# IIIT Hyderabad

COMPUTER SCIENCE AND ENGINEERING

PROJECT

## Prediction of probability of risk given precipitation and temperature data

Prof. Dr. Rehana Shaik

---

*Contributor:*
Sagnik Gupta
Roll No.: 2019201003
M.Tech CSE

# 1 ABSTRACT

Due to rises in temperature and abrupt changes in rainfall in different regions, calculating the probability of a region falling under the risk zone is of utmost importance. In this project, we aim to calculate the probability of risk in different regions in India, based on their historical data of temperature and drought index values over the years between 1951 - 2015. We do so using different combinations of drought and temperature values that are present to us.

# 2 INTRODUCTION

Given the Precipitation data (P), and temperature data (T), the probability of risk can be calculated using Multiple Logistic Regression. The general form of multiple logistic regression (MLR) with two explanatory variables to obtain the predicted value of the response is given as

$$P(Risk/P,T) = e^{(\beta_0+\beta_1*P+\beta_2*T)/(1+e^{\beta_0+\beta_1*P+\beta_2*T})} \qquad (Eq.1)$$

where P(Risk/P,T) is the probability of occurrence of risk, conditioned on precipitation, P and temperature, T. We apply this method to obtain the required probability. The $\beta$ coefficients are estimated from the historical data.
The probability of occurrence of risk conditioned on precipitation and temperature, P(Risk/P,T), obtained from the MLR model can be further used to determine the likelihood of risk, as follows.

$$P(Risk) = \sum_{i=1}^{n} * \sum_{j=1}^{n} P(Risk/P,T) * P(P_i, T_j) \qquad (Eq.2)$$

where P(Risk) is the likelihood of risk for a given forecast of $P_i$ and $T_i$ and $P(Risk/P_i, T_j)$ is the conditional probability obtained from Eq. 1. $P(P_i, T_j)$ is the joint PDF of precipitation and temperature. This can be estimated using Copula methodology.

## 2.1 Copula Methodology

A Copula models the dependence between two extreme events in which the marginal distribution of each extreme event is uniform over (0,1). Such marginal distributions can be used to derive the joint probability distributions of various precipitation and temperature extreme events for a given return period. An n-dimensional copula (C) can be defined as the joint probability of univariate marginal distributions of F1, F2, .., Fn as follows:

$$H(x_1, x_2, .., x_n) = C[F_1(x_1), F_2(x_2), .., F_n(x_n)] \qquad (Eq.3)$$

$$H(x_1, x_2, .., x_n) = C[u_1, u_2, .., u_n] \qquad (Eq.4)$$

Where $u_1, u_2, .., u_n$ are cumulative distribution functions (CDFs) of $x_1, x_2, .., x_n$.

The following properties are valid :

- $C(u_1, .., u_d)$ is non-decreasing in each component, $u_i$.
- The $i^{th}$ marginal distribution is obtained by setting $u_j = 1$ for $j \neq i$ and since it it is uniformly distributed

$$C(1, ..., 1, u_i, 1, ..., 1) = u_i$$

- For $a_i \leq b_i$, $P(U_1 \in [a_1, b1], ..., U_d \in [a_d, b_d])$ must be non-negative. This implies the rectangle inequality

$$\sum_{i_1=1}^{2} ... \sum_{i_d=1}^{2} (-1)^{i_1+i_2+...+i_d} C(u_{1,i_1}, .., u_{d,i_d}) \geq 0$$

  where $u_{j,1} = a_j$ and $u_{j,2} = b_j$.

There are several types of Copula families. The joint probability function ($P(P_i, T_j)$) obtained from Copula methodology can be used in Eq. 2 to estimate the probability of risk for a given forecast of precipitation and temperature.

# 3 EXPERIMENTS

A number of experiments have been performed for measuring the probability of risk in different regions, given their historical data of drought index and temperature. Some of the experiments performed are listed below.

- Different thresholds of maximum temperature and drought index values have been used for the classification of risk. Drought indices are classified as moderate, severe and extreme drought based on SPEI values, and temperature have been classified with different percentile values.
  - Moderate drought (SPEI value between -1.0 and -1.49) and Tmax threshold 75 percentile
  - Moderate drought (SPEI value between -1.0 and -1.49) and Tmax threshold 80 percentile
  - Moderate drought (SPEI value between -1.0 and -1.49) and Tmax threshold 90 percentile

  - Moderate drought (SPEI value between -1.0 and -1.49) and Tmax threshold 95 percentile
  - Severe drought (SPEI value between -1.50 and -1.99) and Tmax threshold 75 percentile
  - Severe drought (SPEI value between -1.50 and -1.99) and Tmax threshold 80 percentile
  - Severe drought (SPEI value between -1.50 and -1.99) and Tmax threshold 90 percentile
  - Severe drought (SPEI value between -1.50 and -1.99) and Tmax threshold 95 percentile
  - Extreme drought (SPEI value less than -2.0) and Tmax threshold 75 percentile
  - Extreme drought (SPEI value less than -2.0) and Tmax threshold 80 percentile
  - Extreme drought (SPEI value less than -2.0) and Tmax threshold 90 percentile
  - Extreme drought (SPEI value less than -2.0) and Tmax threshold 95 percentile
- The risk has been estimated for two time scales :
  - Risk during Water year : June - May
  - Risk during monsoon season: June - September
- Trend analysis have been done on the risk values to test how it changes in the period of 1951 - 2015.

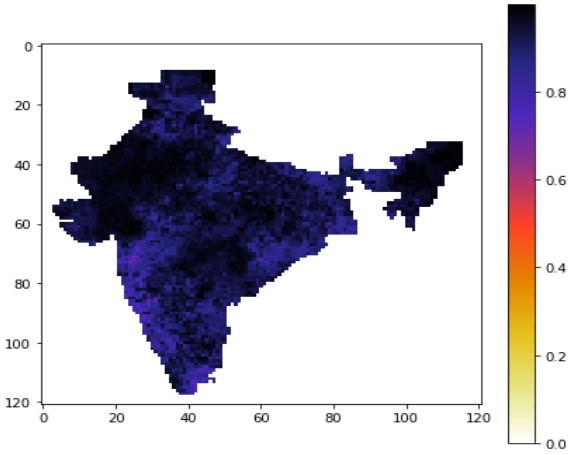
The results and observations on performing these experiments are provided in the following section.

# 4 RESULTS

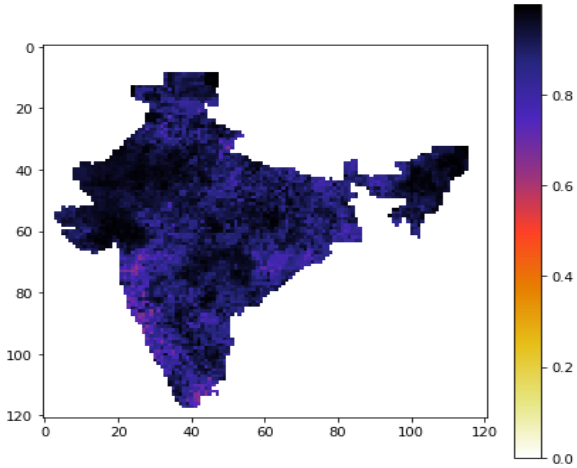
Following the above mentioned approach, the results obtained at different threshold of maximum temperature and drought index values are shown below. In the figures shown below, a higher shade of black signifies a higher probability of risk, whereas lighter shades signifies low risk regions.

