GEOS 5984: Geo Data Science with Python

Final Project:

**Analyzing patterns of Rainfall, Temperature, Drought, and Ground Water Storage in South Asia**

Date: 12/10/2022

Submitted by:

Nazmul Huda

MS in Geography, Virginia tech

Submitted to:

Dr. Susanna Werth

Assistant Professor, Department of Geosciences,

Virginia Tech

Keywords: machine learning, geosciences, python, meteorology, geospatial sciences

**Background:**

In this project, four different variables from three different datasets will be analyzed to look for pattern in the South Asian region. The datasets are,

1. Global Land Data Assimilation System (GLDAS) (Rainfall & Temperature)
2. Standardized Precipitation Evapotranspiration Index (SPEI) (Drought)
3. Gravity Recovery and Climate Experiment (GRACE) (Ground water Storage)

These datasets are large datasets with very high temporal and spatial coverage. All these datasets are global which allows the data to investigate regions that are not accessible. Areas in South Asia includes 8 countries and the Himalayan Mountain range with very diverse geological elements. So, with the availability of global data in this project South Asian region have been studied to look the temporal and spatial variation of rainfall, temperature, drought, and ground water storage.

Chart

Description automatically generated

Figure 1: Map of the study area

Jupyter Notebook and Python programming language have been used to carry out the analysis and the notebook resulting from the analysis and necessary datasets have been uploaded to GitHub for replicability and accessibility.

**Data & Methods:**

Four variables have been used including temperature, rainfall, drought, and ground water storage. Short description for the variables has been listed below.

*Table 1: Description of the datasets*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variable | Dataset | Temporal Resolution | Spatial Resolution | Coverage | Starts From | Ends at |
| Rainfall | GLDAS | 30 days | 1° / 111 km | Global | 2000 | 2022 |
| Temperature | GLDAS | 30 days | 1° / 111 km | Global | 2000 | 2022 |
| Drought | SPEI | 30 days | 0.5°/ 55 km | Global | 1901 | 2020 |
| Ground Water Storage | GRACE | 30 days | 300-400 km | Global | 2002 | 2022 |

To analyze the data, the datasets have been all clipped to the South Asian geographical boundary, where latitude Min = 6, latitude Max = 38, longitude Min = 60.25, and longitude Max = 98 degrees. Then the datasets have been clipped from 2000~2020 to match the common time period except for Ground water storage data which starts at 2002.

The datasets have been then analyzed for max, mean, and minimum values on the monthly basis to find out any anomaly or pattern. Results have been plotted in the XY Graphs and in maps to find out any pattern if it exists.

Finally, principal component analysis has been performed for the three variables excluding the drought index data because that dataset was not producing the singular value decomposition products.

**Results**:

In order to find out the spatial pattern in temperatures, the temperature map has been plotted for January 2000 which is the coldest month.

|  |  |
| --- | --- |
|  |  |
| Figure 2: Pattern of Temperature in South Asia, January 2000 | Figure 3: Pattern of Rainfall in South Asia, January 2000 |

In figure 2, the Himalayan Mountain range is visible and that works as the boundary between very high and very cold air temperature. The southern part of the study area merges to Indian Ocean and the temperature there is as high as 30 degrees Celsius in one of the coldest months of the year. Gradual increase of temperature is observed throughout the region going from north to south. In figure 3, the pattern of rainfall in the month when rainfall is the highest can be observed and again, the mountainous Himalayan range works as an obstacle not allowing the clouds pass through it and as a result, north of the mountain observes 5 times lower rainfall than the southern region. The two spots looking bright yellow are in Bangladesh and India which observes one of the highest annual rainfalls in this region.

|  |  |
| --- | --- |
|  |  |
| Figure 4: Pattern of Drought in South Asia, January 2003 | Figure 4: Pattern of Ground water storage in South Asia, January 2009 |

Another two important factors to consider while looking into the weather and climatological variables are, drought and ground water. As this region is very highly dependent on agriculture, these two factors are very essential to study. While drought is measured from standard precipitation and evaporation indices, the ground water storage is calculated from very sophisticated GRACE satellite data which measures the changes in gravity and allows the study of the ground water storage depletion or accumulation.

From figure 4, a few hot spots of drought are visible and in the southwestern coast of India and one in northern part of India. In this map, the lower index value represents drought. Also, this image confirms the high rainfall and less drought by showing the northeastern India and northern Bangladesh to have very high SPEI value (low drought index) which was observed to have very high rainfall amount. From figure 5, high ground water storage is visible in Bangladesh, and it is expected because this small country has more than 300 rivers and outlets from Ganges, Brahmaputra and Meghna all passes through Bangladesh and meets the Indian Ocean through the Bay of Bengal. This region still uses hand pump for ground water access as the ground water is accessible just a few feet below the ground and ground water recharge is also very frequent as it observes a lot of rain throughout the year.

