Report: Design Credit (CSN1020)

Topic: Handling big datasets for Analyzing long term weather over western India

Student Name: Akshat Jain(B20AI054)
Mentor: Dr. Amit Sharma

Objective: Handling and analyzing various large datasets for long term weather analysis over western India

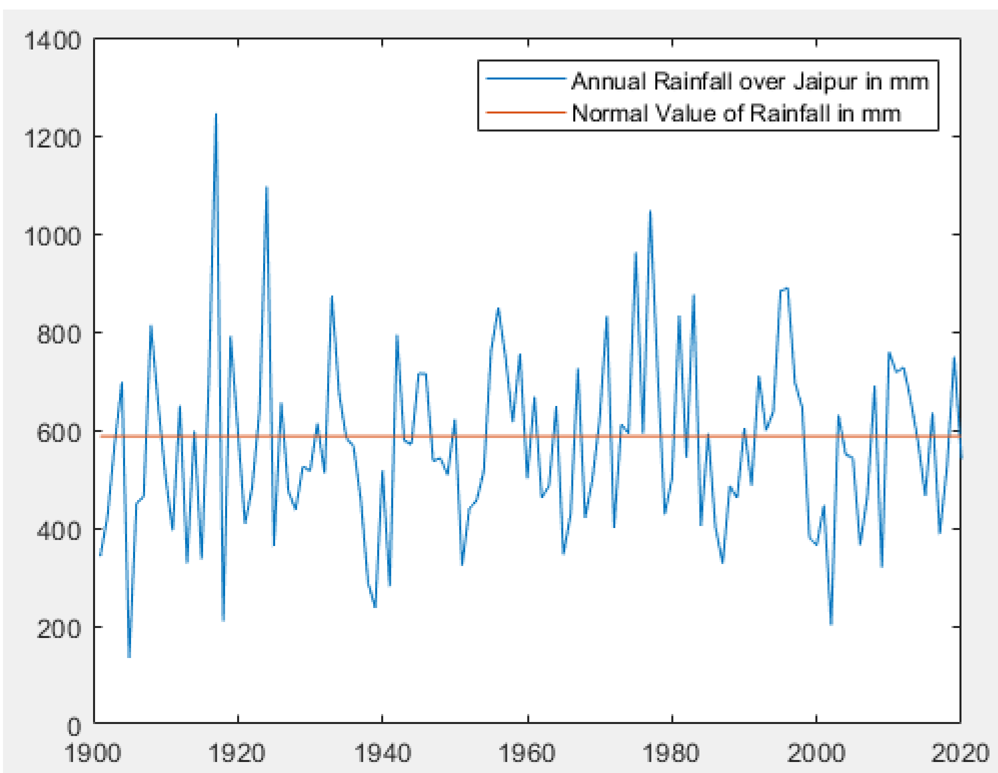
Duration of the Project:

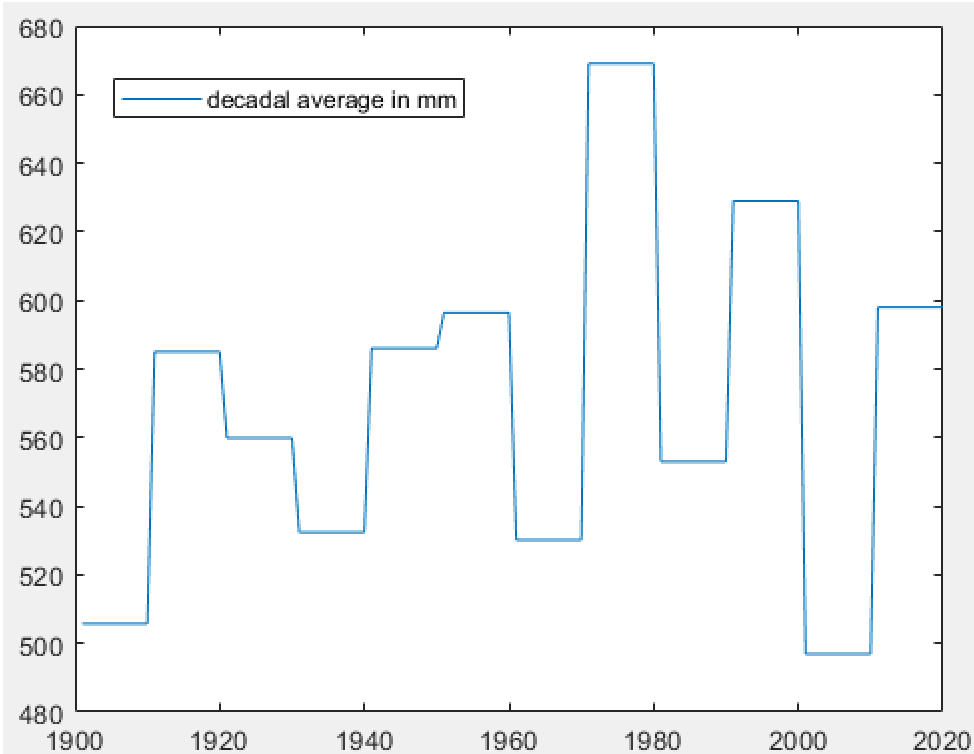
Tools Used: MATLAB, Panoply

Datasets and Resources: WRIS annual and monthly rainfall data (1900-2020), ICRISAT data for irrigated Land, Giovanni data for Evapotranspiration(time-series), Giovanni data for Soil Temperature (time-series), Giovanni air temperature data (time- series), Giovanni precipitation data(time-series), Giovanni precipitation data(time-averaged), Giovanni Air Temperature data (time-averaged)

