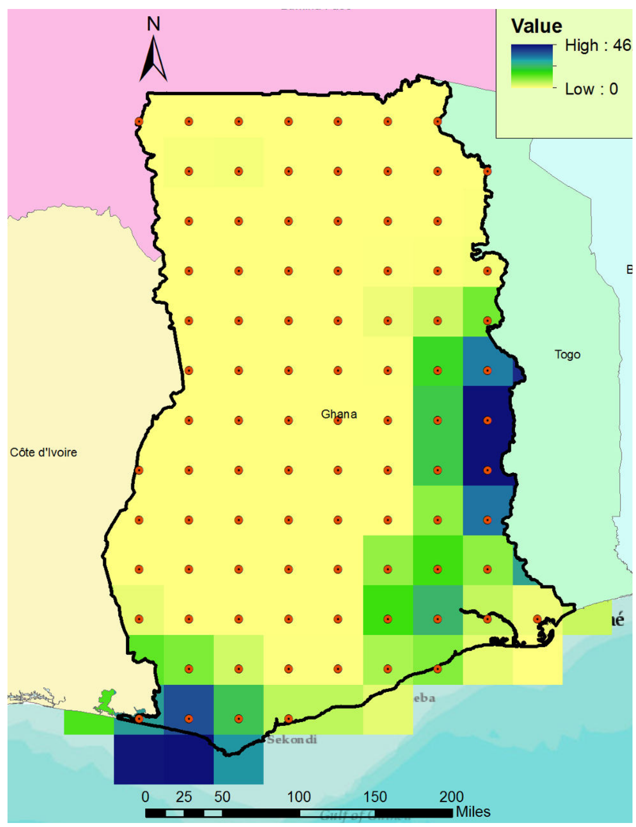

Figure 8 - GPCC Rainfall (mm) January 2009

The small January rainfall amount from GPCC was expected for the middle of the dry season. The high values shown in a portion of the eastern part of the country may be attributed to where the highest peak in Ghana is located.

