INTERCOMPARISON OF THE ESA SENTINEL-5P NO_2 TROPOSPHERIC COLUMN NUMBER DENSITY PRODUCT AGAINST IN-SITU GROUND MEASUREMENTS

MGR INŻ. PATRYK GRZYBOWSKI

INTERDISCIPLINARY DOCTORAL SCHOOL,

INSTITUTE OF GEOPHYSICS, FACULTY OF PHYSICS, UNIVERSITY OF WARSAW

DR HAB. KRZYSZTOF MARKOWICZ, PROF. UW, INSTITUTE OF GEOPHYSICS, FACULTY OF PHYSICS, UNIVERSITY OF WARSAW

DR JAN MUSIAŁ, CLOUDFERRO

PHD RESEARCH TOPIC AND HYPOTHESIS

PhD research topic:	Potential of using Sentinel-5P satellite data to support air pollution monitoring in Poland, using computing cloud
Hypothesis:	Sentinel-5P (S-5P) satellite data combined with meteorological data can be used to support the air quality monitoring system in Poland

MOTIVATION

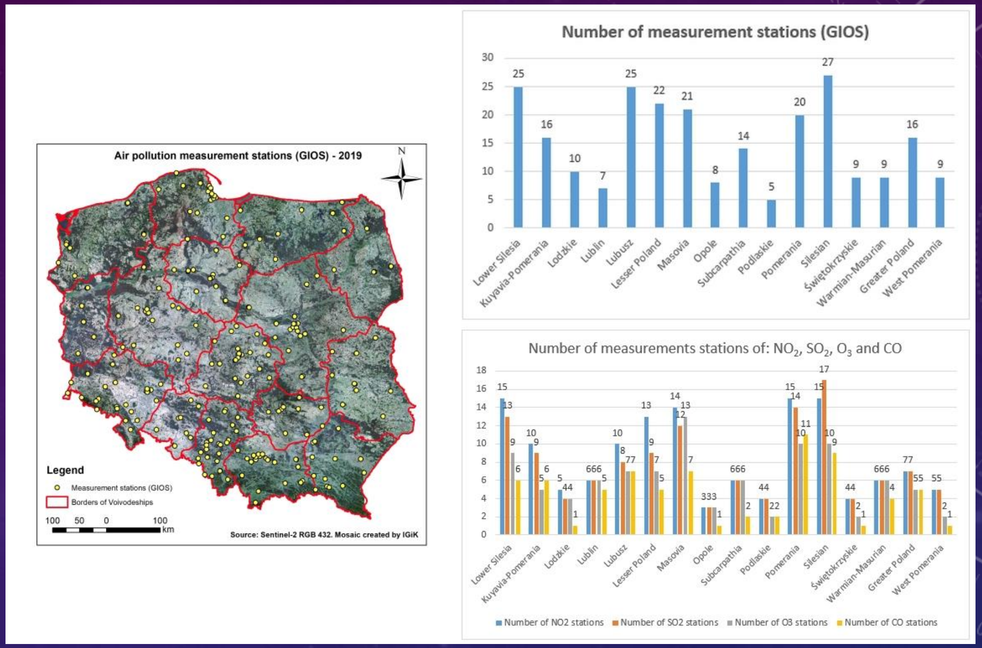
Mission has started in October 2017

Data has beeing provided from July 2018

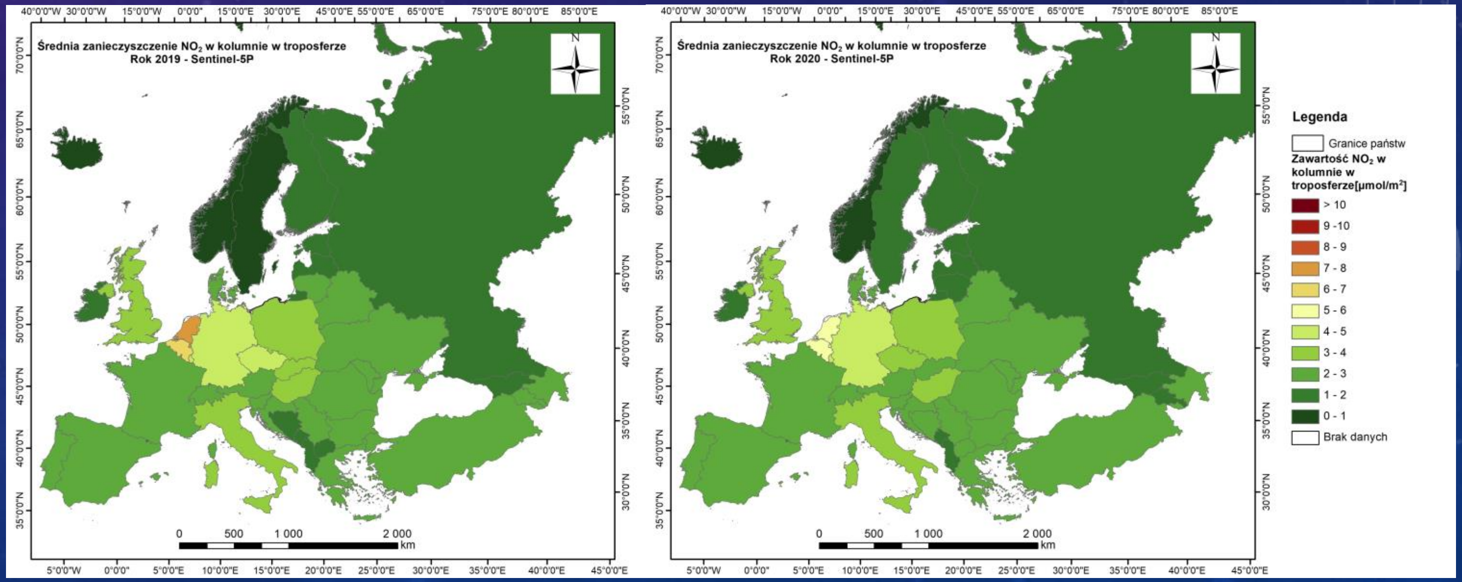
Spatial data

Supplementing monitoring with data from non-urban regions

Poor air quality over Poland



National system of air quality monitoring in Poland.



NO_2 TVCD over Europe – average NO_2 level in 2019 (left) and 2020 (right)

SENTINEL-5P PRECURSOR

Sentinel-5P

Passive sensor

Start mission

- 13th of October 2017

Temporal resolution – 1 day (at the same point in every 17 days)

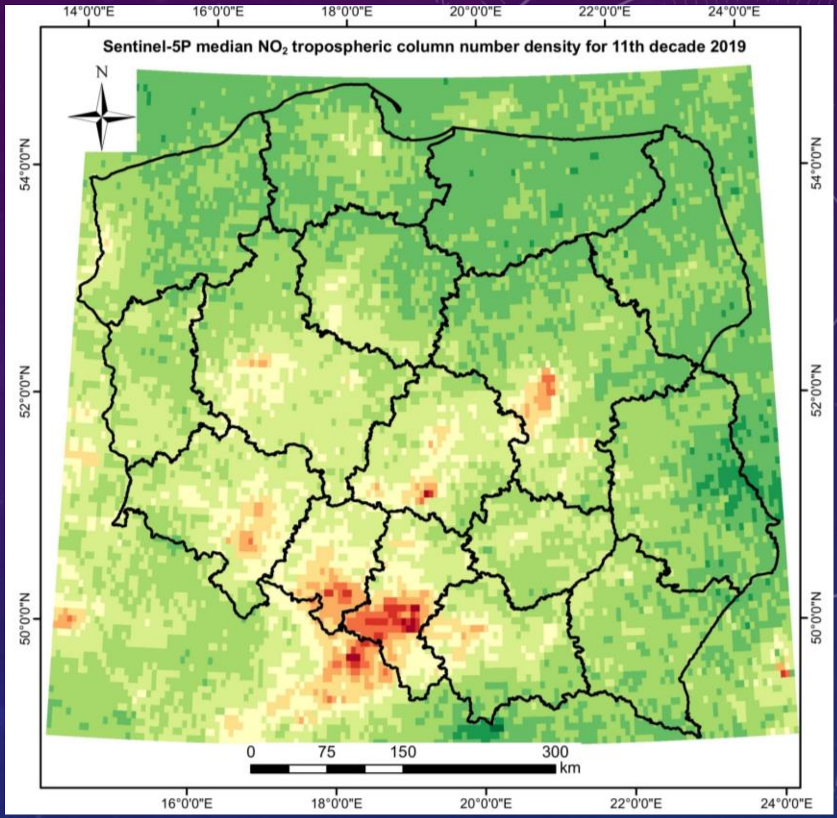
Spatial resolution – 3,5kmx7km, 7kmx7km (od sierpnia 2020 – 3,5kmx5,5km)

Sensor: TROPOMI

Objectives

- Monitoring of the atmosphere and air pollution

	UV		UVIS		NIR		SWIR	
Band	1	2	3	4	5	6	7	8
Range [nm]	270-300	300-320	320-405	405-465	675-725	725-775	2305-2345	2345-2385