Results:

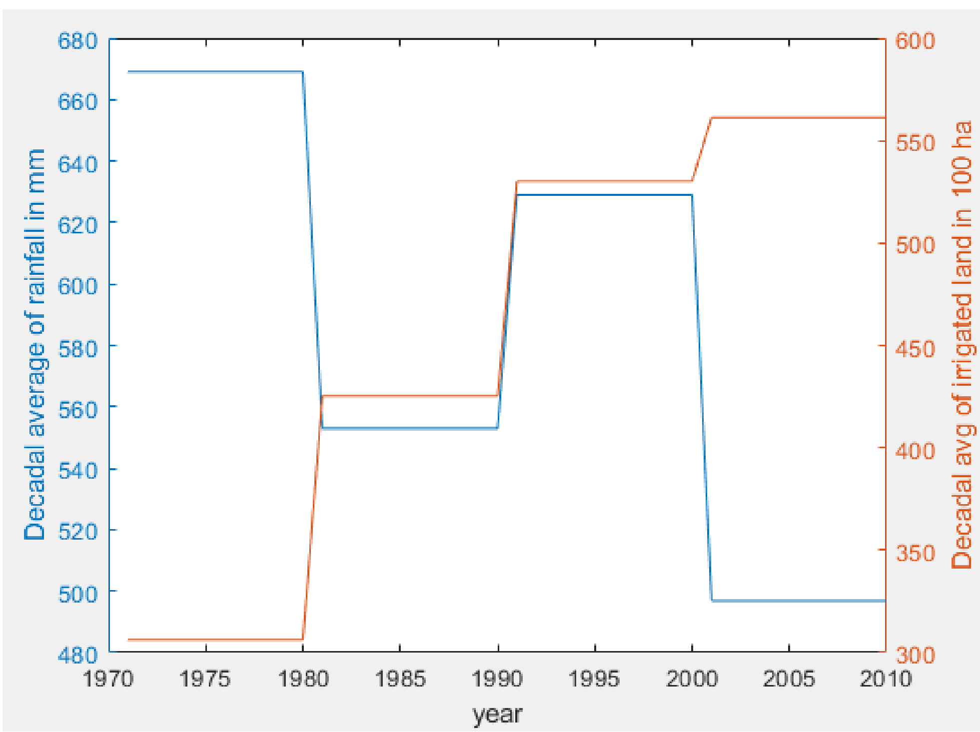
1. **WRIS Data**





Here we observed a decreasing trend in the decadal average of rainfall after 1970s

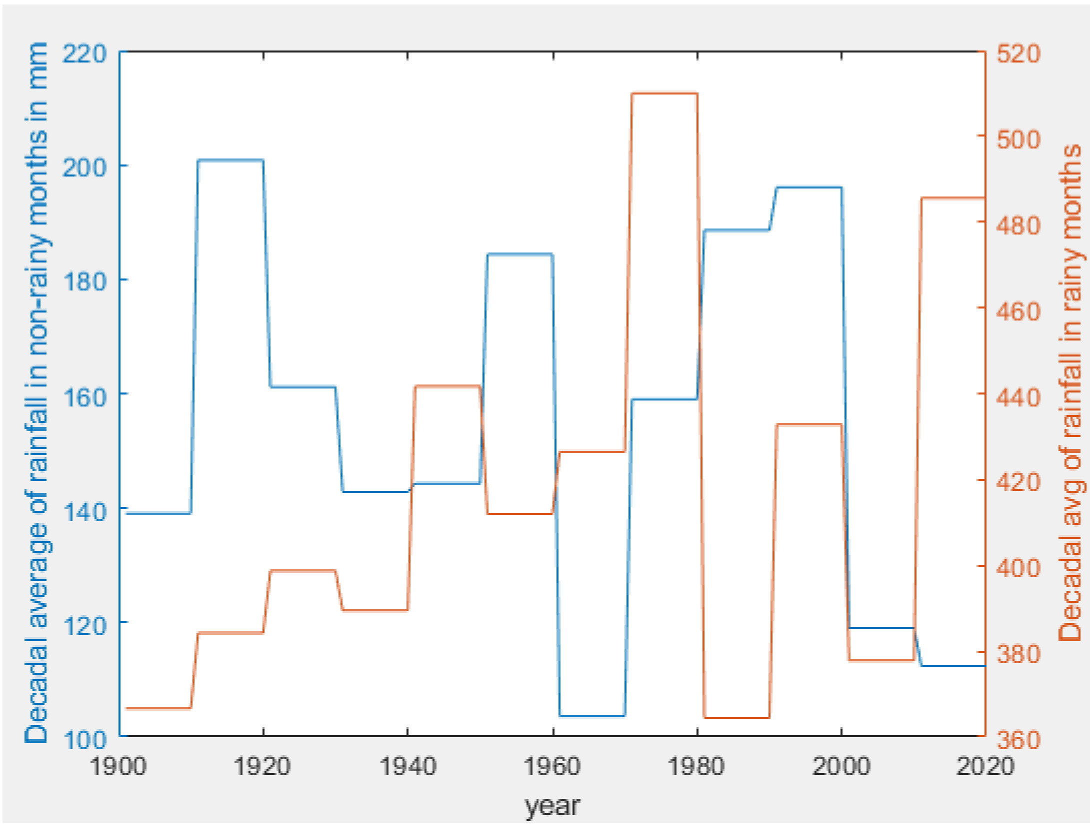
2. Rainfall and Irrigated Land analysis for Jaipur



Correlation Coefficient(R)= -0.6611

Here, a negative correlation coefficient suggests that the rainfall aand irrigated land have an anti dependence.