## 4.1 Moderate drought conditions (SPEI values between -1.0 and -1.49)

### 4.1.1 Using a max temperature threshold of 75 percentile

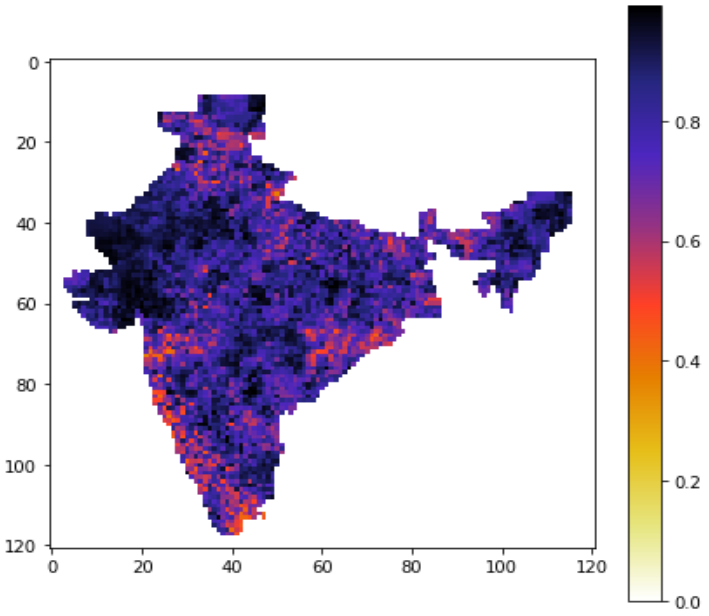



### 4.1.2 Using a max temperature threshold of 80 percentile



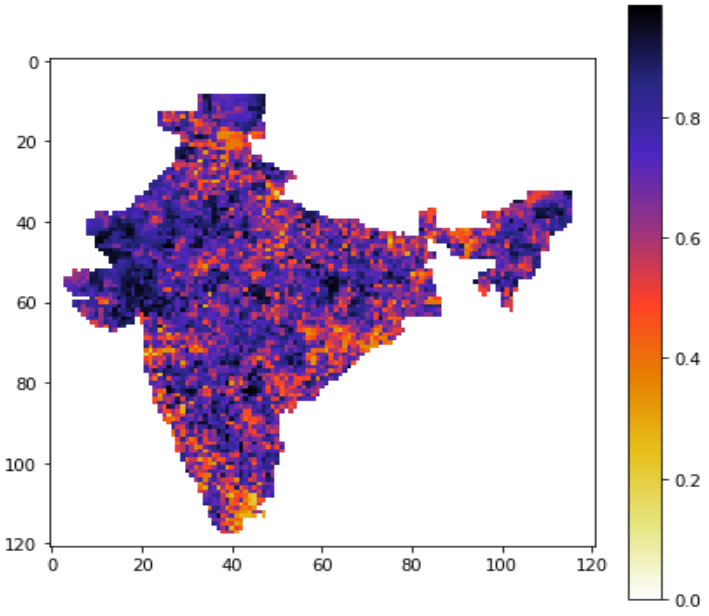
### 4.1.3 Using a max temperature threshold of 90 percentile

The probability of risk in different regions is shown below.



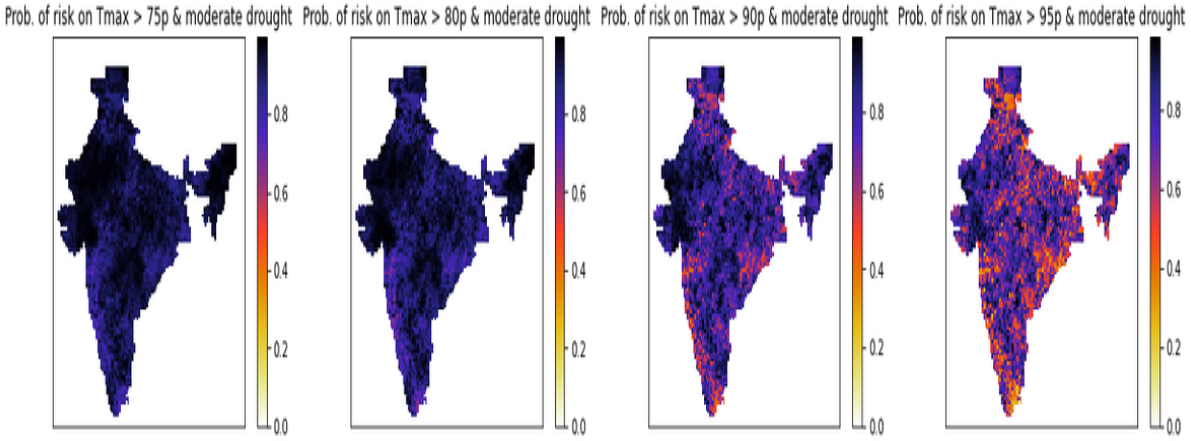

### 4.1.4 Using a max temperature threshold of 95 percentile

The probability of risk in different regions is shown below.



### 4.1.5 Comparison between the plots at different thresholds for moderate drought values

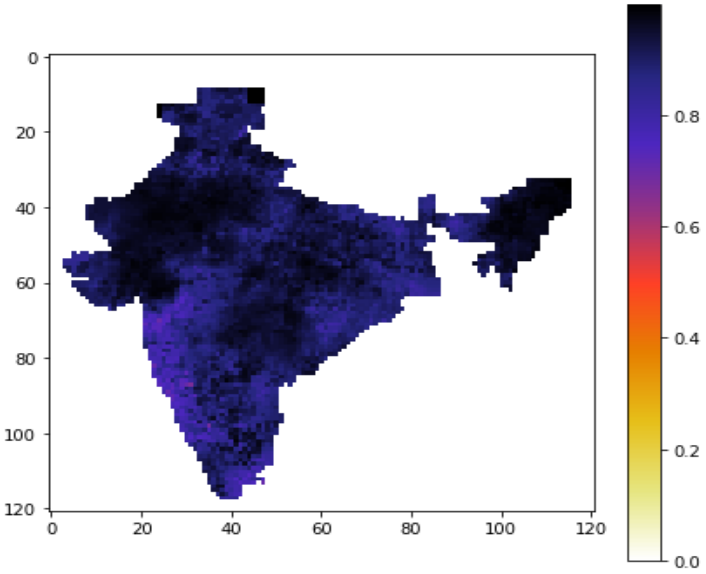
The comparison between the graphs plotted at different thresholds of temperature in moderate drought regions is shown below.



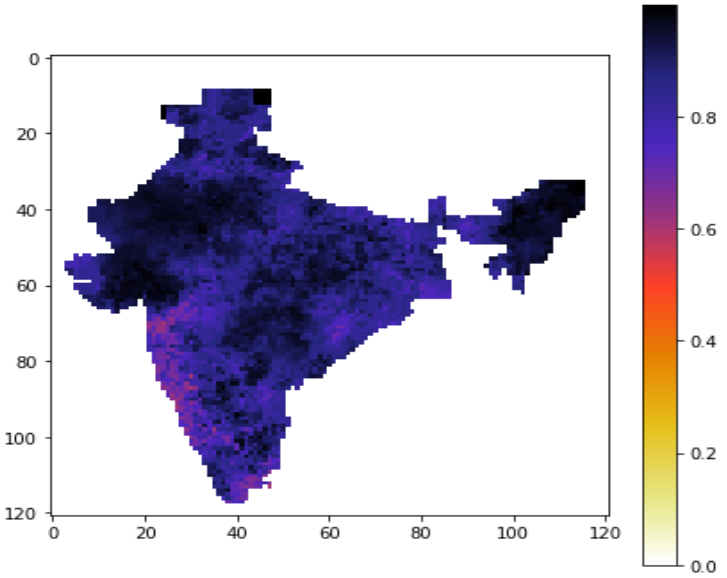

**Inference :** We can see from the above figures that for moderate drought conditions, regions closer to desert areas like Rajasthan having high temperature or places in the North-East having high rainfall have a higher probability of risk as compared to the other areas. This is because temperature or rainfall values are extreme in these areas. Also, areas close to Andhra Pradhesh or Chattisgarh are classifies as high risk zones mainly due to the continental climate they experience because of being far away from the sea. In comparison, regions in the south-west coast can be seen as having lower probability of risk, since these regions experience moderate temperature and rainfall throughout the year. As we increase the maximum temperature threshold, lesser regions show possibilities of risk, since they do not cross the higher temperature threshold needed for it to be classified with risk.