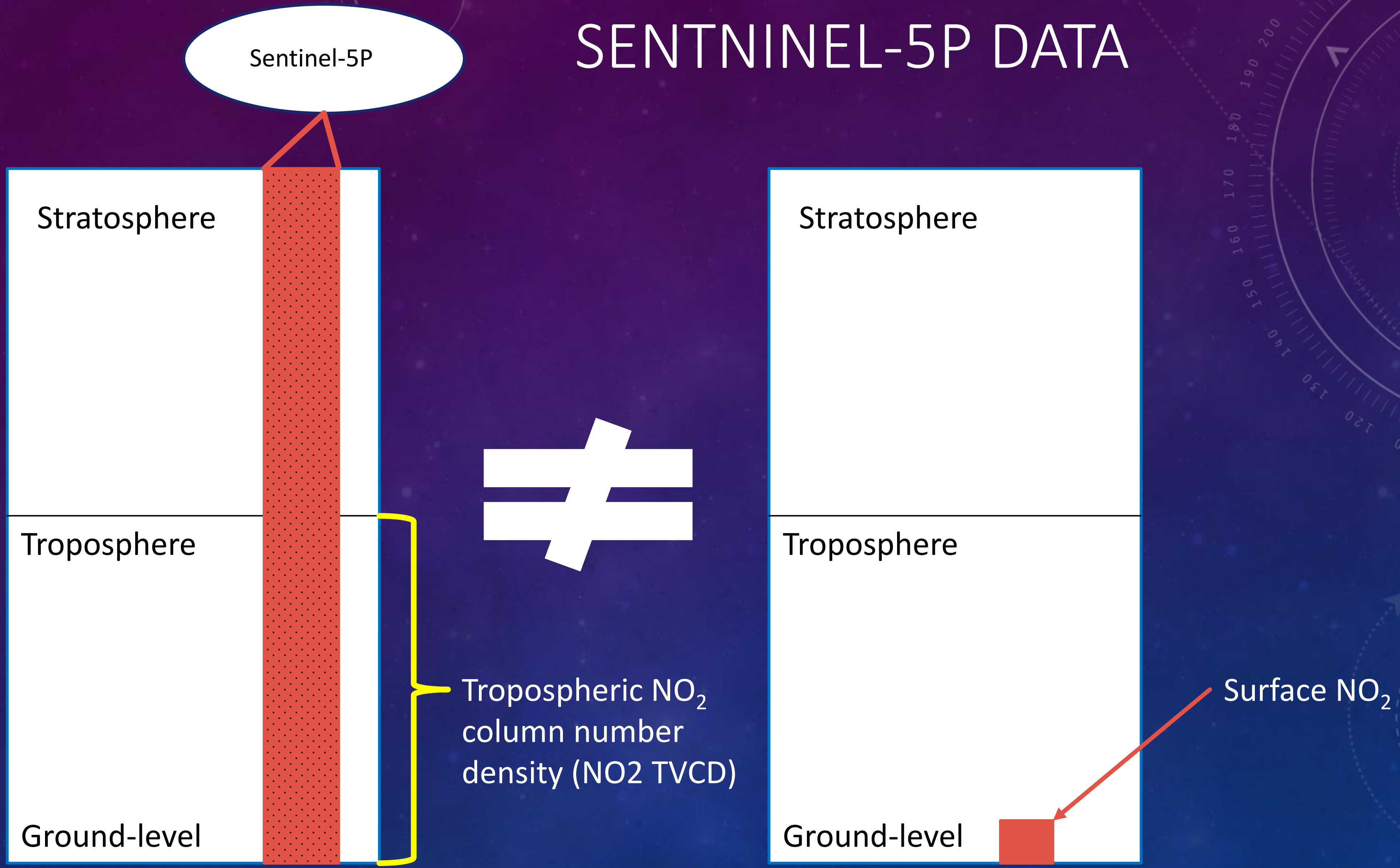


Products	
	Nitorgen dioxide
	Sulphur dioxide
	Carbon monoxide
	Methane
	Aerosols
	Formaldehydes
	Ozone
	Clouds

RETRIEVAL AND GENERATION OF THE PRODUCT

- 400-465 nm used for NO2 retrieval;
- Generated by The Royal Netherlands Meteorological Institute (KNMI);
- Procedure:
  - Using Differential Optical Absorption Spectroscopy (DOAS) method NO_2 slant columns from the measured radiance and irradiance spectra is retrieved;
  - Separation of tropospheric and stratospheric columns. Conversion to tropospheric and stratospheric slant columns;
  - Conversion of tropospheric and stratospheric slant columns into tropospheric and stratospheric vertical column density (Boersma et al., 2011; van Geffen et al., 2015).
  - Vertical profiles of NO_2 are calculated for the center of a pixel featuring a spatial resolution of 1° × 1° and are based on the chemistry transport model (CTM)—TM5-MP (van Geffen et al., 2015; Williams et al., 2017).
  - Finally, filtering of cloud cover is performed by the Fast Retrieval Scheme for Clouds from the Oxygen A-band (FRESCO-S) algorithm (Loyola et al., 2020).
- Within this study, reprocessed Sentinel-5P TROPOMI NO_2 product (S5P-PAL) data was used (version 2.3.1)

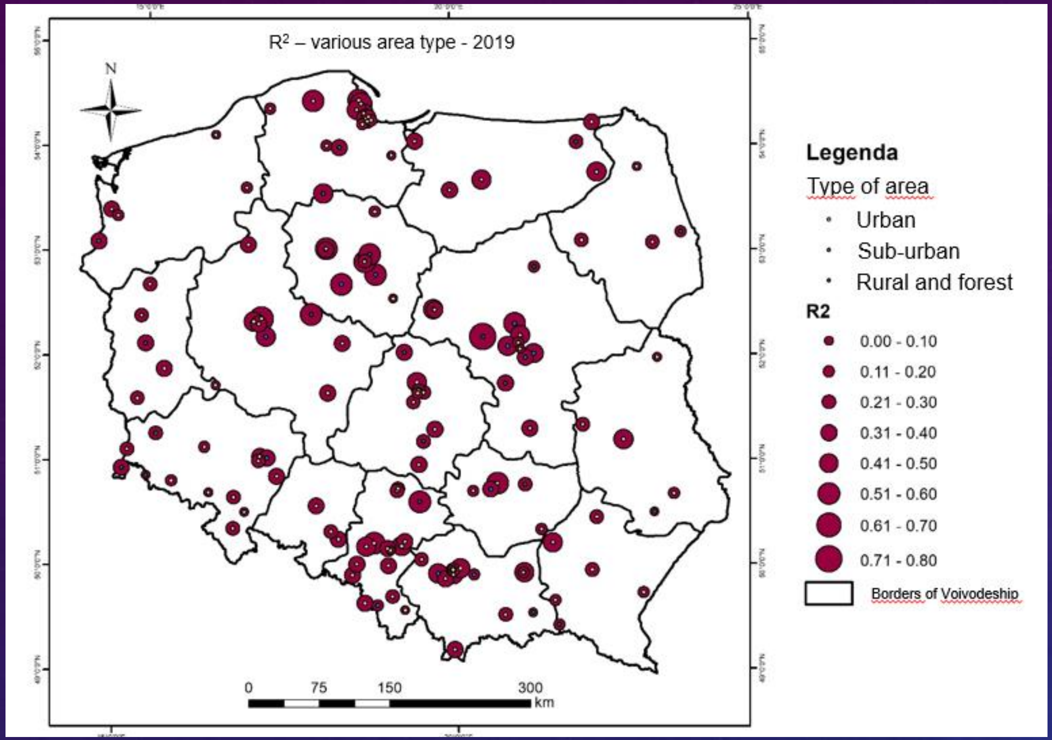
SENTNINEL-5P DATA



Satellite data – Sentinel-5P [μmol/m^2]

Ground measurementes - GIOŚ [μg/m^3]

Problems? Challneges?