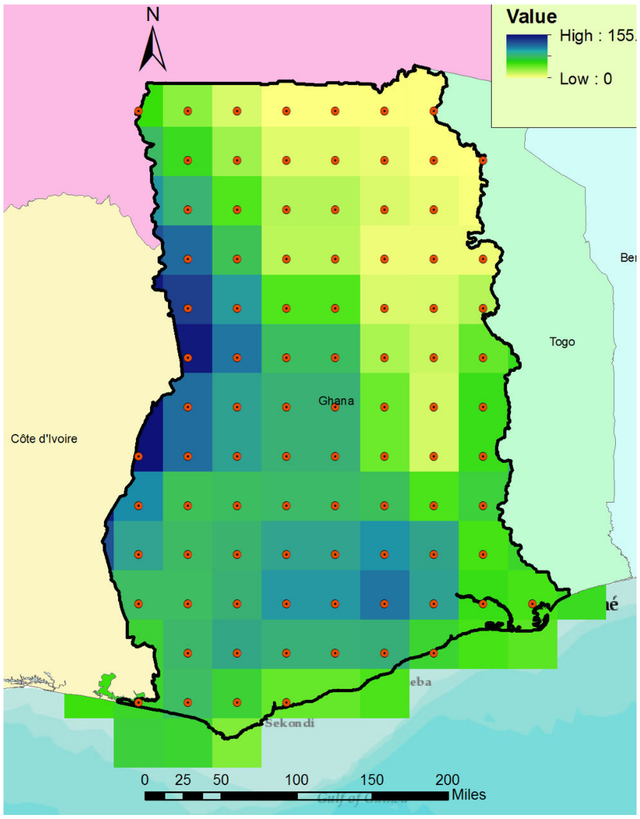

Figure 10 - GPCC Rainfall (mm) March 2009

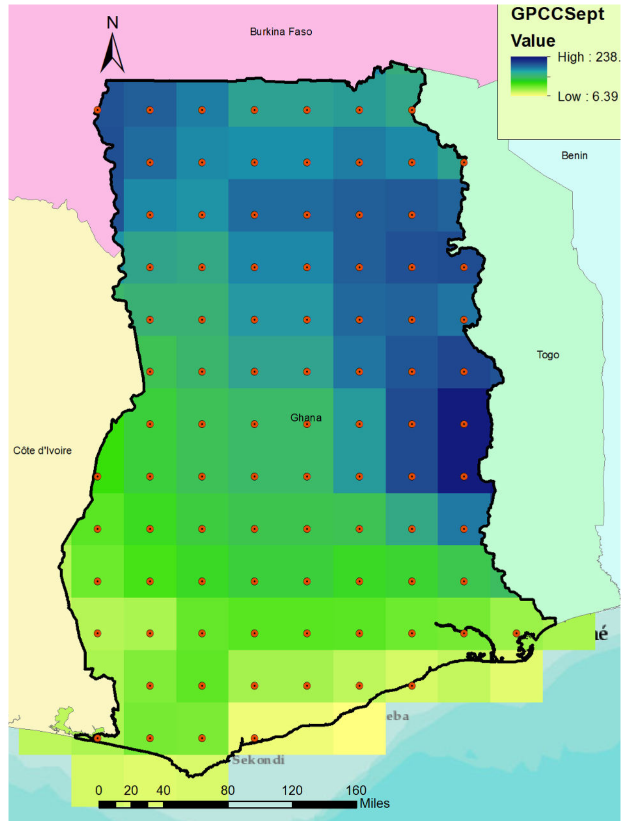

Figure 10b – GPCC Rainfall (mm) September 2009

GPCC rainfall is higher for the month of September, which is one of the highest rainfall months in the country. More rainfall occurred in the north of the country during the month of September.

Satellite Data

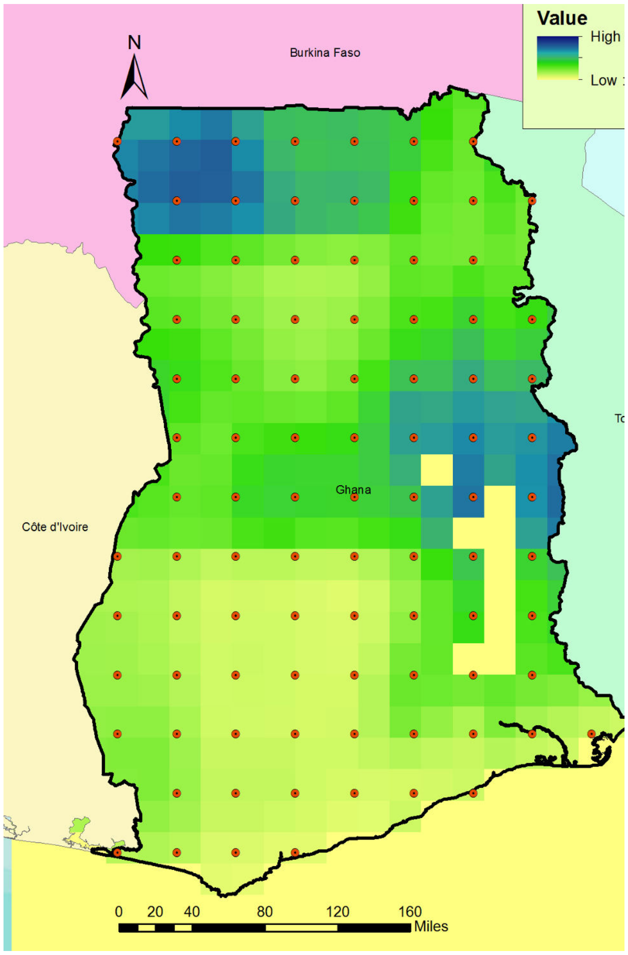

Figure 12 - GLDAS Septemeber 2009

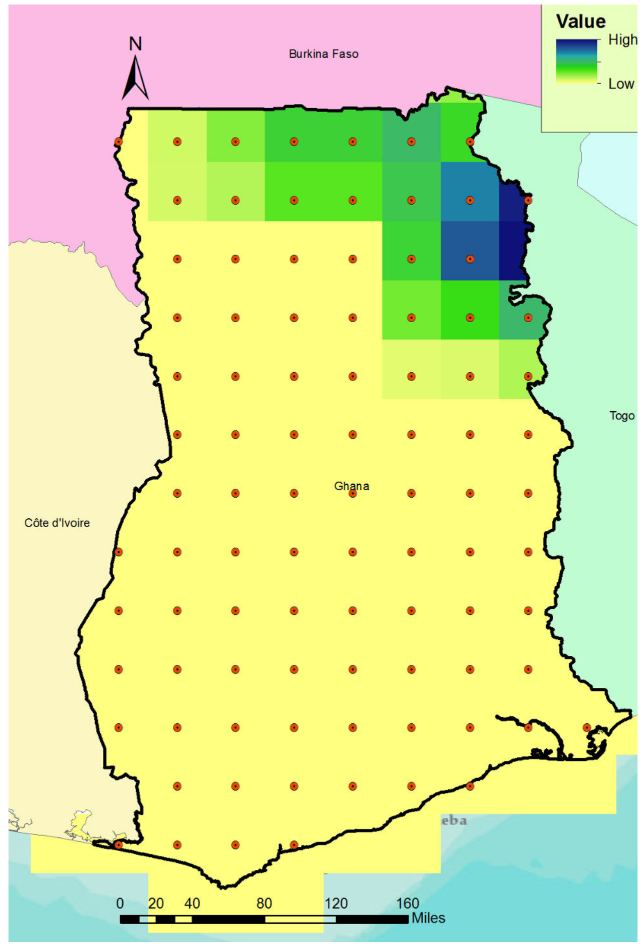

Figure 12b - E2O September 2009

The GLDAS grid size is ¼° while the E2O grid size is ½°, which is consistent with the Guatemala comparison. For GLDAS there are no data for 9 grids in the middle-right of the country. This corresponds to where Lake Volta is located. In terms of surface area, Lake Volta is the largest man-made reservoir in the world. E2O has complete coverage over the portions of Lake Volta, perhaps due to the larger grid sizes.

