Analysis and Forecasting of Precipitation ERA5 land data of India

Omkar Mehta (omehta2)

Milestone#2 (March 2022)

1 Problem Statement

I have chosen ERA5-land hourly data [1] for my project. The objective is to find useful information about the weather in the Konkan area of Maharashtra, which receives the highest rainfall, leading to the floods. If we are able to find useful patterns in the data that could help us forecast the precipitation in the future, even by an hour, it could help the people to better prepare for the flood. The data is in ***netCDF*** file format. Hence, we can use ***xarray*** to read the data.

1.1 Floods in Mahad [2]

- My hometown is Mahad, a small town in the Konkan district of Maharashtra, India. (**lat**: 18.1^oN, **lon**: 73.4^oE).
- On July 21 and July 22, 2021, the torrential rains had left '94' houses completely damaged and '9649' partially wrecked.
- About *3709* houses were damaged.
- The landslide in **Taliye** village in the Mahad tehshil killed at least 84 people.

1.2 The parameters that I have chosen

1. **10m_v_component_of_wind or v10** : northward component of the $10m$ wind, in m/s units, moving at a height of $10m$ above the surface.
2. **2m_temperature or t2m**: the temperature of air at $2m$ above the surface, in *Kelvin* units.
3. **surface_net_solar_radiation or ssr**: the amount of solar radiation that reaches a horizontal plane at the surface minus the amount reflected by the Earth's surface, in W/m^2 units.
4. **surface_pressure or sp**: pressure of the atmosphere on the surface of land, in Pa units.
5. **total_precipitation or tp**: the accumulated liquid and frozen water, comprising rain and snow, that falls to the surface in m units.

2 Data Collection

This has been covered in **Milestone#1** in great details. In summary, **cdsapi** [3] is used to download the data. I have ticked on the parameters that I have chosen to do analysis on. I have selected 2019, 2020, 2021 as the years. I have only considered $14, 15, \cdots, 24$ as the days, since

I need to know what happened on *22 July, 2021*. I have taken the latitudes and longitudes to be that of Maharashtra state (**lat**: 22^oN - 16^oN, **lon**: 72^oE - 79^oE). The code can be found in *'Project/data_collection.ipynb'* [4].

3 Reading Data

I have read the data using *xarray* package, since the data is multi-dimensional.

4 Basic Characteristics of the variables

The data has five variables: **v10, t2m, ssr, sp, tp**. Each variable has three dimensions: **time, latitude, longitude**. The coordinates are **latitude, longitude, time**. The shape of the data is **latitude:** 61**, longitude:** 71**, time:** 792. Total number of data points are 3430152. The direction of the latitude is northward, and of the longitude is eastward. I have separated the **data** into **data_2019, data_2020, data_2021** according to the years.

The missing values in the data are attributed to the locations where there is no land, i.e. ocean.

4.1 Statistics of precipitation

On 22nd July, 2021, the maximum precipitation in Mahad was 97.2 mm. The minimum precipitation in Mahad was 0.46 mm. The standard deviation of precipitation in Mahad was 20.9 mm. The mean precipitation in Mahad was 28.78 mm. The interquartile range was 31.05 mm.

Variable2019	Max	Min	Mean	Std	IQR
tp	0.038647472858429000	1.52513384819031E-05	0.009194524	0.009008652530610560	0.008935155346989630
ssr	14228702.0	1363.0	7849041.5	4669063.0	7711432.875
t2m	303.75189208984400	296.7691650390630	298.95444	1.500083327293400	2.0300369262695300
v10	3.0437960624694800	-1.110135555267330	0.9523016	0.7549567818641660	0.7506721019744870
sp	100424.4453125	99382.5546875	99924.375	274.2175598144530	476.15625

Variable2020	Max	Min	Mean	Std	IQR
tp	0.11406556516885800	1.14366412162781E-05	0.017825091	0.02299460768699650	0.02425815351307390
ssr	12768257.0	668.0	6382346.0	4360974.0	8161690.125
t2m	303.087890625	296.498779296875	298.769	1.321419596672060	1.5544052124023400
v10	3.4909563064575200	-0.4907798767089840	1.2920784	0.856518566608429	1.0630333125591300
sp	100265.7421875	99425.3984375	99900.95	213.41897583007800	299.974609375