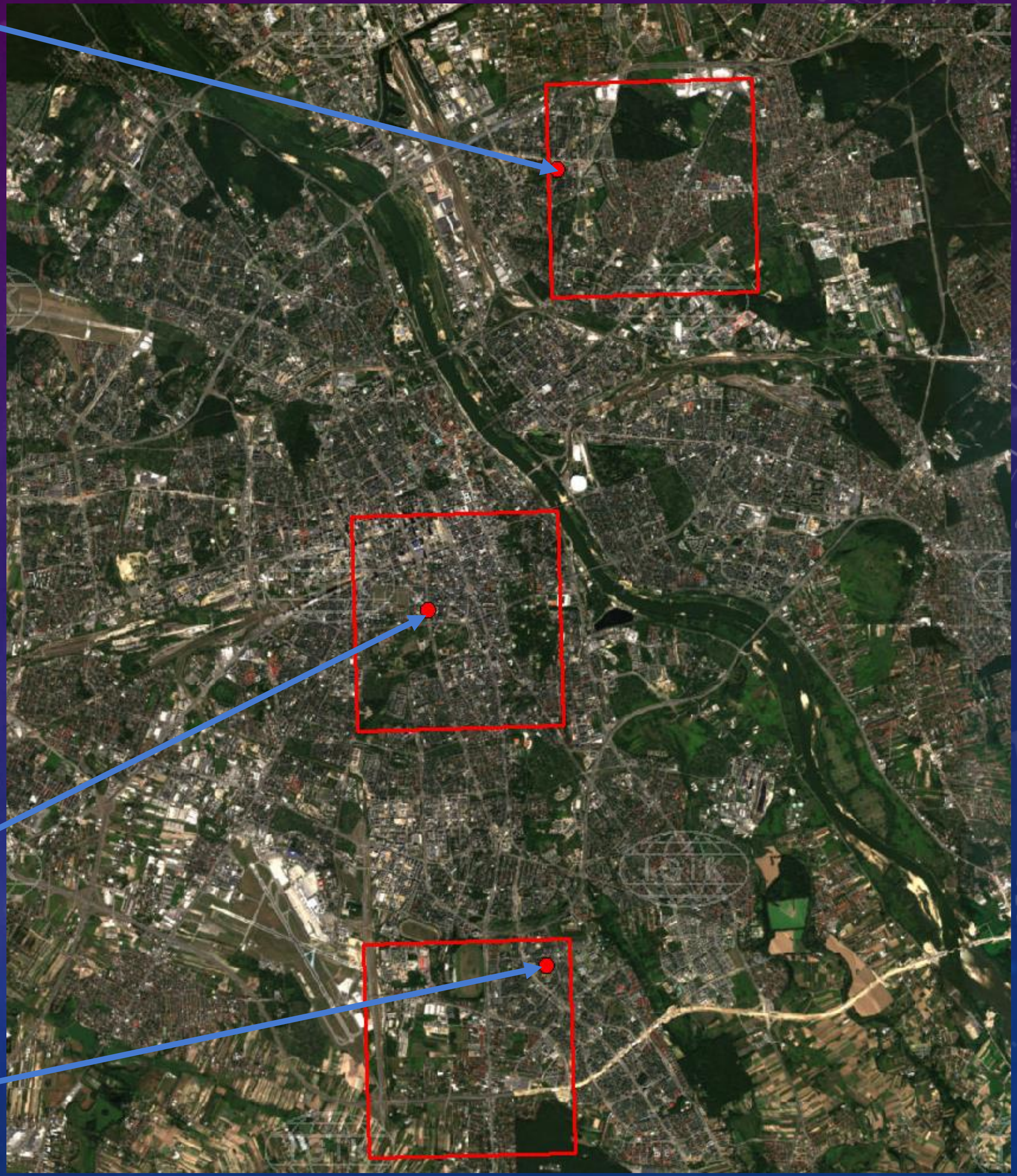


R^2 (Surface NO_2 vs. NO_2 TVCD) for each station in Poland

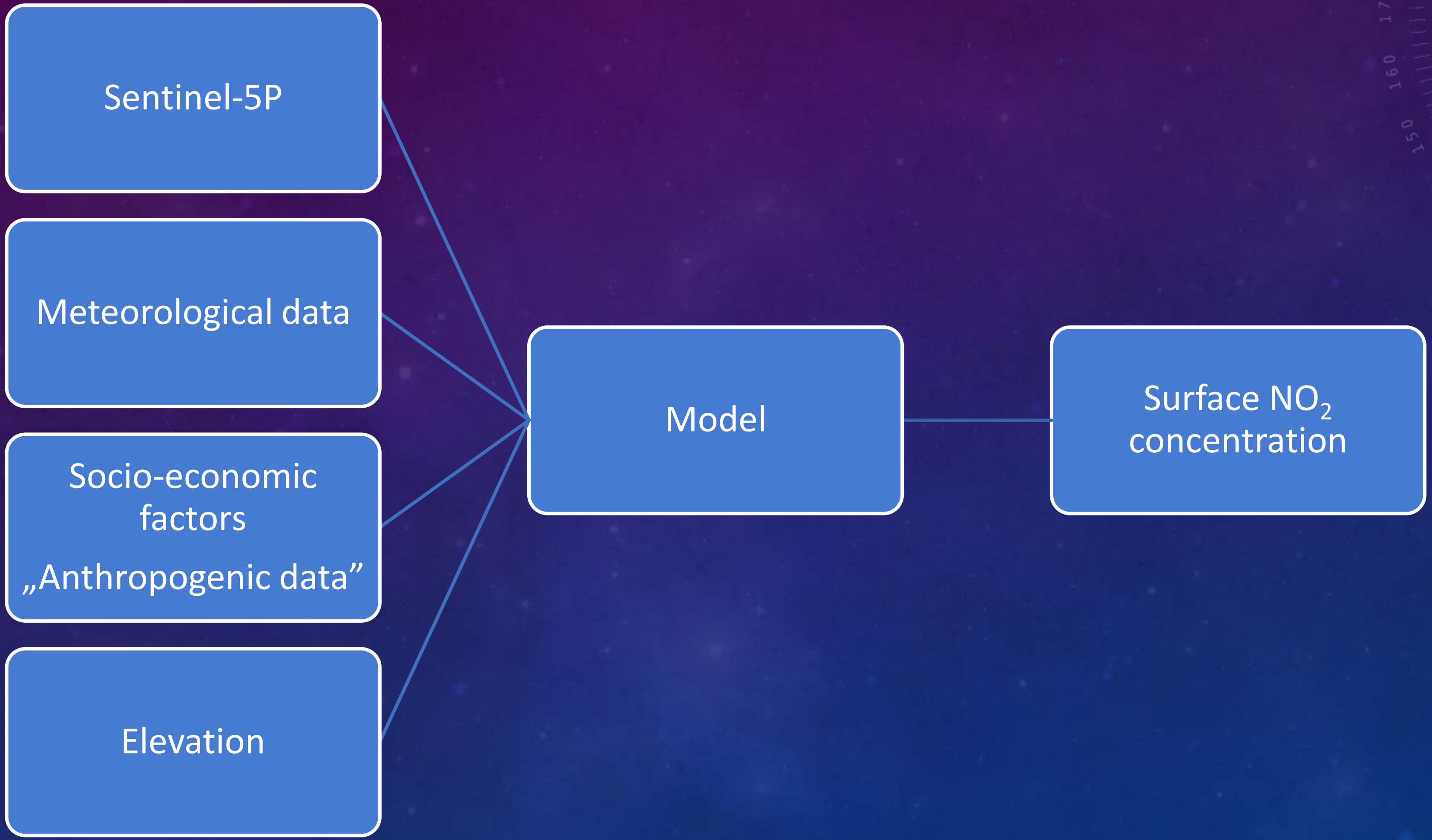
Kondratowicza

Al. Niepodległości

Wokalna



WHAT IS GOING TO BE ACHIEVED?

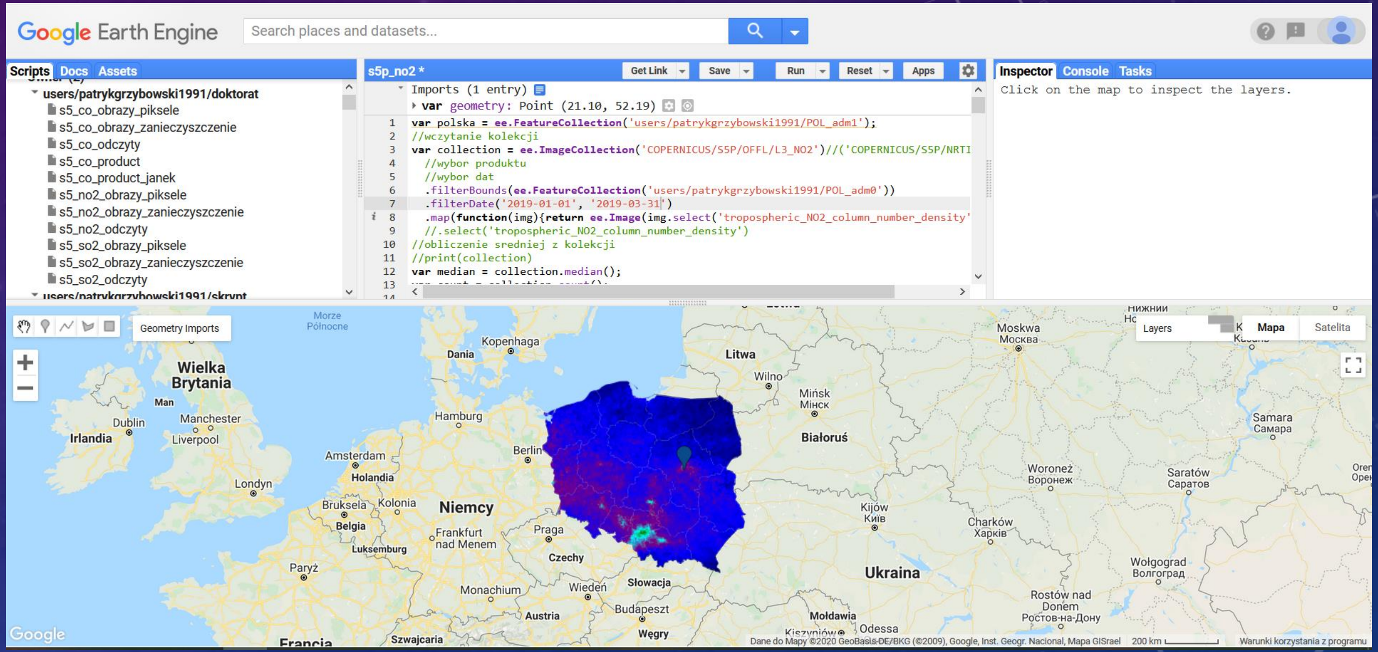


DATA

Kind of data	Source	Data
Satellite data	Senitnel-5P – ESA, European Commission	NO2 TVCD
Ground-based concentration	GIOŚ	Surface NO2
Meteorolgical data	ECMWF	Temperature Dew point temperature Pressure Solar radiation accumulated by 1 hour U component of wind V component of wind Boundary layer height Relative humidity Wind speed Wind direction
	NCEP/NCAR	Sea level pressure
Socio-Economic data	NASA/EC JRC/NOAA	Population Nightlights
Geography	NASA	Elevation

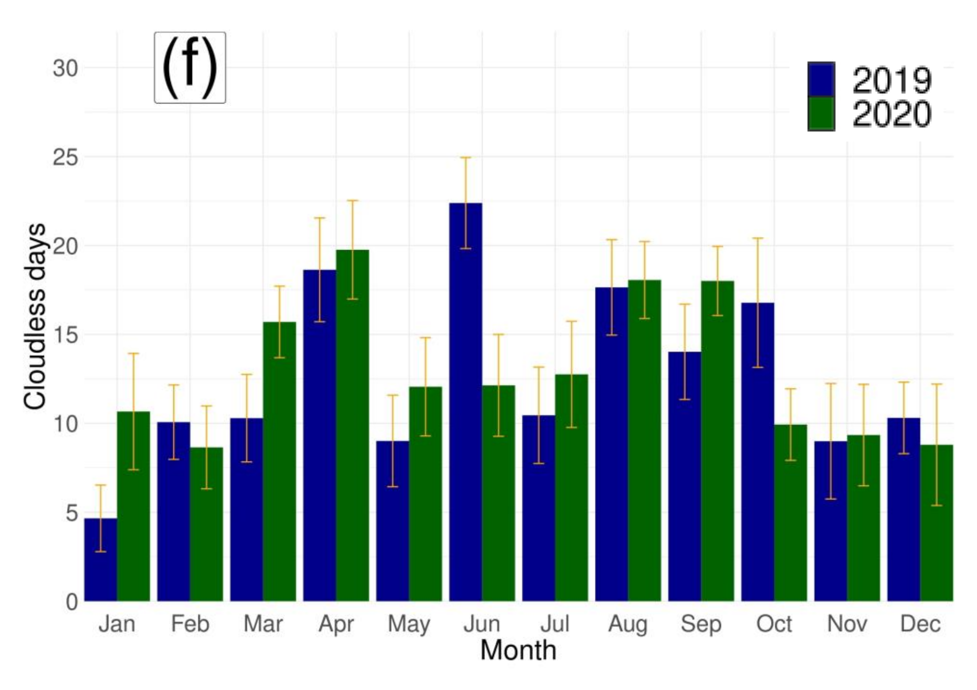
DATA PROCESSING

- Platform based on the JavaScript programming language
- A platform for analyzing and visualizing geospatial datasets for academic, non-profit, business and government users
- It provides satellite data and stores it in a public data archive that includes historical images of the Earth from more than forty years ago. The images that are taken daily are then made available for data mining on a global scale.
- Earth Engine also provides APIs and other tools to analyze big data
- Provides easy, access to a comprehensive catalog of satellite imagery and other geospatial data in an analysis-ready format. The data catalog is connected to the cloud computing supported by Google data centers and flexible APIs that allow you to seamlessly implement existing scripts and schemas. This enables state-of-the-art global analysis and visualization of geospatial data.



DATA PRE-PROCESSING

- Time analysis: 07.2018-06.2021
    - 2019 and 2020 for availability of data
- Data was linearly interpolated to the time of S-5P acquisition
- Data has been resampled to spatial resolution 3.5 km x 5.5. km using bilinear interpolation
- Outliers:
    - Station measuring surface NO2 was closer than 100 meters from the road (2 stations; 1 024 observations; 2% of all observations).
- Training dataset (22 678):
    - January, March, May, July, September, Novemeber
- Testing dataset (27 241):
    - February, April, June, August, October, December

