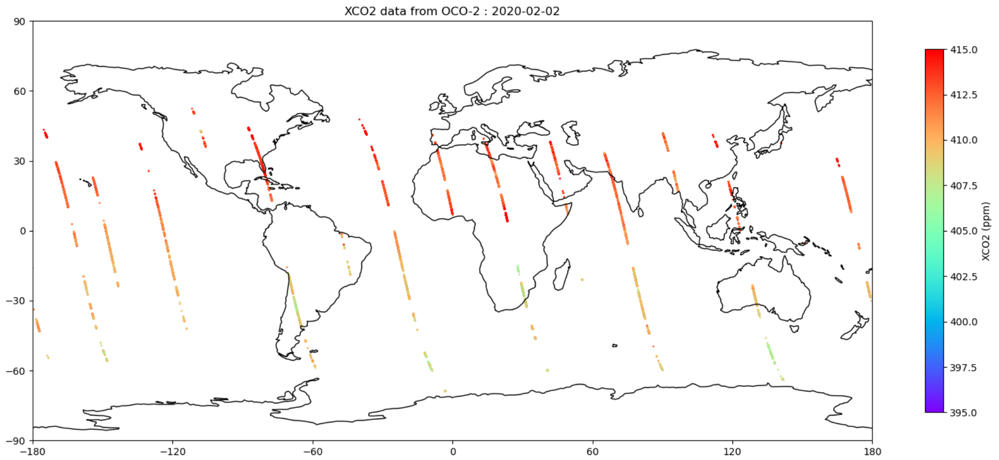
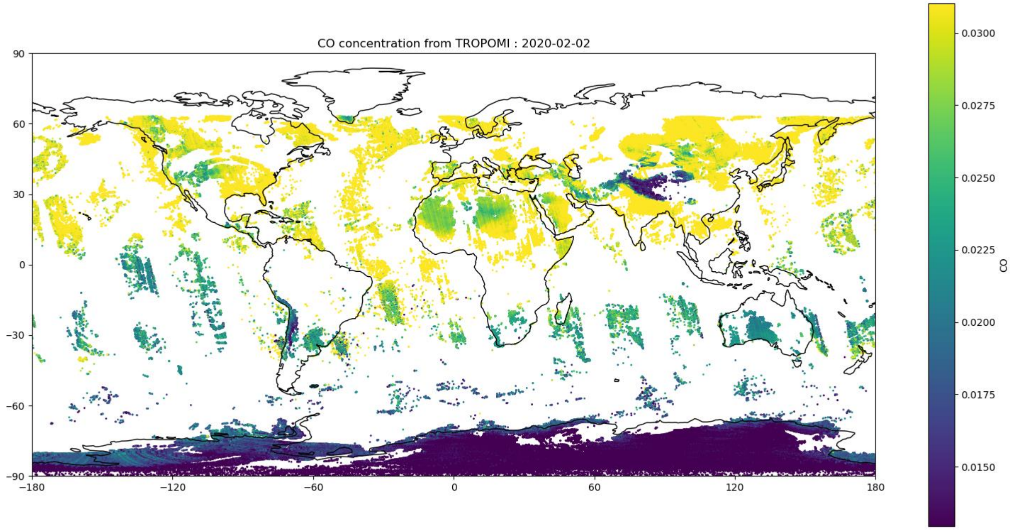
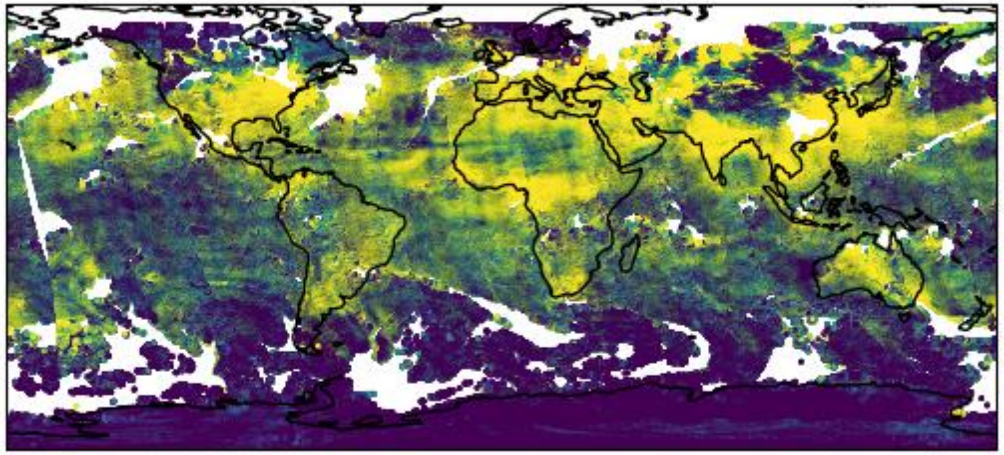
Global mapping of XCO2 retrievals from OCO-2 using machine learning with TROPOMI and ERA5 data : The new way to generate XCO2 data from incomplete sampling