Variable2021	Max	Min	Mean	Std	IQR
tp	0.09719663858413700	0.000457763671875	0.028783046	0.02097214013338090	0.03104749694466590
ssr	9130712.0	320.0	3769202.0	2643788.0	3633109.5
t2m	300.45404052734400	296.5606689453130	298.02362	0.699251651763916	0.7313919067382810
v10	2.316418409347530	-1.4563539028167700	0.7222222	0.7639433145523070	1.0843294262886000
sp	100403.6328125	99528.421875	99943.28	188.17091369628900	271.87109375

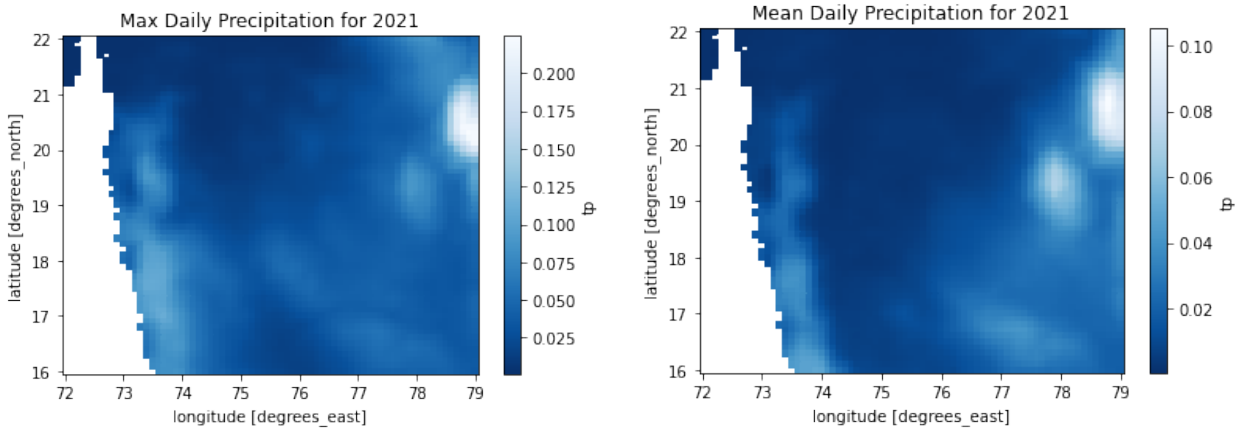
Figure 1: Statistics of all variables in Mahad for each year

The unit of tp is **m**. The maximum precipitation had been 114 mm (maximum of all years) in 2020, but since the mean value was 17.9 mm that is lower than the mean value of 28.8 mm in 2021, the rainfall occurred only for a small period of time in 2020. But, the rainfall in 2021 occurred for

a longer period of time or the maximum rainfall occurred in short periods of time multiple times. The floods would be the result of the accumulation of water in the river **Savitri** in a short period of time. Since the river is narrow, there was flood.

5 Exploratory Data Analysis

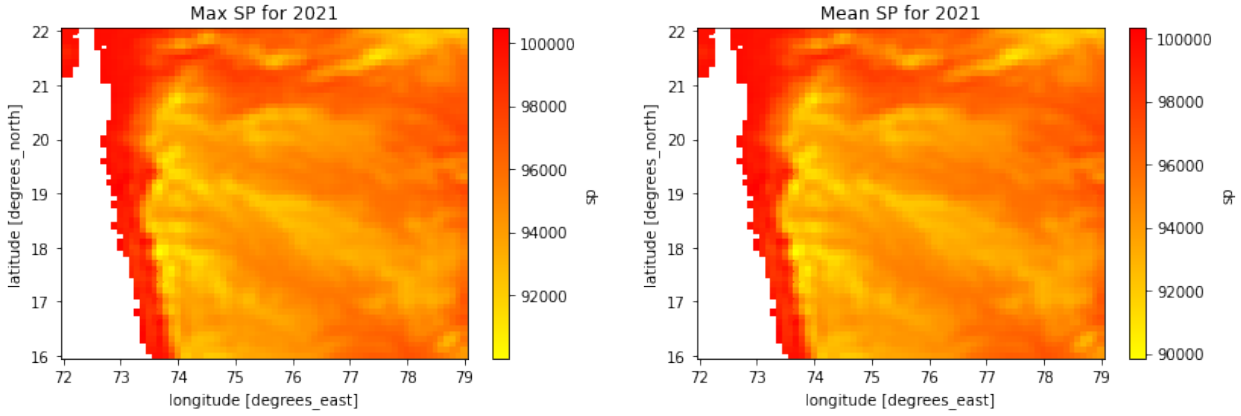
I have plotted the maximum and minimum of all variables for July 22, 2021 in Maharashtra in Figure 2, Figure 3, Figure 4, Figure 5 and Figure 6.