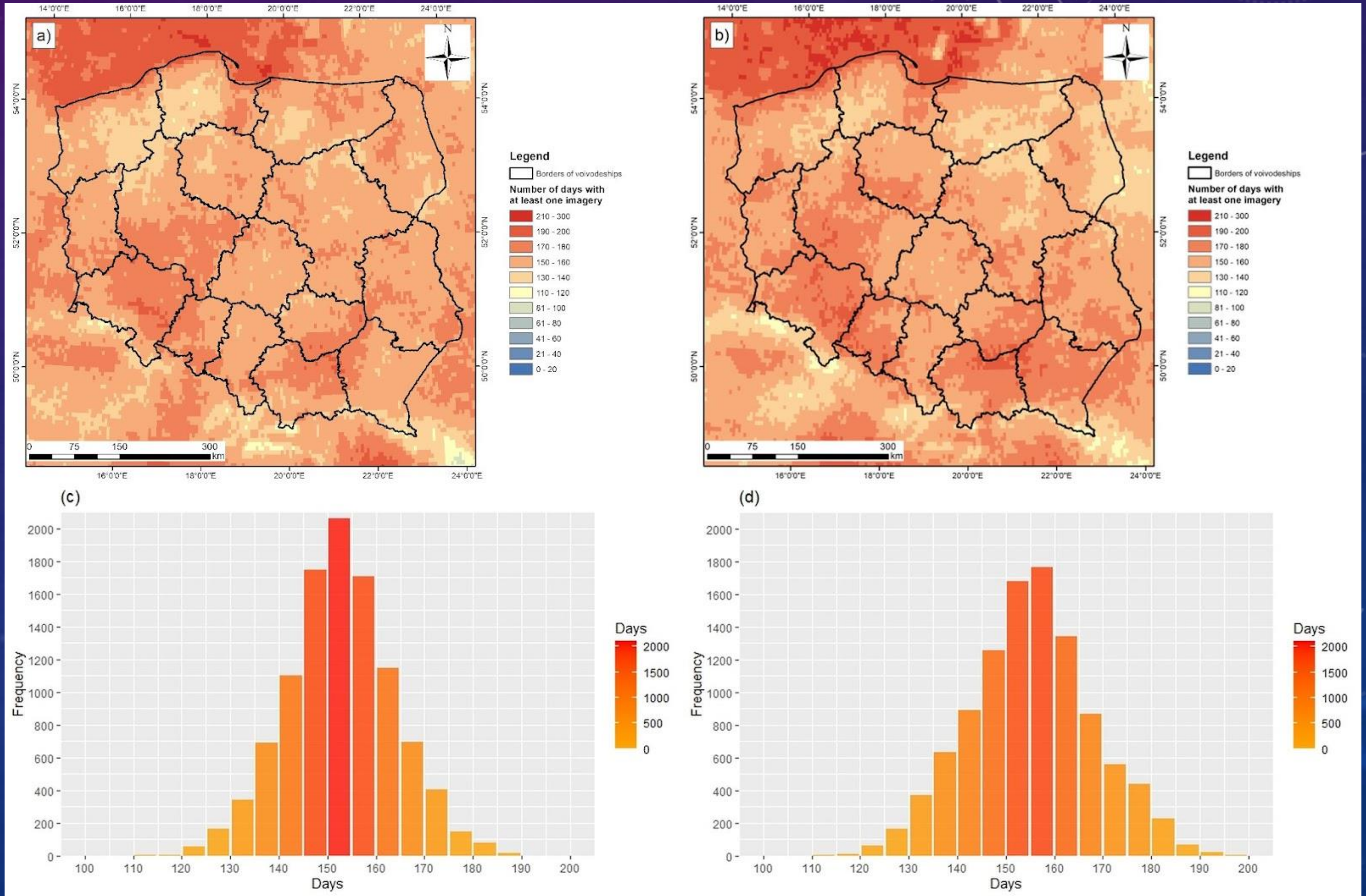


Avearge numer of cloudless days over Poland in each month. Blue – 2019; green – 2020. Yellow bars are minimum and maximum cludless days.

KNOW THE LIMITATIONS

qa_value >= 0.75
cloud fraction =< 0.50

Count of cloudless days in Poland in 2019 (a – map; c – histogram) and in 2020 (b – map; d – histogram)



PREDICTION METHODS

Linear regression with one undependable variable (LM)

- NO_2 TVCD

Multiple linear regression with several undependable variables (MLM)

- NO_2 TVCD
- Air temperature (T)
- Pressure (P)
- Solar radiation (RADNET)
- Wind speed (WS)
- Planetary boundary layer height (PBLH)
- Nightlights (NIGHTLIGHT)
- Population (POP)
- Road density (ROADS)
- Elevation (ELEVATION)

Random forest with several undependable variables (RF)

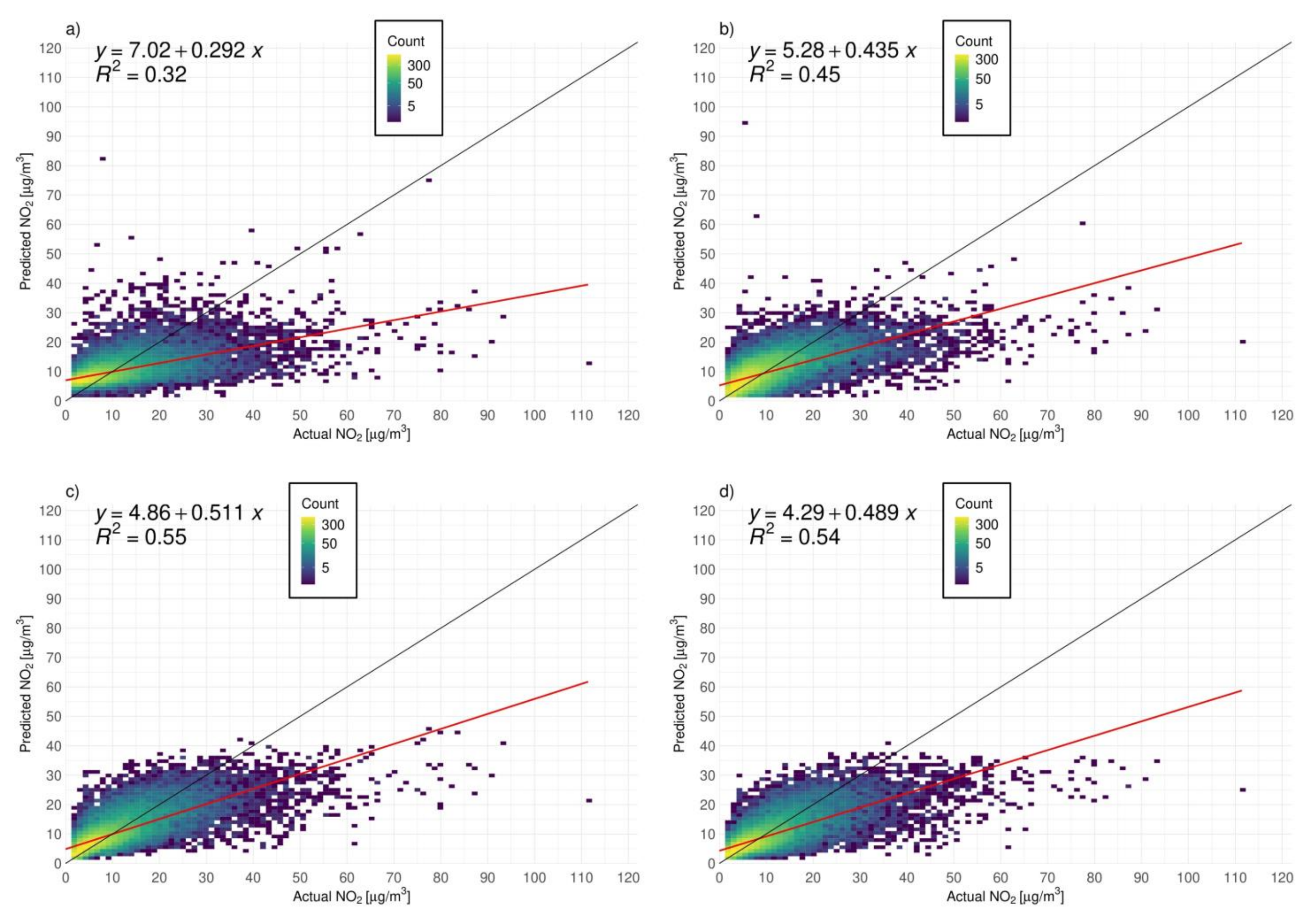
- NO_2 TVCD
- Air temperature (T)
- Pressure (P)
- Solar radiation (RADNET)
- Wind speed (WS)
- Planetary boundary layer height (PBLH)
- Nightlights (NIGHTLIGHT)
- Population (POP)
- Road density (ROADS)
- Elevation (ELEVATION)
- Circulation type (CT)

Radial kernel support vector machine with several undependable variables (SVM)

- NO_2 TVCD
- Air temperature (T)
- Pressure (P)
- Solar radiation (RADNET)
- Wind speed (WS)
- Planetary boundary layer height (PBLH)
- Nightlights (NIGHTLIGHT)
- Population (POP)
- Road density (ROADS)
- Elevation (ELEVATION)
- Circulation type (CT)

