Mason Modelling Days 2014 Climate Dynamics Project: A Drought Prediction System

Muhammad Baqui, Lorianne Donato, Adam Moskey, Nick Oderio, Changkai Sun, Marilyn Vasquez

June 14, 2014

Abstract

Abstract.

1 Introduction

Though the definition of what is considered a drought varies among different countries, it is possible to define a drought in the most general sense. A widely accepted way of defining a drought is to consider it a period of time where a region will receive lower than average precipitation, having detrimental results such as a low supply water to irrigation canals. For how much and for how long the state of lower than average precipitation must occur for the weather event to be officially considered a drought varies among differing regions and people. Notwithstanding disagreements, there are four basic ways in which to categorize droughts; namely, meteorological, agricultural, hydrological, and socio-economical.

Droughts are common throughout the world with a devastating impact upon agriculture, economy, and society. Though they are usually not thought of as one of the most severe natural disasters, they are indeed one of the most costly, with impacts into many different areas. In fact, droughts have impacted the most amounts of people out of all environmental disasters, and have affected more people in North America than any other hazard. Due to the multifaceted nature of its impacts, it is not surprising that the cause of drought is similarly quality. High temperature, global weather patterns, wind, humidity, precipitation, aridity, high pressure, soil moisture, and topography all play a role in the development of a draught. It is clear then, that an accurate prediction or measure of precipitation must play an important role in being able to measure and identify draughts.

2 Existing Draught Indices

In order to properly understand the work outlined in this survey, it is pertinent to discuss some of the existing methods used to identify and predict droughts as well as other relevant methods used to forecast weather events. There exist several drought indices currently used in identification and prediction of drought. This would allow decisions to be made in determining what is the best course of action and when is it necessary to take such action. The simplest index, called the Percent of Normal, is calculated by dividing the actual precipitation by the average precipitation (usually over 30 years). The main disadvantage of using this method is that the use of such a comparison implies that the mean and median precipitation levels are the same, which usually is not the case.

The Palmer Drought Severity Index, which works best with large areas of flat land, is currently used by the U.S. Department of Agriculture, in order to determine when it is best to grant emergency assistance. Though due to the fact that it is based on soil moisture, it is not quite as well suited for other areas, such as mountainous or rocky areas.

Another relevant index used in Australia is the Deciles Drought Index, which groups monthly precipitation into deciles so that the lower than average weather cannot occur more often than twenty percent of the time. The main problem this and many other indices face is that long term accurate climate data is required in order to compute it with a usable amount of accuracy.

3 Importance and Motivation

There is a need for an investigation into the assumptions of the spatiotemporal analysis, for which many of those in environmental fields use commonly in multidimensional analysis. One of the assumptions of interest is the stationarity assumption. Stationarity is used by climatologists to refer to the unchanging nature of weather patterns. One of the main disadvantages of current drought identification and prediction systems is the inherent assumption of stationarity built into the model. Whether or not climate patterns remain stationary over time is still under debate. However, it is clear that some aspects of the climate data are dependent on time, which aspects and to what extent is still unknown. By analysing current methods, while focusing in particular on the accuracy of such assumptions, we are better able to work at the modification or creation of a more accurate system.

This project is focused on constructing a drought prediction system. Our goal is to develop a better indicator of drought than previously identified in the literature and, in particular, analyzing the stationarity assumption of those methods. There are six indicator variables that are used to identify drought; namely precipitation, run off, soil moisture, water vaporization, convergence, and evaporation. Among the variables, precipitation will be considered as our primary focus.

4 Data

We use the NASA-MERRA reanalysis data available at http://gmao.gsfc.nasa.gov/merra/. Our focus is primarily on its water cycle variables (Precipitation, Evaporation, Runoff, Water Vapor, Soil Moisture, Moisture Convergence).

5 Sparcity Analysis

5.1 Temporal-Spatial Analysis

From the precipitation data for one particular location we can scan through the entire 35 years to see the temporal evolution of the precipitation for 12 months. Doing this for multiple locations allows us to identify either possible patterns in our data or the lack thereof. Consider first the frequency of precipitation occurring over all regions over a decade. In 35 years, three intervals were chosen and corresponding histograms were constructed for each month of the entire 35 years. It was observed that the largest precipitation occurs during the spring, from May to July. Given this, the month of June was selected for comparison purposes. Analysis of the histograms shows that they do not necessarily hold the identical patterns. This means nothing can conclusively be said as per the existence of stationarity in the precipitation pattern.

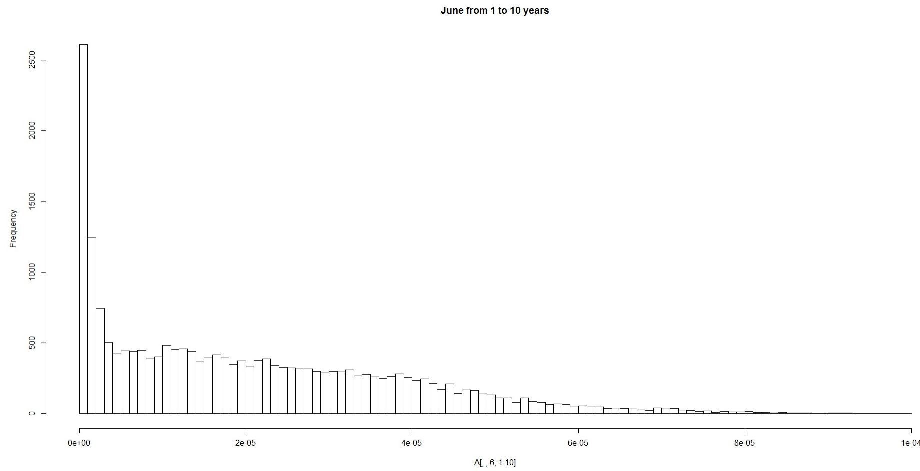

Figure 1

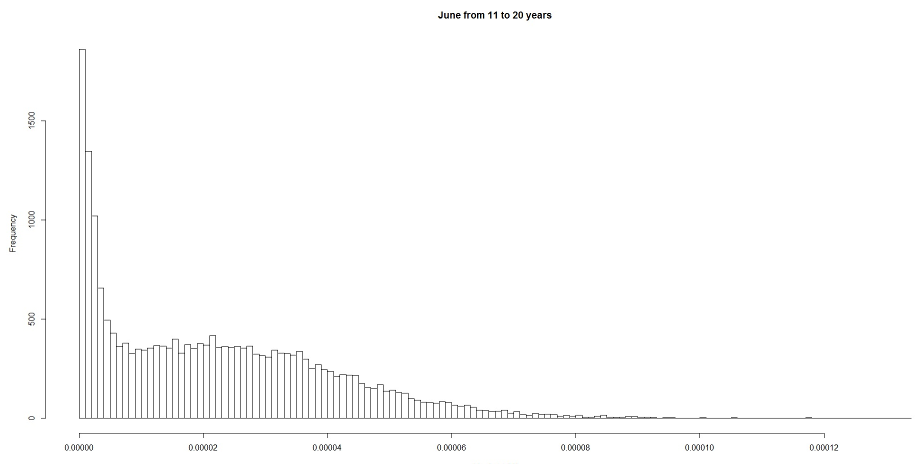

Figure 2

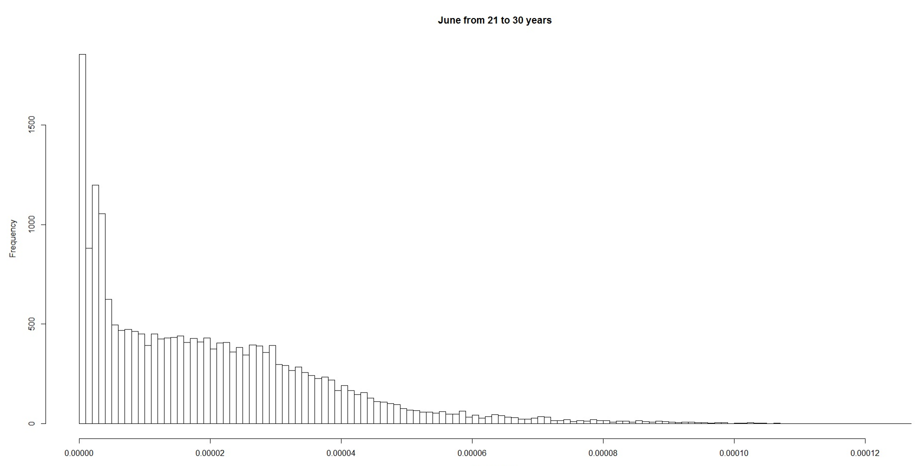

Figure 3

A surface plot was constructed from the average precipitation data for a specific location in the Western region of the United States over the entire 35 years. Similar to the procedure mentioned above, three difference surface plots are constructed for slightly over three decades. A high value of precipitation would indicate a darker red coloured region, whereas a darker blue colour would indicate existence of drought. The three plots indicate smaller regions of high precipitation.