(a) Maximum (b) Mean

Figure 2: Total Precipitation (m), July 22, 2021

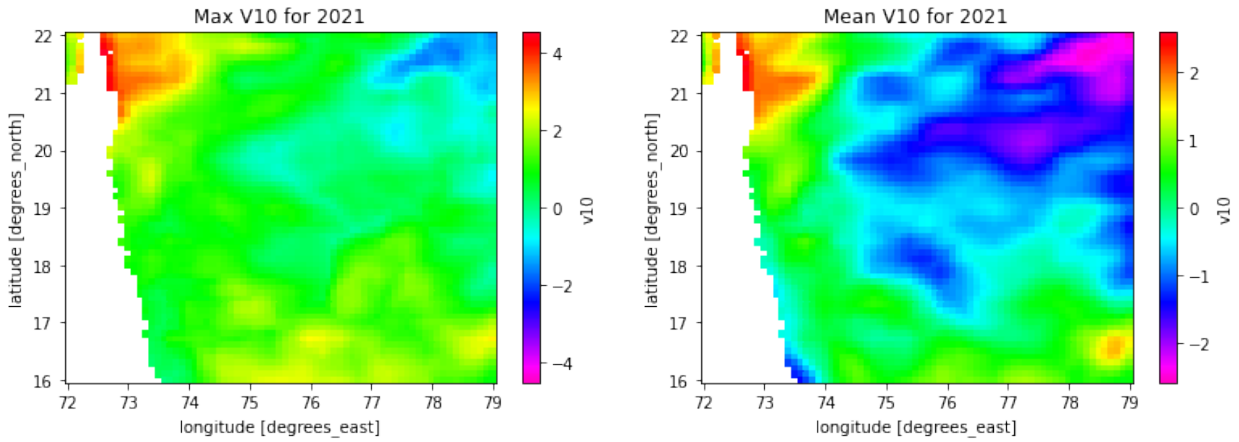
The total precipitation could be seen more on the western part of Maharashtra, where Mahad is.



(a) Maximum (b) Mean

Figure 3: Surface Pressure (Pa), July 22, 2021

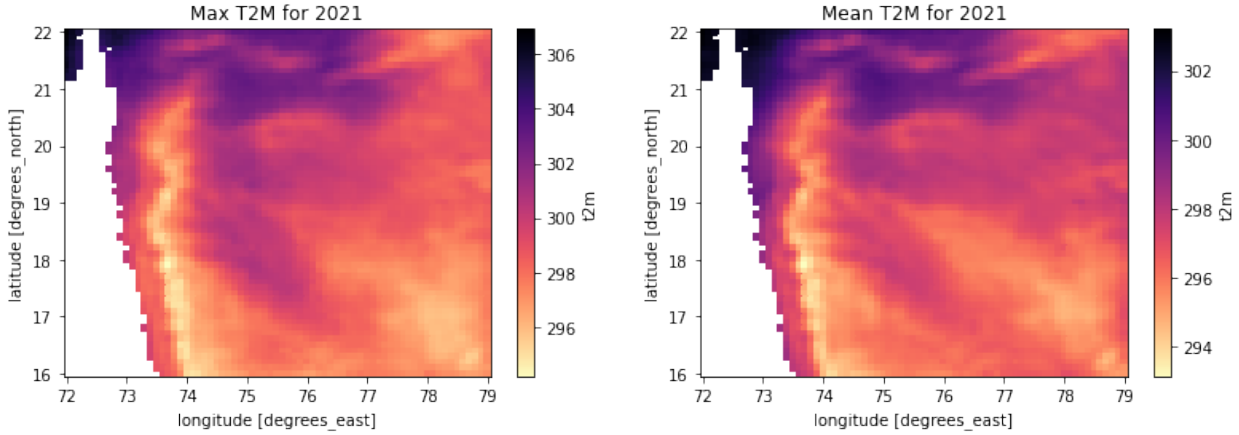
The surface pressure could be seen more on the western part of Maharashtra, where Mahad is.