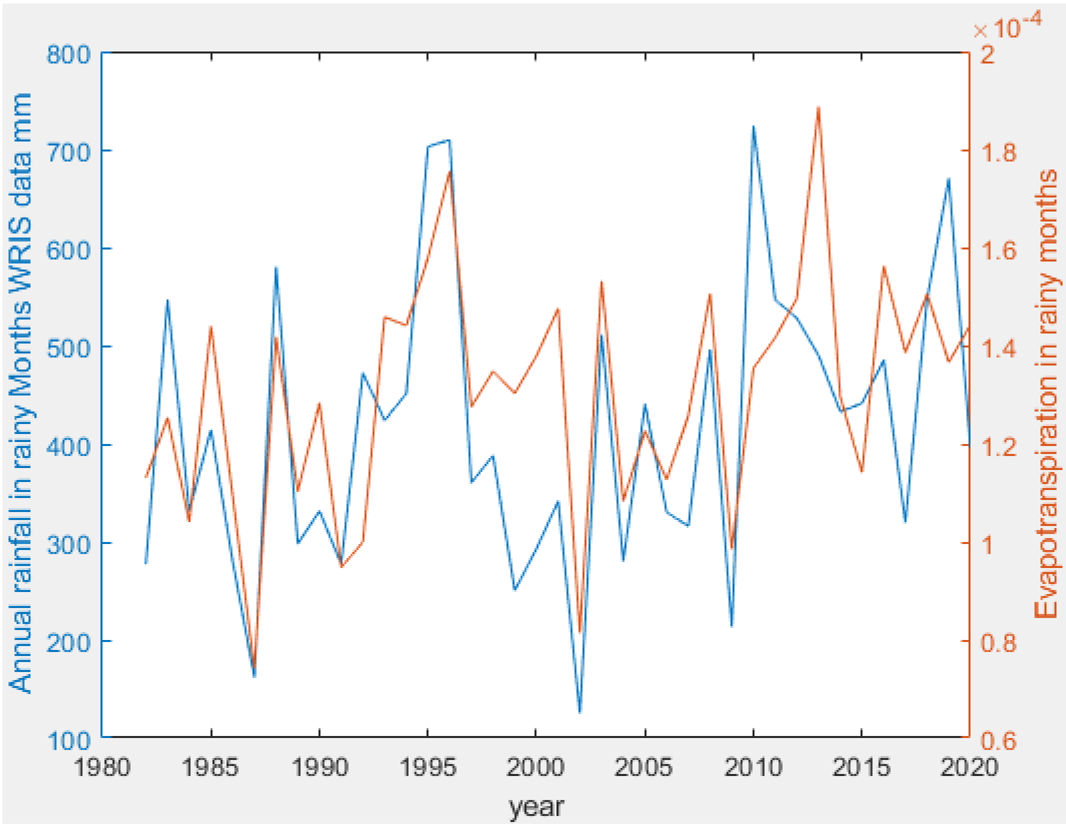
3. **Seasonal Analysis of Rainfall into rainy months(June to September) and non-rainy months(Rest of the year) for Jaipur:**



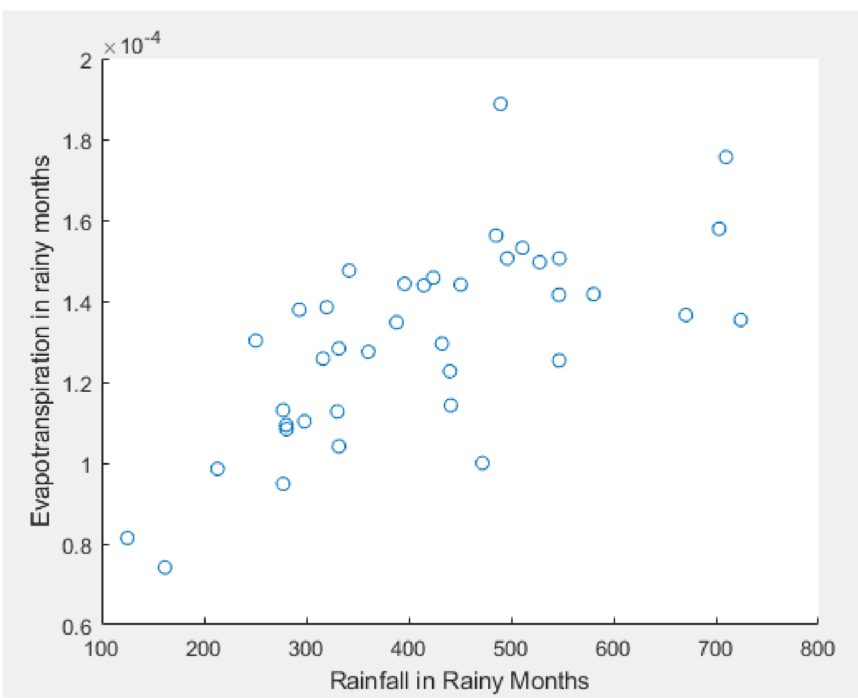
R= -0.2167

Note: Here the y-axis for both are different

4. **Rainfall And Evapotranspiration Analysis (From FLDAS model in Kg/m^2/s) for Jaipur**