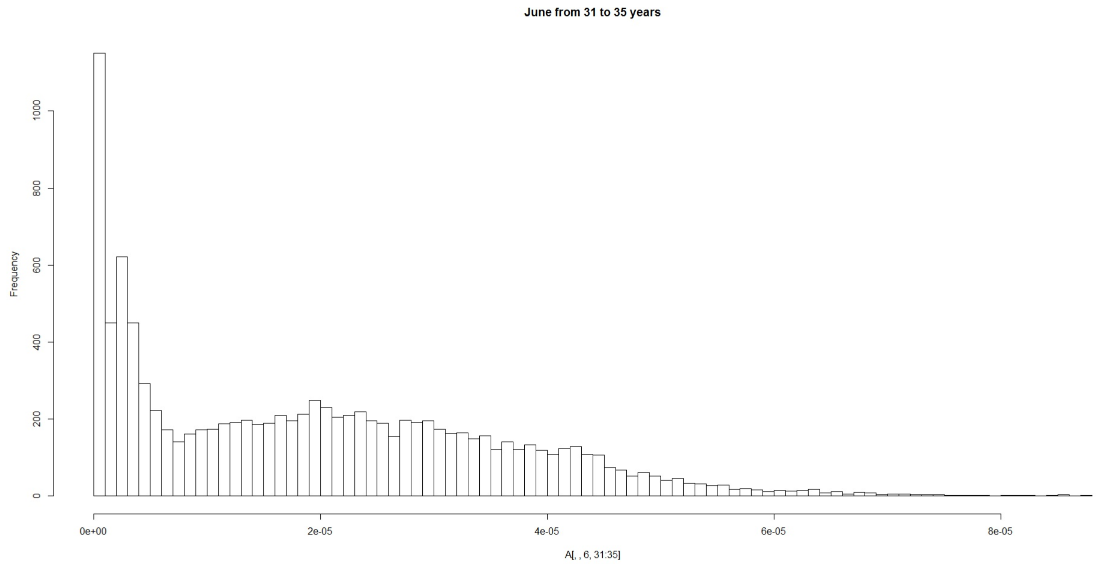

Figure 4

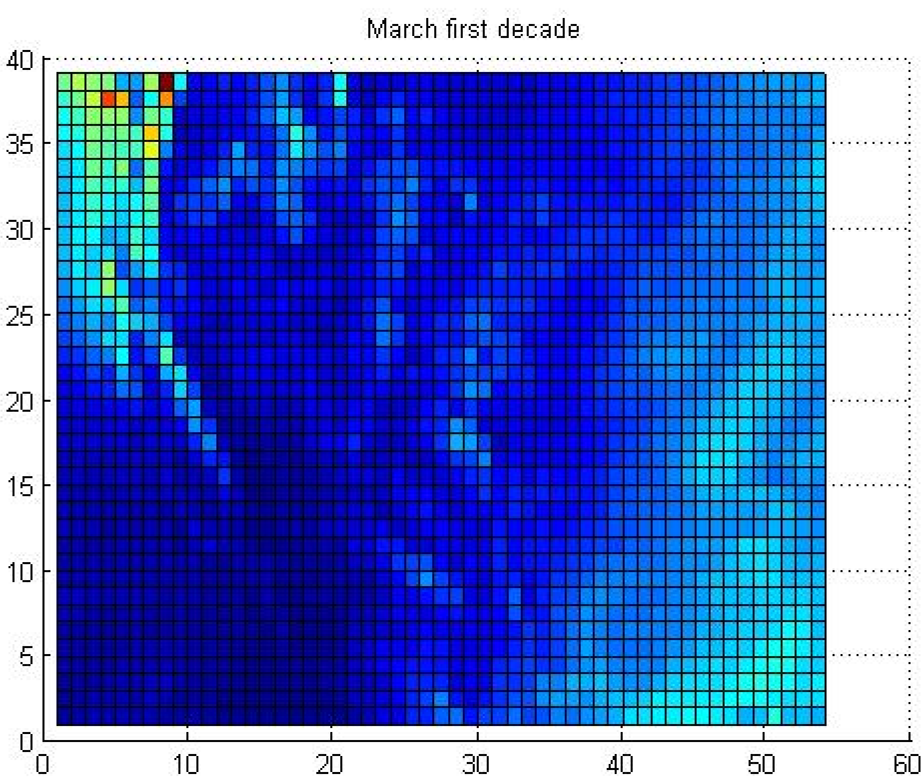

Figure 5

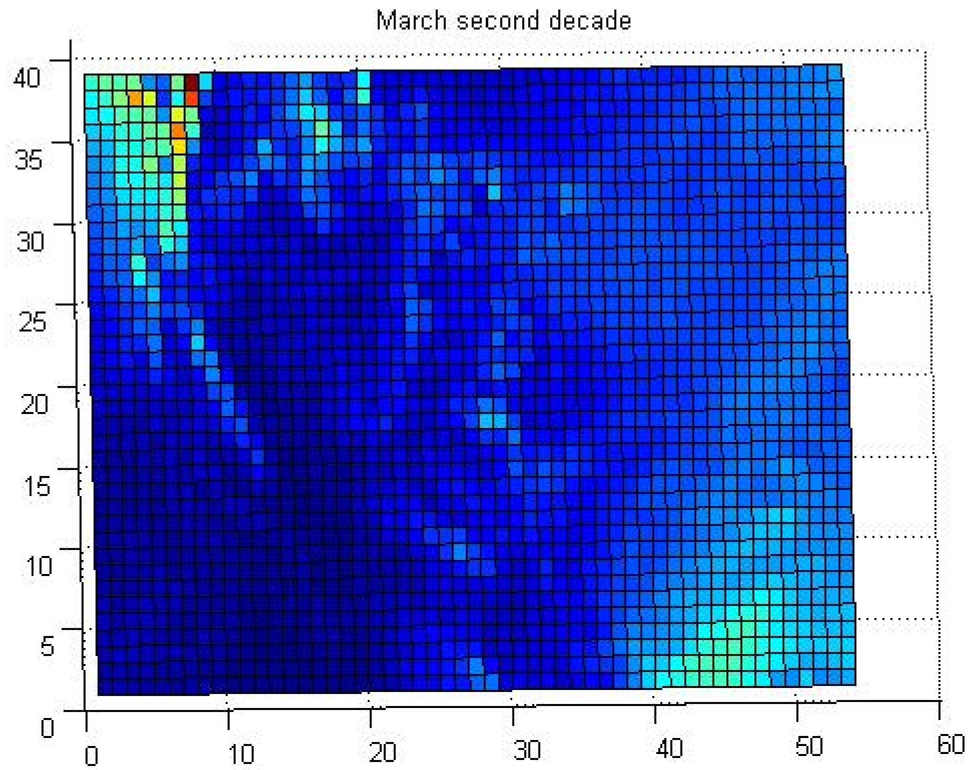

Figure 6

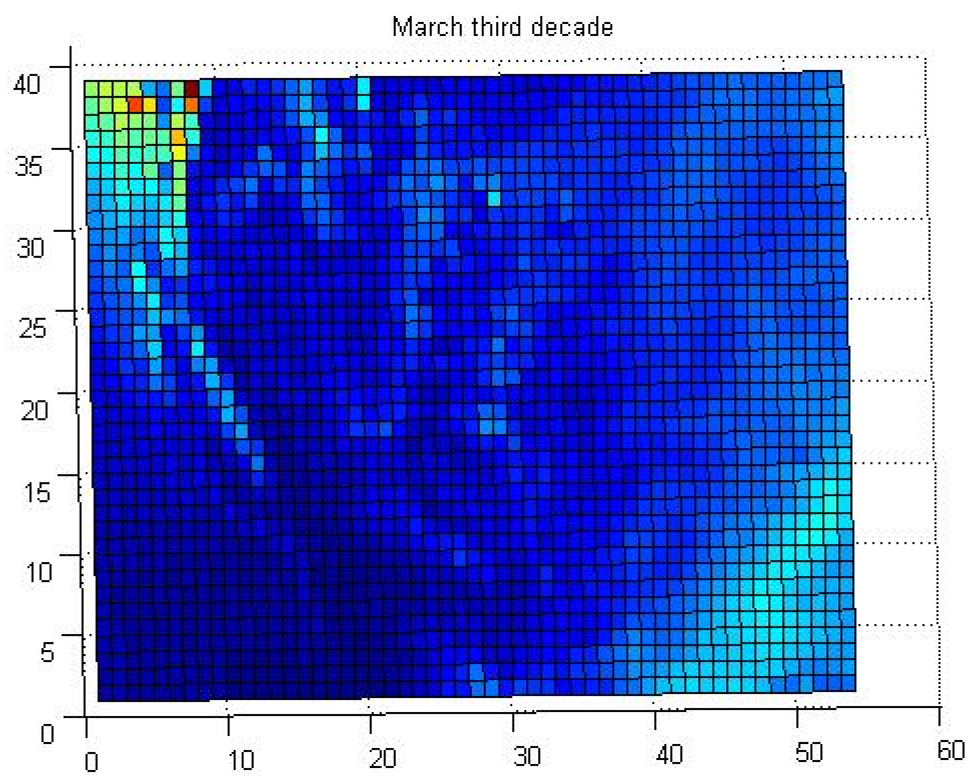

Figure 7

In order to support our assumption of stationarity, we did precipitation spatial analysis. Firstly we use again the data from western U.S. to generate the every pixel's average of precipitation of a certain month over the first decade, the second decade, and the third decade. We also generated the data from 1979 to 1997 and from 1998 to 2013. Then we created 3-D graphs to see if the graphs can support out stationarity assumption. From the figures, we can clearly see the graphs of the average of precipitation in January over different time periods are very similar.