## 4.2 Severe drought conditions (SPEI values between -1.50 and -1.99)

### 4.2.1 Using a max temperature threshold of 75 percentile

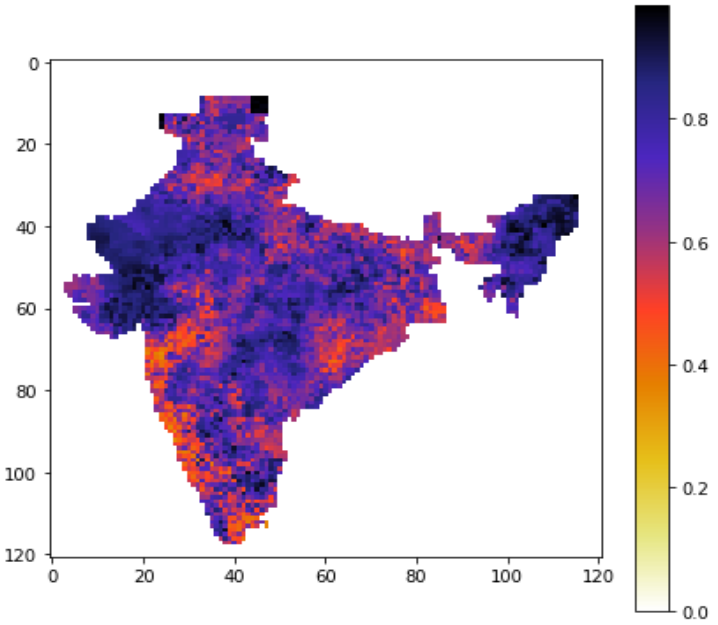



### 4.2.2 Using a max temperature threshold of 80 percentile



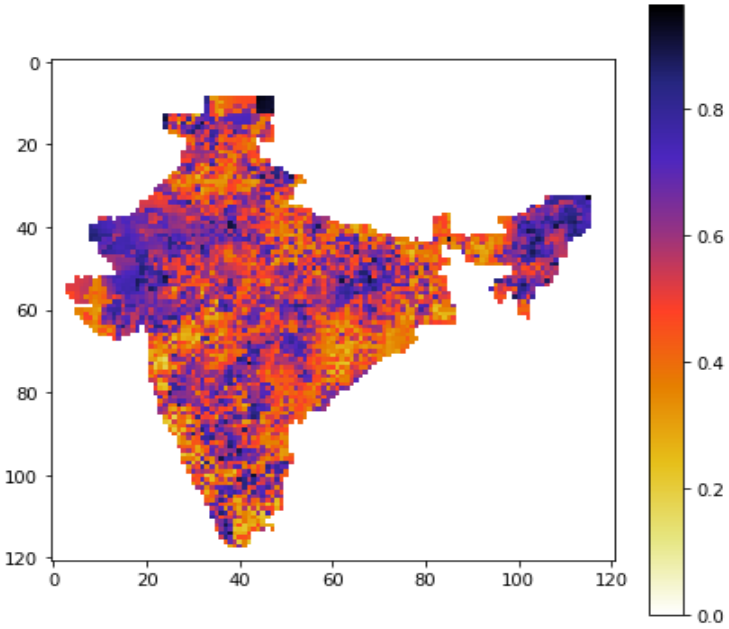
### 4.2.3 Using a max temperature threshold of 90 percentile

The probability of risk in different regions is shown below.



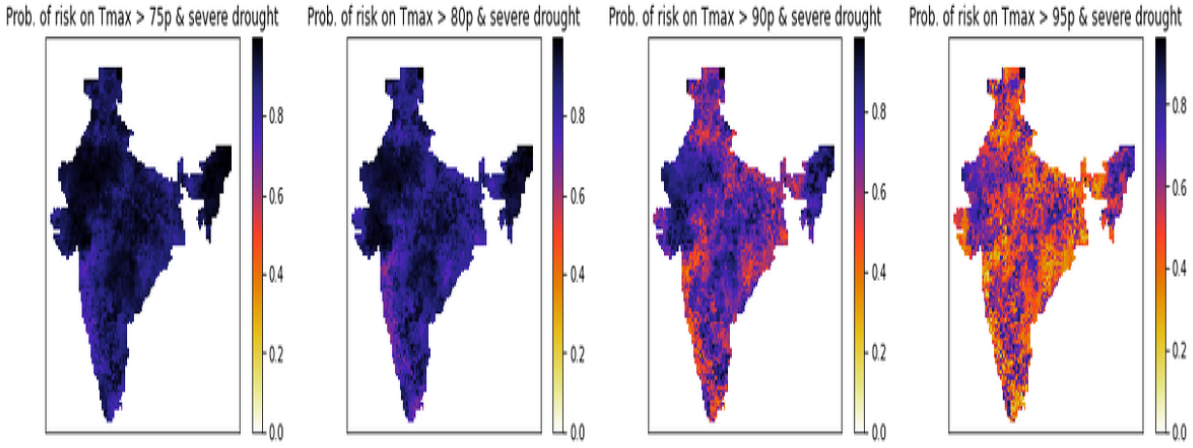

### 4.2.4 Using a max temperature threshold of 95 percentile

The probability of risk in different regions is shown below.



### 4.2.5 Comparison between the plots at different thresholds for severe drought values

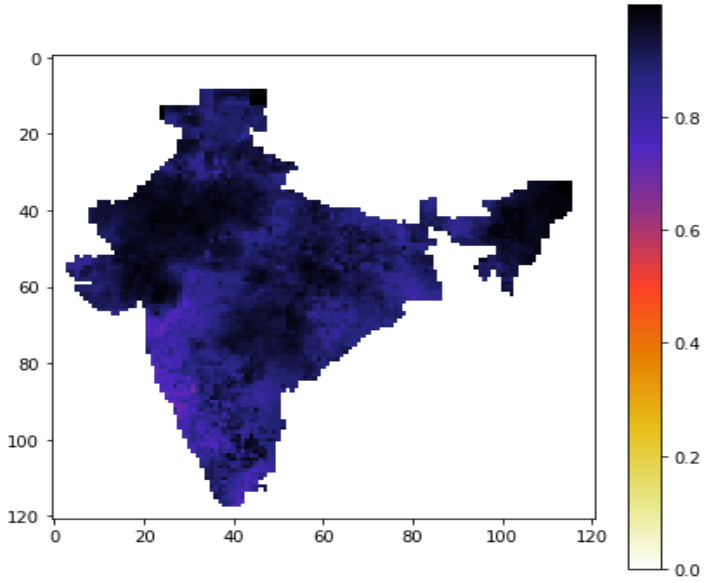
The comparison between the graphs plotted at different thresholds of temperature in severe drought regions is shown below.



