Analyzing the impact of COVID-19 on the environment in the Indian subcontinent

Blue Sky Analytics - DS Take Home Assignment

By Aditya Sengupta

Objective - To analyze the effect of COVID-19 and associate changes in the national and regional air and water quality and regional vegetation in India.

Before analyzing the effects, we need to probe into the exact variability of COVID cases and the lockdown periods, which is shown in figure 1 and figure 2 below.

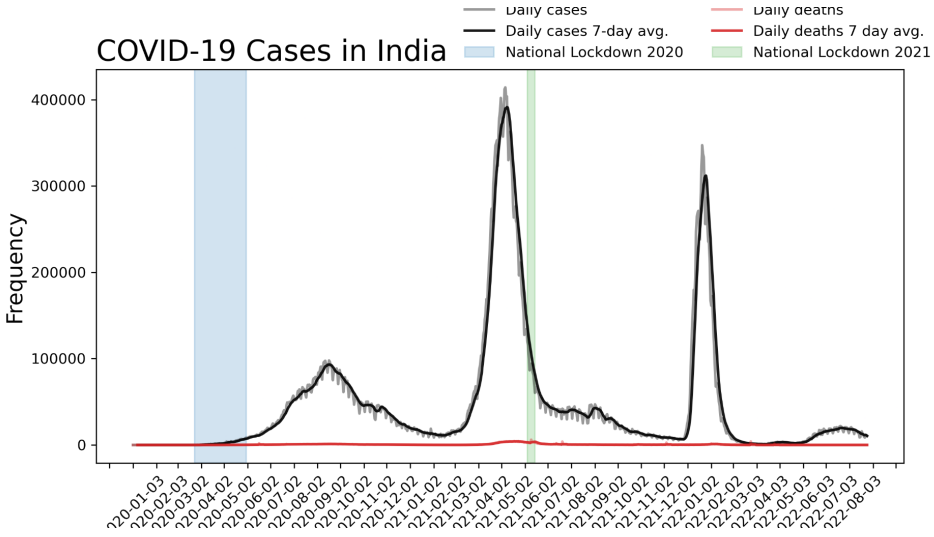

Fig 1. Covid-19 cases in India and the lockdown periods from 2020 to 2022

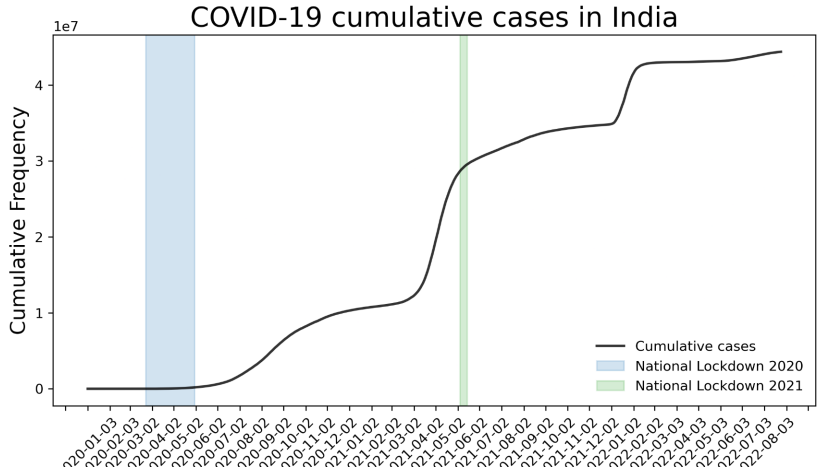

Fig 2. Cumulative cases of Covid-19 in India from 2020-2022

The first lockdown in 2020 lasted in the March-April-May (MAM) months and will be referred from here as the lockdown MAM season. The analysis focuses on the variations between 2019 MAM months (pre-lockdown) and 2020 MAM lockdown season to determine the impact that COVID-19 has had on the various environmental parameters.

AOD Variability

Data Source - LAADS DAAC (Satellite derived geospatial data)

The 2019 MAM and 2020 MAM season of AOD are compared (shown in figure 3) and the results clearly indicate that the AOD declined in 2020 lockdown season.

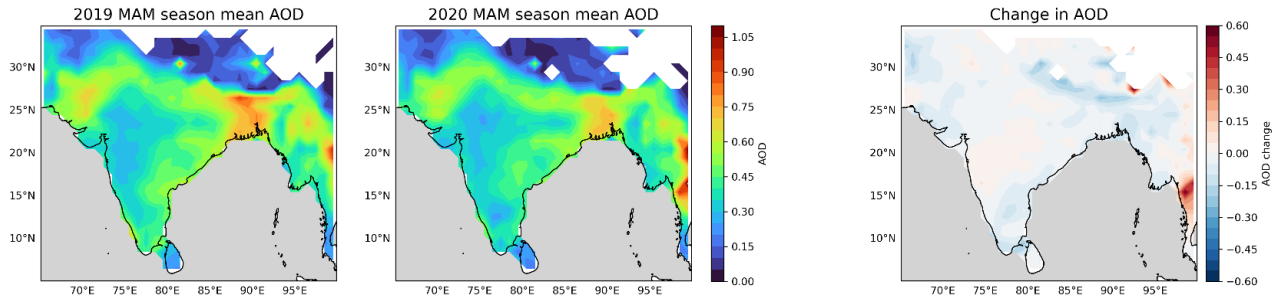

Fig 3. Comparison of AOD in the 2020 and 2019 MAM season.