Both GLDAS and E2O rasters above have the similar color scale, with larger rainfall values in dark blue and smaller rainfall vales in light yellow. GLDAS initially appears more similar to GPCC than E2O for the month of September. Clear differences are still noticeable between GLDAS and GPCC, with lower rainfall values throughout the middle and upper portion of the country.

Statistical Analysis

A variety of basic statistical analysis was performed to analyze which satellite system was most accurate with respects to the land rain-gauge stations. Comparing the statistical accuracy of satellite systems has been previously researched (Bell and Kundu, 2002, Dessu and Melesse, 2013) and is often a complicated process. To focus this report, only basic statistical analysis was performed.

The first analysis used was to graph the distribution between the accurate measurements (rain gauge stations) and the estimated (satellite system) measurements. The 45° line on each graph represents no error in the satellite data. Any points below the line are an underestimate and anything above the line is an overestimate of the true value with respects to the satellite data.

General statistics were performed for each month of 2009. These statistics were based on the data *differences* between the rain gauge data and the satellite system data. The following statistics were performed: absolute minimum, absolute maximum, standard deviation (SD), mean absolute error (MAE), mean signed error (MSE), and a paired t-test. (Anderson, et al.)

Mean Absolute Error is a statistical method defined as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |f_i - y_i|$$

Where f_i is the predicted value (satellite value) and y_i is the actual value (land based station). This is a measure of error that considers negative and positive differences equally.

Another statistical method used was Mean Signed Error (MSE), defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n} f_i - y_i$$

MSE is similar to MAE except for the lack of absolute values. This statistical method was used to potentially indicate a biased towards a negative or positive error.

The last statistical test was the paired t-test. This test is often used to test the hypothesis that the data groups are correlated. An alpha of 0.05 was used and the reported table values are p-values.

Statistics Tables

Guatemala

Statistic	Global Land Data Assimilation System (GLDAS)												
	January	February	March	April	May	June	July	August	September	October	November	December	Overall
Abs. Minimum Difference	1.6	1.3	0.1	0.2	0.5	3.6	1.0	0.6	0.1	2.9	1.1	0.2	0.1
Abs. Maximum Difference	167.3	58.4	69.2	72.4	438.4	655.3	435.8	549.3	471.4	388.3	187.7	74.5	655.3
Standard Deviation	34.0	13.6	15.9	18.5	122.0	142.0	100.8	119.4	103.2	104.4	42.8	20.2	99.7
Mean Absolute Error	22.6	13.0	10.8	21.0	131.9	147.4	128.4	146.2	121.3	101.4	47.2	25.8	76.4
Mean Signed Error	-6.2	-0.1	-5.6	0.9	39.8	69.9	52.8	2.8	42.5	29.0	-4.4	-23.2	16.5
Paired T-Test	0.3416	0.9617	0.0644	0.8436	0.1722	0.0322	0.0406	0.9277	0.0939	0.2203	0.6760	0.0000	-
Statistic	Earth-2-Observe (E2O)												
	January	February	March	April	May	June	July	August	September	October	November	December	Overall
Abs. Minimum Difference	0.0	0.0	1.0	0.0	0.2	51.9	2.3	89.3	0.1	70.2	70.8	0.0	0.0
Abs. Maximum Difference	180.6	519.5	71.6	173.3	632.3	617.8	1369.6	2446.4	349.3	1750.5	4324.3	413.2	4324.3
Standard Deviation	47.1	161.9	16.0	34.6	130.9	128.6	435.1	541.3	66.8	389.9	1166.6	93.4	642.6
Mean Absolute Error	29.7	162.1	24.9	26.1	159.9	255.0	475.2	1203.7	86.4	645.1	1612.2	71.2	395.0
Mean Signed Error	-3.7	160.3	10.7	-26.1	-122.0	-163.7	379.0	1203.7	6.3	645.1	1612.2	4.3	308.0
Paired T-Test	0.6644	0.0000	0.0142	0.0000	0.0000	0.0000	0.0000	0.0000	0.7045	0.0000	0.0000	0.8120	-