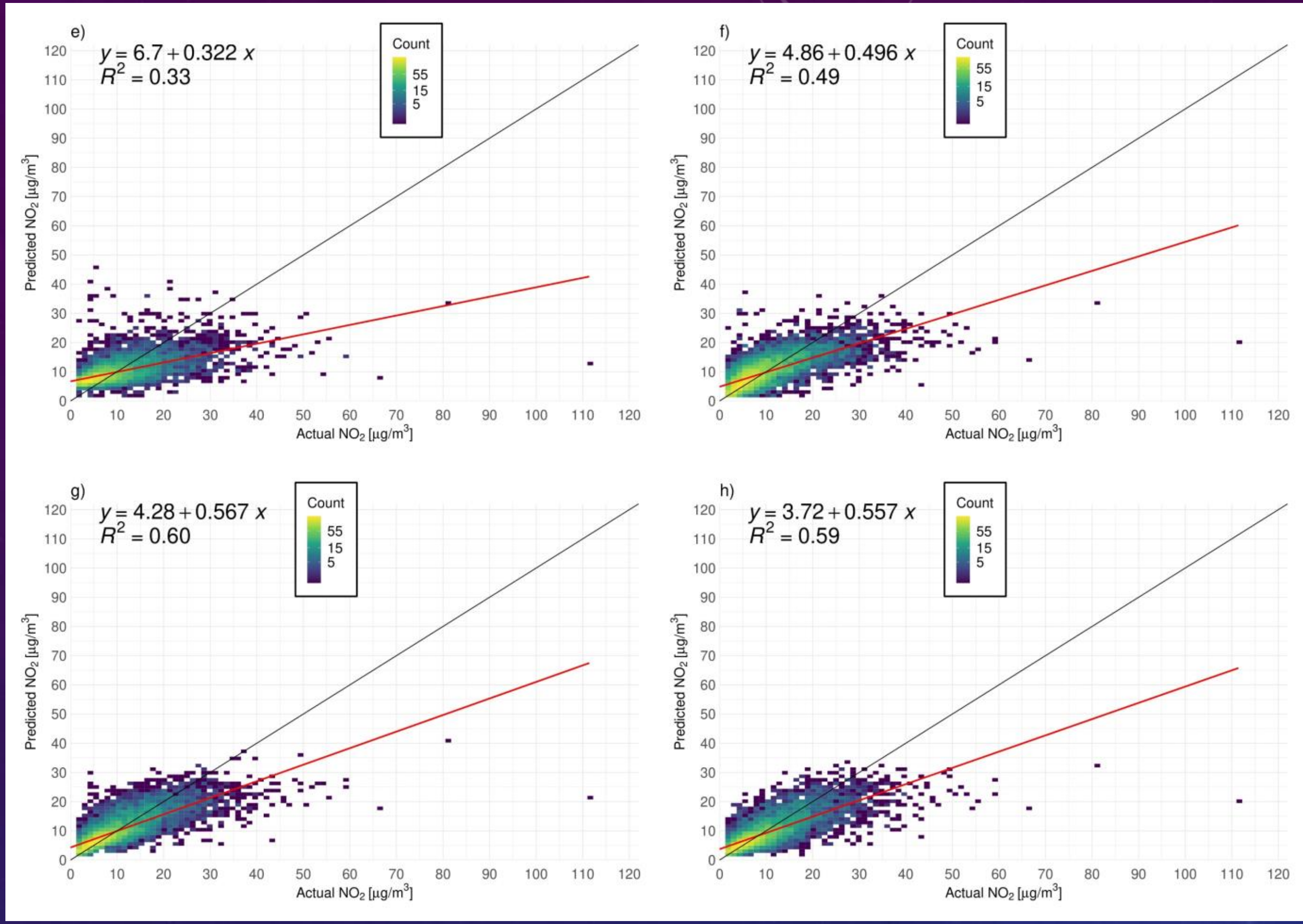
PREDICTION OF SURFACE NO_2 - HOURLY

Surface NO2 mass concentration estimations							
		MEAN	**MIN**	**MAX**	**SD**	**1st Q**	**3rd Q**
Testing dataset (n=27241)		10.07	0.01	111.41	8.89	4.40	13.12
Training dataset (n=22678)		10.44	0.01	100.59	8.61	4.44	12.93
Hourly measurements		**R2**	**MSE**	**RMSE**	**Bias [ug/m3]**	**MAE [ug/m3]**	**MAPE [%]**
	LM - S5P	0.32	53.22	7.30	0.11	4.87	48.38
	MLM	0.45	43.11	6.57	0.42	4.24	42.08
	RF	0.53	34.83	5.90	-0.04	3.74	37.15
	SVM	0.54	37.67	6.09	1.04	3.72	36.94



Results of prediction of surface NO_2 based on four methods: linear regression – LM-S5P (a); multilinear regression - MLM (b); random forest - RF (c) and suport vector machine - SVM (c)

PREDICTION OF SURFACE NO_2 – WEEKLY AVERAGES

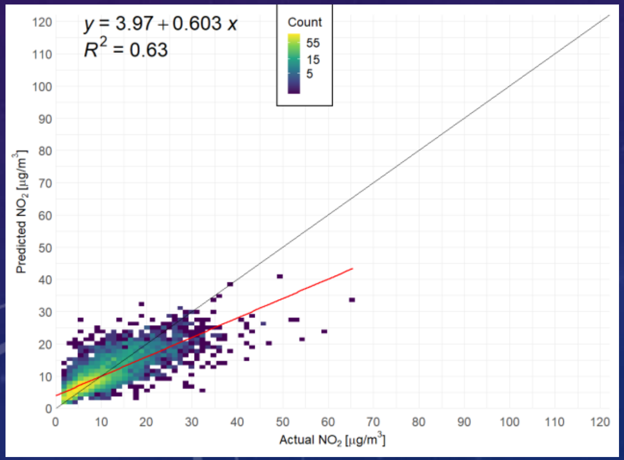


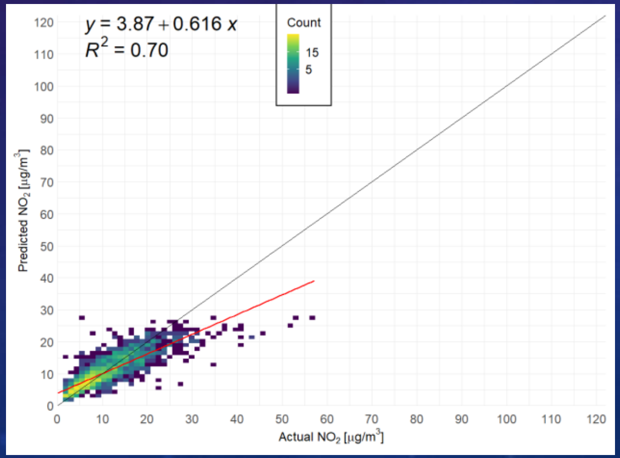
Results of prediction of surface NO_2 based on four methods: linear regression – LM-S5P (a); multilinear regression - MLM (b); random forest - RF (c) and suport vector machine - SVM (d)

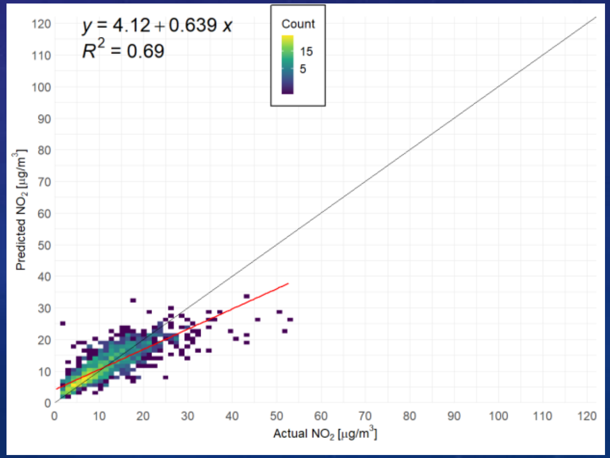
Surface NO2 mass concentration estimations							
		MEAN	**MIN**	**MAX**	**SD**	**1st Q**	**3rd Q**
Testing dataset (n=6639)		10.07	0.01	111.41	8.89	4.40	13.12
Training dataset (n=7126)		10.44	0.01	100.59	8.61	4.44	12.93
Weekly averages		**R2**	**MSE**	**RMSE [ug/m3]**	**Bias [ug/m3]**	**MAE [ug/m3]**	**MAPE [%]**
	LM - S5P	0.33	38.09	6.17	0.14	4.27	42.37
	MLM	0.49	28.96	5.38	0.23	3.59	35.53
	RF	0.60	23.08	4.81	-0.09	3.10	30.75
	SVM	0.59	24.34	4.93	0.76	3.15	31.21

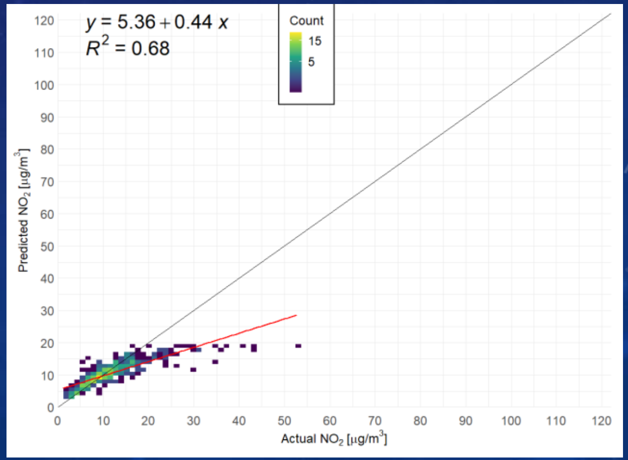
MONTHLY, SEASONLY, YEARLY AVERAGES

	R^2	MSE	RMSE [ug/m3]	BIAS [ug/m³]	MAE [ug/m³]	MAPE [%]
HOURLY	0.53	34.83	5.90	-0.04	3.74	37.15
WEEKLY	0.60	23.08	4.81	-0.09	3.10	30.75
TWO WEEKLY	0.63	19.14	4.38	-0.07	2.96	29.00
MONTHLY	0.71	15.51	3.94	-0.31	2.58	23.49
SEASONLY	0.69	15.72	3.97	-0.08	2.63	23.57
YEARLY	0.68	20.42	4.52	-1.24	2.50	21.22









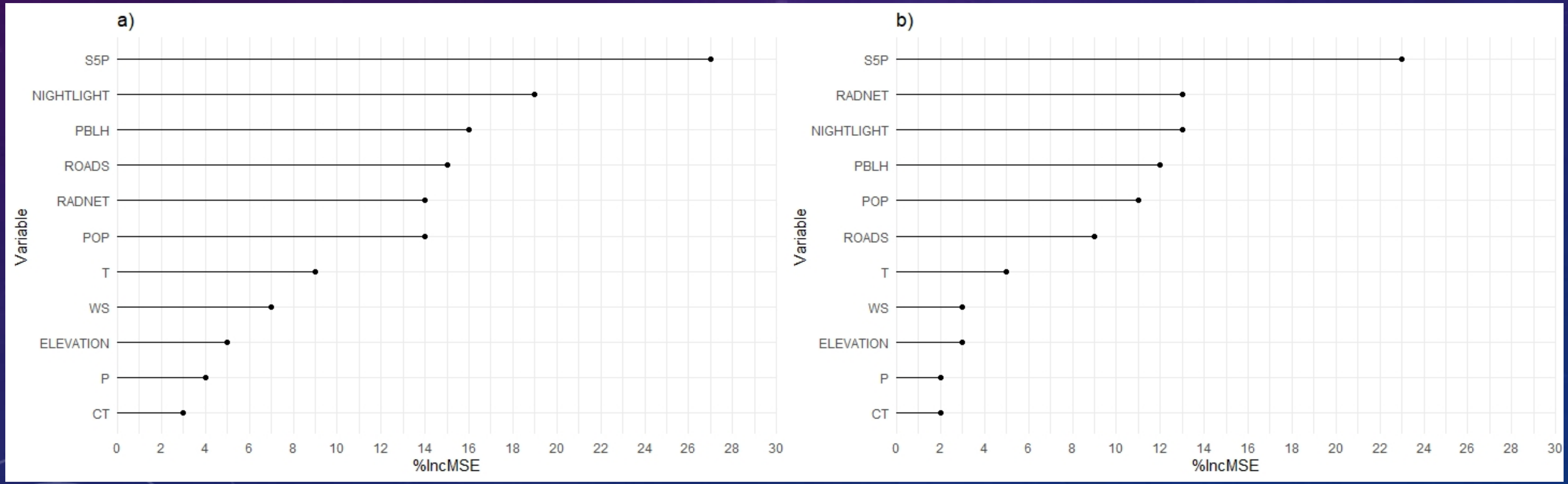
VARIABLES' IMPORTANCE

Multilinear regression

Variable	S5P	T	P	RAD	WS	PBLH	NIGHTLIGHT	POP	ROADS	ELEVATION	INTERCEPT
Coefficient hourly	0.62 (57%)	-0.01 (1%)	-0.03 (3%)	-0.06 (6%)	-0.07 (6%)	-0.08 (7%)	0.05 (5%)	0.03 (3%)	0.06 (6%)	0.01 (1%)	0.07 (6%)
Coefficient weekly	0.69 (53%)	-0.04 (3%)	-0.04 (3%)	-0.06 (5%)	-0.07 (5%)	-0.10 (8%)	0.06 (5%)	0.04 (3%)	0.08 (6%)	0.01 (1%)	0.10 (8%)