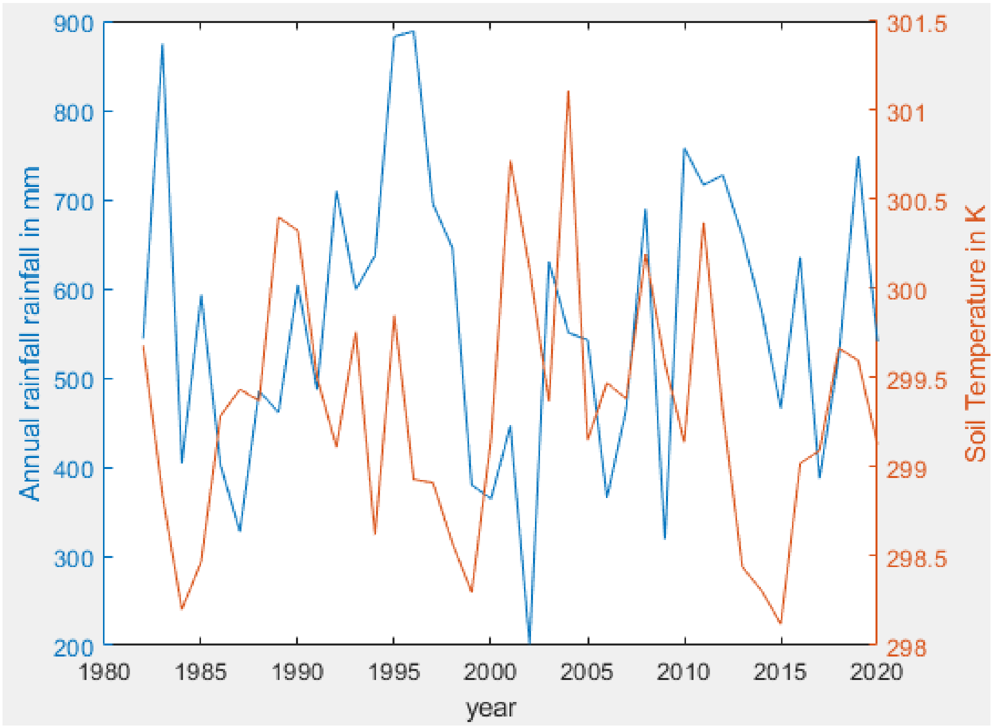


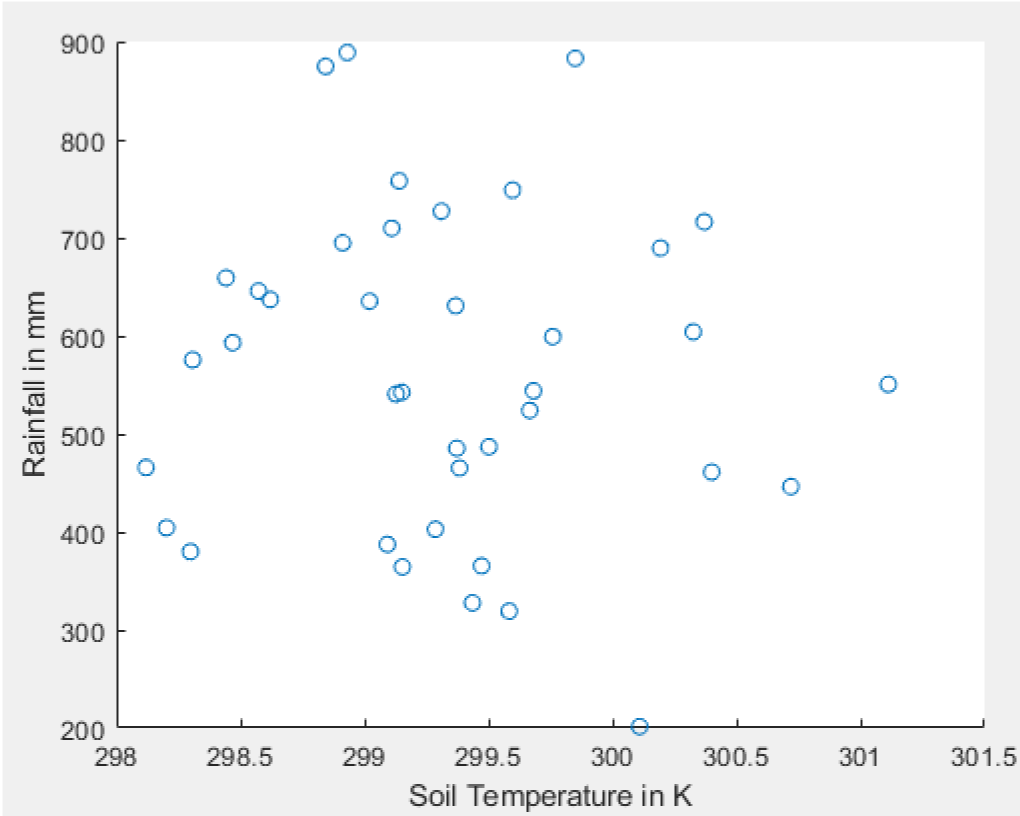
R= 0.6819



The R value and the Scatter plot suggest that their is a strong positive correlation between the Rainfall and Evapotranspiration.