Table 1 - Guatemala Statistics

Ghana

Statistic	Global Land Data Assimilation System (GLDAS)												
	January	February	March	April	May	June	July	August	September	October	November	December	Overall
Abs. Minimum Difference	0.0	0.3	1.1	0.1	0.7	1.8	1.9	0.1	0.4	0.2	0.2	0.0	0.0
Abs. Maximum Difference	150.3	126.5	232.5	429.8	259.7	513.8	212.0	242.8	142.8	136.1	124.7	52.7	513.8
Standard Deviation	20.3	37.8	49.2	98.5	70.1	88.7	82.2	51.2	53.8	56.6	37.0	16.8	68.0
Mean Absolute Error	8.5	28.2	43.9	67.1	55.6	73.3	96.3	46.5	42.6	44.2	30.2	11.4	46.0
Mean Signed Error	7.2	19.8	31.6	-2.6	-14.5	64.2	-69.7	38.7	-8.4	3.2	20.7	-5.9	7.0
Paired T-Test	0.0010	0.0000	0.0000	0.8051	0.0511	0.0000	0.0000	0.0000	0.1404	0.5952	0.0000	0.0012	-
Statistic	Earth-2-Observe (E2O)												
	January	February	March	April	May	June	July	August	September	October	November	December	Overall
Abs. Minimum Difference	0.0	1.9	0.8	5.8	1.1	42.8	93.9	0.2	6.6	11.0	0.2	0.0	0.0
Abs. Maximum Difference	33.6	96.5	145.8	700.4	847.8	299.4	287.8	272.1	618.6	408.2	405.8	65.4	847.8
Standard Deviation	7.0	22.2	37.3	191.1	184.4	42.0	41.8	76.8	88.7	90.8	78.1	18.8	114.6
Mean Absolute Error	3.4	28.8	58.7	187.3	182.1	129.1	209.0	94.3	121.9	124.0	58.3	17.2	101.0
Mean Signed Error	-3.2	-28.8	-58.7	71.6	5.3	-129.1	-209.0	-72.7	-54.5	16.2	31.2	-17.2	-37.4
Paired T-Test	0.0000	0.0000	0.0000	0.0096	0.8458	0.0000	0.0000	0.0000	0.0004	0.3172	0.0018	0.0000	-

Table 2 - Ghana Statistics

Guatemala Statistics

Once again for clarity, the tables above show statistics for the *differences* between the rain-gauge data and satellite data. The tables do not represent actual rainfall amounts for either GLDAS or E2O.

Initial observations of the E2O data show extreme data differences. In the month of November the rainfall differences between the rain gauges and E2O show a value of over 4,000 mm. This is highly unlikely, indicating a large error in satellite measurements. Overall SD for E2O was 640 mm, compared with 100 mm for GLDAS.

Overall for 2009, comparisons between GLDAS and E2O show lower MAE for GLDAS (76 mm) than E2O (395 mm). MSE was also lower overall for GLDAS (16.5 mm) than E2O (308 mm).

The paired t-test was not found to be useful in this particular statistical analysis. It is often in direct opposition to other statistics reported. This may be contributed to the limitations of requiring the data to be a normal distribution. Looking at the month of November, E2O has an absolute maximum difference of over 4,000 mm of rainfall, yet the t-test reported a $p<0.0001$.

Below are graphs showing the distribution between rain-gauge data and satellite data for three months.