Figure 1. (a) XCO2 retrievals by OCO-2 on February 2, 2020. (b) XCO measurements by TROPOMI on February 2, 2020. (c) Tropospheric Column nitrogen oxide measurements by TROPOMI on February 2, 2020.



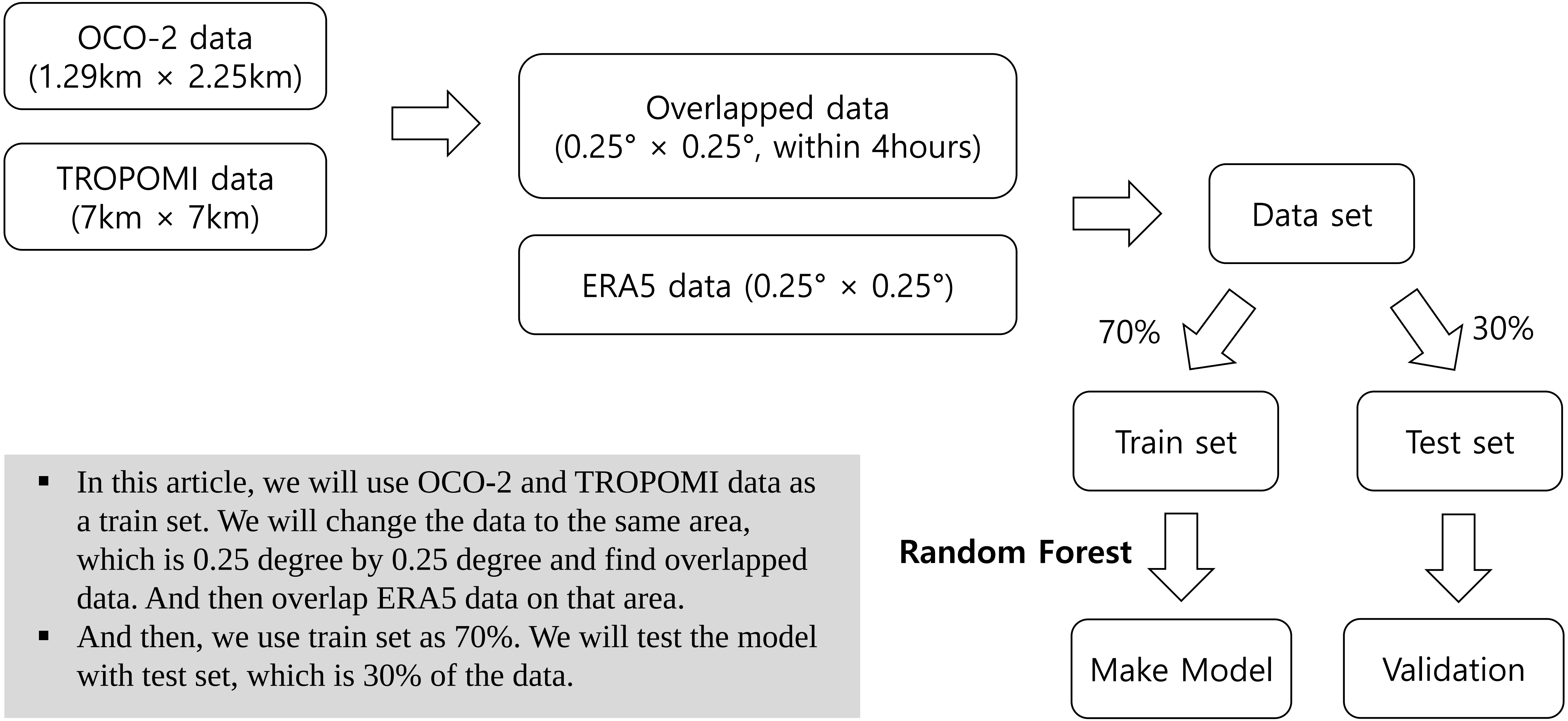
- This incomplete sampling of atmospheric CO2 prevents us from understanding the global carbon cycle.
- By using TROPOMI data which can detect more space than OCO-2, we can broaden XCO2 data spatially and temporally.