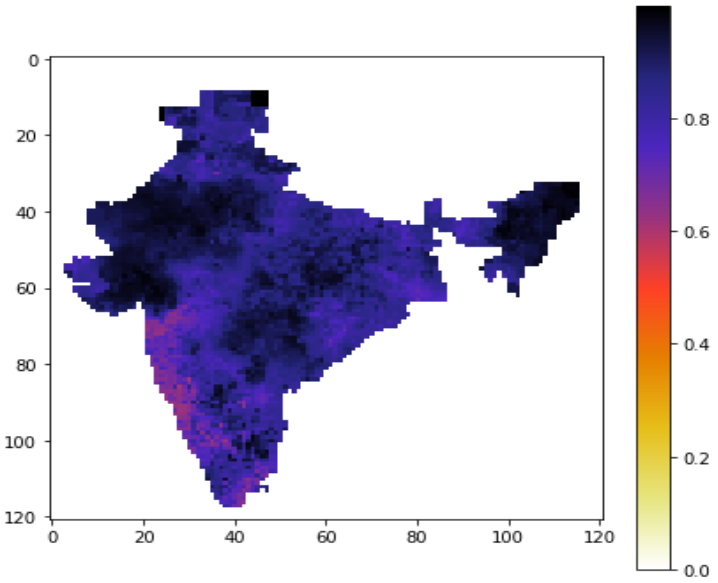

**Inference :** We can see from the above figures that for severe drought conditions lesser regions come under high risk zones, which is because of the low value of SPEI that needs to be there for it to be classified as severe drought. Regions closer to desert areas like Rajasthan having high temperature or places in the North-East having high rainfall have a higher probability of risk as compared to the other areas. This is because temperature or rainfall values are extreme in these areas. In comparison, regions in the south-west coast can be seen as having lower probability of risk, since these regions experience moderate temperature and rainfall throughout the year. As we increase the maximum temperature threshold, lesser regions show possibilities of risk, since they do not cross the higher temperature threshold needed for it to be classified with risk.

## 4.3 Extreme drought conditions (SPEI values less than -2.0)

### 4.3.1 Using a max temperature threshold of 75 percentile

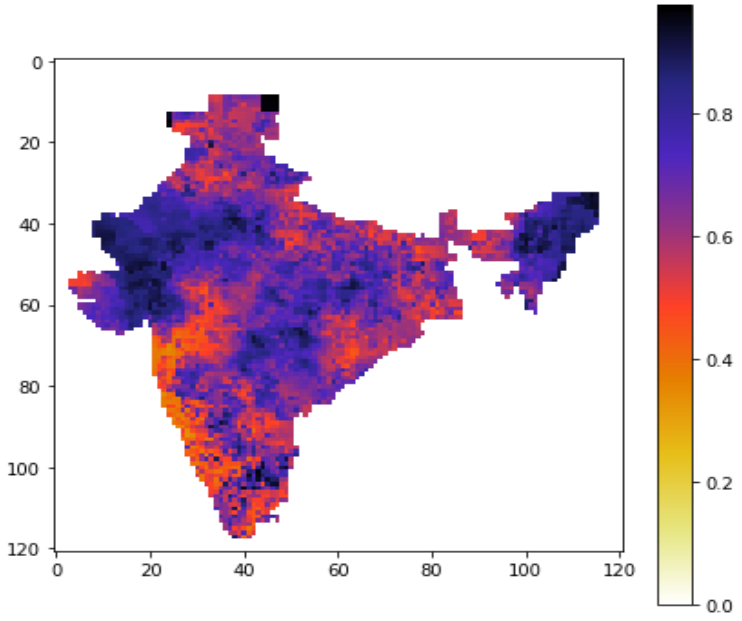



### 4.3.2 Using a max temperature threshold of 80 percentile



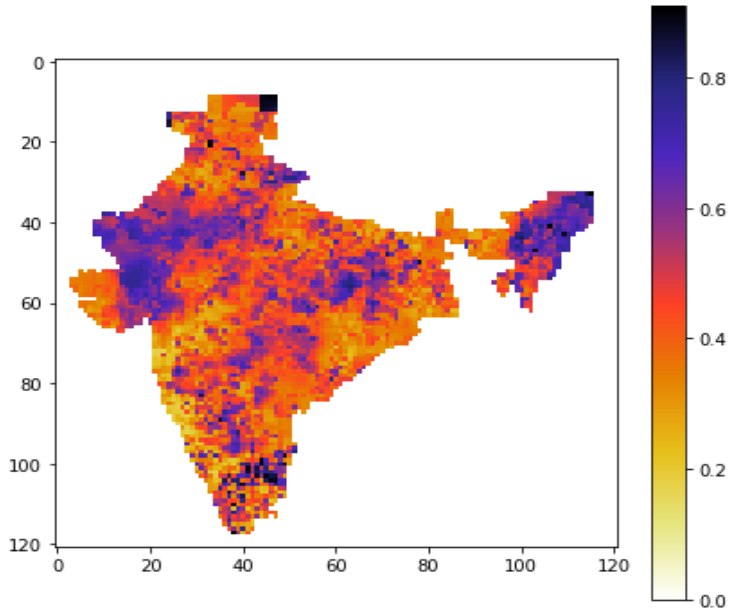
### 4.3.3 Using a max temperature threshold of 90 percentile

The probability of risk in different regions is shown below.



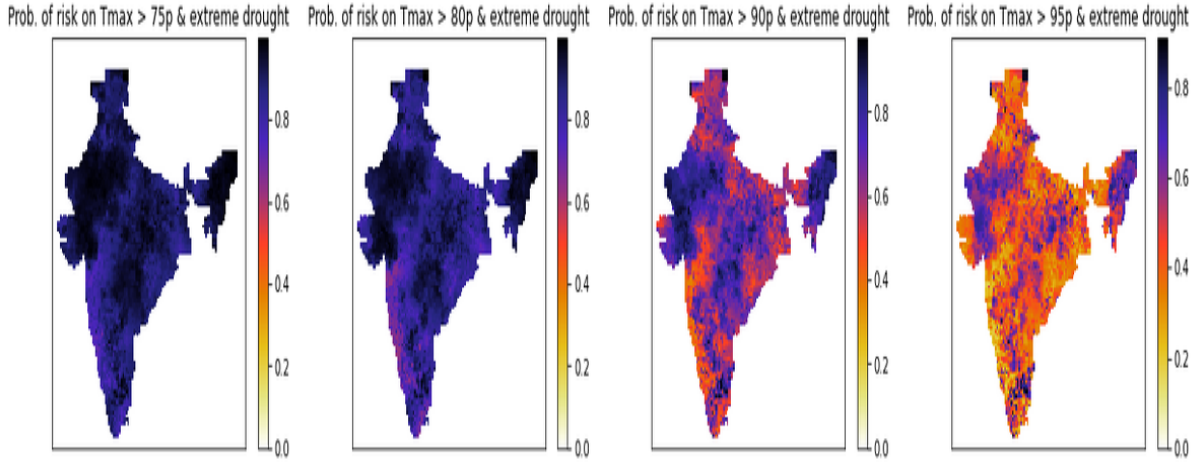

### 4.3.4 Using a max temperature threshold of 95 percentile

The probability of risk in different regions is shown below.



### 4.3.5 Comparison between the plots at different thresholds for extreme drought values

The comparison between the graphs plotted at different thresholds of temperature in extreme drought regions is shown below.