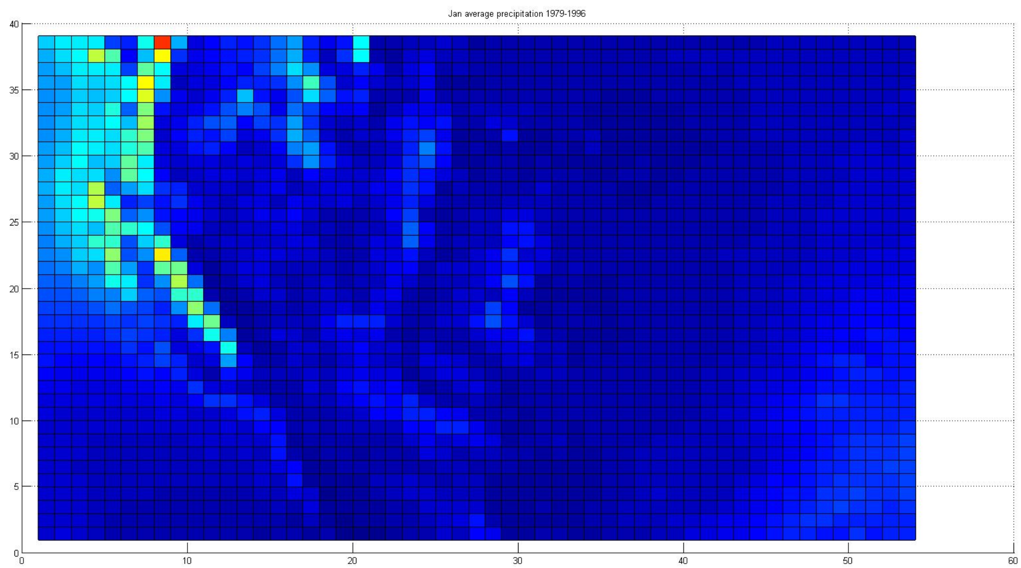

Figure 8

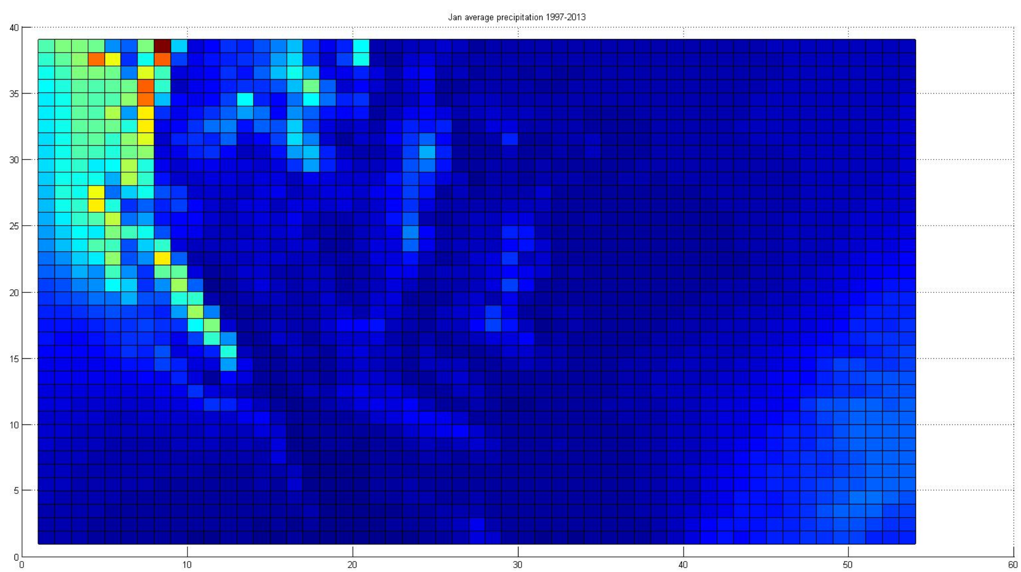

Figure 9

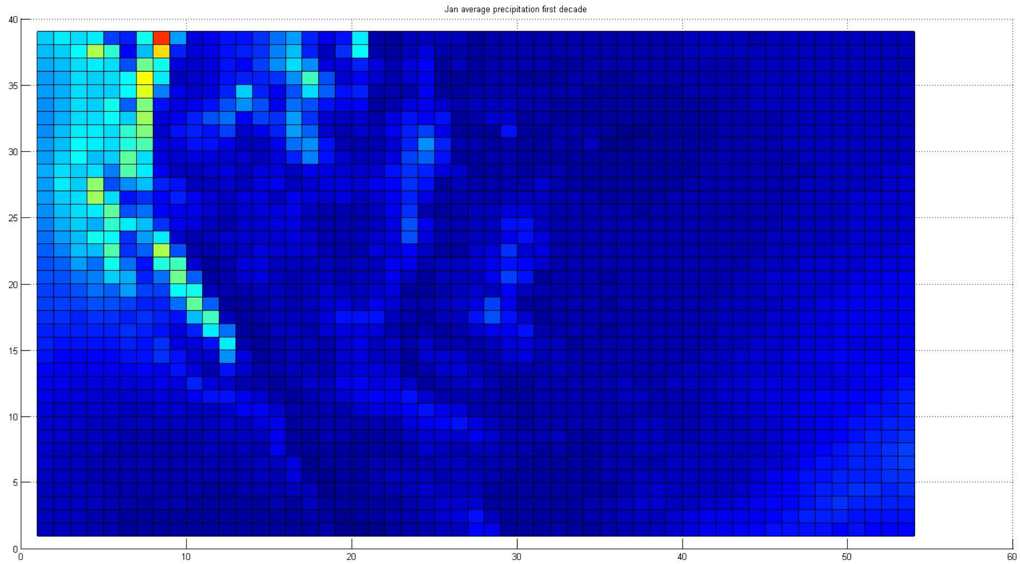

Figure 10

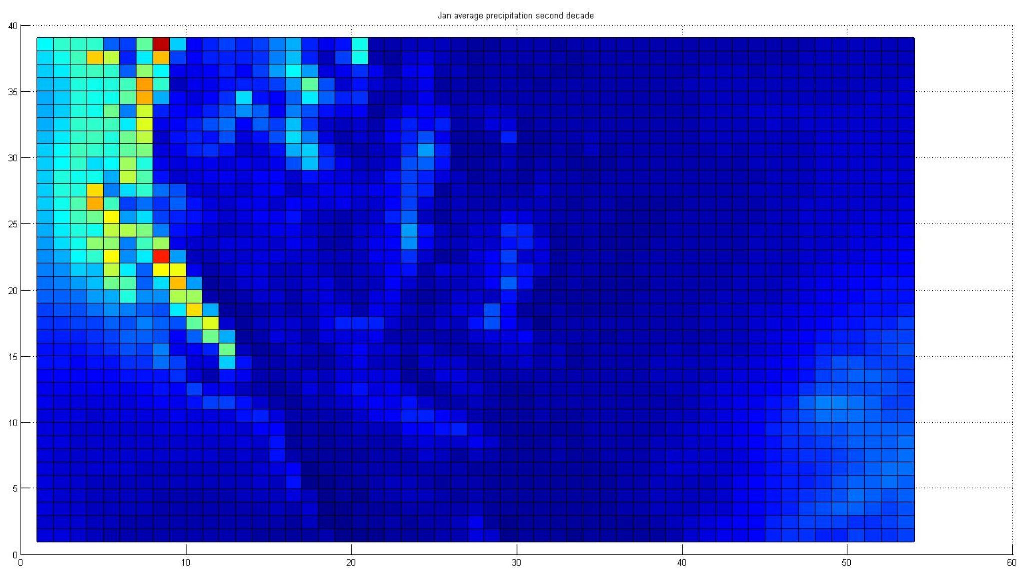

Figure 11

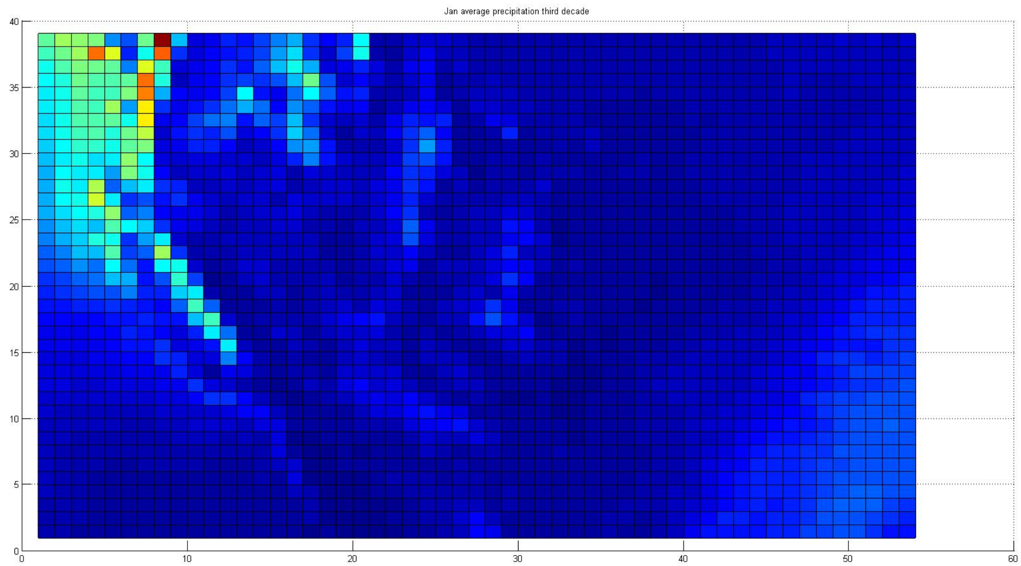

Figure 12

However, in the northwestern area, there is a significant change. Situations are different from month to month. In August, the northwest remains the same, while the south and the southeast areas happen to have some significant changes.