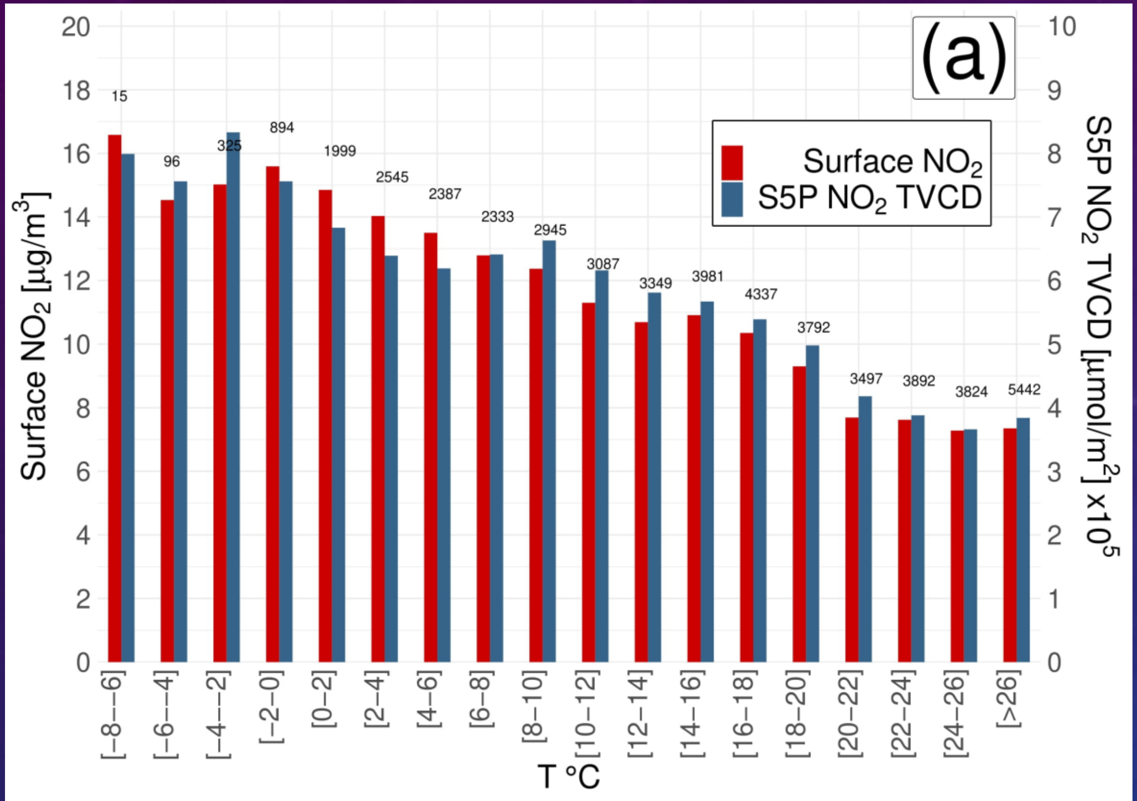
1. Standarization of variables
2. Sum of coefficeints (absulute values)
3. Percentage of each coeffcient within the sum

Random Forest

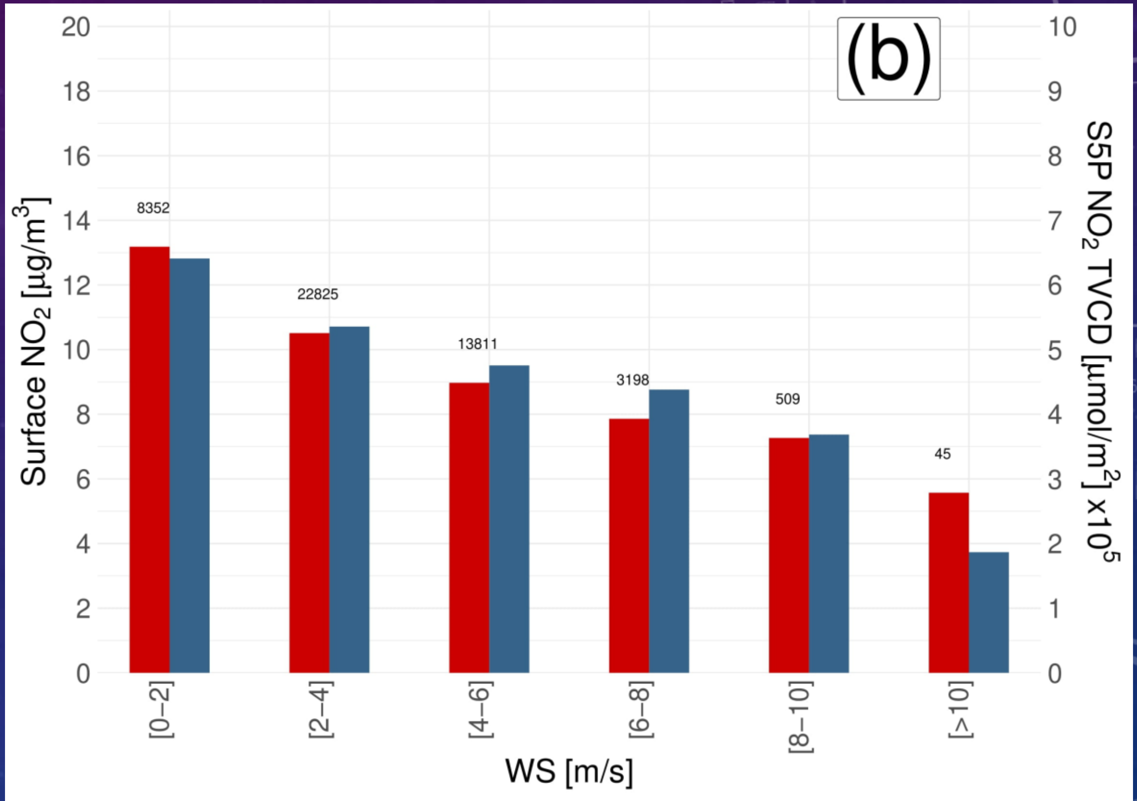


1. Compute model MSE;
2. For each variable in the model:
   1) Permute variable
   2) Calculate new model MSE according to variable permutation
   3) Take the difference between model MSE and new model MSE

CHANGES IN NO2 TVCD AND SURFACE NO2 MASS CONCENTRATION IN RESPECT TO METEOROLOGICAL CONDITIONS



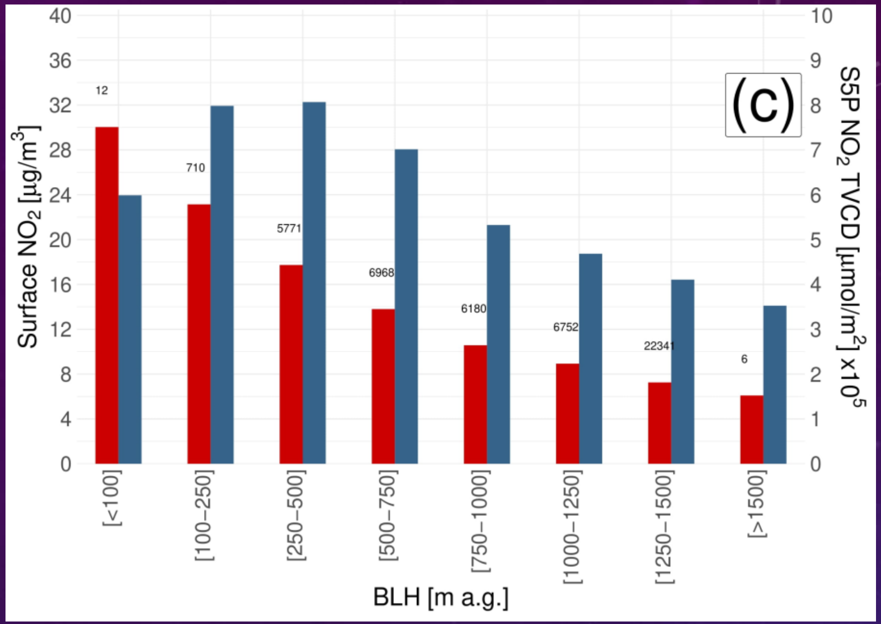
Change of air pollution in respect to air temperature



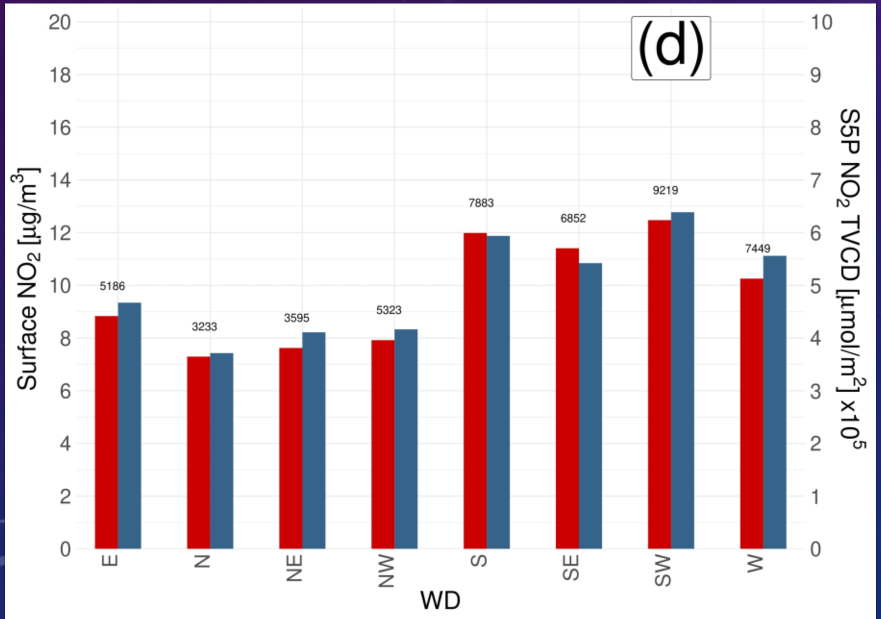
Change of air pollution in respect to wind speed

Left y axis, red font and red bars correspond to surface NO_2 mass concentration while the right y axis, blue font and blue bars correspond to TVCD of NO_2.