To further highlight, this is not just by proxy of comparison with the year 2019, the anomalies of both the years were determined against the long-term average (2015-2021) of AOD and the results show that the negative anomalies during the 2020 lockdown season is much higher and consistent over the Indian subcontinent.

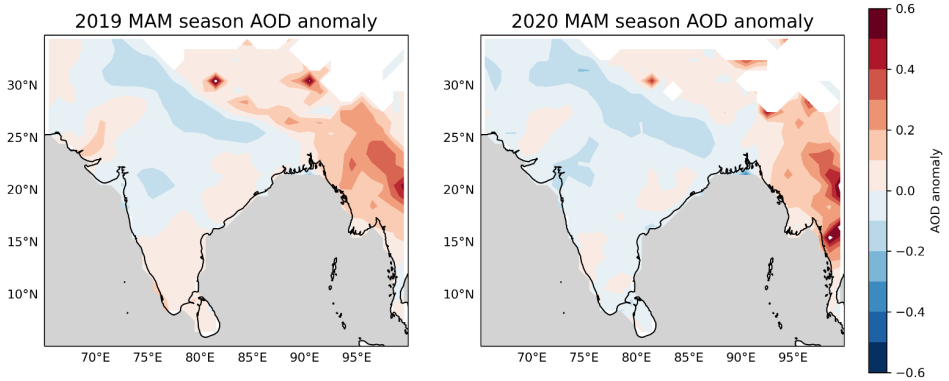

Fig. 4. AOD Anomaly of 2019 and 2020 MAM season against the 2015-2021 long-term mean

Following this, the time-series of AOD values (over the entire Indian subcontinent) and the associated kernel density estimates were analyzed -

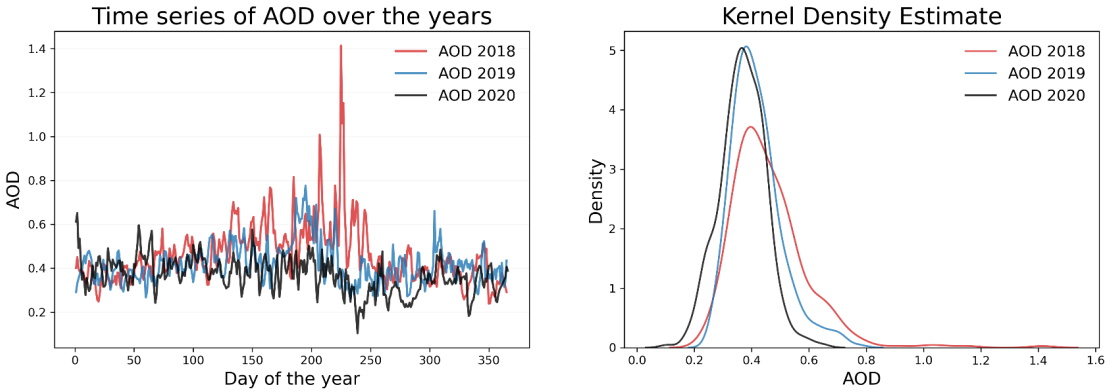

Fig. 5. Tim series and Kernel Density Estimates of the AOD values for 2018, 2019 and 2020

From the results in figure 5, it is evident from the analysis of AOD retrievals that the post-covid year of 2020 and the lockdown period has been associated with lower AQI when compared with the AOD values of pre-covid years, be it 2019 or 2020. To further confirm that this reduction in the AQI of the 3 sample years is not by change, a non-parametric Friedman's Chi square test is employed on the three annual AOD samples of 2018, 2019 and 2020. The p-value of the test was determined to be 2.623809645595713e-25. Hence, the p-value << 0.001 indicating rejecting the hypothesis that this change in AOD variability is by chance and is statistically significant. The Friedman' test is similar to the ANOVA test, but it is advantageous in the sense that it is non-parametric and doesn't assume any particular distribution of the samples beforehand.

CO, HCHO and NO2 variability

A similar analysis if further carried out for individual pollutants of CO, HCHO and NO2. The Source of the data is from the TROPOMI satellite retrievals which have been preprocessed and available at [TROPOMI QA4ECV]. The CO and HCHO data were available in NETCDF formats, but the NO2 data was in ESRI grid format for download as monthly means.