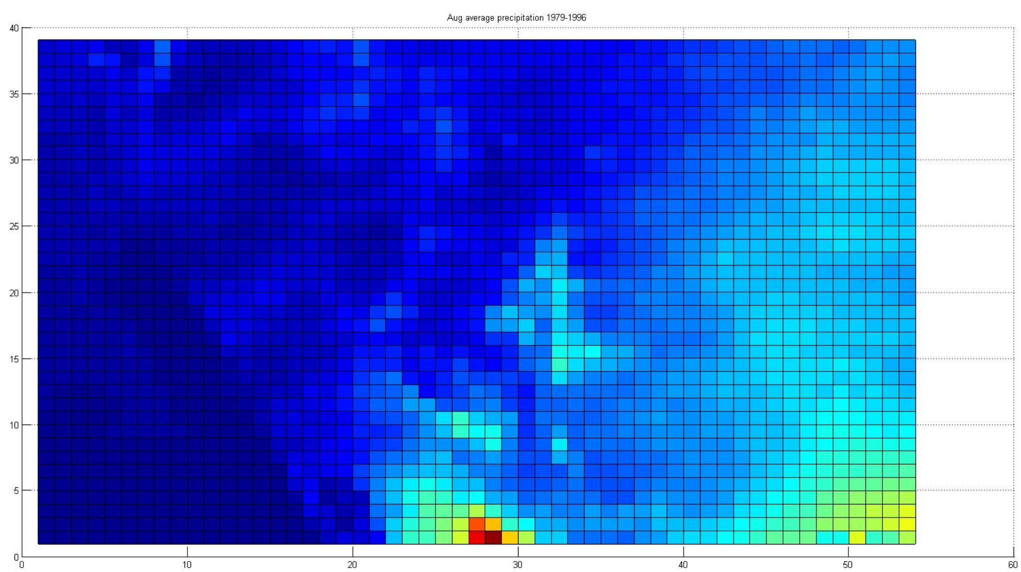

Figure 13

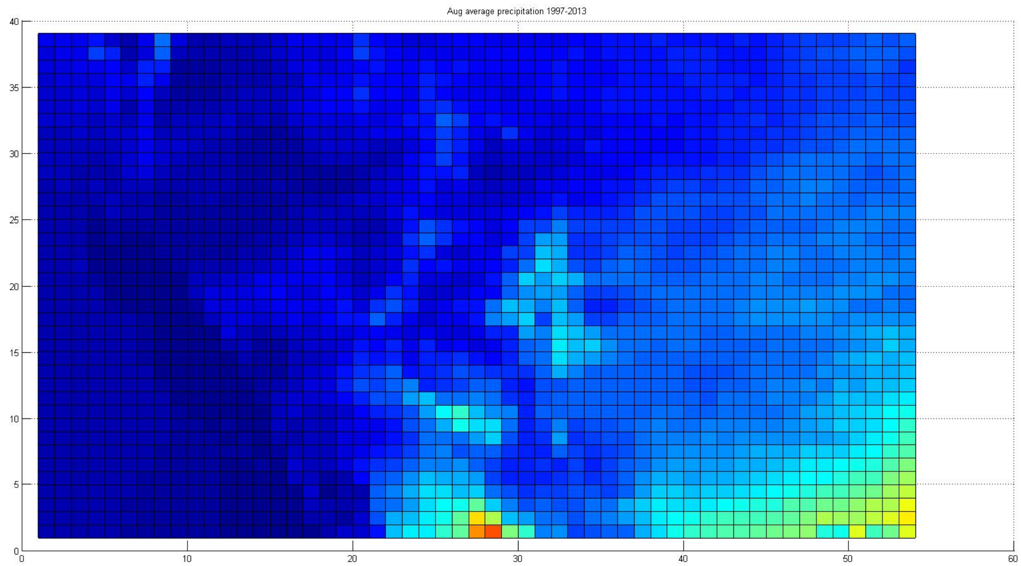

Figure 14

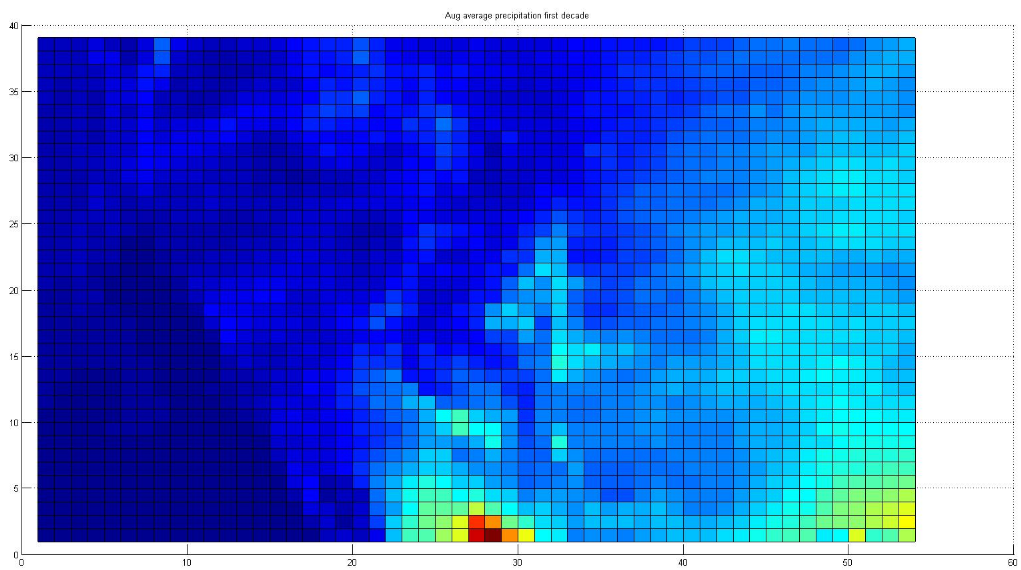

Figure 15

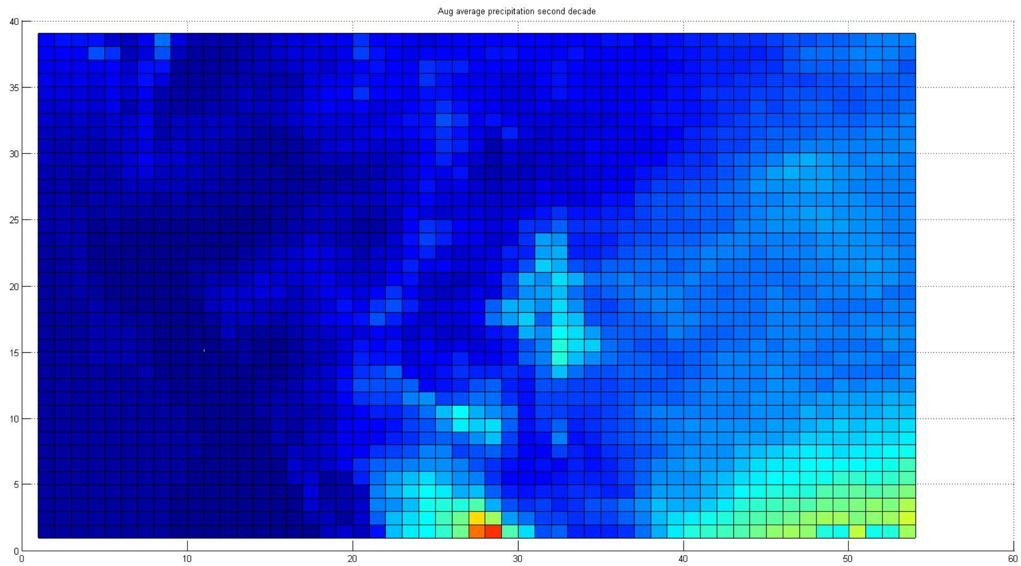

Figure 16

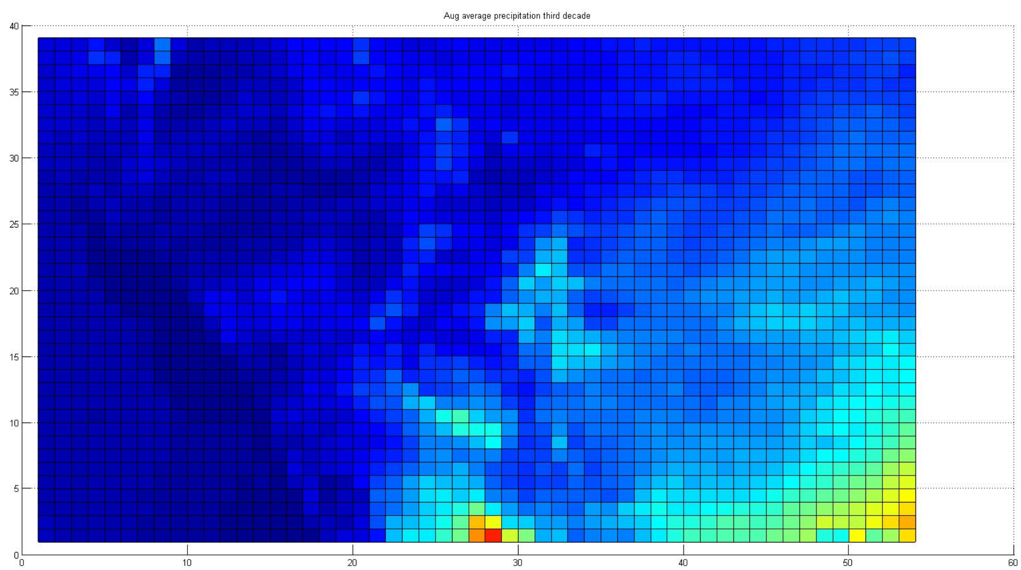

Figure 17

Furthermore, we calculated and graphed the relative difference of each pixel between different time periods, trying to find the trend of the change. In January, the precipitation of the northwest area increases overtime, while other areas roughly remain the same.