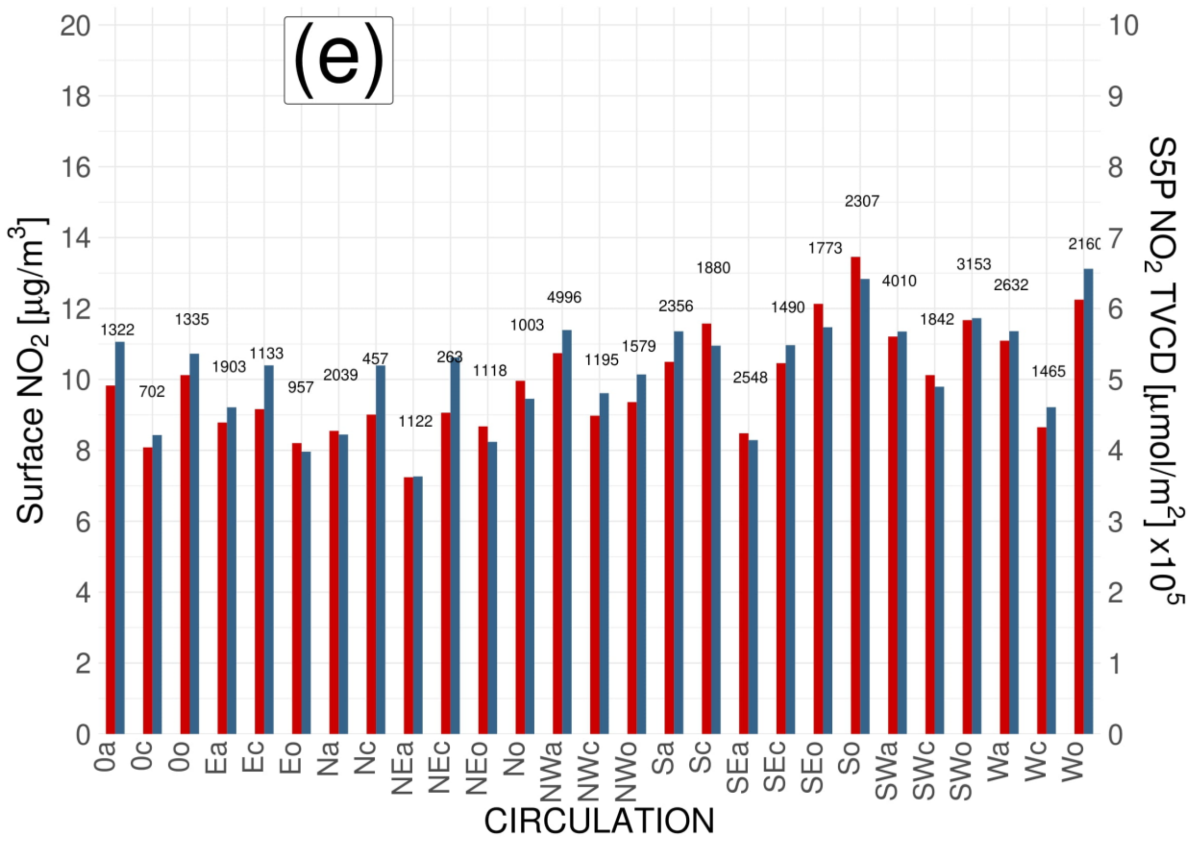
CHANGES IN NO2 TVCD AND SURFACE NO2 MASS CONCENTRATION IN RESPECT TO METEOROLOGICAL CONDITIONS AND OTHER FACTORS



Change of air pollution in respect to PBLH



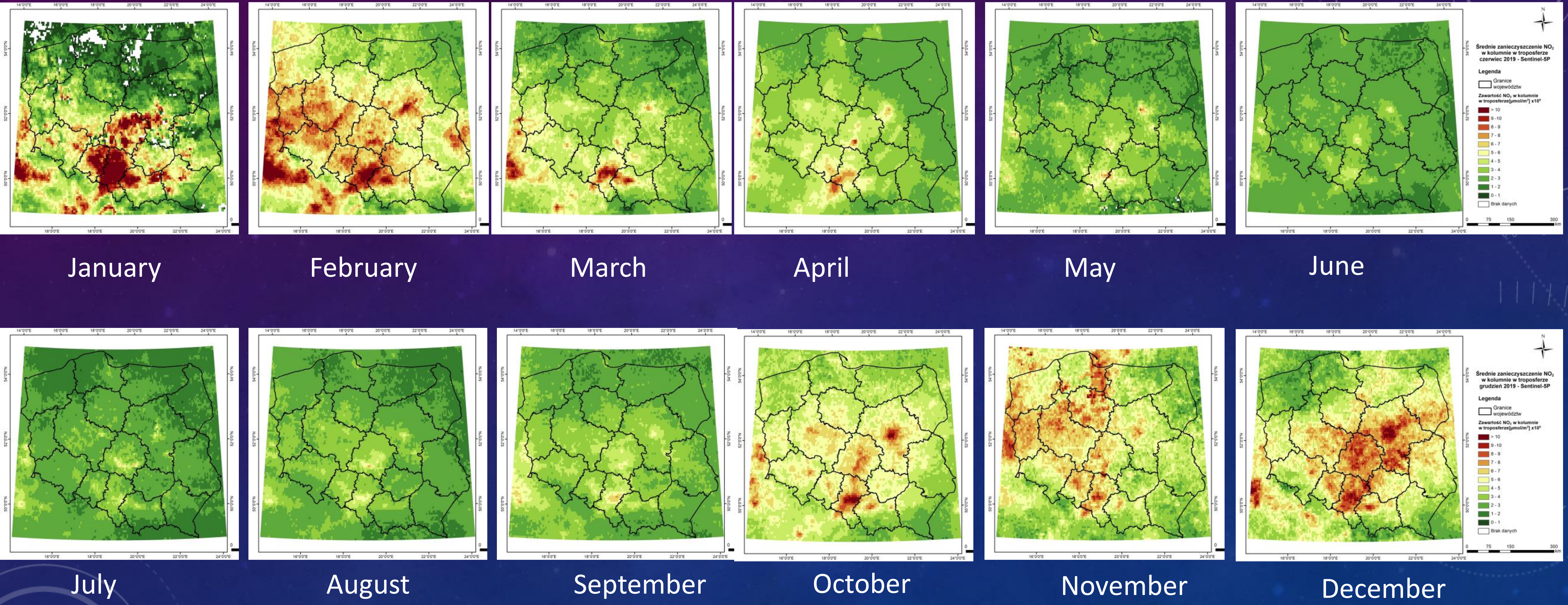
Change of air pollution in respect to wind direction



Change of air pollution in respect to type of atmospheric circulation

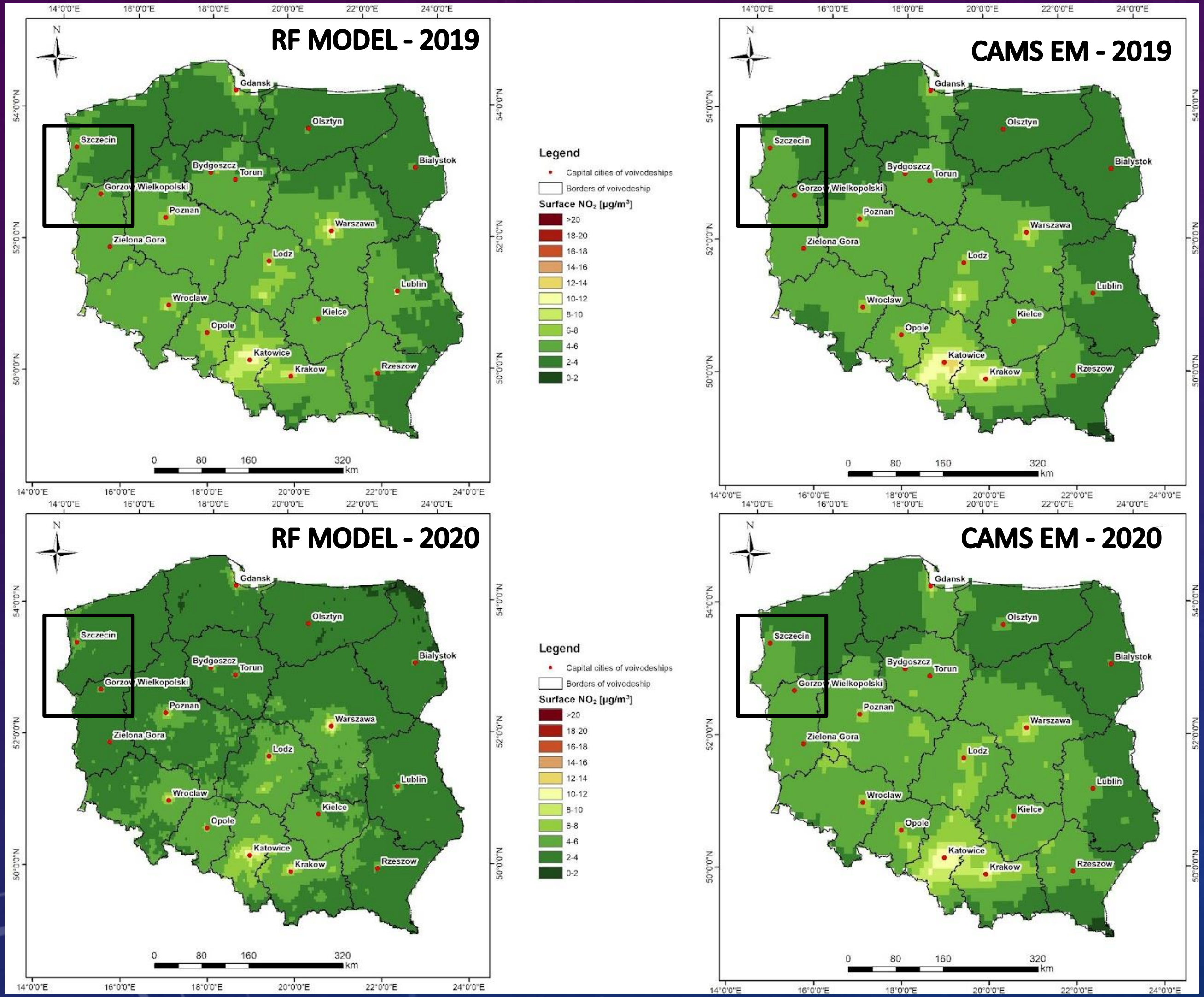
Left y axis, red font and red bars correspond to surface NO_2 mass concentration while the right y axis, blue font and blue bars correspond to TVCD of NO_2.

SPATIO-TEMPORAL CHANGES

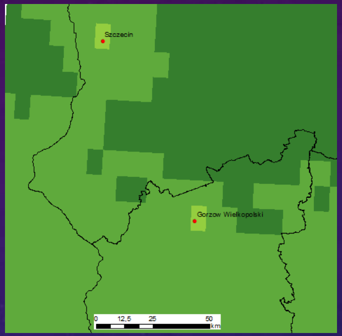


Spatio-temporal variabilities of NO2 TVCD over Poland in 2019.