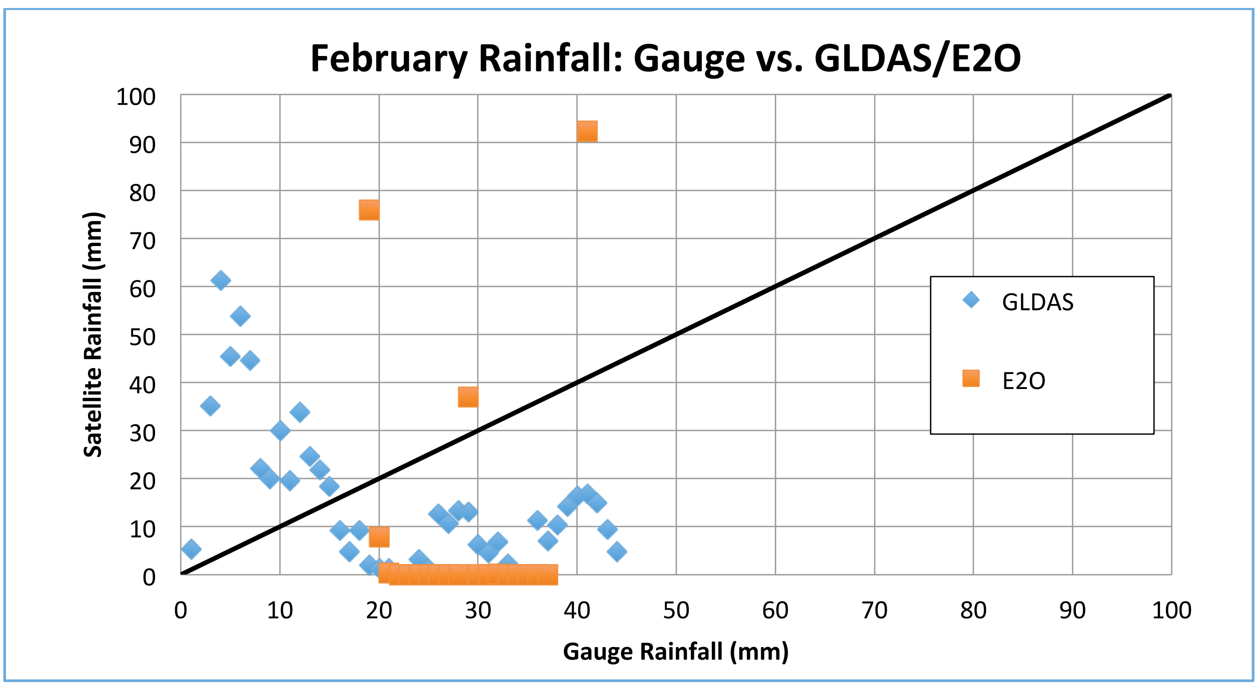

Figure 13 - February Comparison for Guatemala

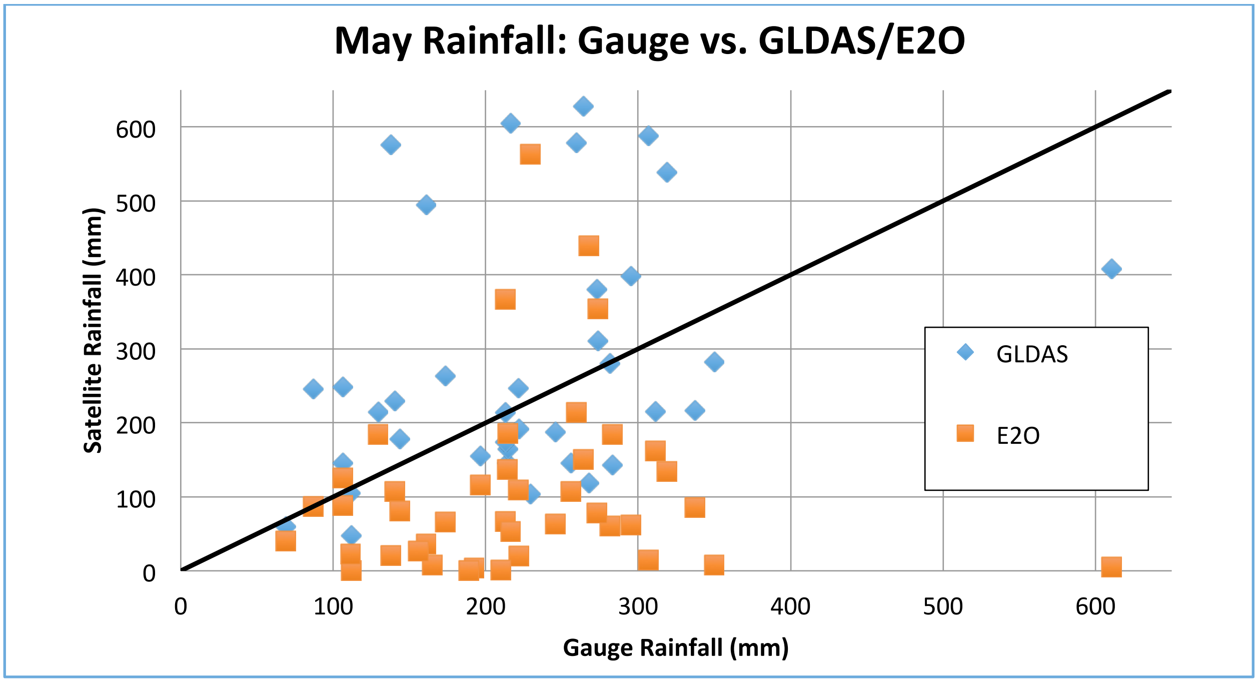

Figure 14 - May Comparison for Guatemala

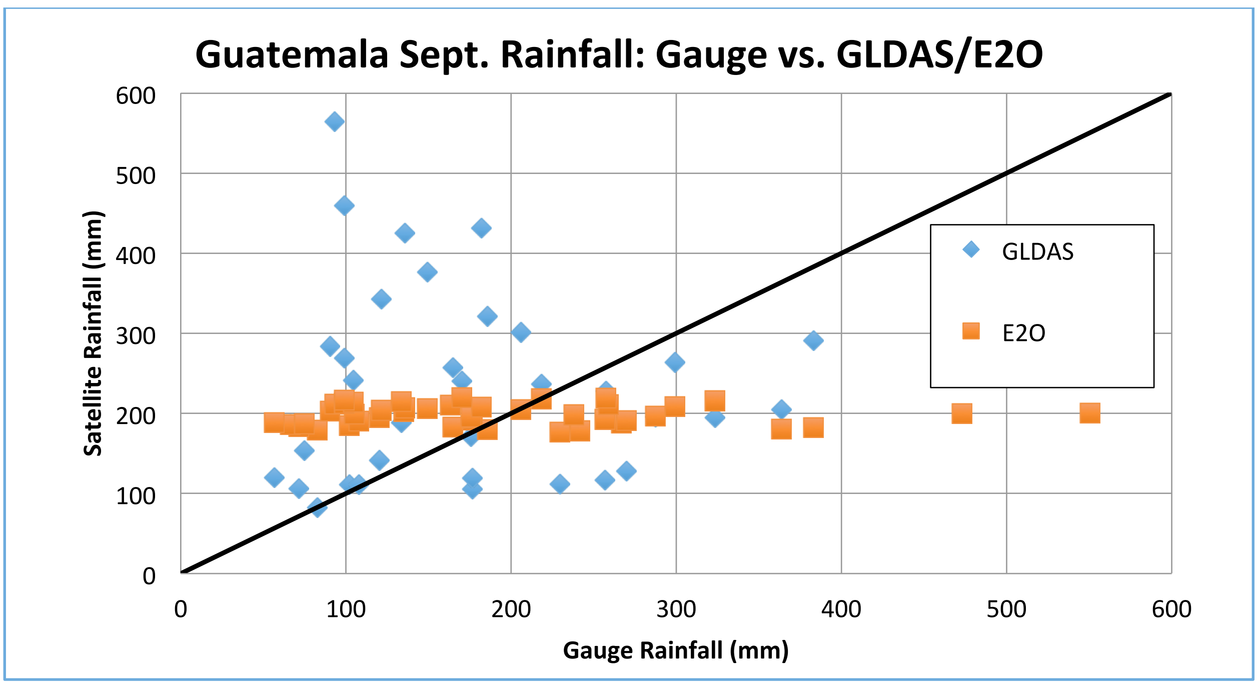

Figure 15 - September Comparison for Guatemala

The month of February has a wide variation of over and underestimation for GLDAS, yet the MSE was -0.1 mm. For rainfall under 15mm GLDAS overestimated, while for rainfall over 15mm GLDAS underestimated.

E2O has a much larger variation in data for February, with a MSE of 160 mm. It should be noted that many of the large positive points were not included in the graph to be able to view the GLDAS data better. Overall GLDAS's maximum difference for February was the smallest out of all months, at 58 mm.

The month of May, compared to February, shows more variation for GLDAS, with a MSE of 40 mm. For E2O there was less variation in May, compared to February, with a MSE of -122 mm. E2O is consistently underestimated while GLDAS is overestimated.

September showed a MSE of 42 mm for GLDAS and a MSE of 6 mm for E2O. The E2O data is strangely consistent around 200mm of rainfall, while the actual gauge rainfall varies from 65 – 550 mm.

Ghana Statistics

When comparing the overall MAE and MSE, GLDAS has significantly less error than E2O. The error is still significant though, as the overall maximum monthly differences for both GLDAS and E2O are above 500 mm.