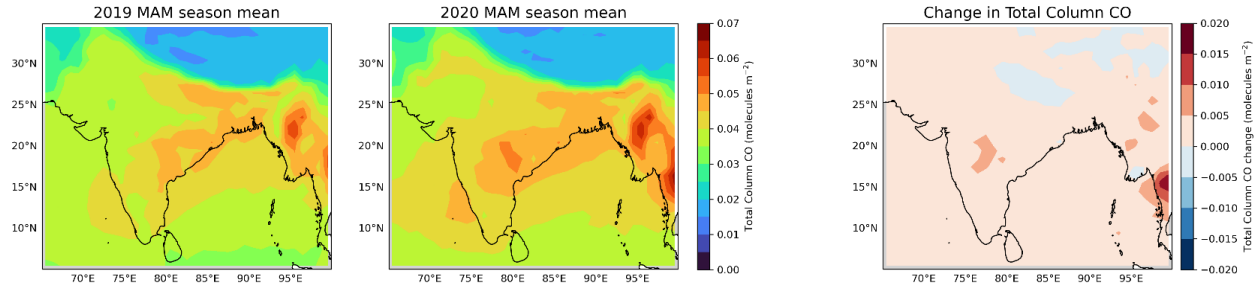

Fig. 6. Comparison of CO in the 2020 and 2019 MAM season.

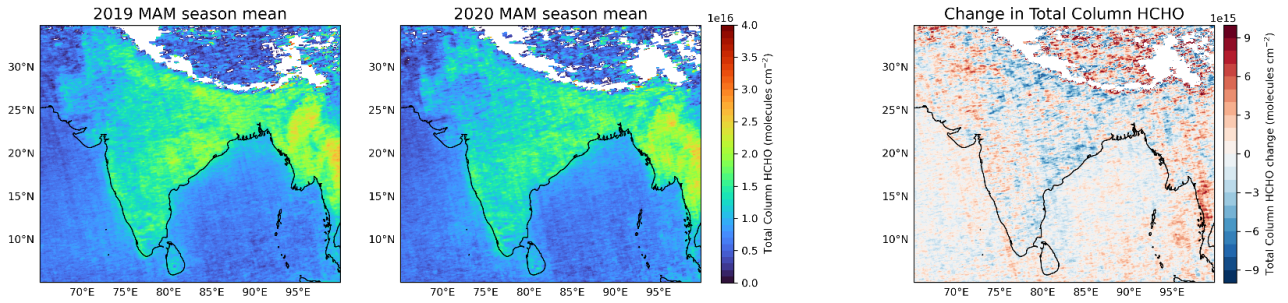

Fig. 6. Comparison of HCHO in the 2020 and 2019 MAM season.

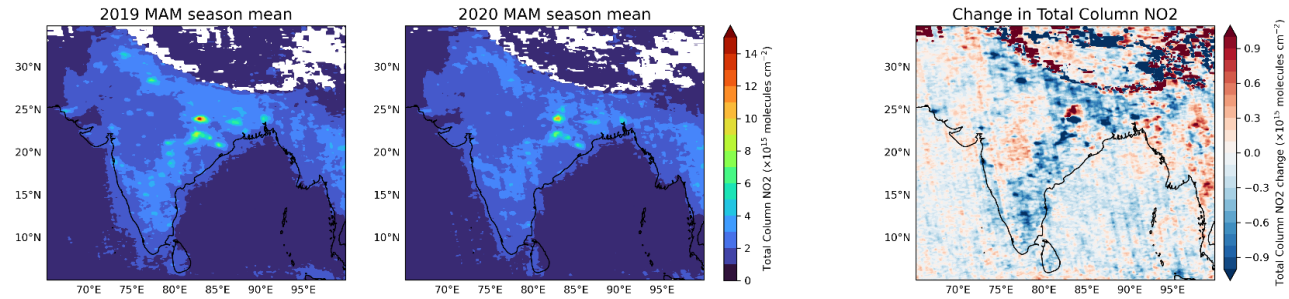

Fig. 6. Comparison of NO2 in the 2020 and 2019 MAM season.

It can be concluded here that the concentrations of tropospheric column CO have marginally increased, while the changes in HCHO concentrations are not clearly significant. But for NO2 there is a significant reduction in the concentrations, especially in the Southern Indian Peninsula and the Indo-Gangetic Plain.

Using Google Earth Engine, a similar analysis to compare the 2019 and 2020 concentrations of NO2, SO2 and Aerosol Index with the TROPOMI Sentinel-5P product is done.