You can see my code here : https://github.com/Taeye-Kwack/Predict_CO2

Figure 2. Schematic diagram

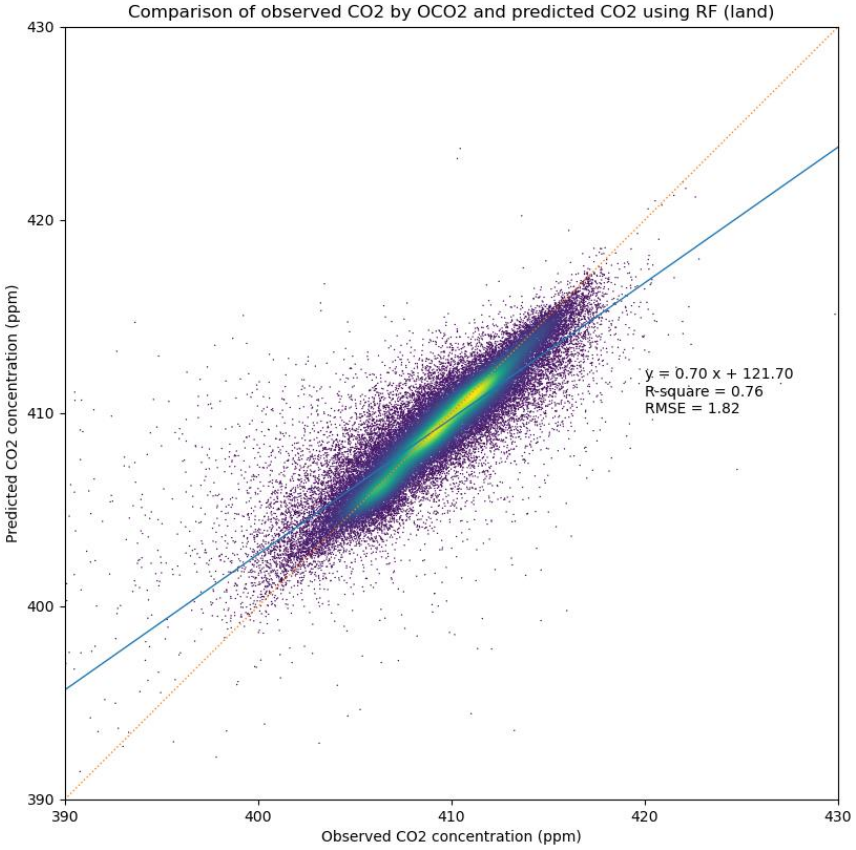


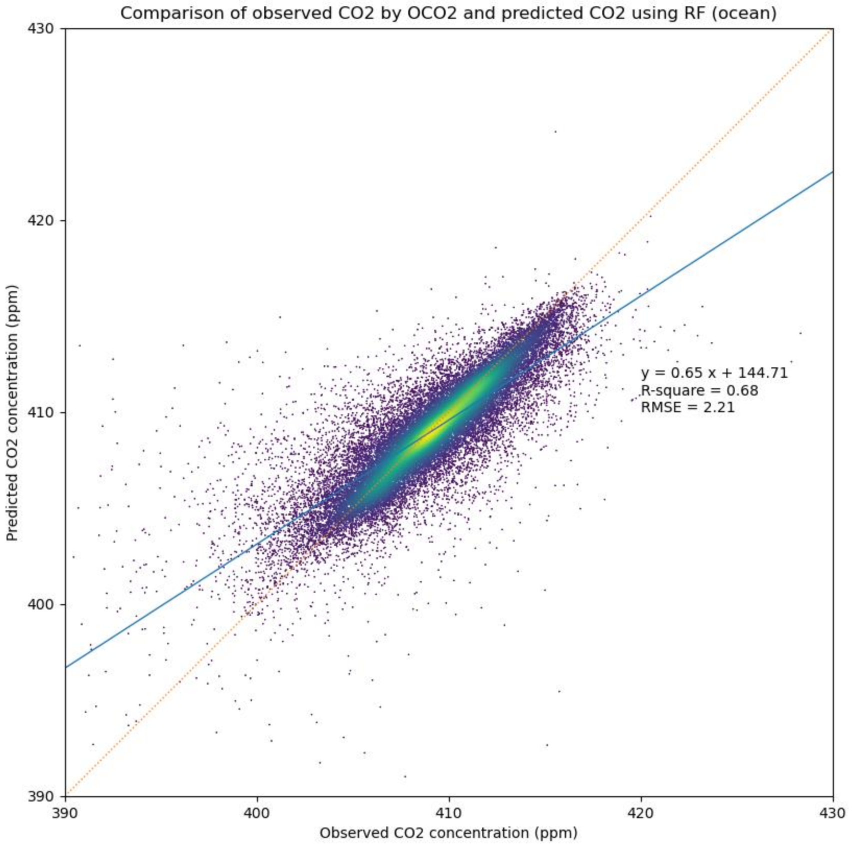
- In this article, we will use OCO-2 and TROPOMI data as a train set. We will change the data to the same area, which is 0.25 degree by 0.25 degree and find overlapped data. And then overlap ERA5 data on that area.
- And then, we use train set as 70%. We will test the model with test set, which is 30% of the data.

Table 1. Data used for training. Temporal and Spatial resolution is described.

Data	Variables	Temporal resolution	Spatial resolution
Meteorological data (ERA5 Reanalysis)	2m Temperature	Hourly	0.25° × 0.25°
	Sea Surface Temperature (for ocean)		
	Skin Temperature		
	Surface Pressure		
	10m U component of wind		
	10m V component of wind		
	100m U component of wind		
	100m V component of wind		
	500hPa U component of wind		
	500hPa V component of wind		
	500hPa Vertical velocity		
TROPOMI data	Carbon monoxide total column	Real-time	7km × 7km
	Nitrogen dioxide tropospheric column		
	Solar Induced chlorophyll Fluorescence (for land)		
	Observation time		-
OCO-2 data	Carbon dioxide total column	Real-time	1.29km × 2.25km
	Observation time		-
Gridded data	Longitude	-	0.25° × 0.25°
	Latitude		