5. Soil Temperature (From MERRA-2) and Rainfall(WRIS data) for Jaipur:



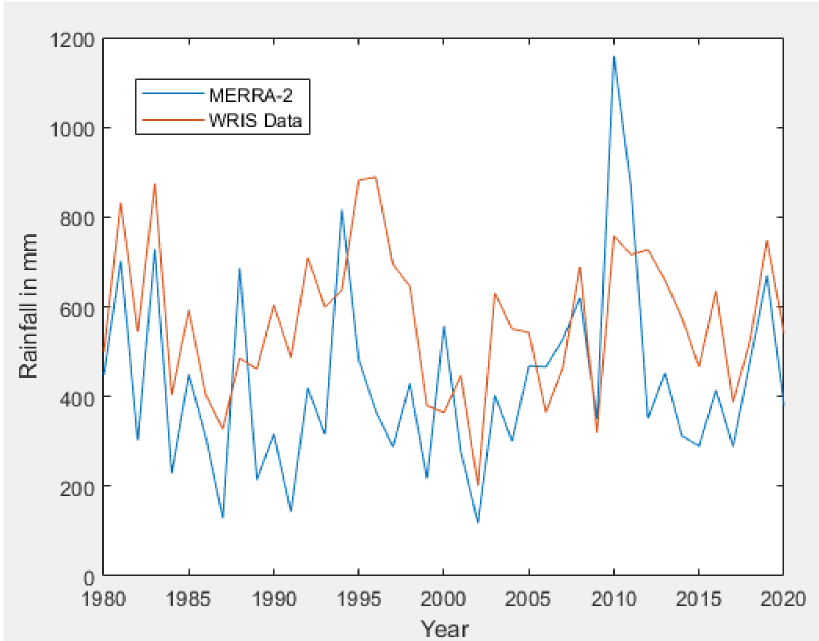


R Value= -0.06

Here, the negative correlation suggests that rainfall decreased with an increase in temperature but the dependence is not that strong.

6. **Model Analysis of Rainfall data**:

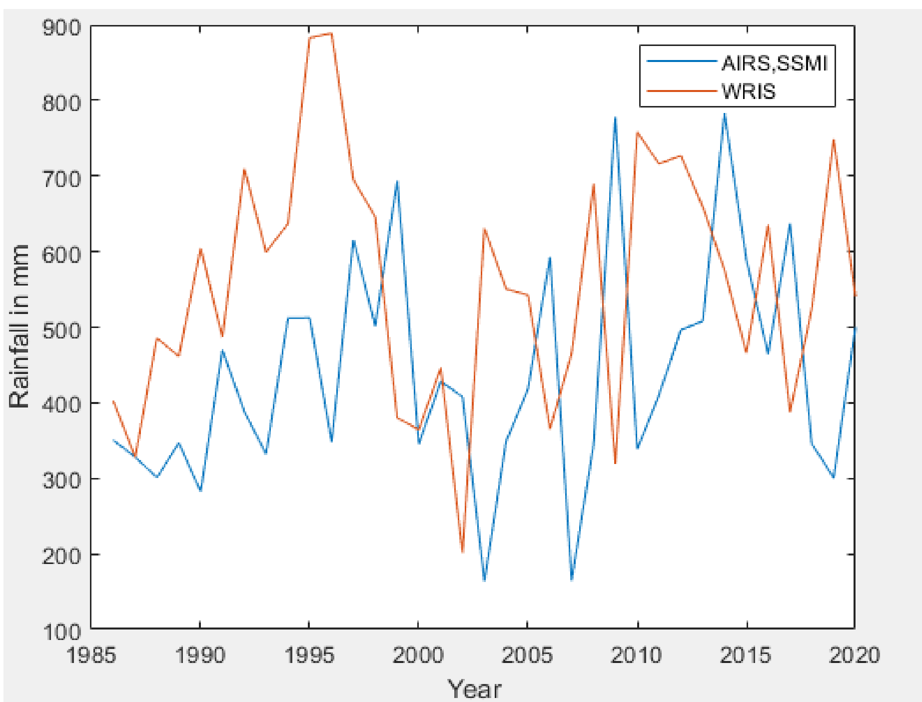
   A. **WRIS data vs MERRA-2 data**



R= 0.5367

The positive correlation coefficient implies that MERRA-2 is satisfactorily consistent with the ground observation of WRIS data.

B. AIRS SSMI observation vs WRIS Data:



R Value: 0.678

Again here we observe a decent positive correlation between AIRS, SSMI observations and the WRIS data.

7. Standardised Precipitation Index (SPI) and Droughts:

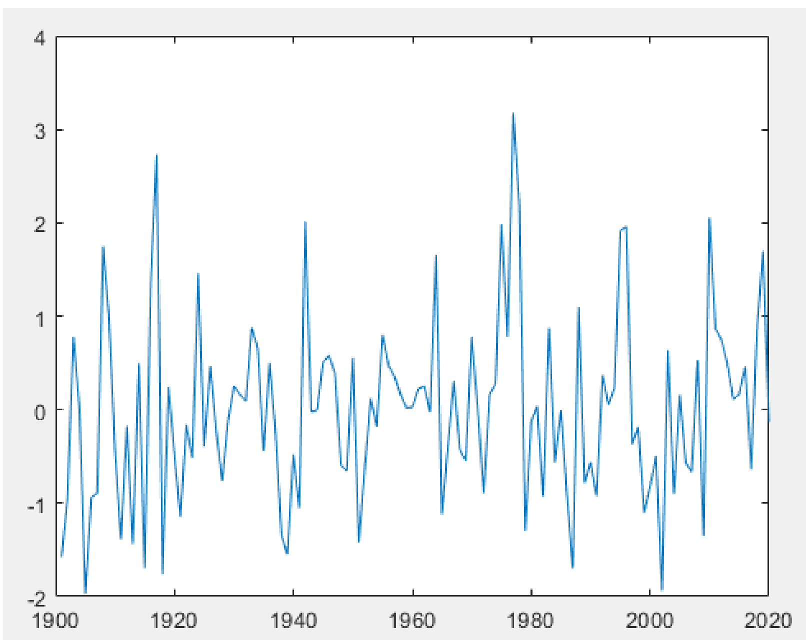
Reference: Global Nest Research Paper (Link)
SPI = (X – Xm)/ σ
Where X is precipitation for the station, Xm is mean precipitation and is standardized deviation.
SPI can be a very useful tool to analyse drought frequency in a particular region.