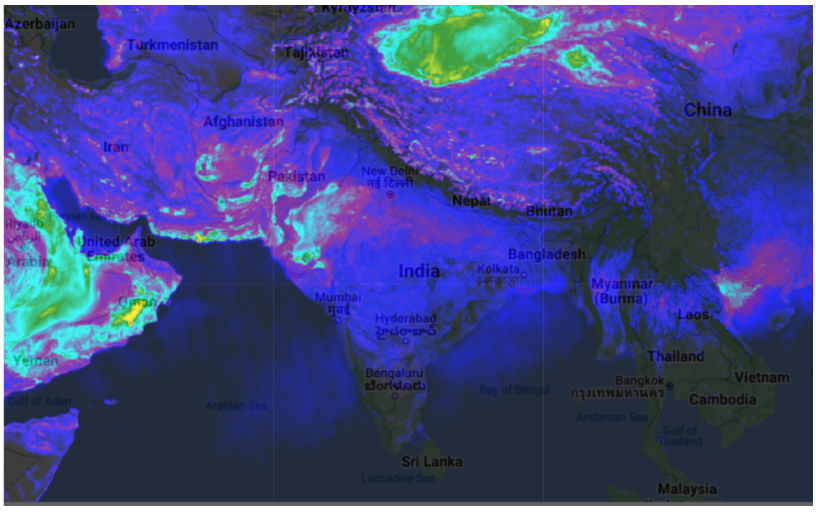

Fig. 7. TROPOMI S5P Aerosol Index 2019.

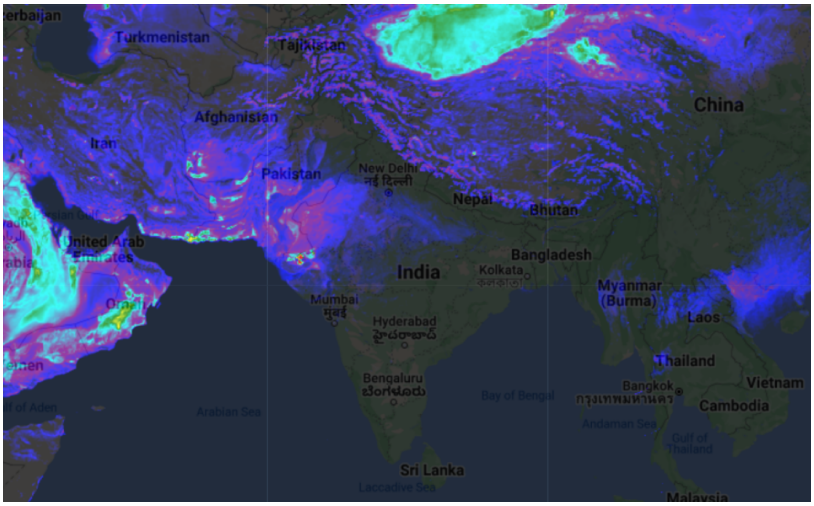

Fig. 8. TROPOMI S5P Aerosol Index 2020.

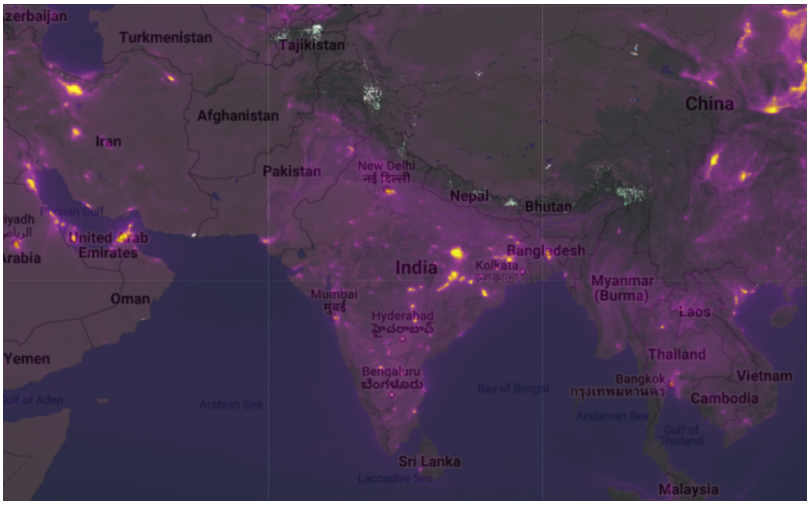

Fig. 9. TROPOMI S5P NO2 2019.