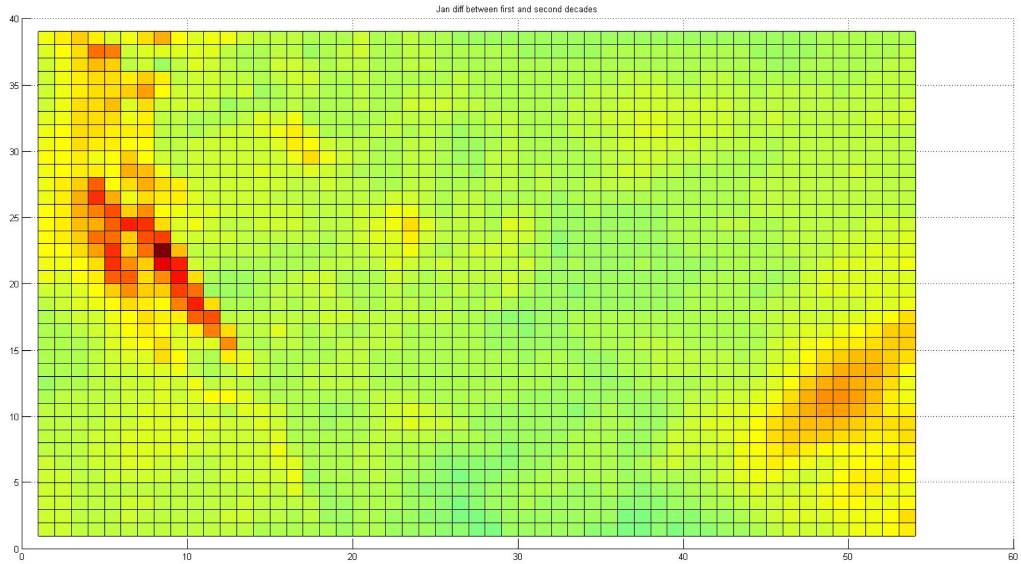

Figure 18

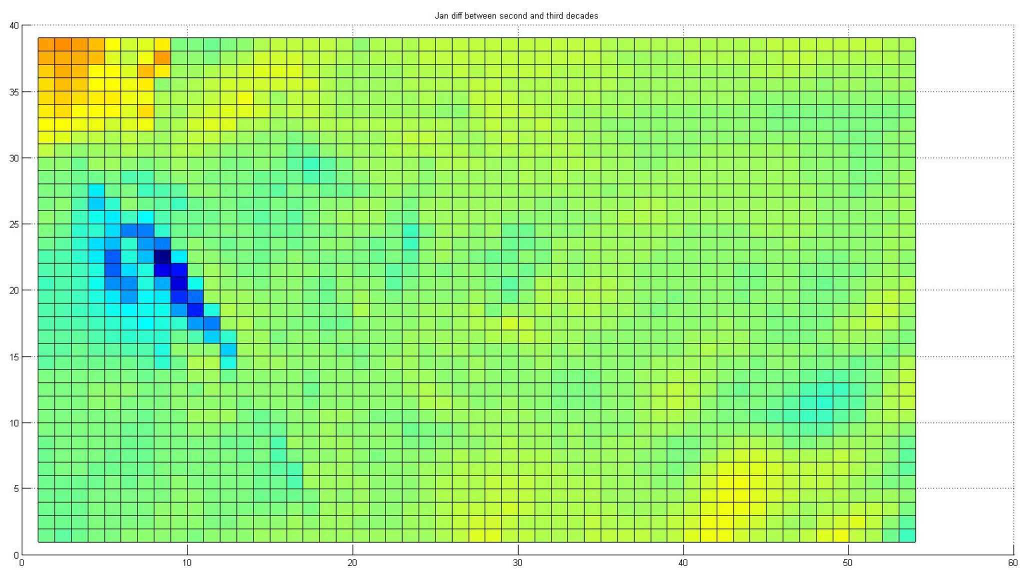

Figure 19

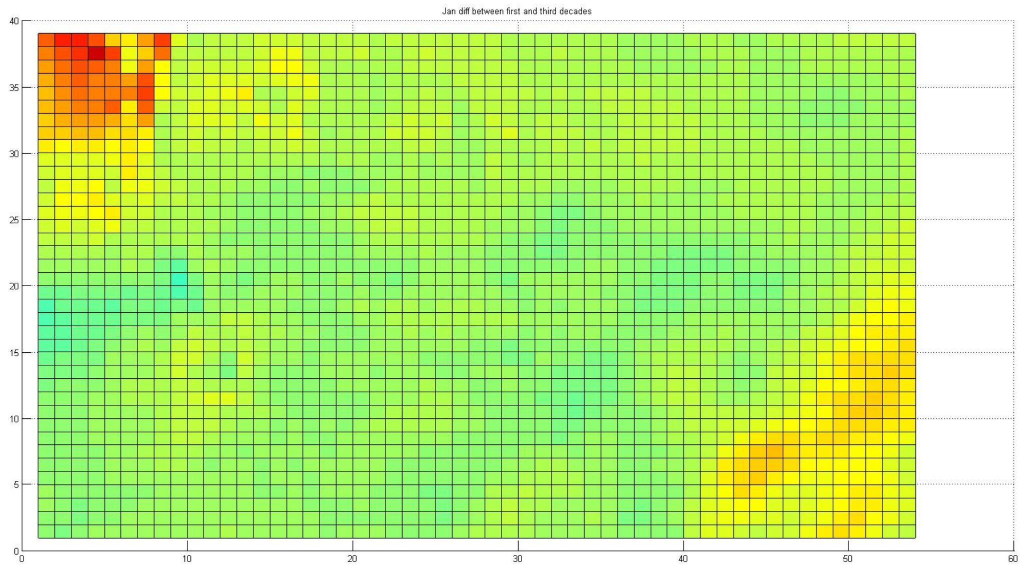

Figure 20

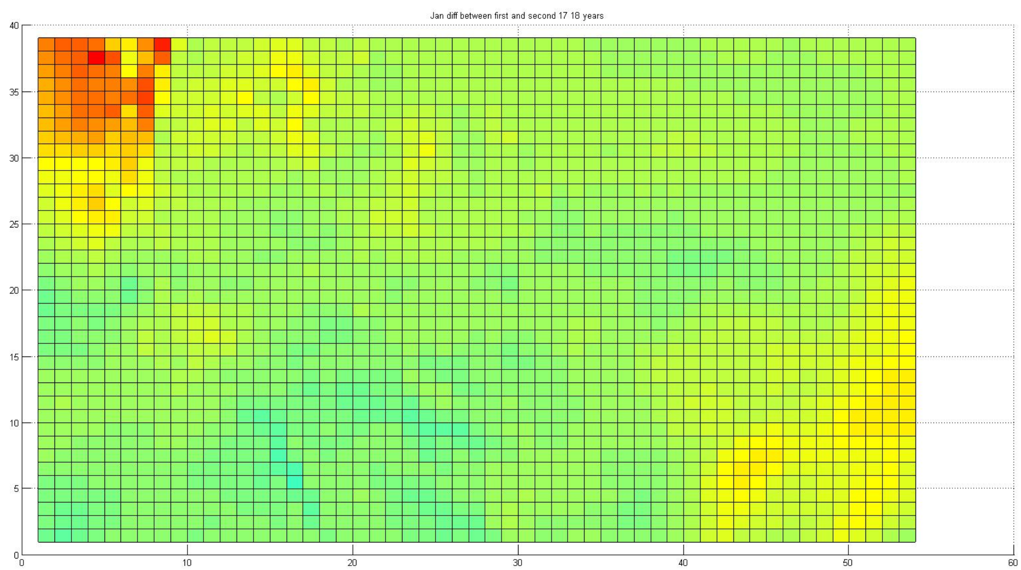

Figure 21

Then in August, most of the areas remain no difference, however, in the south and southeast area, difference changes dramatically. The observation contradicts the assumption of stationarity. However, after graphed the yearly average precipitation over three decades, we can clearly tell from the graphs that the precipitation is relatively stationary.