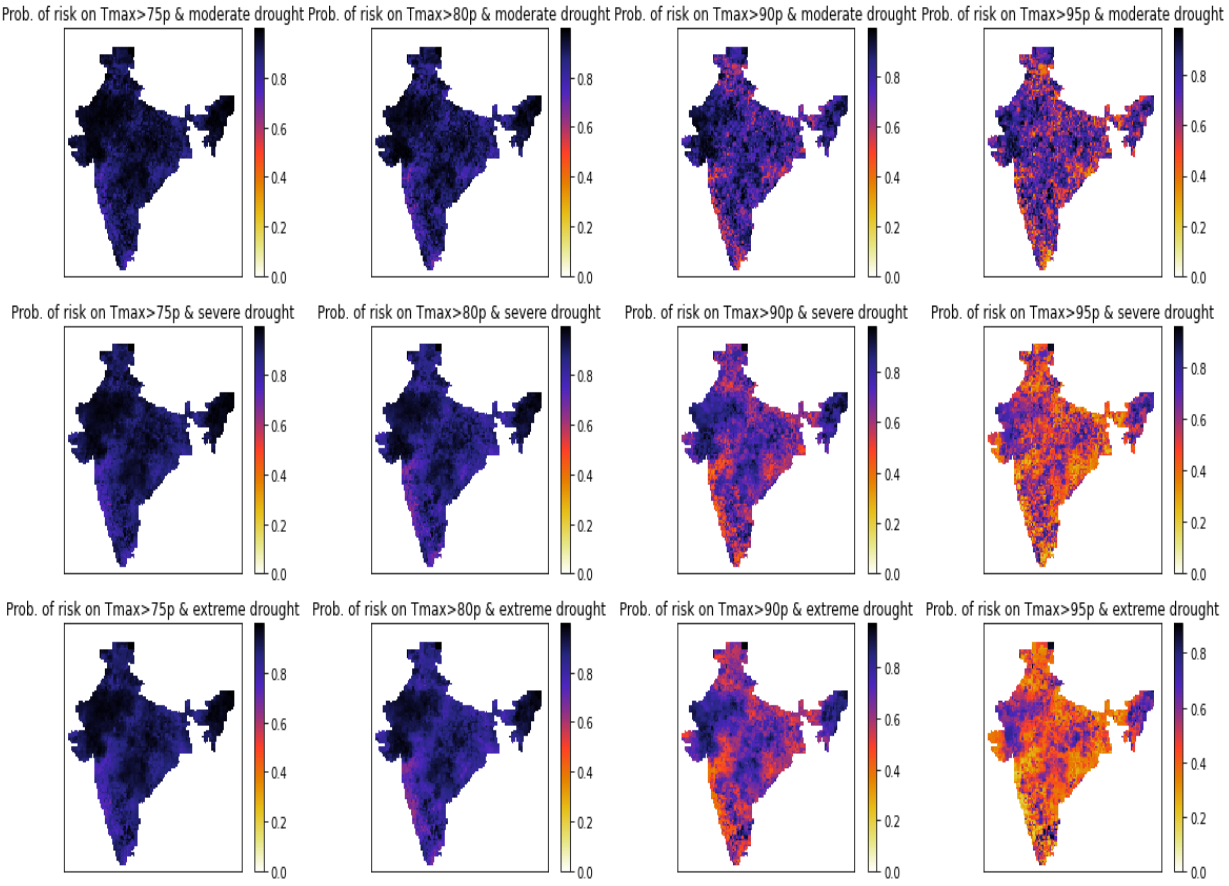

**Inference :** We can see from the above figures that for extreme drought conditions, even lesser regions come under high risk zones. This s because of the high negative value of SPEI that needs to be there for it to be classified as extreme drought. Even with such strict conditions, regions closer to desert areas like Rajasthan having high temperature or places in the North-East having high rainfall have a higher probability of risk as compared to the other areas. This is because temperature or rainfall values are extreme in these areas. In comparison, regions in the south-west coast can be seen as having lower probability of risk, since these regions experience moderate temperature and rainfall throughout the year. As we increase the maximum temperature threshold, lesser regions show possibilities of risk, since they do not cross the higher temperature threshold needed for it to be classified with risk.

## 4.4 Comparison of probability of risk at different temperature and drought index thresholds

The results obtained at different threshold of maximum temperature and drought index values are shown below. The thresholds used are classified as :

- Moderate drought (SPEI value between -1.0 and -1.49) and Tmax threshold 75 percentile
- Moderate drought (SPEI value between -1.0 and -1.49) and Tmax threshold 80 percentile
- Moderate drought (SPEI value between -1.0 and -1.49) and Tmax threshold 90 percentile
- Moderate drought (SPEI value between -1.0 and -1.49) and Tmax threshold 95 percentile
- Severe drought (SPEI value between -1.50 and -1.99) and Tmax threshold 75 percentile
- Severe drought (SPEI value between -1.50 and -1.99) and Tmax threshold 80 percentile
- Severe drought (SPEI value between -1.50 and -1.99) and Tmax threshold 90 percentile
- Severe drought (SPEI value between -1.50 and -1.99) and Tmax threshold 95 percentile
- Extreme drought (SPEI value less than -2.0) and Tmax threshold 75 percentile
- Extreme drought (SPEI value less than -2.0) and Tmax threshold 80 percentile
- Extreme drought (SPEI value less than -2.0) and Tmax threshold 90 percentile
- Extreme drought (SPEI value less than -2.0) and Tmax threshold 95 percentile

In the figures shown below, a higher shade of black signifies a higher probability of risk, whereas lighter shades signifies low risk regions.

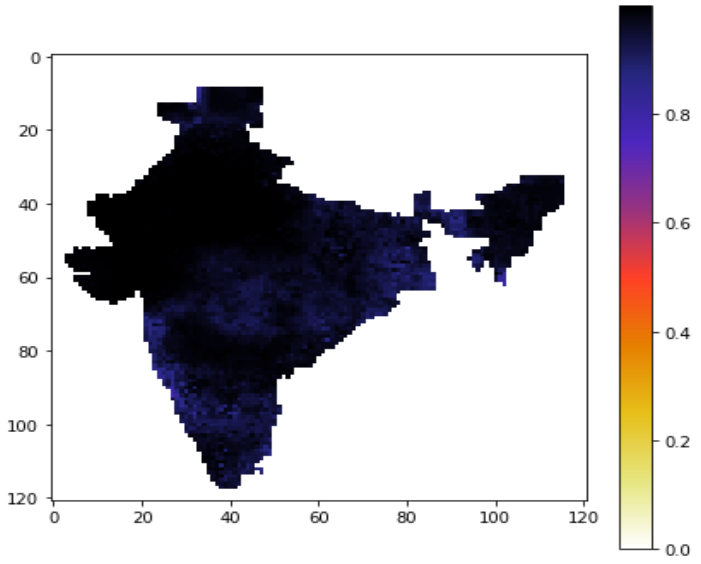



**Inference :** The comparison of the plots at different temperature and drought index thresholds is shown in the figure. One observation that we can see is that areas like Rajasthan which experience high temperature and North-East India which experience high rainfall, comes under high risk zones all the time, i.e. starting from moderate to extreme drought zones. But areas in the Central India like the states of Chattisgarh or western parts of Andhra Pradhesh comes under high risk zone when classified as moderate to severe drought, but not in extreme drought cases. This is because of the continental climate they experience because of being far away from the sea. Regions in the western coast of India comes under low risk zones because of the moderate temperature and rainfall they experience almost throughout the year.

## 4.5 Calculating risk only during monsoon season (June - September)

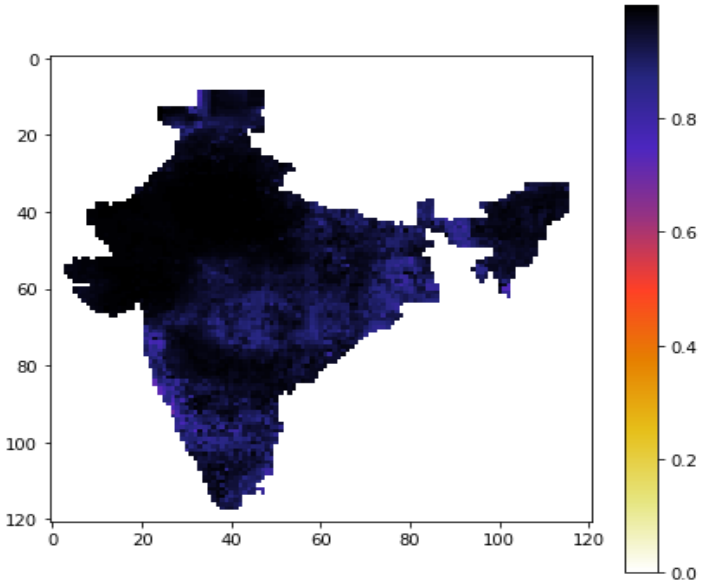
### 4.5.1 Using a max temperature threshold of 75 percentile

The probability of risk in different regions is shown below.