(a) Maximum (b) Mean

Figure 4: 10m_v_component (m/s), July 22, 2021

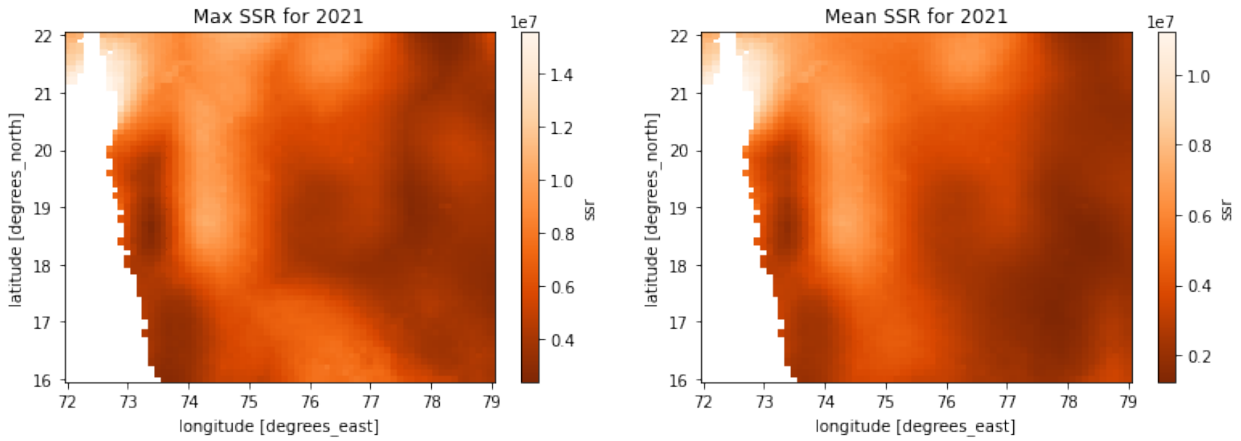
The wind speed was almost close 0 m/s on the western part of Maharashtra, where Mahad is.



(a) Maximum (b) Mean

Figure 5: 2m_temperature (K), July 22, 2021

The temperature of air at 2m above the surface of the western part of Maharashtra was between $296 - 298$ K, expected since it was colder (told by my parents).



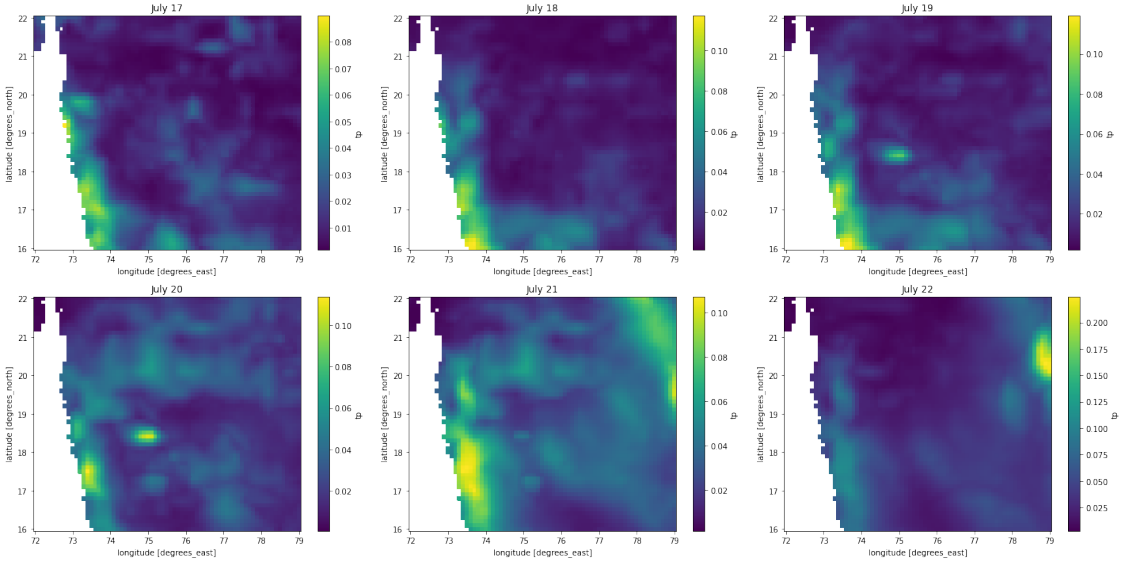
(a) Maximum (b) Mean

Figure 6: surface_net_solar_radiation (W/m^2), July 22, 2021

The net solar radiation was around $0.2 - 0.4\ W/m^2$, which was not much, as compared to the other parts.

It's clear that some anomaly had taken place on July 22, 2021. The people experience "Cloudburst" [5]. Did cloudburst occur on prior days to July 22, 2021?

