Name: Mansi Uniyal
Roll No: 19EE10039

Dataset:-

International Comprehensive Ocean-Atmosphere Data Set (ICOADS)
ICOADS Data Online at the NOAA Earth System Research Laboratory

For data starting in 1960, the netCDF files are further subdivided by grid box size (2°x2° or 1°x1° boxes) and spatial domain, with the 1°x1° summaries available in global and equatorial formats. The equatorial products cover the latitude band 10.5°N to 10.5°S and are global with respect to longitude. In contrast to the global 1°x1° boxes, the equatorial 1°x1°boxes are shifted half a degree in latitude (only) in comparison to the global domain, such that the center-latitude of the central row of boxes is the equator (e.g., 0-1°E, 0.5°S-0.5°N). Both the global and equatorial formats have a standard and enhanced product.
The netCDF files are subdivided into "standard" and "enhanced" products, which reflect the "trimming" (quality control) procedure used and the data mixture. In the standard products, the data have been edited using climatological 3.5 standard-deviation (sigma) limits, with the observations limited (as nearly as practical) to those from ships. In the "enhanced" products, in contrast, the data have been edited using broader 4.5 sigma limits to better represent extreme climate events, with observations from ships plus from other in situ marine platform types (e.g., drifting and moored buoys).

The netCDF files are further subdivided by spatial domain, with the 1°x1° summaries available in global and equatorial formats. The equatorial products cover the latitude band 10.5°N to 10.5°S and are global with respect to longitude. In contrast to the global 1°x1° boxes, the equatorial 1°x1° boxes are shifted half a degree in latitude (only) in comparison to the global domain, such that the center-latitude of the central row of boxes is the equator (e.g., 0-1°E, 0.5°S-0.5°N). Both the global and equatorial formats have a standard and enhanced product.

Spatial Coverage:
1.0 degree latitude x 1.0 degree longitude equatorial grid (21x360).
10.0N - 10.0S, 0.5E - 359.5E.
1.0 degree latitude x 1.0 degree longitude global grid (180x360).
89.5N - 89.5S, 0.5E - 359.5E.
Levels:
Sea Surface

DATABASE LINK: https://psl.noaa.gov/data/gridded/data.coads.1deg.html

Solution:-
Initially, we determine the maxima and minima of temperature values from the hourly values of the temperatures that the dataset provides us. A one-year timeframe was formed for this, and the maximum and minimum temperatures were recorded. The block extremes for each of the sites are these maxima and minima. Note that the locations are real-world locations. Each pixel in the map corresponds to a 1^{o} x 1^{o} area. Since the original dataset provides information for the 0.5^{o} x 0.5^{o} region, the resolution was upsampled by taking the mean of the 4 0.5^{o} x 0.5^{o} to form the 1^{o} x 1^{o} region. The GEV distribution would then be fitted to each of the locations. Each location has a time series of temperature associated with it. Now, in order to fit the GEV, Scipy's built-in function gen extreme was utilized. It took roughly 2 hours to fit GEV distribution for each of the 64800 sites on the map (because the number of latitudes is 180 and the number of longitudes is 360). Furthermore, return levels were estimated for return periods of 5, 10, 20, 50, and 100 starting in the year 2000. It was discovered that when the return duration lengthened, the temperature returned at higher levels. This could be due to the fact that the temperature distribution is monotonically growing, with higher temperatures having higher return times. The next step is to illustrate which locations saw an increase in extreme events with slow return frequency after the year 2000. As a result, a manual selection of frequency 10 was picked to test this. This implies that if an extreme temperature happens more than 10 times in a specific region, that region will be displayed separately on the world map. Furthermore, the precise temperature value may not occur multiple times. To make the findings more realistic, a leave of 1K was allowed, which means that if the temperature is within 1K of the extreme value of the temperature, that region will also be segregated. The pixelated map is shown in Figure 1.

The final step is to cluster the locations based on the parameters (shape, location, and scale) of the GEV distribution. The K Means clustering technique was implemented with the feature vectors comprising of shape, location, and scale associated with each location. It is based on the assumption that locations belonging to the same cluster will have comparable values of shape, location, and scale. A total of 20 clusters were created. Because the silhouette score was not computed and verified for different numbers of clusters on different sites, this may not be the optimal number of clusters

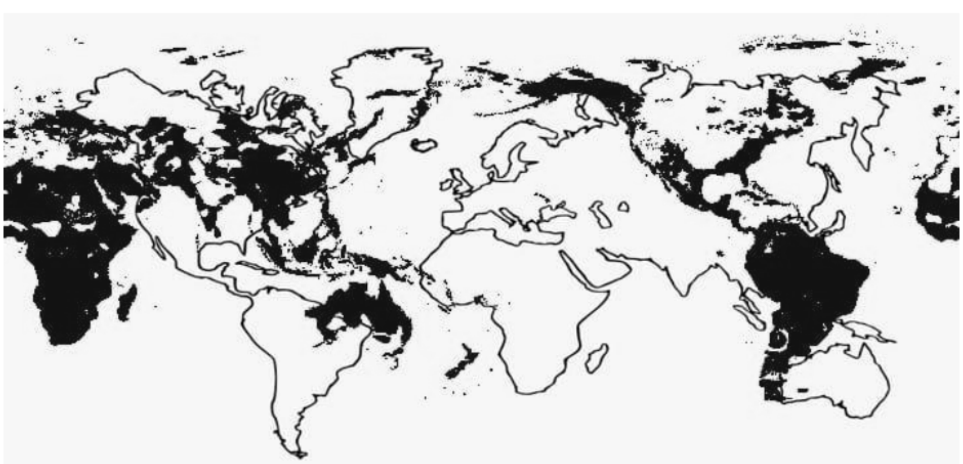

Figure 1: Map with locations having an extreme event with a frequent high return period.