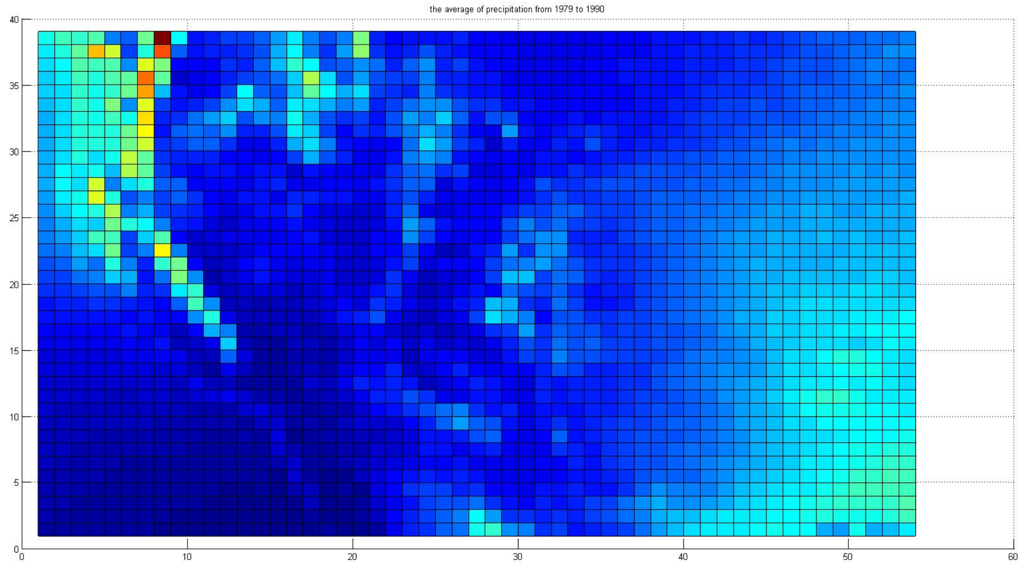

Figure 22

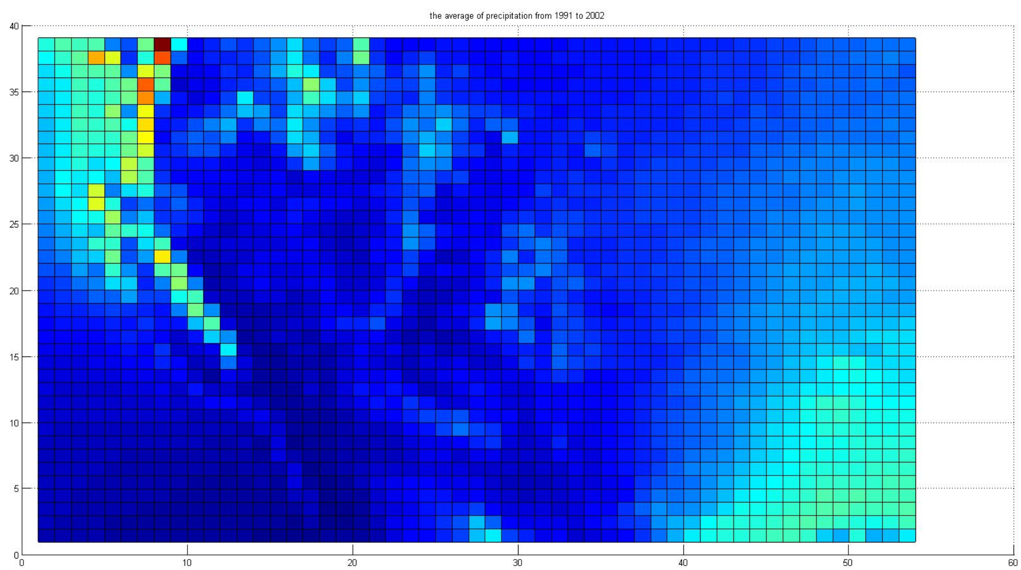

Figure 23

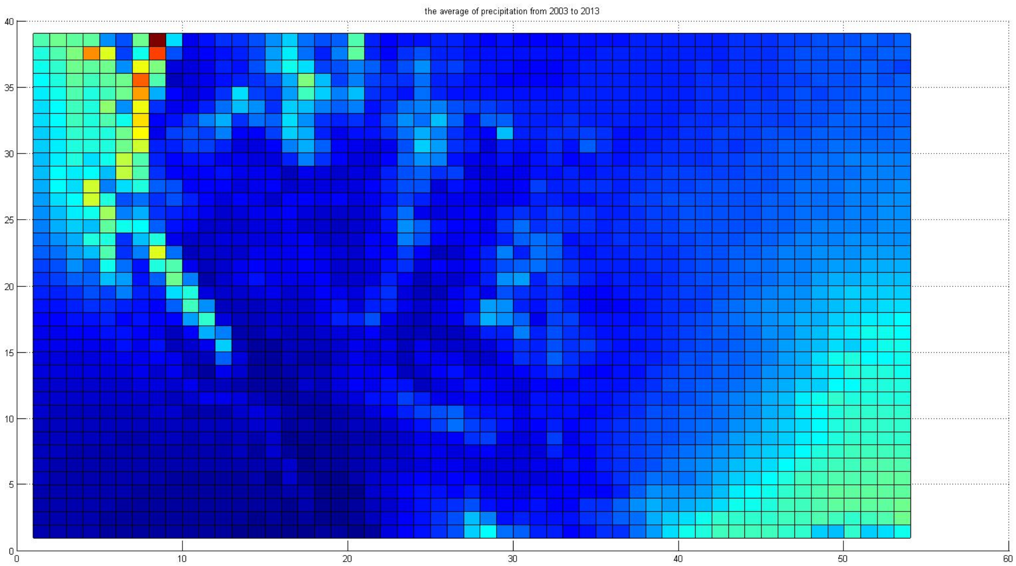

Figure 24

The reason for this contradiction is that since the value of precipitation is around 10^-4, the difference of each point is around 10^-5. Due to the scaling, the graphs of difference do not show the stationarity because a slightly change in the 10^-5 will cause a huge change in the graph. We can reduce the oscillation by using yearly average and the graphs proved it (Figure 22-24). Also it is important to note that only one variable was used to test the stationarity. The result may be violated by the change of other variables. Therefore, in order to increase the level of accuracy, we should apply more variables to see if the graph is more consistent. Overall, these graphs reinforce our assumption.

The second aspect of the stationarity analysis was to check the temporal patterns of precipitation. Two kinds of analysis were done in this regard i.e. monthly and yearly analysis. In the monthly case, the precipitations on a monthly basis were considered with for entire 35 years. Comparison was done in a 12 year long period except for the last one where 11 years were considered. From the monthly case it can be seen that the precipitation pattern varied for some months. In Figure 25 the precipitation for month of February is plotted. In the x axis the actual amount of precipitation can be seen and in y axis the frequency of particular precipitation can be seen. It can be seen that there are some variations in the pattern indicating a non-stationary trend. The differences are made further visible in Figure 26 where the precipitation has been put in logarithmic scale. Similar variation can also be seen for month of September in Figure 27 and Figure 28.

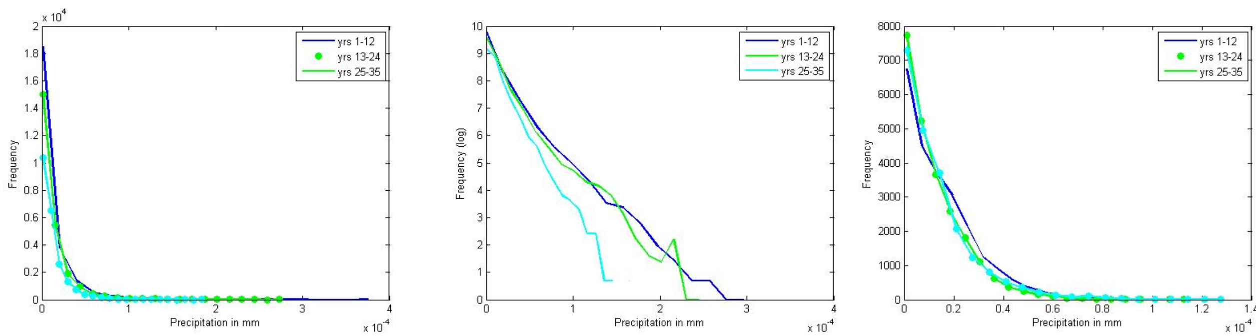

Figure 25-27

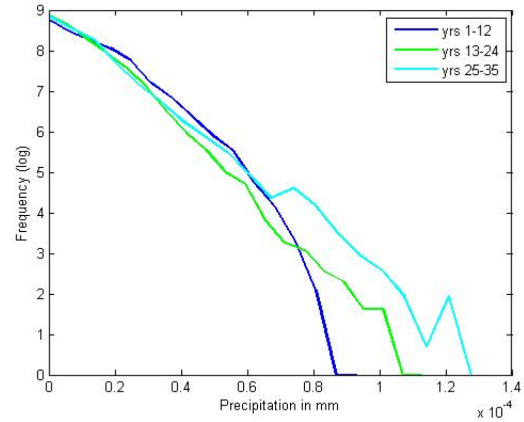

Figure 28

After the monthly trend analysis, the yearly trends are also observed as can be seen in Figure [14], here the precipitation trends resemble very close to each other. From the overall analysis it can be said although some inconsistencies are found in the monthly analysis the overall precipitation pattern is stationary.

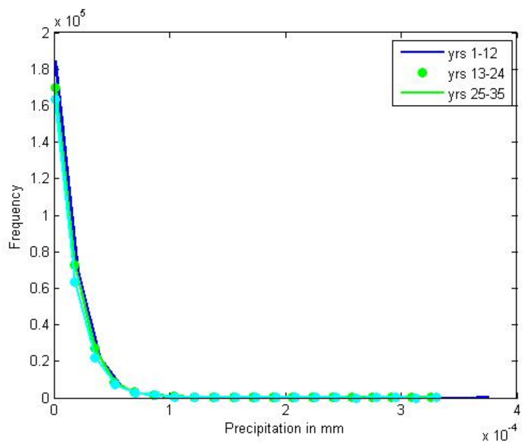

Figure 29

5.1.1 Cluster Analysis

The first step the cluster analysis was to determine the number and size of each drought cluster, and assign an appropriate value to each cluster identified. The number of spatially continuous droughts, as well as the number of pixels in each cluster was identified. Comparing the graphs of drought and the graphs of average precipitation, we can see the drought areas all within low precipitation areas, but there is not much similarity between each pair of graphs. This means the low precipitation does not represent drought. This does not contradict out assumption since we define drought is lower than 10th percentile. Low precipitation is the sufficient condition of drought, but not the necessary condition.