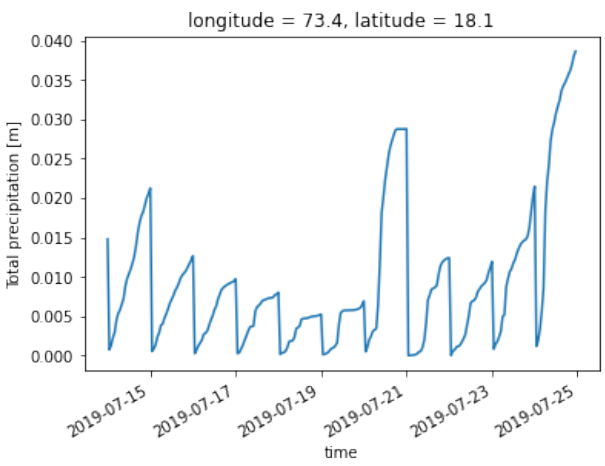
I have plotted the maximum precipitation on 5 days prior to July 22, 2021 on the same plot in Figure 7



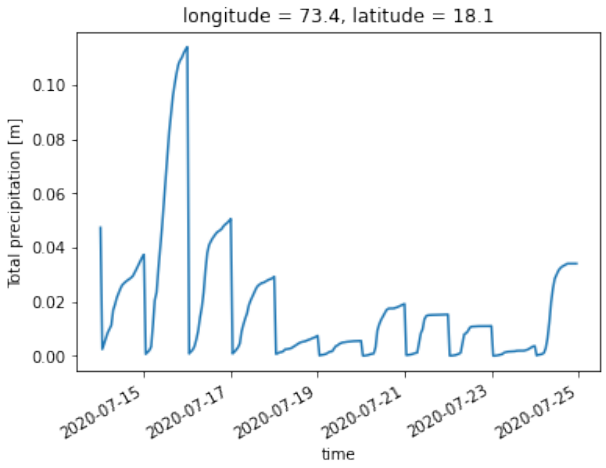
(a) Maximum

Figure 7: Maximum precipitation on days prior to July 22, 2021

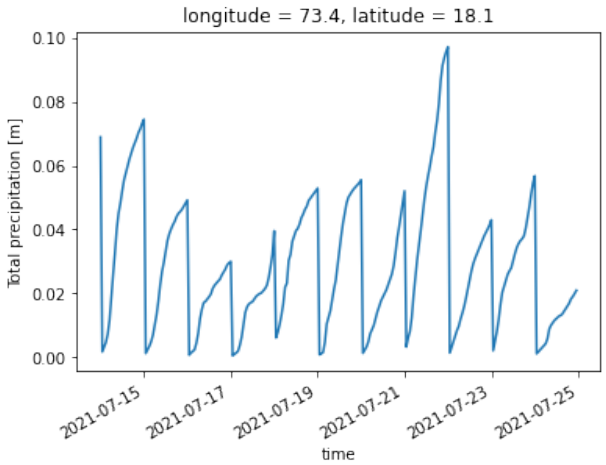
As we can see that maximum rainfall occurred on days prior to July 22, 2021. This could be the reason that the river **Savitri** got flooded sooner than observed in the previous years.



(a) 2019



(b) 2020



(c) 2021

Figure 8: Line Plot of Precipitation in Mahad in 2019, 2020, 2021

As we can see from the Figure 8, in 2021, the rainfall was consistently more than what we could see in 2019 and 2020. If the rainfall is more, the river gets flooded pretty sooner. So, once we know that the rainfall has been consistently more for a few days, we can warn the people to start taking precautions.

6 Challenges/Concerns

I had initially downloaded the data for the entire country of India. But, it would be irrelevant to the aim of this project, and the size was 400 mb in total. So, I downloaded the data pertaining to the Maharashtra state, where my hometown is.

The next steps would be sub-grouping data to find similarities between Mahad and Mahabaleshwar (where **Savitri** river originates). We will also be transforming the data for deep neural networks for forecasting. We will also see if we could find similar time series clusters.

References

[1] "Reanalysis era5 land hourly data," https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=form.

[2] "Maha: Over 9,700 houses, 45 buildings damaged in mahad due to july floods," https://www.business-standard.com/article/current-affairs/maha-over-9-700-houses-45-buildings-damaged-in-mahad-due-to-july-floods-121081301002_1.html#:~:text=A%20total%20of%2045%20buildings,occurred%20on%20July%2021%2D22.

[3] "How to use the cds api," https://cds.climate.copernicus.eu/api-how-to.

[4] "Climate data science github," https://github.com/OmkarMehta/climate-data-science.git.

[5] "Cloudburst wikipedia," https://en.wikipedia.org/wiki/Cloudburst.