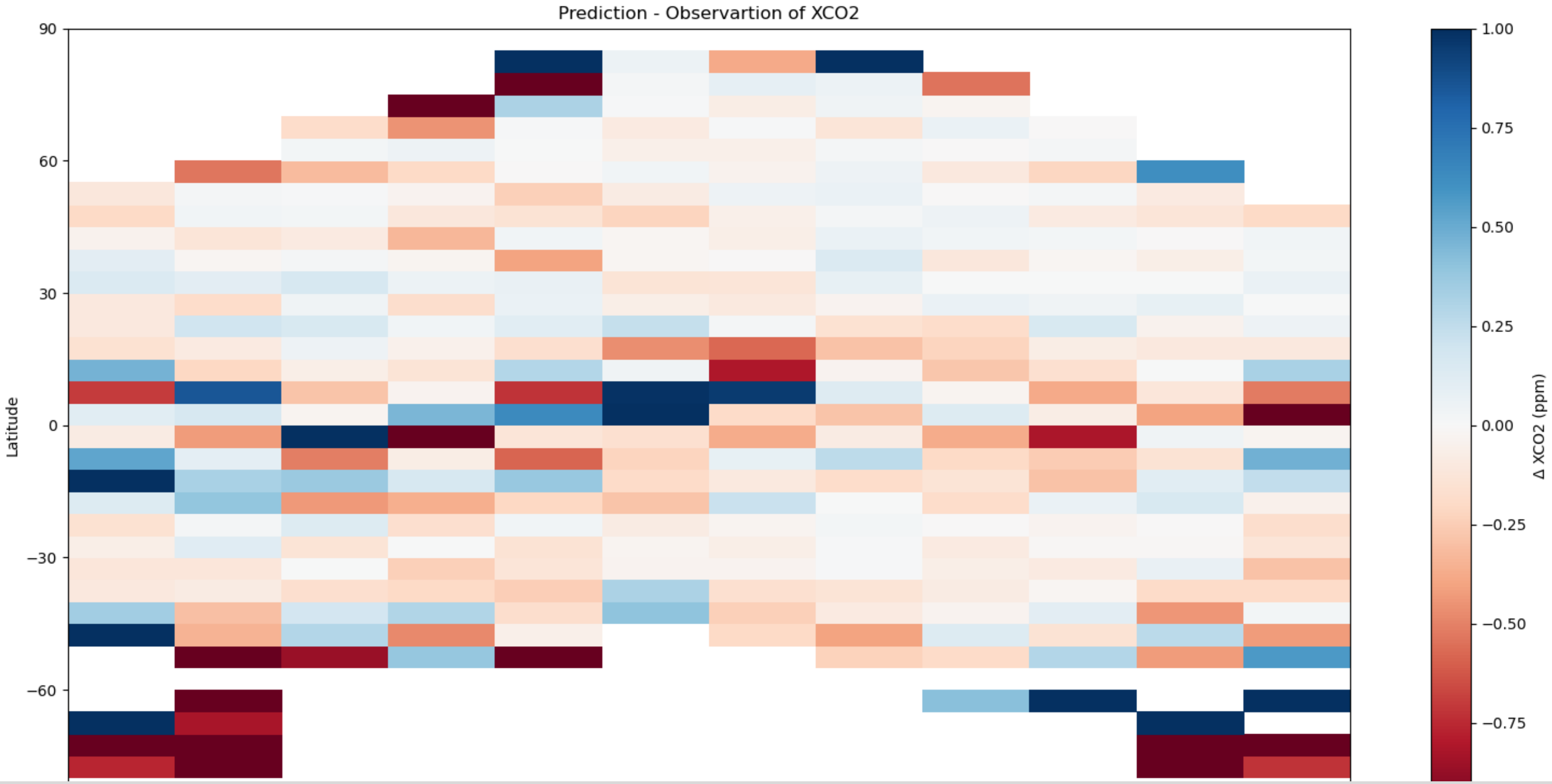
Figure 3. Model performance. X-axis is XCO2 measurements by OCO-2, Y-axis is predicted XCO2 data by model.