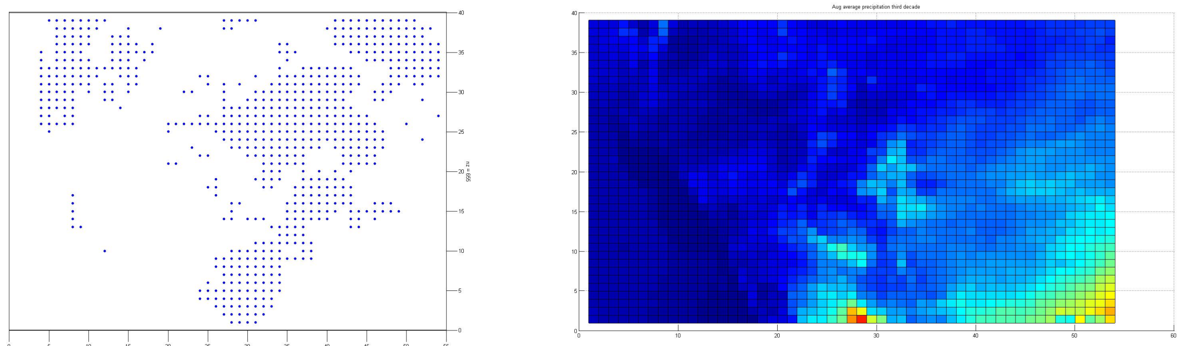

Figure 25-26

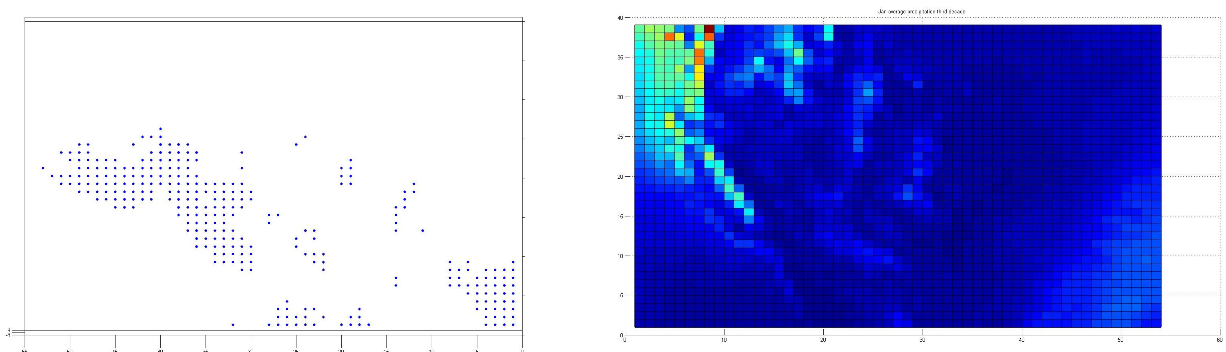

Figure 27-28

5.2 Correlation Analysis

5.2.1 Correlation Length Scale

Fix a year, and formulate a vector V^* indicating if the place experienced drought for a given month (there are 12 months) i.e.

$$V^* = (v^1, ..., v^{12}) \tag{1}$$

where V^* is .... and v^i is .... for i = 1...12.

For any particular pixel, we have 8 nearest neighbouring pixels. For each neighbour we will construct the same vector indicating the presence or absence of drought for a particular month. There will be 8 such vectors,

$$V^i = (v^i_1, ..., v^i_{12})$$

for i = 1..8

(2)

For each of the neighbours, we will calculate the correlation with neighbours. We will compute,

$$[r^j, p^j] = corr(V^*, V^j)$$

where j = 1,..8

(3)

where r^j is the correlation of V^* to V^j and p^j is the confidence.

The value of $r^j \leq 1.0$ will indicate strong correlation having $p^j \leq 0.05$, from the values of r^j obtain $r^* = \sum r^j/8$.

5.2.2 Results

The main objective in the correlation analysis was to use the Drought Indicator Data over a region in order to observe patterns or the lack thereof that show up. The region of interest was chosen to be the Western US, due to the frequency of droughts in the area. For each pixel, the correlation between itself and its eight surrounding neighbors was found for each time instance. The average of these correlation values was assigned to the pixel. The result is a shaded grid of the region that shows how well each pixel is correlated to each of its neighbors (Fig. 29 shows the correlation between each pixel and its neighbors for 1979). Note that each pixel was a matrix and columns/rows consisting of all zeroes were given a value of zero since there was nothing to correlate. Dark red indicates a strong correlation and blue indicates a weak or no correlation. This grid represents the correlations for one year.

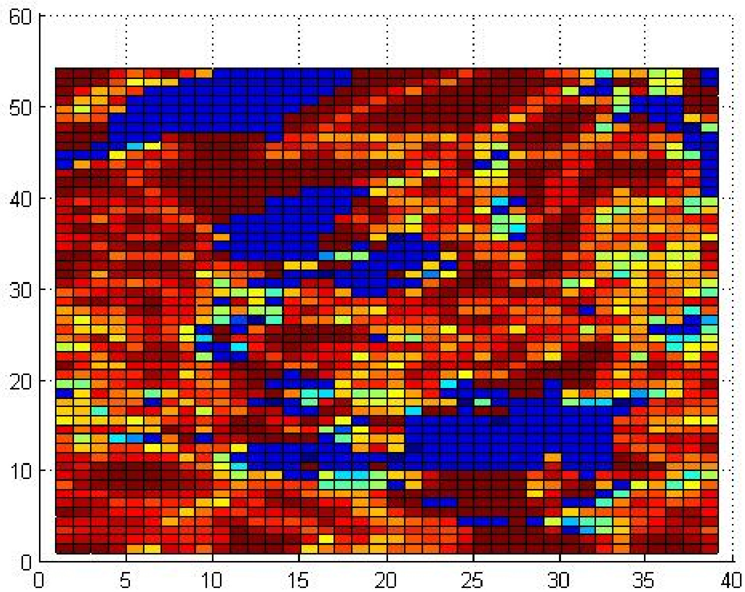

Figure 29

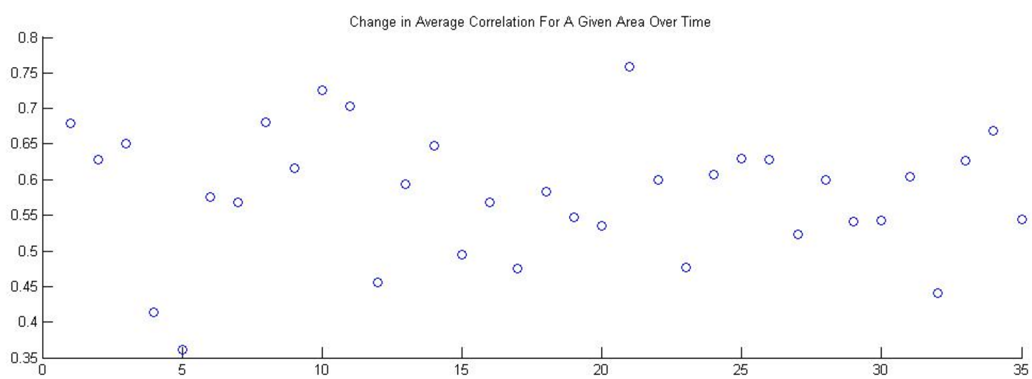

Figure 30

We then observed how these correlations changed over the 35 year period. The correlation values were averaged for each grid, and that value was assigned to the grid for that year. This process was repeated for all years and plotted (Fig. 2). This graph looked as though it had some noise, but it was an interesting result. This plot may suggest that stationarity does not hold true, at least in terms of patterns. However, this plot looked at the averages year by year. This may be the reason for such an irregularly shaped plot, as a more linearly shaped plot was expected. Given this, rather than considering each year for all 35 years, the correlation values for the first 18 years and the last 17 years were averaged (Fig. 3). These values were compared, and it was observed that they are relatively close (.579 and .58125 respectively). It may be the case that data for a larger time period is needed so that we can look at more subintervals of similar length.

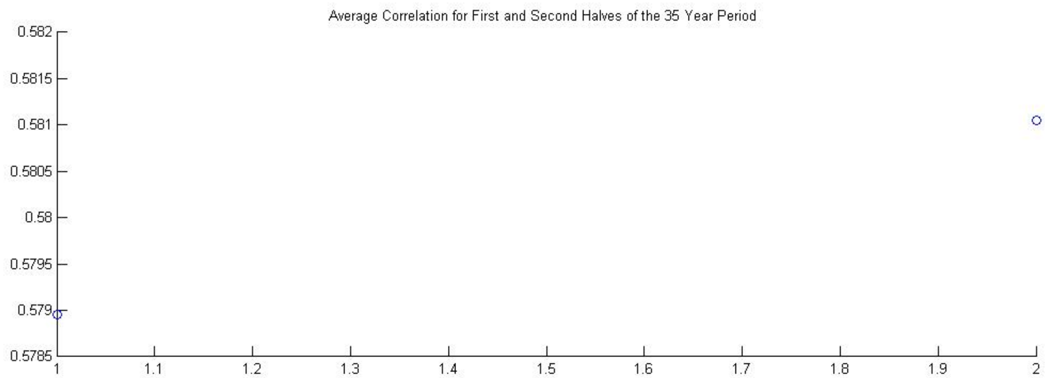

Figure 31

6 Prediction of Drought Areas

6.1 The 10^{th} Percentile Drought Identification Method

First, the data corresponding to the United States was extracted. We began with a basic estimate, by defining a **drought** to be the lowest 10% of precipitation in an a given area. Given this, the lowest 10th percentile of the total amount of precipitation over all time was identified. In order to determine drought in one location, we began with a specific location a **pixel**; that is to say, a particular coordinate for which the area of a quadrangle was bounded by a longitude of length 0.667 degrees and latitude of 0.5 degrees . For that location one must then identify which months out of the entire 35 years, where we have precipitation that falls in the lower ten percent of the average value. This will give us a matrix M for each pixel location that confirms the existence or non-existence of a drought for a given month in a given year. We identify the months that experience lower than ten percent of average precipitation values by a "1" (indicating existence of drought), and use "0" to denote greater amounts of precipitation (indicating that there is not a drought). We have $M_{x,y} = (m_{i,j})_{12\times 35}$ where x and y denote latitude and longitude respectively, $m_{i,j} \in \{0,1\}$, $i = 1, ..., 35$ years and $j = 1, ..., 12$ months.

$$M_{x,y} = \begin{pmatrix} m_{1,1} & m_{1,2} & m_{1,3} & ... & m_{1,34} & m_{1,35} \\ m_{2,1} & m_{2,2} & m_{2,3} & ... & m_{2,34} & m_{2,35} \\ m_{3,1} & m_{3,2} & m_{3,3} & ... & m_{3,34} & m_{3,35} \\ ... & ... & ... & ... & ... & ... \\ m_{12,1} & m_{12,2} & m_{12,3} & ... & m_{12,34} & m_{12,35} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & ... & 0 & 0 \\ 0 & 1 & 0 & ... & 0 & 1 \\ 1 & 0 & 1 & 1 & ... & 0 \\ ... & ... & ... & ... & ... & ... \\ 1 & ... & 0 & 0 & 0 & 1 \end{pmatrix}$$

These matrices can be used as indicators to see if changes in drought changed with respect to time and space. .

