# Overview

Two types of datasets are provided: raster data in geotiff format (satellite and weather data) and csv files with data for power plants.

* All data is provided for the time period 1 Mar 2019 - 30 Sep 2023.
* All geospatial data is in the WGS 84 coordinate system (=EPSG 4326).
* All raster data is provided at latitude-longitude grid with 0.025° resolution, for a domain covering Taiwan (119°E, 21°N; 123°E, 27°N).
* For raster data: 1 file/day

All files except Taiwan\_nox\_emissions.csv can be used as input data for your model/AI application, and Taiwan\_nox\_emissions.csv provides the training/testing/validation data.

A Jupyter notebook is provided with simple examples for reading and visualising the data.

# Datasets

## Taiwan power plants

List of power plants in Taiwan, for which NOx emissions should be estimated. Provides the name, location and an unique identifier (“facility\_id”) for each power plant.

File: Taiwan\_powerplants.csv

## Ground truth for NOx emissions

This is your training, testing and validation dataset. Hourly NOx emissions are provided for power plants in kg per hour. To get total emission over a time period (e.g. day or month), simply sum up the given hourly emissions.

File: Taiwan\_nox\_emissions.csv

## Satellite: NO2

NO2 measurement from ESA’s state-of-the-art air quality satellite TROPOMI. The satellite measurements of NO2 give the total amount of NO2 in the column from the surface through the atmosphere (referred to as *column density*) in units of kilograms per squaremeter (kg/m²). The satellite data is filtered for bad quality data - for example clouds interfere with the measurements - so that large and persisting gaps can be expected. There are a few days without any valid data for the region of interest, for these days there are no data files.

Note, that ‘NOx’ refers to both NO2 and NO, and the NOx emission (see *Ground truth for NOx emissions*) is the total emission of all NOx. However, only NO2 satellite data is available. The NOx-to-NO2 -ratio depends on atmospheric conditions.

Daily files: satellite/NO2/no2\_kgm2\_<YYYYMMDD>.tiff

## Satellite: Cloud fraction

Since clouds impact the satellite measurements and are related to many weather phenomena, we also provide the cloud fraction derived from the satellite measurements. Cloud fraction is unitless, with 0 indicating no clouds and 1 indicating full cloud cover in the pixel.

Daily files: satellite/cloud\_fraction/cloud\_fraction\_<YYYYMMDD>.tiff

## Weather data

Since emitted pollutants are transported by wind away from the source, and can be also removed from the air through different processes, we provide a set of meteorological variables that could be useful. You may choose to use only wind, or all the provided data in your solution.

Note, that to fully describe wind both wind speed and direction need to be considered. To avoid any confusion related to wind direction definitions, the wind data is provided as the eastward and northward components of the wind vector (standard meteorological convention).

For each weather variable, a daily file with values representative for the satellite overflight time is provided.

| **Variable** | **Units** | **Notes** |
| --- | --- | --- |
| Eastward wind component (u wind) at 100 m | meter/second [m/s] | Wind speed can be calculated as sqrt(u\*\*2 + v\*\*2).  Wind direction can be calculated by metpy.calc.wind\_direction(u, v) or other available library |
| Northward wind component (v wind) at 100 m | meter/second [m/s] |
| Temperature | Kelvin [K] |  |
| Relative humidity | Percentage [%] |  |
| Boundary layer height | Meter [m] | Determines the height in the atmosphere in which pollutants are dispersed |
| Solar radiation at the surface | Watt/squaremeter [W/m²] |  |

Daily files: weather/<variable>/<var>\_<units>\_<YYYYMMDD>.tiff   
(for example weather/temperature/temp\_K\_20200101.tiff)

## NO2 Flux

[Beirle et al.](https://doi.org/10.1126/sciadv.aax9800) developed a concept to estimate NOx emissions from satellite data based on the fact that NOx transport is related to NOx emission (the more NOx is emitted, the larger is also the transport of NOx away from the source). By using basic equations for atmospheric physics and making some simplifying assumptions, power plant emissions can be related to the net transport of NOx.

We provide the NO2 net transport, formally referred to as NO2 flux divergence, calculated from the satellite NO2 and wind data. The NO2 flux is given as kilograms per square meter per second (kg/m²s). The use of this dataset is optional.

Daily files: NO2flux/no2flux\_kgm2s\_<YYYYMMDD>.tiff

## Data Sources

* Satellite NO2 & cloud fraction: Sentinel 5P TROPOMI NO2 L2 Collection 03, available from [Copernicus Data Space Ecosystem](https://dataspace.copernicus.eu/)
* Weather data: ERA5 reanalysis, available from [Copernicus Climate Data Store](https://cds.climate.copernicus.eu/)
* NO2 Flux: Calculated from satellite NO2 and wind data by CREA, following [Beirle et al. (2023)](https://essd.copernicus.org/articles/15/3051/2023/)