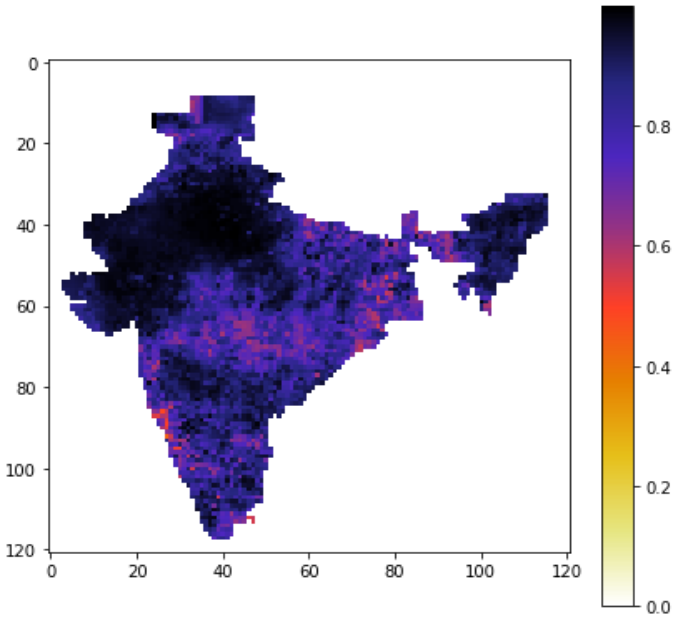

### 4.5.2 Using a max temperature threshold of 80 percentile

The probability of risk in different regions is shown below.



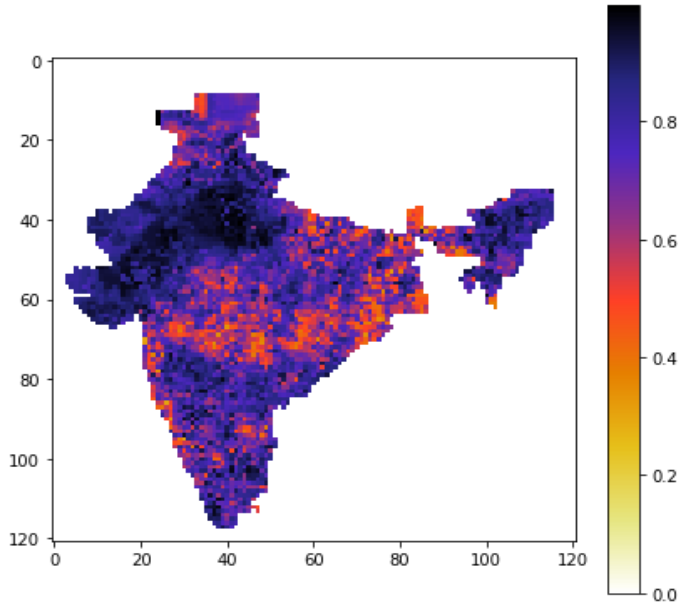
### 4.5.3 Using a max temperature threshold of 90 percentile

The probability of risk in different regions is shown below.



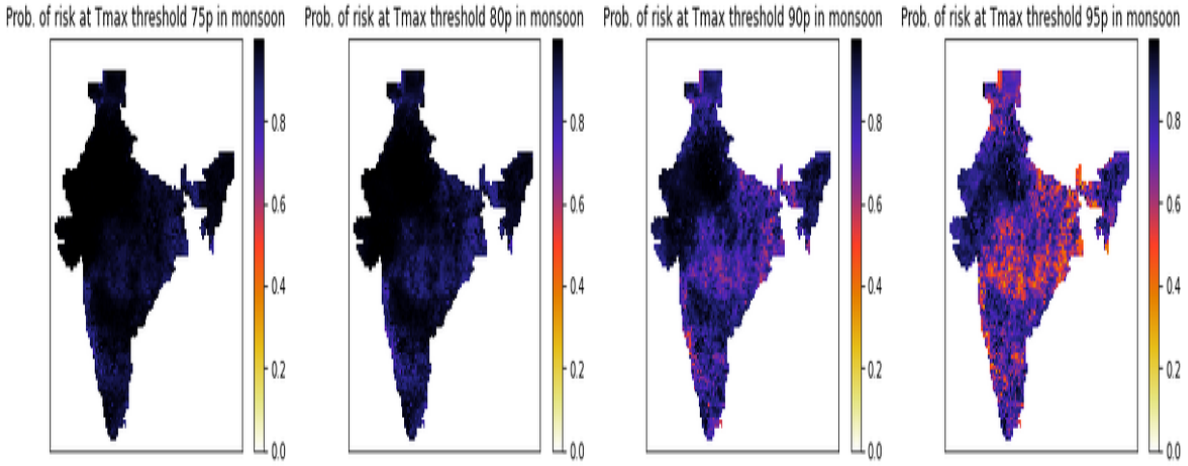

### 4.5.4 Using a max temperature threshold of 95 percentile

The probability of risk in different regions is shown below.



### 4.5.5 Comparison between the plots at different thresholds during the monsoon season

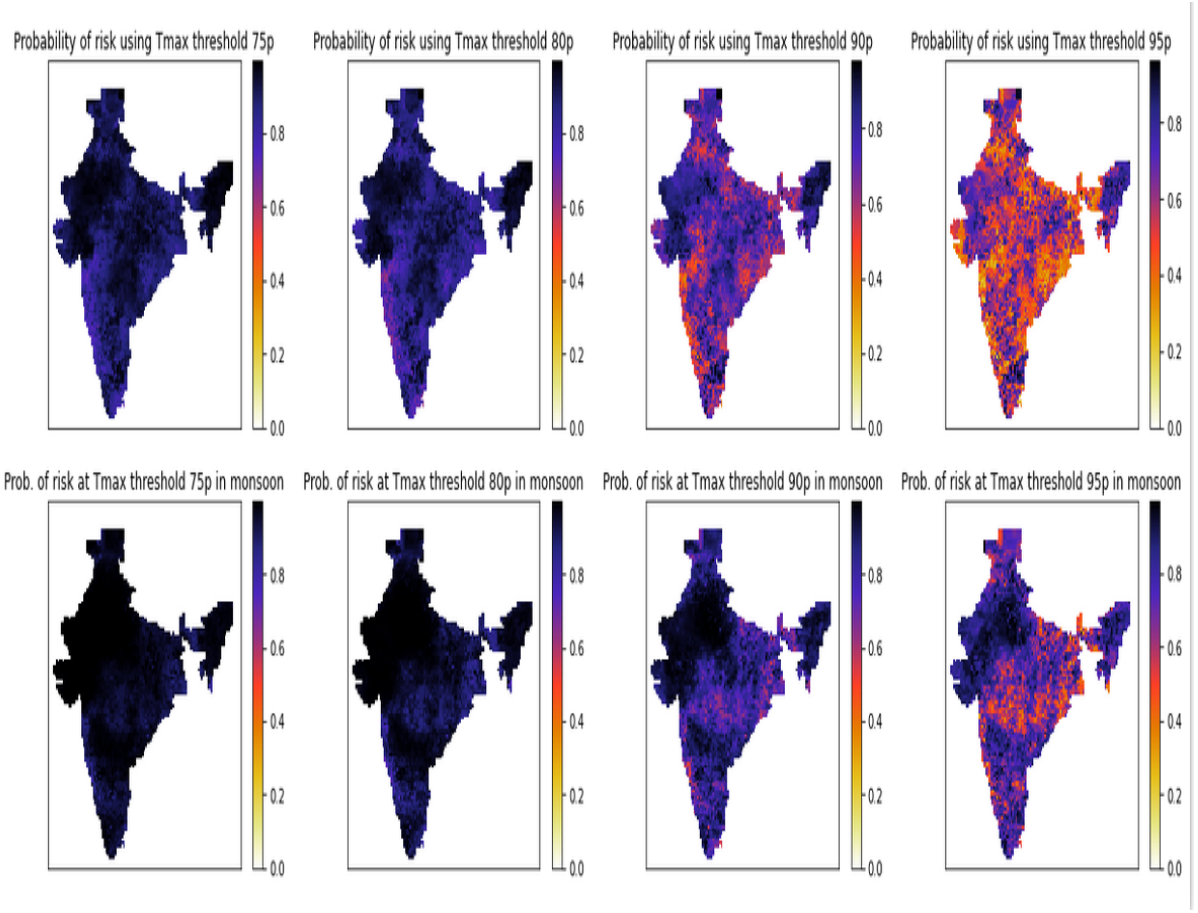
The comparison between the graphs plotted at different thresholds during the monsoon season is shown below.