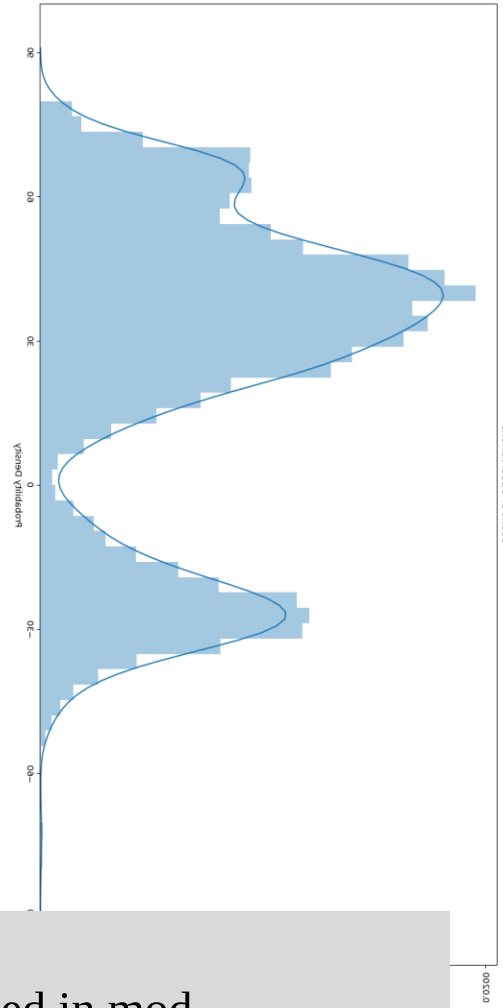




- As SIF(SST) is not exist in Land(Ocean), we have divided data land and ocean and make model separately. The results is shown Fig3. In case of land and ocean, R-square is 0.76 and 0.68, root-mean-square error (RMSE) is 1.82, 2.21.
- The closer the data is to the median, the less error there is. On the contrary, the further away f rom the median, the greater the error.

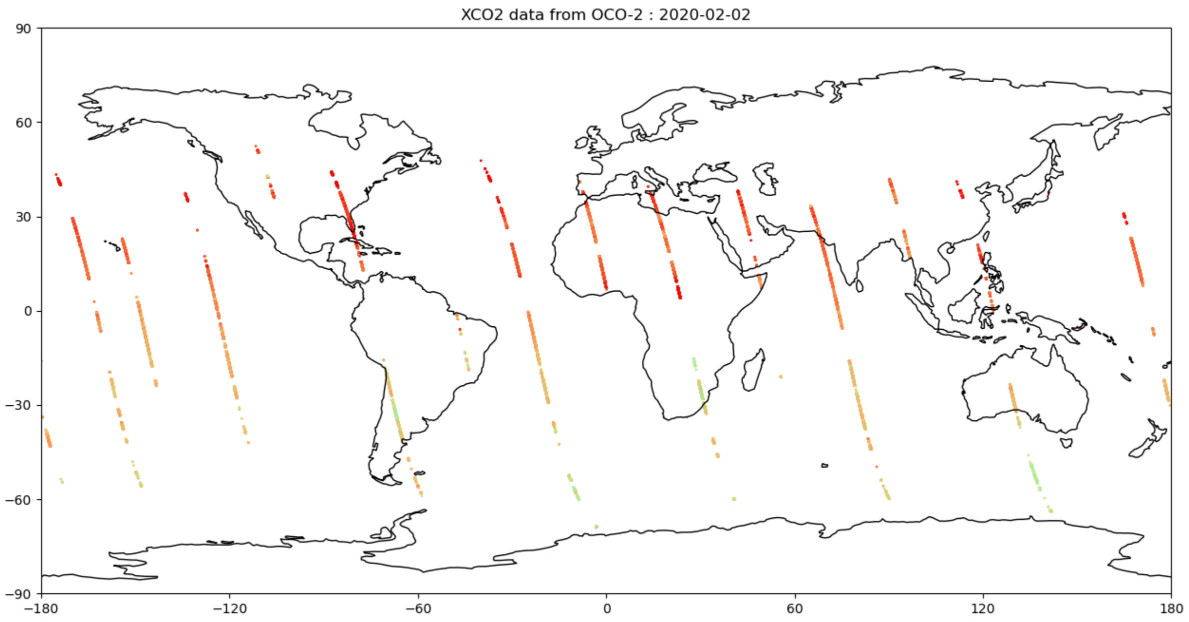
Figure 4. (a) Monthly averaged difference between the prediction of XCO2 data by model and the XCO2 measurements by OCO-2 by latitude. Latitude interval is 5 degrees. (b) Histogram of data counts by latitude. Bins are 100 and Kernel Density Estimation is also drawn.