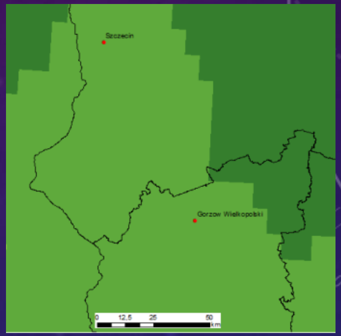
COMPARISON WITH CAMS ENSEMBLE MEDIAN (CAMS EM)



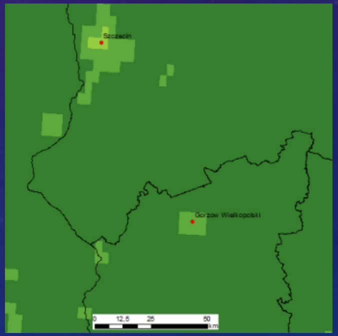
RF MODEL - 2019



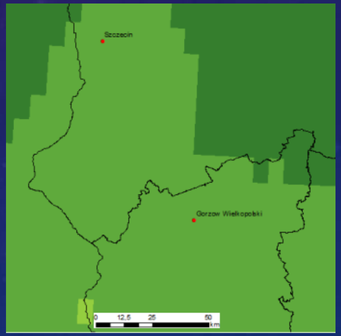
CAMS EM - 2019



RF MODEL - 2020

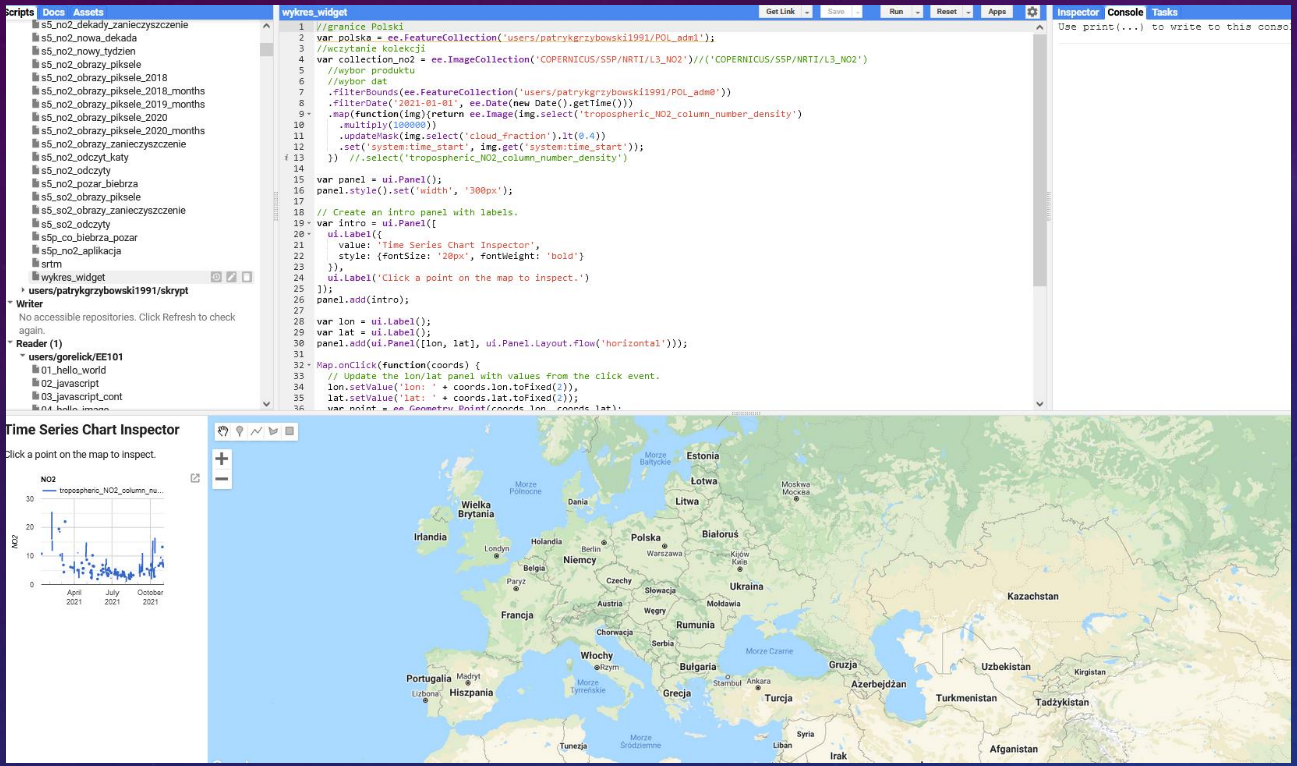


CAMS EM - 2020

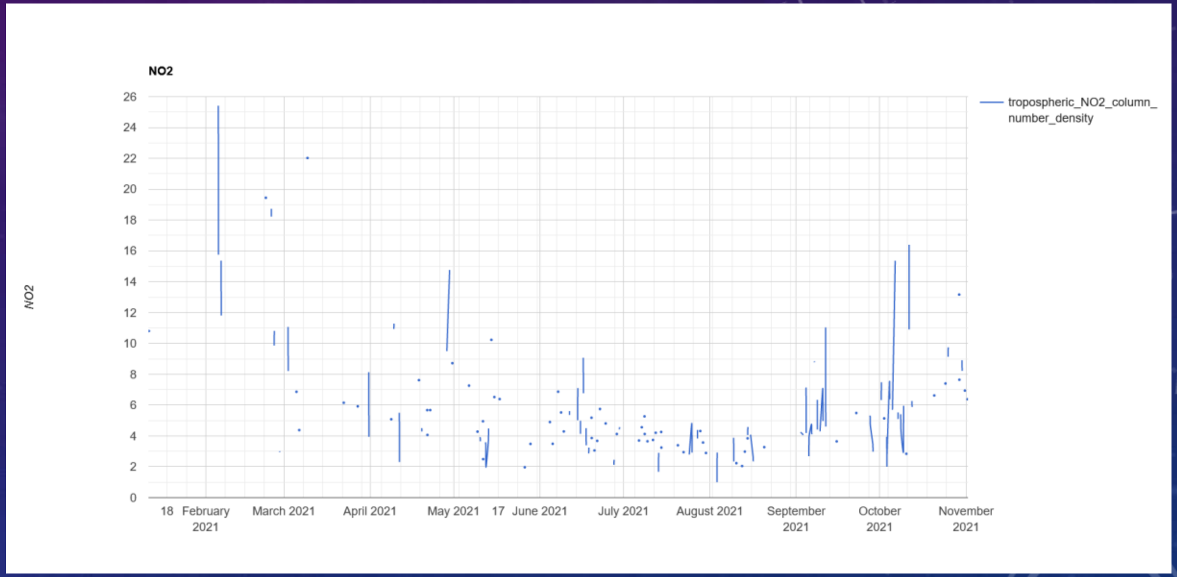


	MAE [ug/m3]	MAPE [%]	R2
RF	3,87	37,96	0,54
CAMS	5,37	52,74	0,46

APP



Scripts
Docs
Assets
s5_no2_dekady_zanieczyszczenie
s5_no2_nowa_dekada
s5_no2_nowy_tydzien
s5_no2_obrazy_piksele
s5_no2_obrazy_piksele_2018
s5_no2_obrazy_piksele_2018_months
s5_no2_obrazy_piksele_2019_months
s5_no2_obrazy_piksele_2020
s5_no2_obrazy_piksele_2020_months
s5_no2_obrazy_zanieczyszczenie
s5_no2_odczyt_katy
s5_no2_odczyty
s5_no2_pozar_biebrza
s5_so2_obrazy_piksele
s5_so2_obrazy_zanieczyszczenie
s5_so2_odczyty
s5p_co_biebrza_pozar
s5p_no2_aplikacja
srtm
wykres_widget
users/patrykgrzybowski1991/skrypt
Writer
No accessible repositories. Click Refresh to check again.
Reader (1)
users/gorelick/EE101
01_hello_world
02_javascript
03_javascript_cont
wykres_widget
Get Link
Save
Run
Reset
Apps
//granice Polski
var polska = ee.FeatureCollection('users/patrykgrzybowski1991/POL_adm1');
//wczytanie kolekcji
var collection_no2 = ee.ImageCollection('COPERNICUS/S5P/NRTI/L3_NO2')//('COPERNICUS/S5P/NRTI/L3_NO2')
//wybor produktu
//wybor dat
.filterBounds(ee.FeatureCollection('users/patrykgrzybowski1991/POL_adm0'))
.filterDate('2021-01-01', ee.Date(new Date().getTime()))
.map(function(img){return ee.Image(img.select('tropospheric_NO2_column_number_density')
.multiply(100000))
.updateMask(img.select('cloud_fraction').lt(0.4))
.set('system:time_start', img.get('system:time_start'));
}) //.select('tropospheric_NO2_column_number_density')
var panel = ui.Panel();
panel.style().set('width', '300px');
// Create an intro panel with labels.
var intro = ui.Panel([
ui.Label({
value: 'Time Series Chart Inspector',
style: {fontSize: '20px', fontWeight: 'bold'}
}),
ui.Label('Click a point on the map to inspect.')
]);
panel.add(intro);
var lon = ui.Label();
var lat = ui.Label();
panel.add(ui.Panel([lon, lat], ui.Panel.Layout.flow('horizontal')));
Map.onClick(function(coords) {
// Update the lon/lat panel with values from the click event.
lon.setValue('lon: ' + coords.lon.toFixed(2)),
lat.setValue('lat: ' + coords.lat.toFixed(2));
Inspector
Console
Tasks
Use print(...) to write to this conso
Time Series Chart Inspector
Click a point on the map to inspect.
NO2
tropospheric_NO2_column_nu...
April 2021
July 2021
October 2021
Estonia
Łotwa
Litwa
Dania
Wielka Brytania
Irlandia
Polska
Białoruś
Niemcy
Ukraina
Kazachstan
Francja
Rumunia
Bulgaria
Włochy
Grecja
Turcja
Hiszpania
Portugalia
Gruzja
Azerbejdżan
Uzbekistan
Turkmenistan
Tadżykistan
Afganistan
Kirgistan
Syria
Irak
Tunezja



NO2
tropospheric_NO2_column_number_density
NO2
26
24
22
20
18
16
14
12
10
8
6
4
2
0
18 February 2021
March 2021
April 2021
May 2021
17 June 2021
July 2021
August 2021
September 2021
October 2021
November 2021

CONCLUSIONS

- There were at least 120 days per year when it was possible to perform model for Poland using TROPOMI observations
- Results revealed machine learning methods (RF and SVM) gave accuracy 63% for hourly estimation and 69% for weekly averages
- Implementation of meteorological and anthropogenic variables extremely improved the quantity of models. MLM approach gives 6% (hourly) and 7% (weekly) lower MAPE than LM-S5P approach while RF and SVM approaches gives 11% (hourly) and 12% (weekly) lower MAPE than LM-S5P
- NO_2 TVCD, Planetary boundary layer height, solar radiation, nightlights, roads density and population influenced on estimations the most
- Air temperature, wind speed, surface pressure, elevation and type of atmospheric circulation influenced on estimation the least
- Trends for surface NO2 and TVCD of NO2 were negative for raising air temperature, PBLH and wind speed while they were positive for rising of population and nightlights
- The RF model created within the study was better fitted to actual values than CAMS ensemble median

FURTHER ACTIVITIES

- Station's representativeness in respect to area of TROPOMI pixel

Thank you for your attention!