**Inference :** We can see from the above figures that regions closer to desert areas like Rajasthan having high temperature or places in the North-East having high rainfall have a higher probability of risk as compared to the other areas. This is because temperature or rainfall values are extreme in these areas. In comparison, regions in the south-west coast or Central India can be seen as having lower probability of risk, since these regions experience moderate temperature and rainfall throughout the year. As we increase the maximum temperature threshold, lesser regions show possibilities of risk, since they do not cross the higher temperature threshold needed for it to be classified as risk.

## 4.6 Comparison between different plots during water year and monsoon season

The comparison between the graphs plotted at different thresholds during the entire year and monsoon season is shown below.

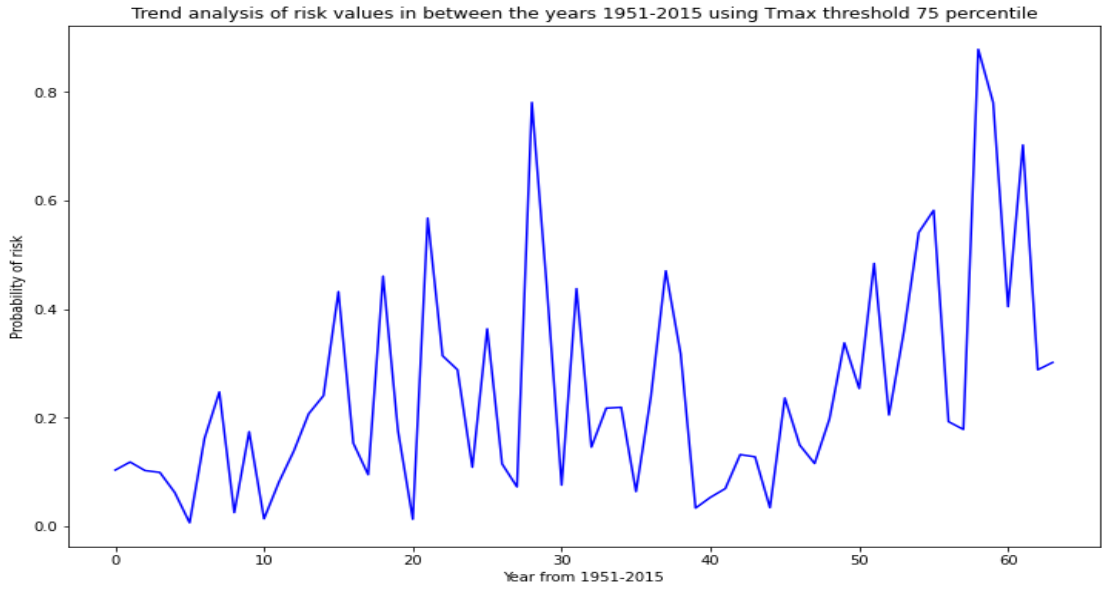



**Inference :** The regions having darker shades of blue signify higher risk, and those with lighter shades signifies low risk. We can see from the above figures that as compared to the graphs using data of the entire water area, the graphs using monsoon season data have more regions having high risk. The reason behind this is because larger areas show abrupt differences in rainfall during the monsoon season. The regions with moderate rainfall are seen to be having lesser probability of risk, whereas the regions showing extreme rainfall cases have high risk.

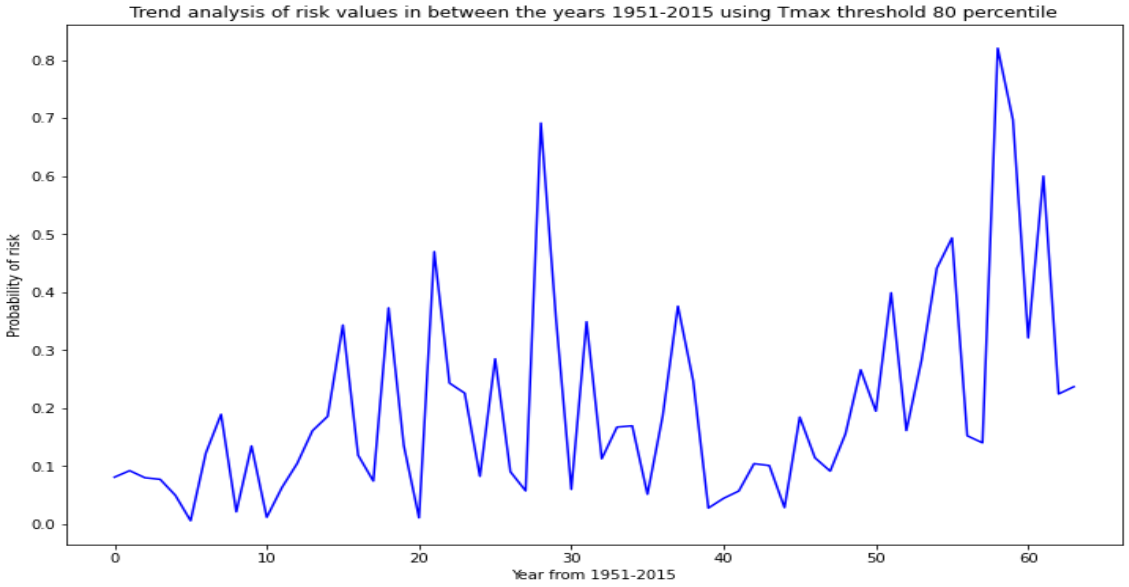
## 4.7 Trend analysis of risk over the years 1951-2015 in one grid

The trend analysis of risk have been done on one grid. Different experiments have been performed using different values of maximum temperature thresholds. Some of the observations are shown below.