Table 2: Maximum and minimum rainfall and temperature recorded in South Asia

|  |  |  |
| --- | --- | --- |
| **Variable** | **Maximum Recorded** | **Minimum Recorded** |
| **Temperature** | June 2014 | January 2001 |
| **Rainfall** | July 2016 | January 2018 |

*Trend Analysis for the variables:*

|  |  |
| --- | --- |
|  |  |
| Figure 6: Trend of temperature in South Asia | Figure 7: Trend of rainfall in South Asia |
|  |  |
| Figure 8: Trend of drought in South Asia | Figure 9: Trend of ground water in South Asia |

From figure 6-9, it is visible that the temperature is increasing, and drought is also increasing at a very similar rate. As the higher temperature brings in torrential rain the amount of annual rainfall is also seemed to be increasing. The only trendline that is going downward instead of upward is ground water storage. With a very highly agricultural region, ground water is very important for the crops, and it is the primary drinking water source. So, this very basic study can imply that it is now very important to analyze the ground water in this area from different sources as the GRACE data is a coarse resolution dataset.

It is also important that the variables studied here should be looked into more closely regarding the time frame. Many variables have a temporal variation and due to seasonality, some environmental factor is more important in a specific time window. Such as, rainfall becomes rainfall when it is cultivating season when irrigation is not available, and if the ground water depletes to lower level, then the rainfall becomes more important.

|  |  |
| --- | --- |
|  |  |
| Figure 10: Monthly average rainfall in South Asia | Figure 11: Monthly average temp in South Asia |
|  |  |
| Figure 12: Monthly average GWS in South Asia |  |

From figure 10-12, the annual variation is observed and from this annual variation it is clear which season is important to analyze when extreme values are needed to analyze.

|  |  |
| --- | --- |
|  |  |
| Figure 13: Temperature trend in January | Figure 14: Temperature trend in June |
|  |  |
| Figure 15: Rainfall trend in January | Figure 16: Rainfall trend in January |

From figure 13-16, the trend change in month can be observed. The months with highest and lowest temperatures can both be seen observing higher temperature as the trendline is facing upward. For precipitation, the scenario is complex as June and July are two consecutive months with maximum rainfall and they are facing different pattern in the trend. While correlation is not causation, it might indicate the shift in weather pattern too.

**Principal Component Analysis:**

Principal component analysis is a great way to identify any pattern in the data. Below are the principal component analysis of the Ground Water Storage, Rainfall, and Temperature.

Chart

Description automatically generated

Figure 17: PCA for Ground Water Storage

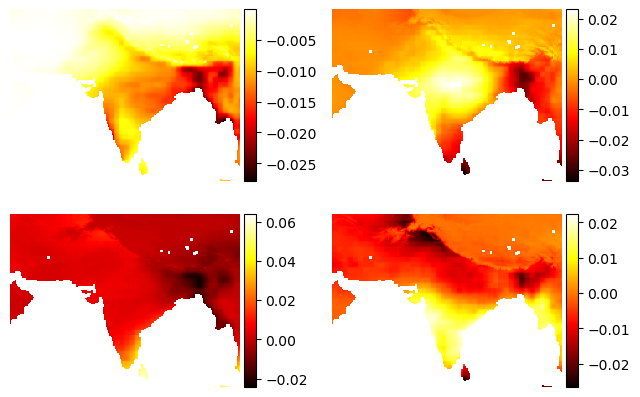


Figure 18: PCA for Rainfall

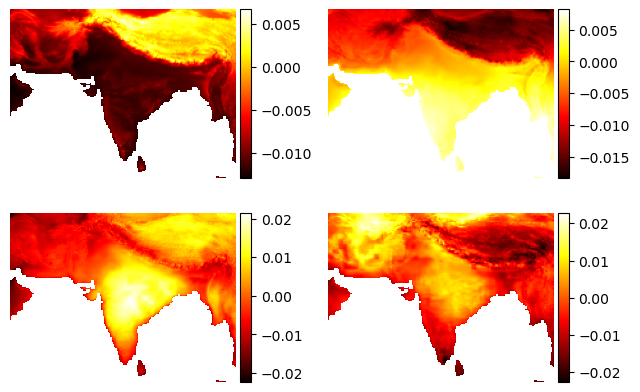


Figure 19: PCA for Temperature

|  |  |
| --- | --- |
|  |  |
| Figure 20: Cumulative of SV (Ground Water) | Figure 21: Cumulative of SV (Rainfall) |
|  |  |
| Figure 22: Cumulative of SV (Temperature) |  |

From the Figures 17-22, the principal component analysis and Singular value decomposition of the three variables can be seen. For temperature, the second component could explain the 70% of the data and the most important feature. PCA can be evaluated more to find out more information about the data and analyze such a high dimensional dataset.

**Conclusion:**

This project looks into the variables of high dimensional datasets and analyzes the value in different dimensions to allow deeper understanding of the data. Once a summary report like this one is generated, more in depth analysis can be undertaken to look into the components of the datasets.