6.1.1 Precipitation Indicator

One of the newer indices currently used by the National Drought Mitigation Centre to measure moisture supply conditions is called the Standardized Precipitation Index. It is based on the probability of precipitation for any time scale, usually calculated up to a 3-year period. SPI is the number of standard deviations that an observed value deviates from its long-term mean, for a normally distributed random variable. This index is a good indicator of meteorological drought i.e. one that experiences a long term lack in precipitation. This index not only allows users to gain an understanding of the severity of the drought and aid in predicting an oncoming drought, but also identifies emerging droughts much sooner than any of the other indices. Given this, upon evaluation of the evolution of the drought over time, we have designed a probability indicator in order to predict areas that are more likely to have a drought. The indicator works in much the same way as the Standardized Precipitation Index, but is calculated on a much longer time scale than the index, which is usually calculated for at most a 3 year period. The probability indicator is calculated by taking the monthly average at each location or pixel for the 35 year span. We average the indicator for the first three decades. After scaling the color appropriately for display comparison purposes, we were able to plot the month of August, and were able to identify some patterns over the first and last decade. The indicator uses the percentile idea previously presented using only the precipitation variable. To make this comparison, we define the following two matrices:

$$Hp_i^{(1)}(x, y) = \frac{1}{8}\sum_{j=1}^{18}(M_{x,y})_{i,j}$$

$$Hp_i^{(2)}(x, y) = \frac{1}{7}\sum_{j=19}^{35}(M_{x,y})_{i,j}$$

$Hp_i^{(1)}(x, y)$ gives a likelihood that for the first 18 years, pixel (x, y) experienced a drought during the jth month, and $Hp_i^{(1)}(x, y)$ gives the same information but for the last 17 years. We found that for the Western USA, $|\max(Hp_i^{(1)} - Hp_i^{(2)})| = 0.0672$ happened in the month of November and $|\min(Hp_i^{(1)} - Hp_i^{(2)})| = 0.0362$ happened in April. This seems to be a relatively large difference, and hence we wouldn't be able to support stationarity of drought only based on precipitation. To have an idea on how the patterns changed over space, we plotted $Hp_i^{(1)}(x, y) \geq 0.17$ versus $Hp_i^{(2)}(x, y) \geq 0.17$. This gave some very interesting results since the recurrent pattern was that if a pixel was identified as being very likely to be in a drought for the first 18 years, it would not be likely to be in a drought for the next year span. This also would not let us support stationarity.

6.1.2 Soil Moisture

We decided to see if another variable would give the same conclusion, so we did a similar analysis using soil moisture. We were hoping that soil moisture would capture a different concept of drought since it contains information highly correlated to the previous time step. Using the same percentile idea as above, define the indicator analogous to $M_{x,y}$ in the following manner:

$$SM_{x,y} = (sm_{i,j})_{12\times 35}$$

Using this SM indicator, the likelihood for a pixel to be in drought in the first 18 years of our data is defined as:

$$Hsm_j^{(1)} = \frac{1}{8}\sum_{i=1}^{18}(M_{x,y})_{i,j}$$

Similarly, for the last 17 years:

$$Hsm_j^{(2)} = \frac{1}{7}\sum_{i=19}^{35}(M_{x,y})_{i,j}$$

From these two, we found that $|\max(Hsm_i^{(1)} - Hsm_i^{(2)})| = 0.2438$ happened in the month of September and $|\min(Hsm_i^{(1)} - Hsm_i^{(2)})| = 0.0464$ in the month of October. As above, this seems to be very large magnitudes and hence we cannot support the stationarity hypothesis. Using the same thresholds, $Hp_i^{(1)}(x, y) \geq 0.17$ versus $Hp_i^{(2)}(x, y) \geq 0.17$, we plotted the Western USA and although there seem to be more instances when a pixel had a high likelihood of drought for the first 18yrs and in the last 17 years, it didn't seem to be that many. And the ones with most overlap were pixels representing the ocean. Only few had many overlaps inland. Hence, with our data, we couldn't support stationarity only using soil moisture as an indicator.

6.1.3 Deficit Indicator

We have also designed a deficit indicator that illustrates how the amount of precipitation each month changes in relation to the monthly average. Analysis using this indicator gives a better understanding of how exactly the patterns of precipitation can change, which allows us to better understand the stationarity assumptions inherent in the current drought prediction and identification systems. The indicator d_k is calculated,

$$d_k = p_k - \bar{p}_m + \alpha d_{k-1} = \sum_{n=1}^{k} \alpha^{k-n}(p_n - \bar{p}_m) \tag{4}$$

where $d_o = 0$, d_k for $k = 1, 2, \cdots, 420$ is a measure of how far away the actual level of precipitation p_k for time k at location (x,y) is from the monthly average p_m, and α is an impact measure of how much effect last months deficit effects the current months. Note that $k = 12(j) + i$, and for our experiments, we chose $\alpha = 0.5$.

Using this idea, we can draw a threshold and find a binary matrix analogous to M in the following way:

$$D_{x,y}^{(i,j)} = \{ \text{ 1if } d_{12(j)+i} \leq 10 \text{ percentile for the } i\text{th month0otherwise}$$

Hence, we can define analogous matrices to Hp_j in the same manner:

$$Hd_j^{(1)} = \frac{1}{8}\sum_{i=1}^{18}(D_{x,y})^{(i,j)}$$

$$Hd_j^{(2)} = \frac{1}{8}\sum_{i=19}^{35}(D_{x,y})^{(i,j)}$$

Again, the largest magnitude of the difference of these is $|\max(Hd_i^{(1)} - Hd_i^{(2)})| = 0.3249$ and the smallest is $|\min(Hd_i^{(1)} - Hd_i^{(2)})| = 0$. This doesn't seem to support stationarity either, but the thresholding at $Hd_i^{(1)}(x,y) \geq 0.17$ versus $Hd_i^{(2)}(x,y) \geq 0.0004$, seems to show that if a pixel was likely to be in drought for the first 18 years, it would also be the case for the last year span.

6.2 Verification

Once the potential drought prone areas are identified using the precipitation indicator, the result was compared with the national drought monitor data made available in the web in order to properly verify our results. Comparison with the Drought Monitor showed that our results differ slightly for a finite number of time instances. For example, in June 1978, though a small drought in New Mexico was registered by the monitor, our results show no such occurance. These types of differences were expected, due to the differences in definition and variables used to determine where and when there was a drought. As outlined earlier, the Drought Monitor uses several identifying variables, combined with individual reporting to develop their maps.

7 Future Work

Given the limitations of specialized drought indicators such as the PDSI, the newest drought indicies that are being developed with the idea being to incorportate as much information as made available. A measure of drought that incorporates both precipitation and soil moisture, for example, would be a better indicator than one that uses precipitation alone. Combining multiple drought indicators would both maximize the amount of information used out of that made available, as well as increase the accuracy of any conclusions or predictions made. Another point of interest for future work is to consider the nature of the datasets used in drought modelling. Drought modelling and simulation use large multidimensional data sets that include information on variables such as precipitation, temperature, groundwater, water storage and supply, agriculture, runoff, soil moisture, pressure, and irrigation on a global scale. For example the NASA-MERRA dataset used includes global montly data taken over 35 years on all of these variables as well as evaporation, water vapor and moisture convergence. Due to the large size of such datasets, it would be useful to develop a method of compressing or reducing the order. This would make calculations and analysis more practical as well as speed up and simplify computations, also making more advanced analysis possible. Using a CVT or SVD data compressing technique outlined below would be a useful tool in achieving this. A further extension of the work done here would be to use the clustering analysis and the work done on the probability indicator to develop a Drought Risk Index which would take into account both the likelihood of a drought occurring in a given area, as well as how severe such a drought might be for a given target area. Such an index would be able to give decision makers a better idea of how much a given region is vulnerable to drought. The accuracy of such an index would be improved by combining multiple variables such as precipitation, soil moisture, and convergence. It is also important to note that much of the work done was based on a simplistic definition of what a drought is, and so the next step in the analysis process would be to allow for droughts of varying severity.

8 Conclusion

Due to the intensely multifaceted nature of droughts, there are many interconnected variables at play when it comes to predicting or even simply attempting to identify which areas have suffered from a drought. Given this, it is incredibly difficult to predict or identify them, and new methods and drought indicators are constantly being created, building new technologies and ideas upon the old ones. Despite this, it is still the case that most national organizations still rely on the additional support from farmers to notify them of any problems occurring. The purpose of this study was to draw upon current temporal-spatial analysis techniques in order to analyse one of the key assumptions inherent in most current drought indicators and predictors, that of stationarity, with the hopes of building a foundation towards to creation of new methods to identify and predict droughts. One of the most important aspects in the development process is in being able to understand what is or is not an accurate descriptor of droughts in the current methods, and what sort of evolution occurs over time in drought patterns. For if even a slight evolution or change occurs over any length of time, short or long, the methods currently in use will become less and less accurate as time goes by.

9 References

- National Drought Mitigation Centre. ¡http://drought.unl.edu/DroughtBasics.aspx¿
- "Global Integrated Drought Monitoring and Prediction System (GIDMaPS)." Hydroclimate Research Lab and Center for Hydrology and Remote Sensing. University of California, Irvine.