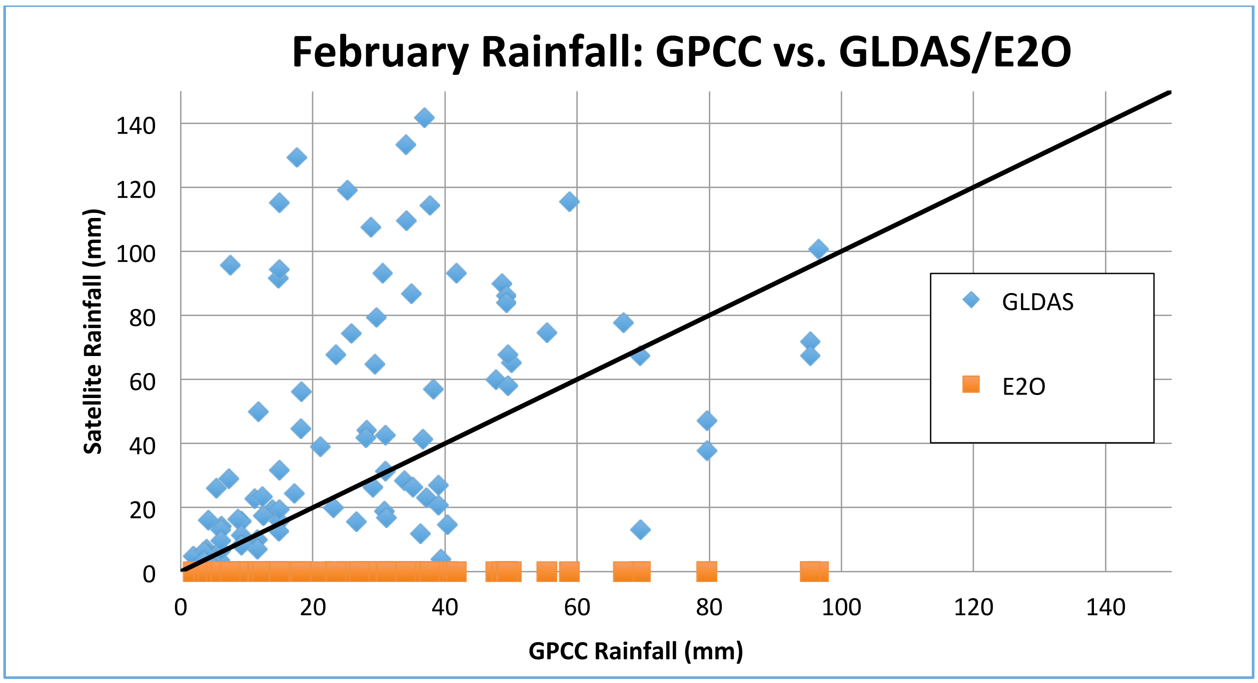

Figure 16 - February Comparison for Ghana

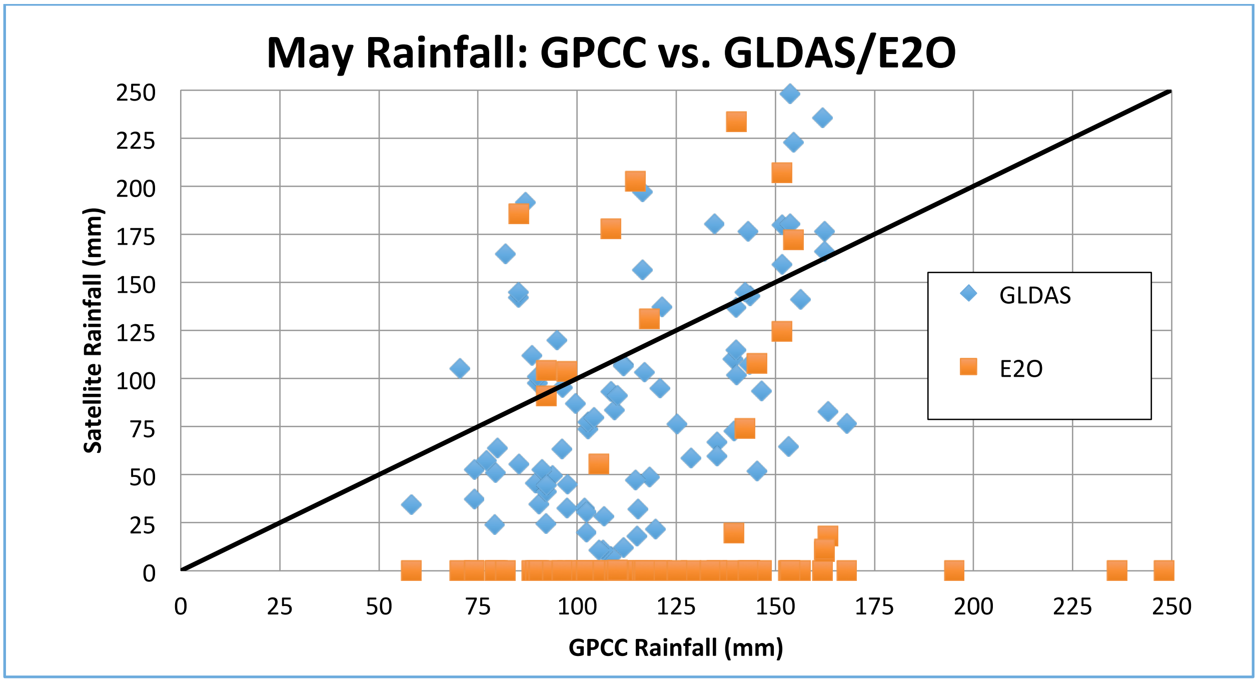

Figure 17 - May Comparison for Ghana

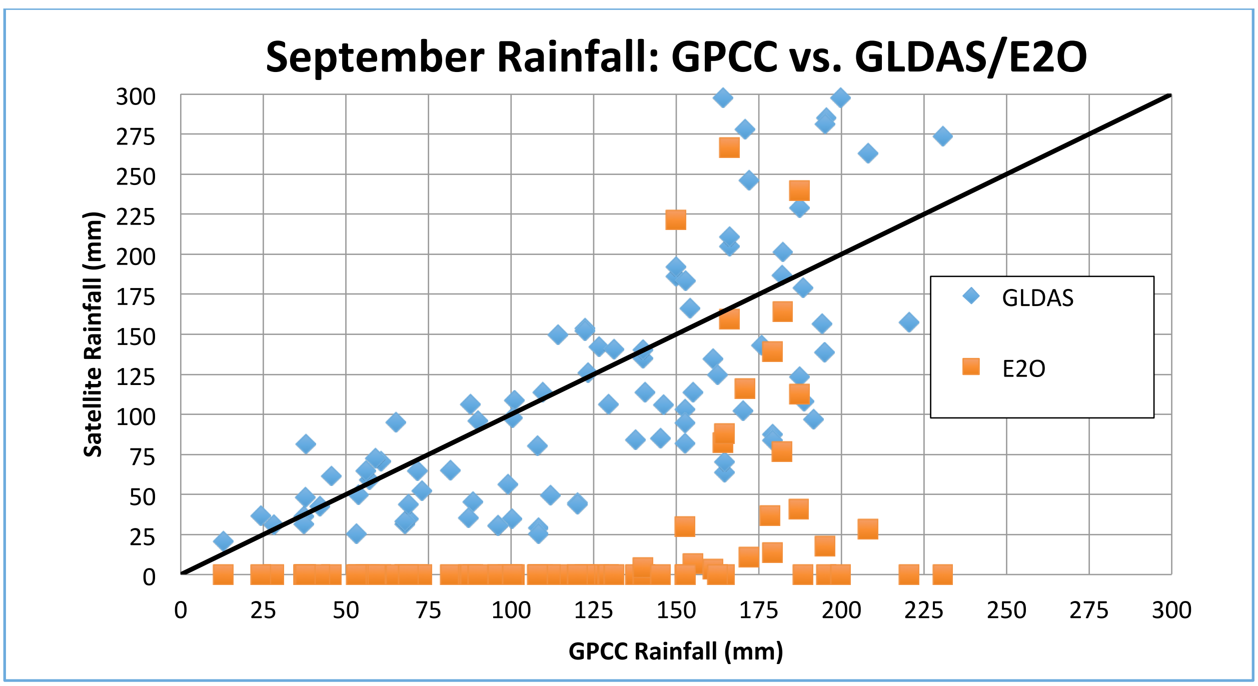

Figure 18 - September Comparison for Ghana

Comparisons for three months were graphed above to show the data distributions. Unique observations can be made for each month, which were selected to correspond to unique rainfall conditions throughout the year.

During February, which typically has low rainfall, GLDAS was overestimated, while E2O reported no rainfall throughout the entire country and was significantly underestimated.

The month of May has higher average rainfall and showed an underestimation for GLDAS and an underestimation for E2O (when removing outliers). May showed a MSE of -15 mm for GLDAS and a MSE of 5 mm for E2O. Even with a lower MSE for E2O the maximum difference was over 800 mm, compared to a maximum difference of 260 mm for GLDAS.

September is underestimated for both GLDAS and E2O, and much larger of an underestimation for E2O. September showed a MSE of -8 mm for GLDAS and a MSE of -55 mm for E2O.

Consistently for E2O there is no reported rainfall for a wide range of GPCC rainfall. For the month of September this happens almost exclusively when actual rainfall is under 150mm.

Conclusions

Satellite systems have the potential to provide easily accessible land and atmospheric data on a global scale. Two satellite systems, GLDAS and E2O were compared to land based data to measure potential error. Guatemala was compared to the satellite systems using rain-gauge stations while Ghana was compared using GPCC data. Monthly rainfall data was compared for the full year of 2009.

Both GLDAS and E2O have significant error over Guatemala and Ghana. There was a decrease in error for both GLDAS and E2O for Ghana, compared to Guatemala. This may be contributed to the widely varying geography in Guatemala and the fairly homogenous geography for Ghana. Additionally, storm patterns and paths seem to have a significant influence in large variations in rainfall between geographically unique regions in Guatemala (coastal region compared to the mountainous highlands) where monthly rainfall can vary from 15 mm to 450 mm.