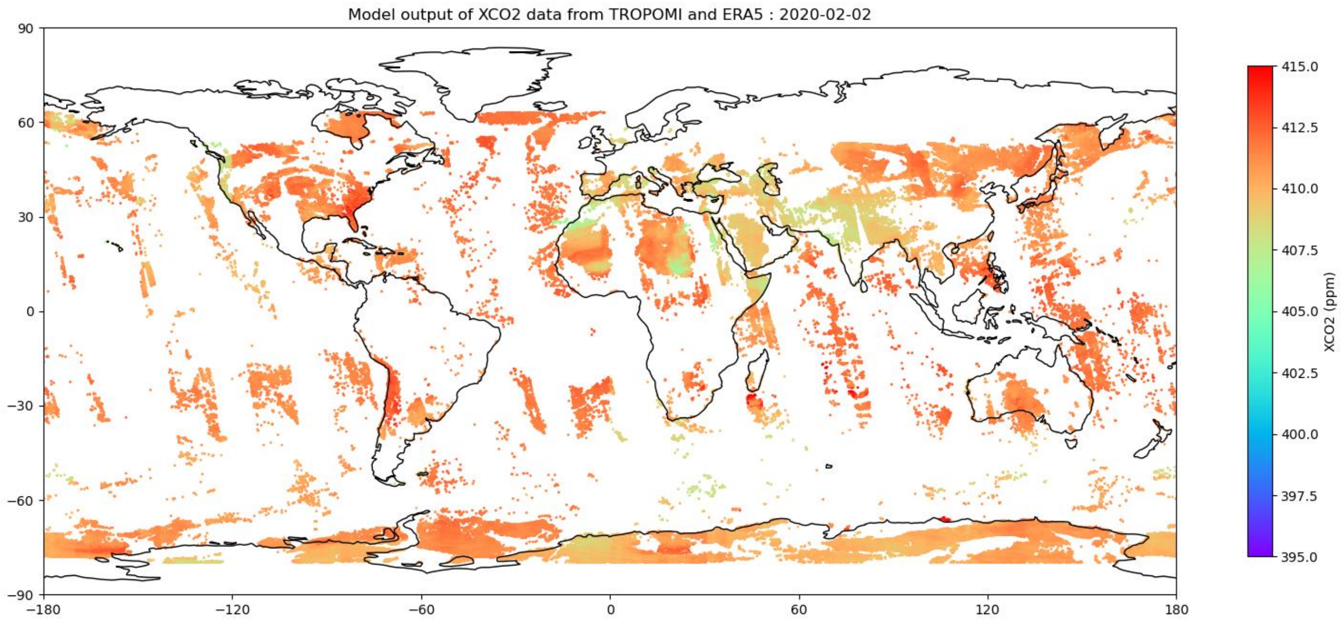




- In general, model error is within 1 ppm which is estimated error by OCO-2.
- In case of high-altitude, the model results were poor due to the small number of data used in mod el learning. This shows the need for observations in high-altitude areas.

Figure 5. Spatial distribution of (a) XCO2 measurements by OCO-2 (b) model output on February 2, 2020.





- Fig 5 shows spatial extension of XCO2 data on February 2, 2020. By using random forest model, we can generate more data due to broad measurements by TROPOMI.