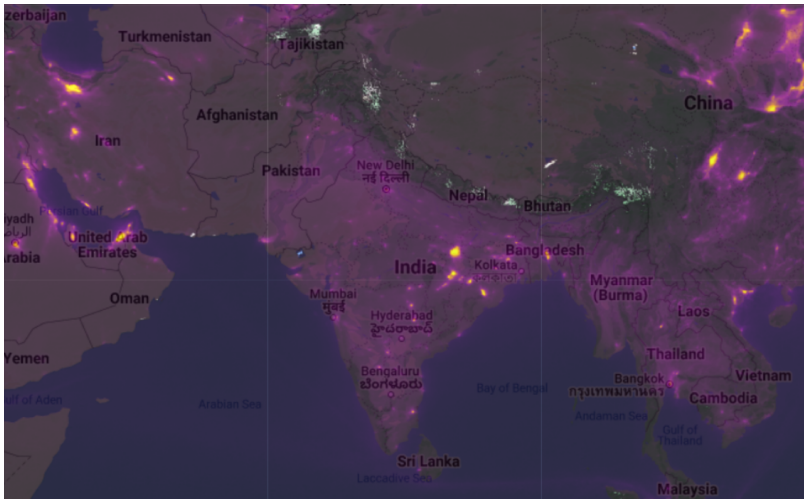

Fig. 10. TROPOMI S5P NO2 2020.

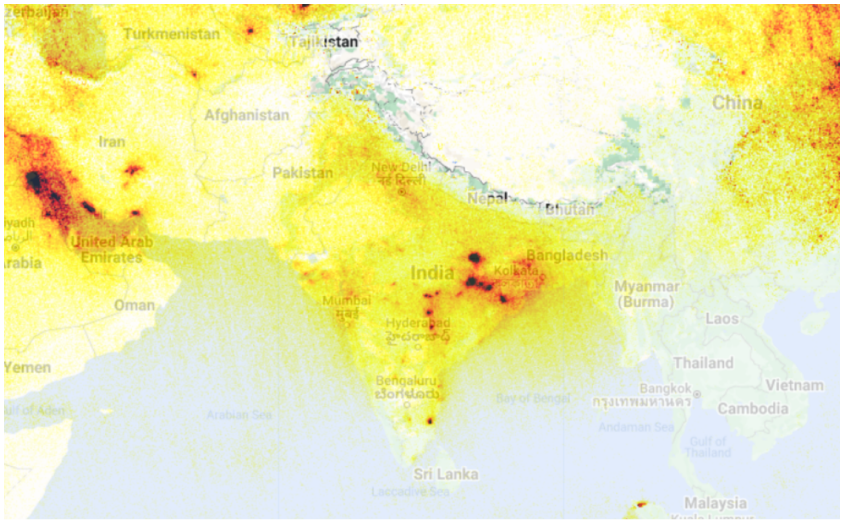

Fig. 11. TROPOMI S5P SO2 2019.

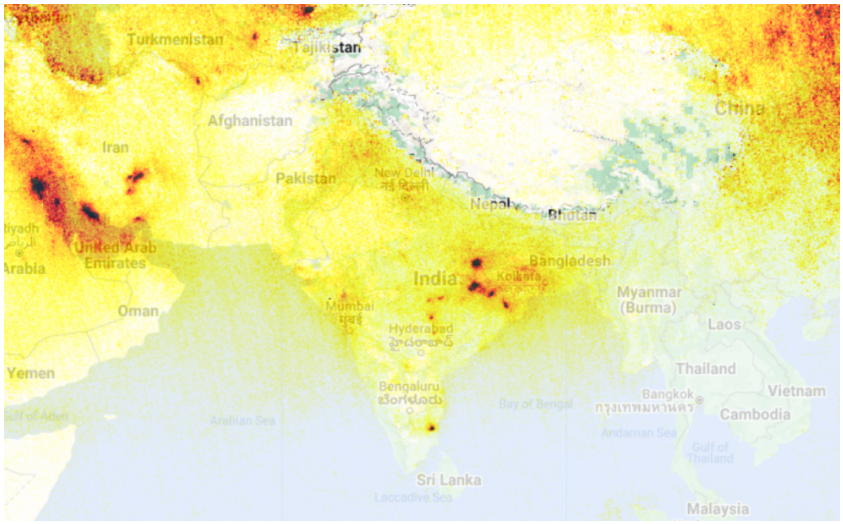

Fig. 12. TROPOMI S5P SO2 2020.

The satellite imagery for the 2020 MAM season for all three variables SO2, NO2 and Aerosol Index are have many regions with lower concentrations. It is evident that air quality variables were lower in concentrations during the lockdown owing to lower industrial activity and vehicular emissions.

Water Quality

Water quality was analyzed using the Google Earth engine and Sentinel-2 MSI : MultiSpectral Instrument, Level-2A bands to calculate the value of Normalized Difference Chlorophyll Index (NDCI), which varies between -1 and 1 and higher the value (reddish color), higher is the turbidity and poor is the quality of water.

NDCI = (band5 - band4) / (band5 + band4)

The analysis is done for the Ganges in the Bhagalpur region for pre-lockdown (March 2019) and post-lockdown (April-May 2019) period.