For both countries GLDAS was more statistically accurate than E2O, yet still had significant errors. In Ghana the maximum rainfall difference was over 500 mm in the month of June. Overall for Ghana GLDAS MSE was 7 mm and MAE was 46 mm. Overall for Guatemala GLDAS MSE was 17 mm and MAE was 76 mm.

When considering satellite data integration for rainwater catchment system design, accurate and long term data is paramount. While satellite data provides sufficient long term data, the accuracy is far below any level to even be considered for system design.

The increased grid resolution (decreased grid size) with GLDAS appears to result in increased accuracy for the two countries compared. Caution should be given in drawing conclusions from this report due to the lack of extensive temporal scope. Only a single year was compared while decades should be considered, along with more extensive statistical analysis, for a more complete comparison.

References

Anderson, James, Guy Ash, and Hailey Wright. *A Statistical Comparison of Weather Stations in Carberry, Manitoba Canada*. Publication. World Meteorological Organization, Web.

Bell, Thomas, and Prasun Kundu. *Comparison Satellite Rainfall Estimates with Rain-gauge Data: Optimal Strategies Suggested by a Spectral Model*. J. Geophys. Res., 2002.

Dessu, Shimelis, and Assefa Melesse. *Evaluation and Comparison of Satellite and GCM Rainfall Estimates for the Mara River Basin, Kenya/Tanzania*. Climate Change and Water Resources, 2013.

Rodell, M., P.R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C.-J. Meng, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich, J.K. Entin, J.P. Walker, D. Lohmann, and D. Toll, The Global Land Data Assimilation System, Bull. Amer. Meteor. Soc., 85(3), 381-394, 2004