Table 1. SPI drought classes

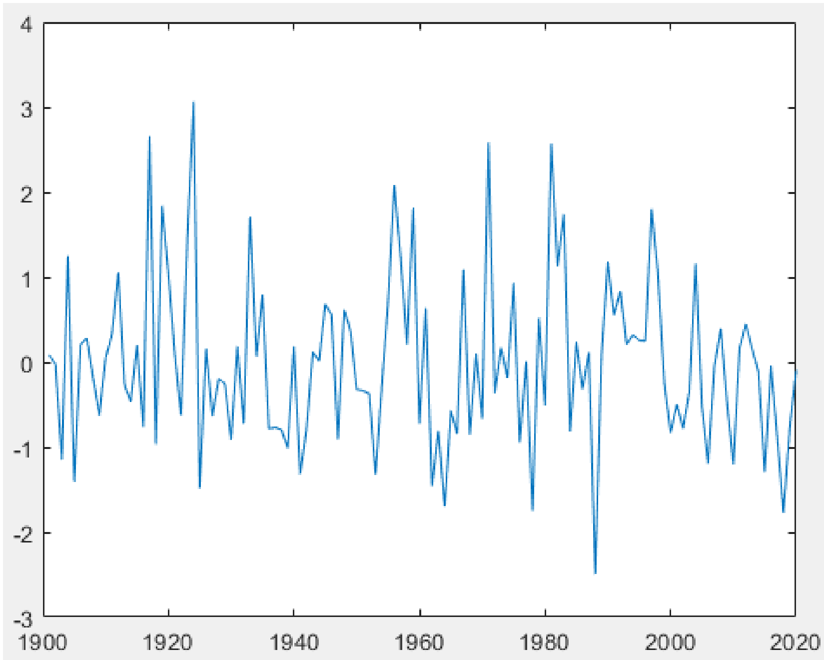
S. No	Criteria of SPI values	Type of drought
1.	0.00 to -0.99	Mild drought
2	-1.0 to - 1.49	Moderate drought
3.	-1.5 to -1.99	Severe drought
4.	-2 and less	Extreme drought
5.	More than 0	Above normal drought

SPI Analysis for Jaipur:

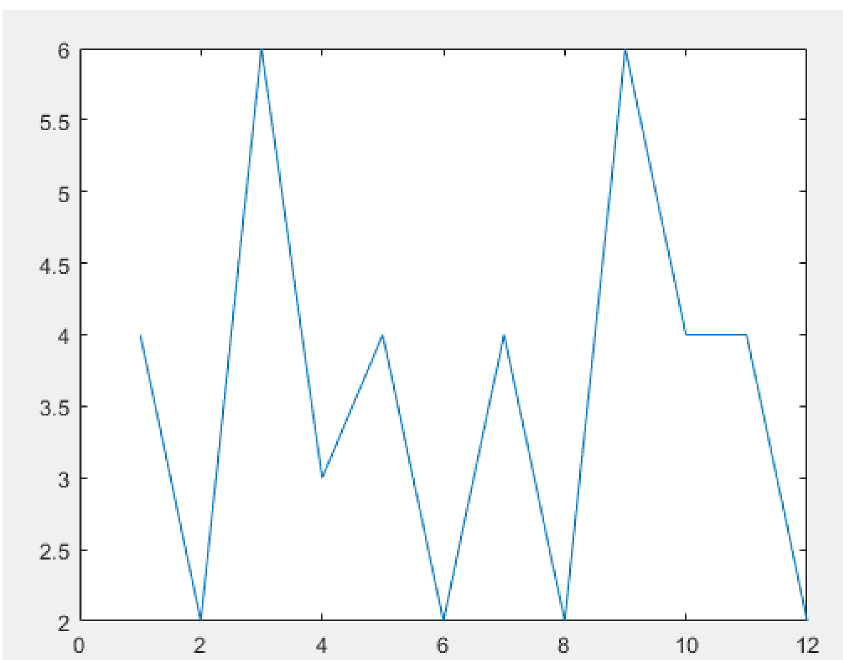
A. Annual SPI for Rainy Months:



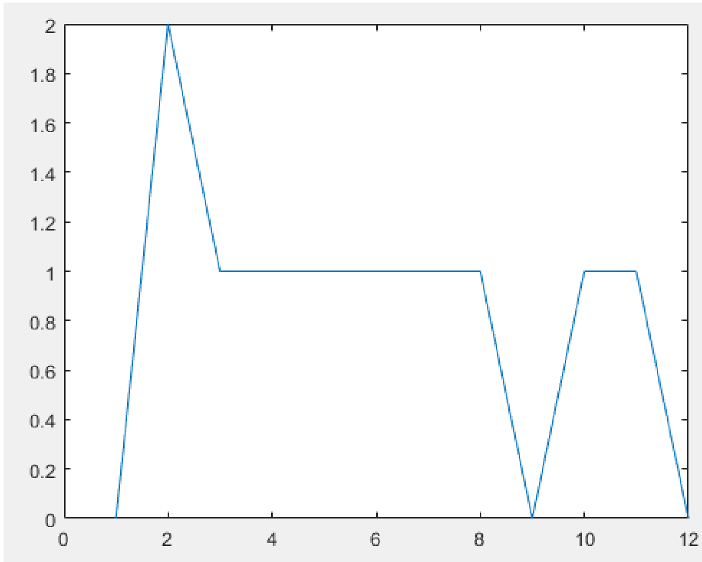
B. Annual SPI for Non-Rainy Months:



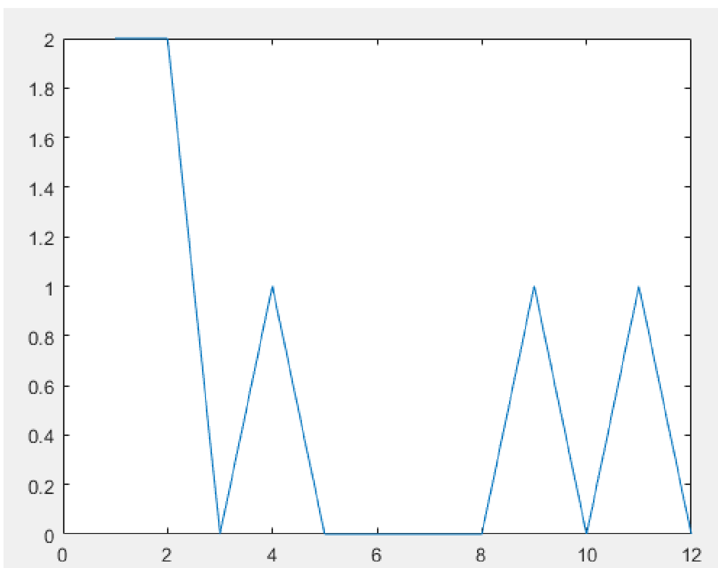
C. The number of Mild droughts in each decade:



D. The number of Moderate Droughts in Each Decade:



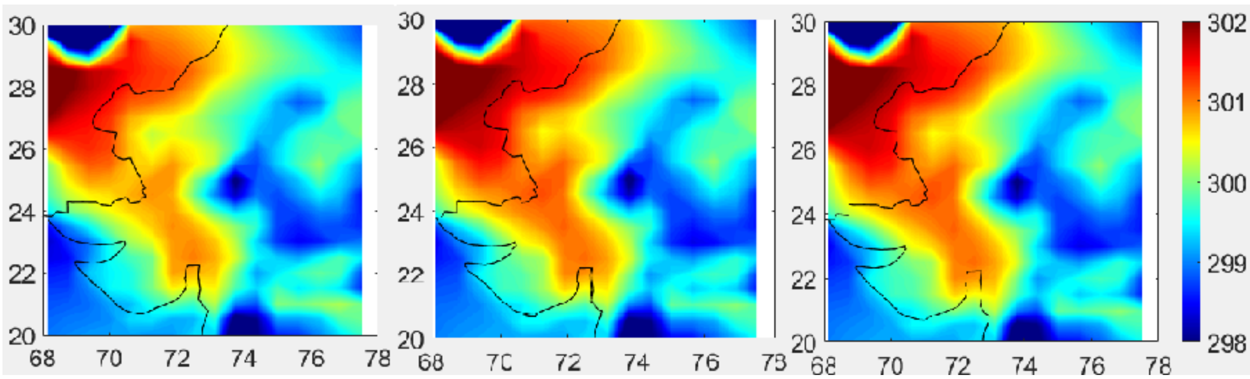
E. The number of Severe droughts in each decade:



Total Number of Mild Droughts: 43
Total number of Moderate Droughts: 10
Total number of Severe Droughts:7
Total number of Extreme and above Normal Droughts: 0

8. Decadal Contour Plots over Western India:

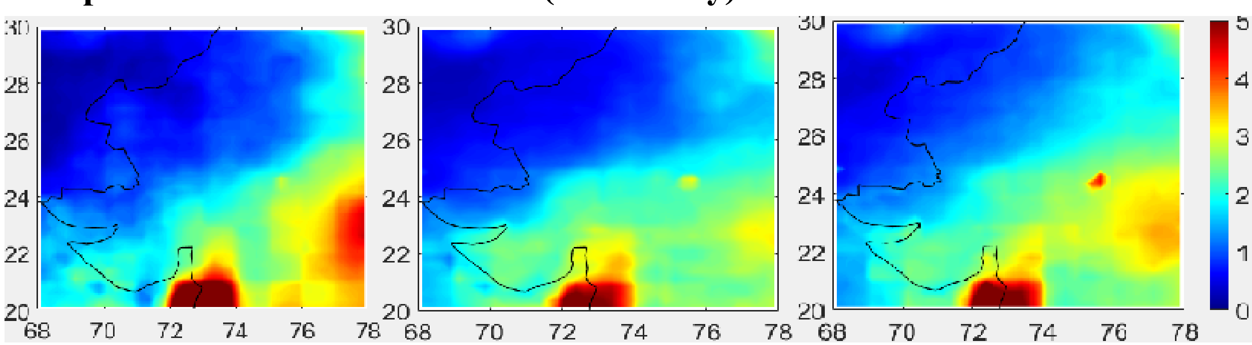
A. Air Temperature in K from Merra-2 model:



1980-1990 **1990-2000** **2000-2010**

Here, we can clearly see that the Air Temperature is increasing over most of the parts of Western India with time.

B. Precipitation Over Western India (in mm/day) from TRMM data:



1990-2000 **2000-2010** **2010-2020**

Here we observed something interesting, i.e. Rainfall first decreased from (1990-2000 decade) to (2000-2010) and then again increased in (2010-2020) decade.