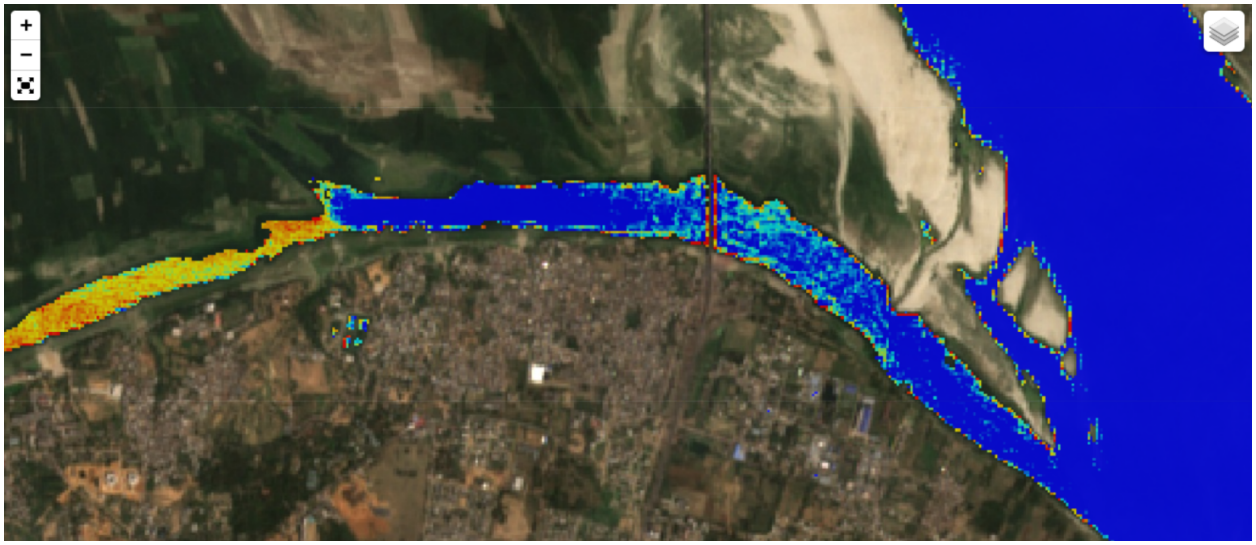

Fig. 13. NDCI values in the pre-lockdown phase in the Ganges sector near Bhagalpur district

Fig. 14. NDCI values in the post-lockdown phase in the Ganges sector near Bhagalpur district

Further analysis over other branches of the Ganges and other rivers can be carried out but are left out from this analysis largely due to time constraint.

Vegetation

For analysis of impact of CoVID lockdown on vegetation, we used MERRA tropospheric reanalysis data for the satellite era produced by NASA Global Modeling and Assimilation Office (GMAO) using the Goddard Earth Observing System Model (GEOS).

The monthly mean data used to analyze the differences in the 2019 and 2020 MAM seasons, which was the lockdown period in the year 2020. The variables of Surface soil wetness, available soil water storage and soil infiltration rate were analyzed to determine the vegetation productivity by proxy. Higher values of each indicate better agricultural productivity and higher fertility for vegetation.

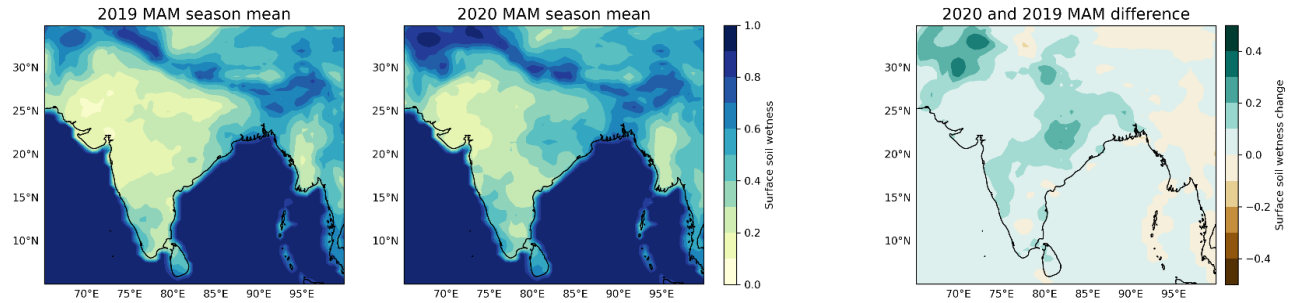

Fig. 15. Comparison of Surface soil wetness in the 2020 and 2019 MAM season.