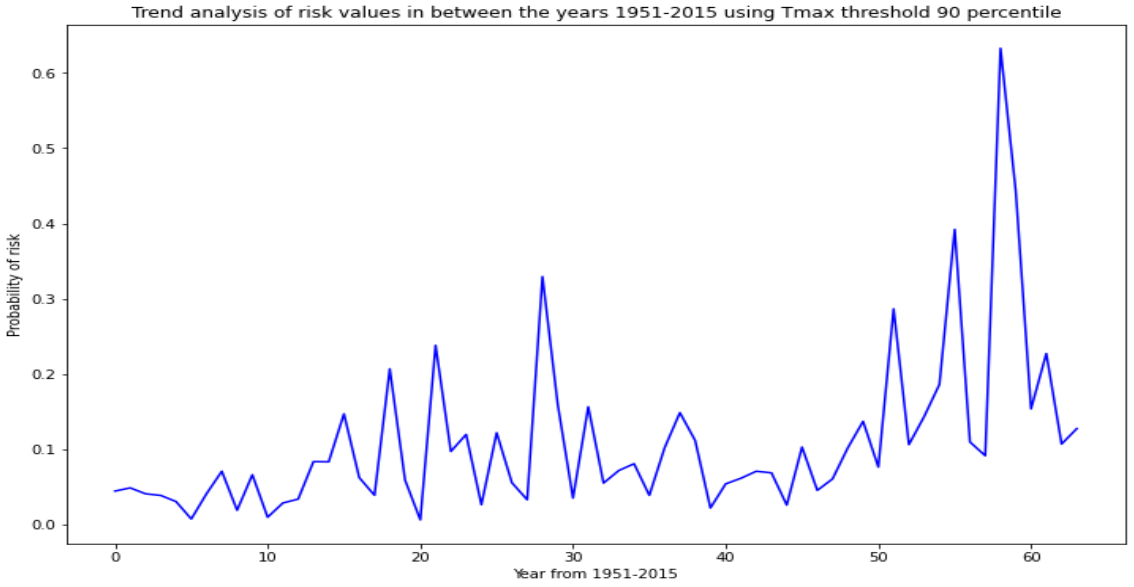
1. Using a max temperature threshold of 75 percentile



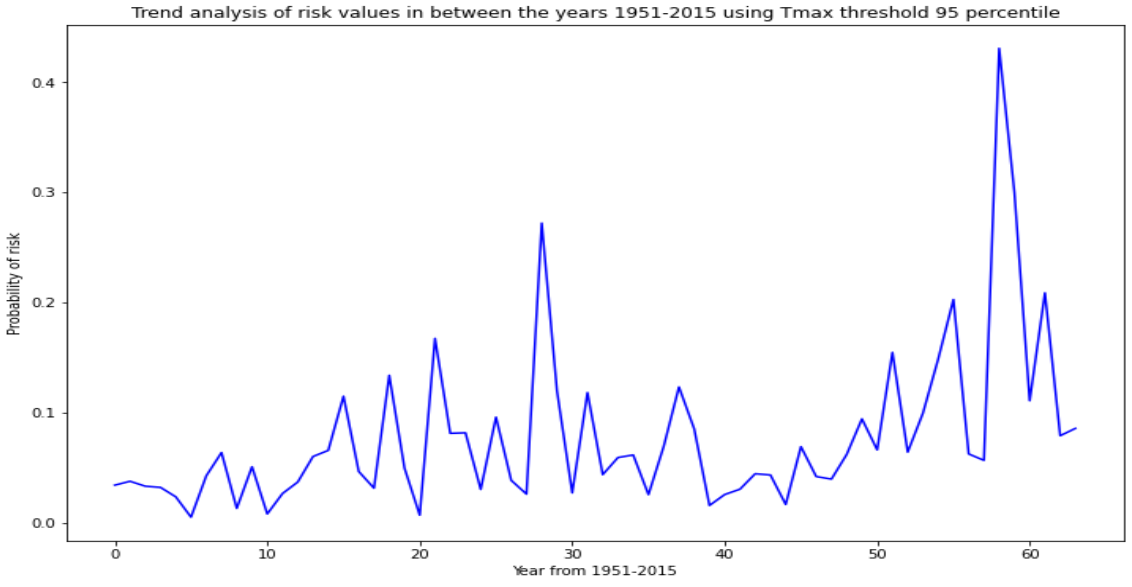

2. Using a max temperature threshold of 80 percentile



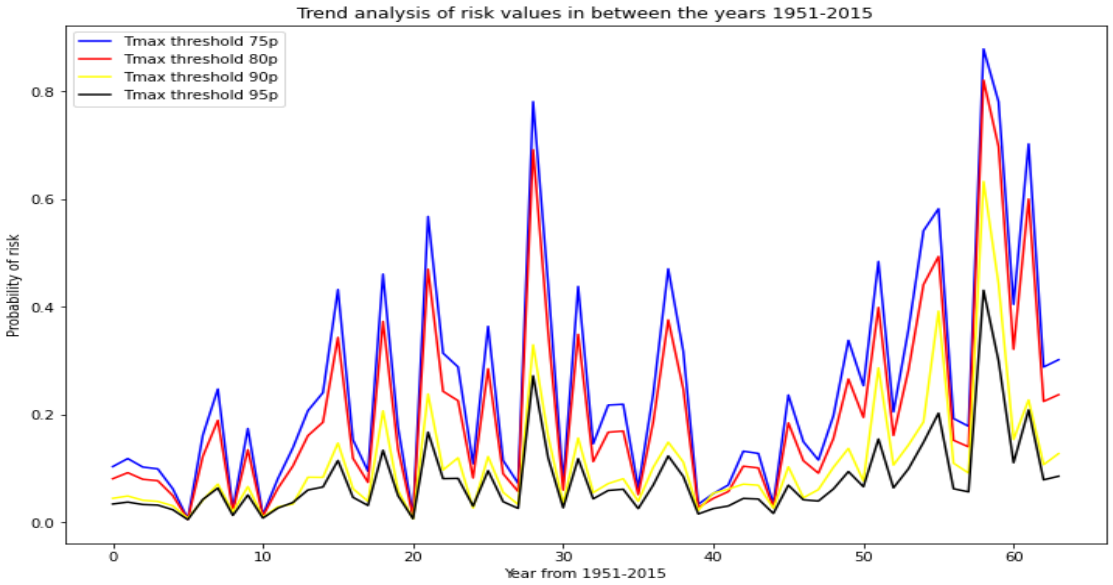
3. Using a max temperature threshold of 90 percentile




4. Using a max temperature threshold of 95 percentile



## 4.8 Comparison between the plots at different thresholds




**Inference :** We can see from the above plots that the risk in general have a slightly upward trend over the years. This is because of the changes in temperature and rainfall that we are experiencing due to global warming. Temperature and precipitation values often leads to extreme values in most of the regions. The changes with respect to the maximum threshold value can also be seen. As we increase the maximum temperature threshold, lesser possibilities of risk are observed, since they do not cross the higher temperature threshold needed for it to be classified as risk.

# 5 CONCLUSION

We have experimented using different drought index values and maximum temperature thresholds to predict the probability of risk in different regions in India. We see that regions having extreme values of temperature and rainfall, like that of the desert regions in Rajasthan or the wetty regions in North-East India are classified as high risk zones. In comparison, areas in the south-western coast of India or central India do not fall under risk since they experience moderate ranges of temperature and rainfall throughout the year. We also see that judging only by the monsoon season, more areas come under the high risk zone, mainly due to abrupt differences in rainfall. The trend in risk values have been seen to exhibit an upward growth with time.

# 6 REFERENCES


- A. Bárdossy and G. G. S. Pegram, "Copula based multisite model for daily precipitation simulation", December 2009
- Martin Haugh, "An Introduction to Copulas", Spring 2016
- Python Copula, "https://pypi.org/project/copulas/"
- Johannes Krouthén, "Extreme joint dependencies with copulas", 2015
- Wenjun Yu, Yanzhong Li, Yanping Cao and Tayler Schillerberg, "Drought Assessment using GRACE Terrestrial Water Storage Deficit in Mongolia from 2002 to 2017", 2019
- Syed Muhammad Aqil Burney, Osama Ajaz, Shamaila Burney, "Multivariate Copula Modeling with Application in Software Project Management and Information Systems", 2018