Mathematical Analysis of Rainfall data:

Resources:

Sen's Slope Estimation,

Mann Kendall's test

1. **Mann- Kendall Test:**

   Mann-Kendall statistic is used to as certain a significant trend in climatically data. To trend analysis Mann-Kendall statistic is one of the best method to detect trend.

$$S = \sum_{k-1}^{n-1} \sum_{j-k+1}^{n} \operatorname{sgn}(x_j - x_k) \quad (1)$$

$$\mathrm{VAR}(S) = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_{p-1}^{g} t_p(t_p - 1)(2t_p + 5)\right] \quad (2)$$

$$Z_{MK} = \frac{S-1}{\sqrt{VAR(S)}} \text{if} S > 0$$

$$= 0 \text{ if} S = 0$$

$$= \frac{S+1}{\sqrt{VAR(S)}} \text{if} S < 0$$

(3)

A positive (negative) value of Z_{MK} indicates that the data tend to increase (decrease) with time.

Normal distribution table:

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995
3.9	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99996	.99997	.99997

A) Mann Kendall's test for annual Rainfall Data(n=120 years):

For complete 120 years: S= 414

Number of tied groups=0

Therefore, Zmk= 0.9368

For confidence interval, let **α**= 0.05 (95% confidence interval)

This corresponds to Z value of 1.65

Since Z>Zmk: for the 120 year interval, trend isn't significant.

B) For last 40 year interval (1980-2020):
S=6
Zmk= 0.0583
Z>Zmk: Trend is not Significant

C) For last 20 year interval (2000-2020):
S= 52
Zmk= 1.6546
Zmk>Z: Therefore we can say with 95% confidence that there is an increasing trend.

2. Sen's Slope test

A measurement is obtained at n points in time; $t_1, t_2, \ldots, t_n$ at a specified location. Compute the $N = n(n-1)/2$ slope estimates

$$Q = \frac{y_j - y_k}{t_j - t_k} \quad (1)$$

The non-parametric slope for the curve would be the median of these values.

Nonparametric Intercept, $\hat{\beta}_o$, of the Linear Trend

$$\hat{\beta}_o = y_{MED} - \hat{\beta}_1 \times t_{MED} \quad (9)$$

where

y_{MED} = median of the n measurements $y_1, y_2, \ldots, y_n$

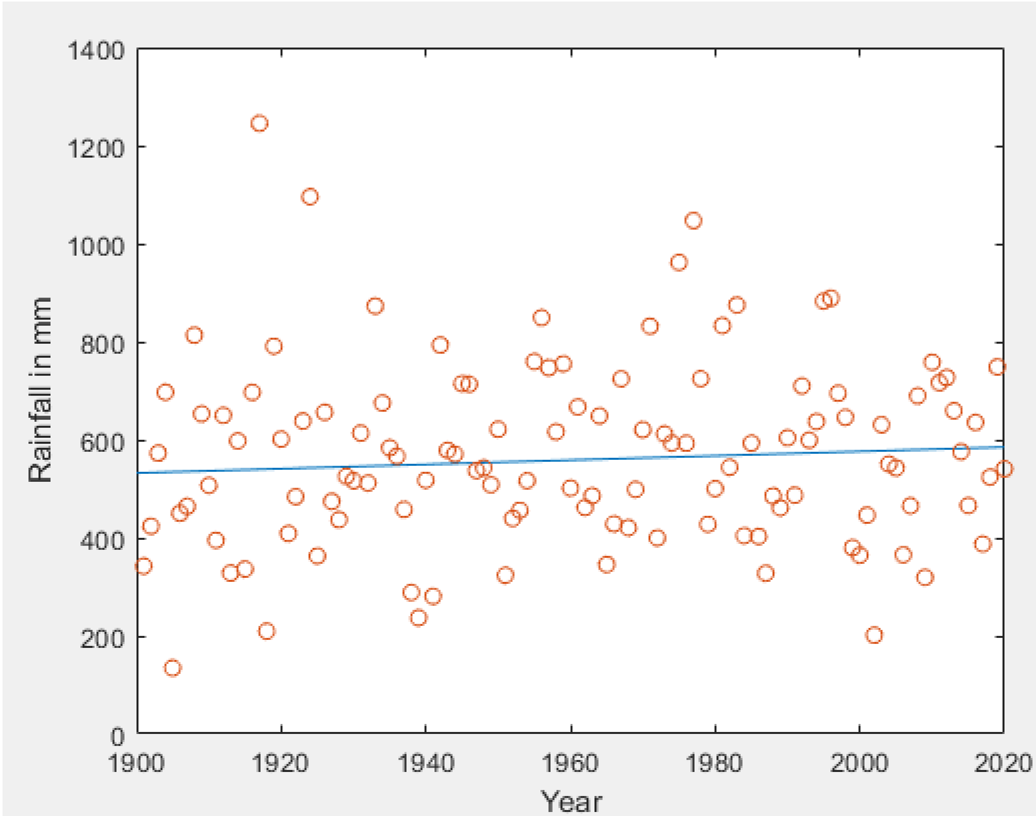
t_{MED} = median of the n times $t_1, t_2, \ldots, t_n$

and

$\hat{\beta}_1$ = nonparametric slope estimate.

i) **For 120 years:**

Non parametric slope= 0.4402
Ymed= 558.57
Tmed= 60
Therefore, the intercept= 532.158



ii) For last 40 years:

Years 1980-2020:

Non parametric slope= 0.2069

Ymed= 563.13

Tmed= 100

Therefore, the intercept= 542.44

iii) Last 20 years:

Years 1980-2020:

Non parametric slope= 7.2169

Ymed= 546.5700

Tmed= 110

Therefore, the intercept= 483.13

Conclusion: This project made us handle a variety of datasets using MATLAB,Panoply and many other tools. We applied a number of data analysis techniques and obtained a number of different types of plots.