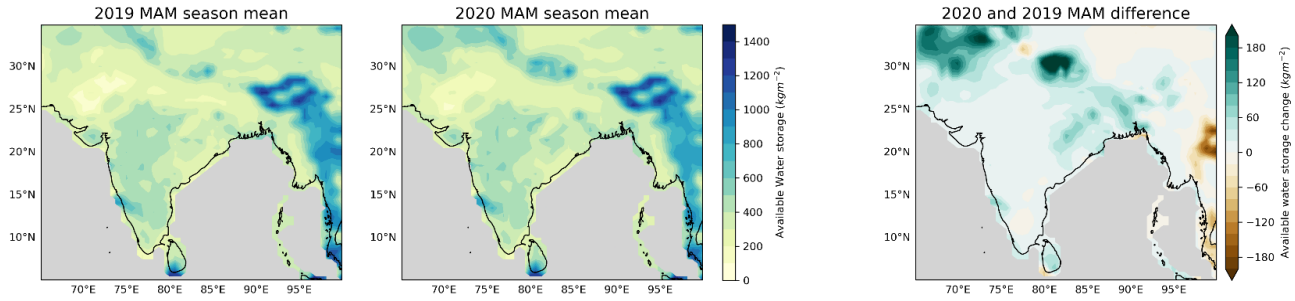

Fig. 16. Comparison of Available Water storage in the 2020 and 2019 MAM season.

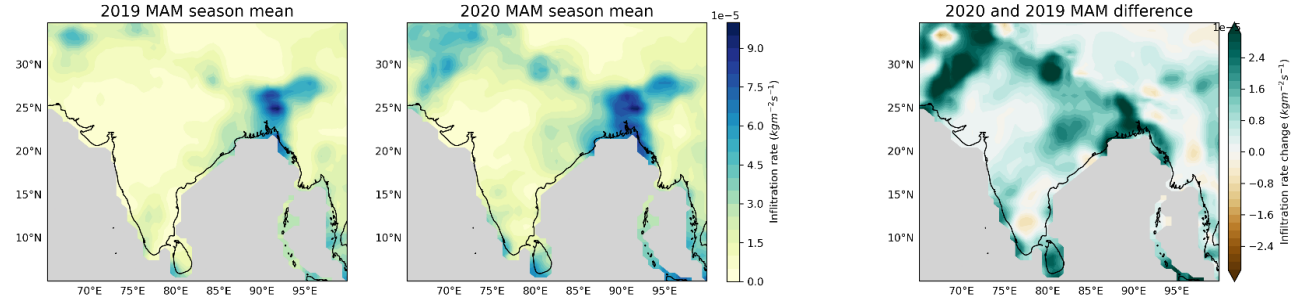

Fig. 17. Comparison of Soil infiltration rate in the 2020 and 2019 MAM season.

The results of the analysis clearly indicate an increase in all three proxies in the lockdown season of 2020. Hence, an improvement in soil fertility owing to lockdown is evident.

NDVI analysis was also carried out in the Sundarban region (dense forest) and Haryana and Punjab (known for agricultural productivity) to determine the impact of CoVID on vegetation in these regions using satellite imagery.

Data source - the satellite imagery was downloaded from the USGS Earth Explorer website, and we used Landsat-8 Bands 4 (red) and 5 (nir) for the analysis and calculation of NDVI.

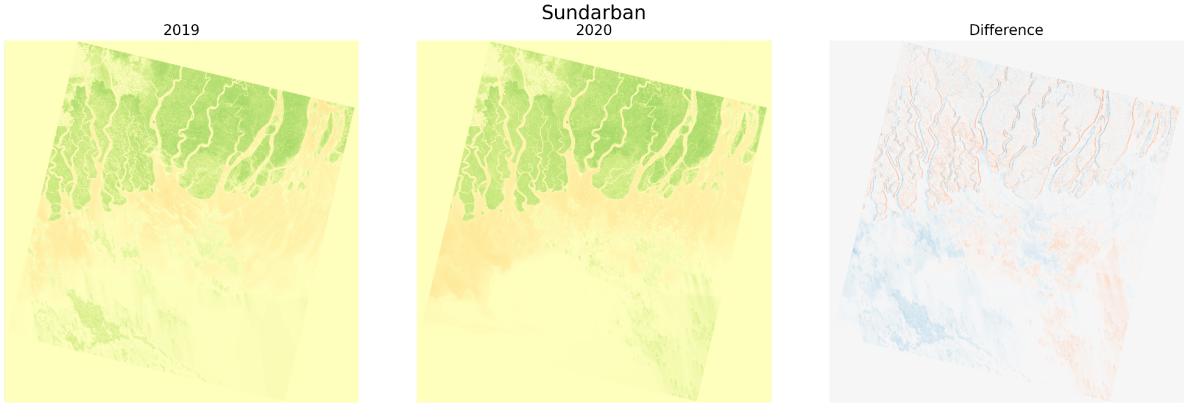

Fig. 18. NDVI variability in Sundarban forests.

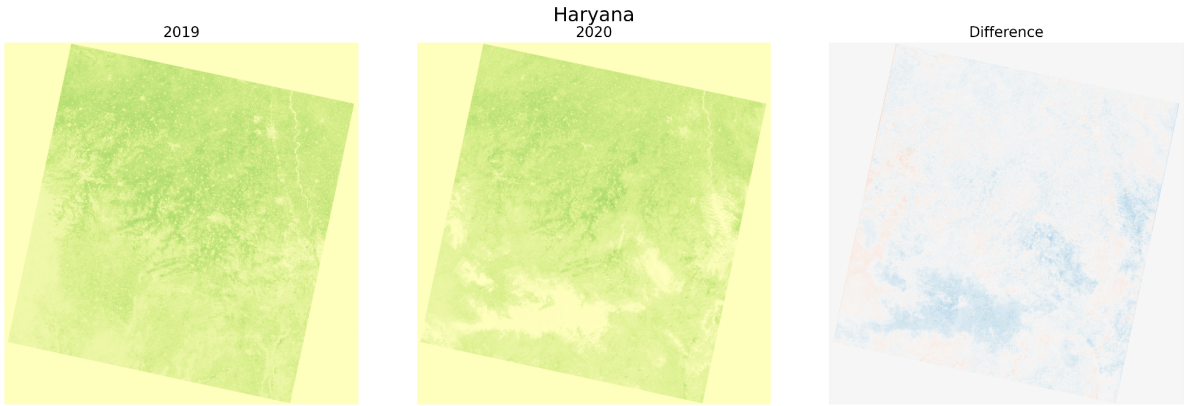

Fig. 19. NDVI variability in Haryana.

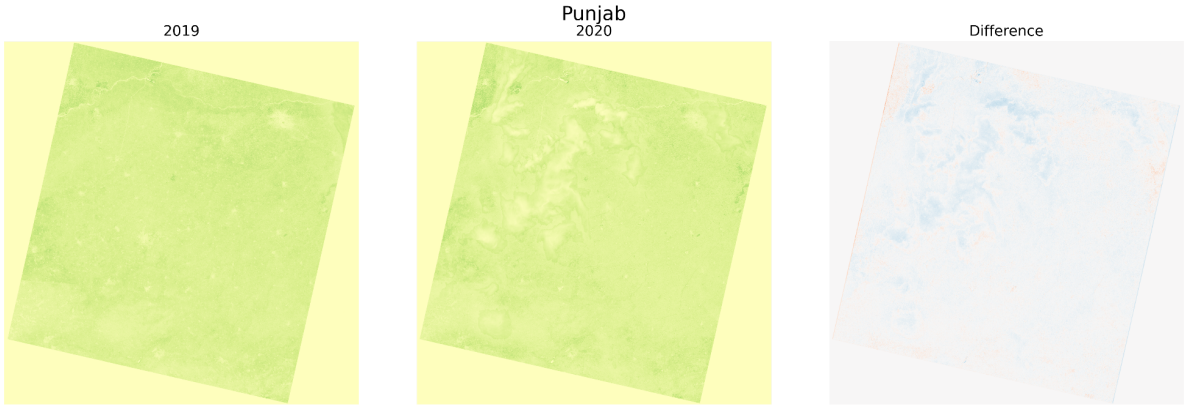

Fig. 18. NDVI variability in Punjab.

The results of our analysis compare the 2019 MAM and 2020 MAM seasons and highlight the Sundarban densely forested area NDVI increased during lockdown in 2020, while the agricultural lands in Haryana and Punjab actually saw a marginal decline in NDVI.

The overall results of the analysis -

- It is clear that there has been a significant decline in AOD and aerosol concentrations in the Indian subcontinent during and post-lockdown. But, not all pollutant species show a decline, such as CO and HCHO.

- Water quality in the Bhagalpur region improved post-lockdown, owing largely to reduced industrial activities. Further analysis over other sectors of the Ganges or other rivers can be carried out to further improve the results

- The soil quality and fertility proxies indicate an improvement in soil water storage and infiltration during CoVID lockdown.

- While the vegetation density seems to improve in the dense Sundarban region, the agricultural lands of Haryana and Punjab actually show a decline in NDVI and vegetation density.

Use of Geospatial data -

- We used Sentinel 5P and Sentinel-2 MSI satellite data for analysis of Air and Water quality.

- We utilized VIIRS AOD retrievals for analysis of AOD variability in the Indian subcontinent

- We used Landsat-8 satellite imagery data from USGS Earth explorer to determine NDVI.

- We used MERRA Reanalysis netcdfand QA4ECV netcdf data which is 3-D geospatial data fro the analysis of vegetation and air quality respectively.

Scope for improvement -

- Using statistical technique for determining causality in the case of COVID-19 is difficult due to lack of high quality spatial data on ground observations such as traffic and pollutant emissions. This is a data limitation, due to which more robust statistical techniques cannot be applied for determining correlation or causal relations.

- While satellite imagery data is great for local or regional analysis, a large scale analysis requires big volume of satellite data for merging into a single raster. This requires large amounts of memory and computing power, which was a major limitation in this study.