Summary of input data: raw data

Category	Name	Source	Format	N. files	Total size	Original projection	Metadata, units, acquisition parameter
AOI	Switzerland boundary	DIVA-GIS (link)	shapefile	4	11.8 MB	EPSG4326 - WGS84	
ground-[NO_2] monitoring	Monitored NO2 values	Immissionsmesswerte Luft der Schweiz: https://www.arias.ch/ibonline/ib_online.php More metadata: link	.csv	1	298 KB	/	encoding: ISO-8859-1 unit: ≈ µg/m^3
	Monitoring sites metadata		.csv	1	12 KB	EPSG: 2056 (CH1903+/ LV95)	
satellite-based NO_2 product	OMI Tropospheric column NO2	NASA EarthData GES DISC (link) OMI_MINDS_NO2d (OMI/Aura NO2 Tropospheric, Stratospheric & Total Columns MINDS Daily L3 Global Gridded 0.25 degree x 0.25 degree)	(original) netCDF	365	(original) ~ 7.8 MB * 365	+proj=longlat +a=6378137 +f=0.003352810562417 4 +pm=0 +no_defs	- Acquisition parameter: 2019/01/01—2019/12/31 - Band: ColumnAmountNO2TropCloud Screened - Sp. res.; 0.25 ° x 0.25 ° (~13*25km) - Temp. res.: 1 day - Unit: molec/cm^2
			(clipped) .tif		(clipped) ~ 2KB * 365 = 1.1 MB		

	TROPOMI Tropospheric column NO2	ESA Sentinel-5P Pre-operations data hub (link)	netCDF	747	~ 320MB * 747	EPSG:4326 - WGS 84 - Geographic	- Acquisition parameter:
							  - Sensing period: 2019/01/01—2019/12/31
							  - Product type: L2__NO2
							  - Processing level: L2
							  - Timeliness: Offline
							  - Polygon
							- Band: nitorgendioxide_tropospheric_colum
							- Sp. res.: 7*3.5km
							- Temp. res.: 1 day
							- Unit: (original) mol/m^2; (preprocessed) molec/cm^2

			GeoTIFF	365	6KB * 365 = 3.1 MB		

Category	Name	Source	Format	N. files	Total size	Original projection	Metadata, units, acquisition parameter

modeled NO_2	Modeled total column amount NO2	CAMS global reanalysis (EAC4) (link), ECMWF	netCDF	1 (but more than one time points)	765.8 KB	? (Guessing: EPSG: 4326 Geographic Latitude/Longitude)	- Acquisition parameter:
							  - Variable: Total column nitrogen dioxide, Total column nitrogen monoxide
							  - Date: 2019/01/01—2019/12/31
							  - Time: 00, 03, 06, 09, 12, 15, 18, 21
							  - Area: 3.50(W), 11.11(E), 44.52(S), 48.44(N)
							  - Format: netCDF
							- Sp. res.: 0.75° * 0.75° (~80*80km)
							- Temp. res.: 3 hour
							- Unit: kg/m^2

	CORINE land cover 2018	Copernicus Land Monitoring Service (link)	GeoTIFF	1	(original) 196.6 MB	EPSG:3035	Sp. res.: Minimum Mapping Unit (MMU) 25 ha Sp. res: 100 * 100m
			GeoTIFF		(clipped) 55.5 MB		
	EU-DEM v1.1 elevation	Copernicus Land Monitoring Service (link)	GeoTIFF	1	(original) 4.91 GB	EPSG:3035 IGNF:ETRS89LAEA - ETRS89 Lambert Azimutal Equal Area	Tile: DEM-v1.1-E40N20 Sp. res.: 25 * 25 m
			GeoTIFF		(clipped) 472.6 MB		
	GHS population 2016	EU Open Data Portal (link)	GeoTIFF	1	497.4 MB	IGNF:ETRS89LAEA - ETRS89 Lambert Azimutal Equal Area	Sp. res.: 100 * 100m
			GeoTIFF		(clipped) 58.2 MB		
	VIIRS light at night	Earth Observation Group (EOG) Annual VNL V2 (link)	(original) .tif.gz	1	(original) 9.3 GB	EPSG:4326 (Geographic Latitude/Longitude)	15 arc second (~500m at the Equator)
			GeoTIFF		3.7 MB		

Category	Name	Source	Format	N. files	Total size	Original projection	Metadata, units, acquisition parameter
Spatial predictors	NOx emissions 2015	*Meteotest*	.adf	5	6.5 MB	OBLIQUE swiss oblique Mercator by Rosenmund 1903 (link) = CH1903_LV03 (EPSG: 21781)	- agriculture, household, industry, traffic, wood smoke - Sp. res.: 200 * 200m
	30m-NDVI		GeoTIFF	1	70.7 MB	EPSG:21781 - CH1903 / LV03 - Projected	- agriculture, household, industry, traffic, wood smoke - Sp. res.: 100 * 100m (resampled from 30m) - pixel values scaled to -10000~10000
	road network	*Sonbase database*	shapefile	1	540.8 MB	EPSG:21781 - CH1903 / LV03 - Projected	
	number of intersections	(road network)	GeoTIFF	1	5.6 MB	EPSG:21781 - CH1903 / LV03 - Projected	- Sp. res.: 100 * 100m
	traffic intensity	*Daily traffic volume * road length Rasterized*	.adf	7	224 MB	EPSG:21781 - CH1903 / LV03 - Projected	- Buffer sizes: 300, 500, 1000, 2000, 5000, 10000m - Sp. res.: 100 * 100m
	road density	(road network)	GeoTIFF	2	16.5 MB * 2	EPSG:21781 - CH1903 / LV03 - Projected	- mjrdens (major roads density) rdens (all roads density) - Sp. res.: 100 * 100m
	distance to nearest major road	(road network)	GeoTIFF	1	33 MB	EPSG:21781 - CH1903 / LV03 - Projected	- nearmjrds (distance to nearest major road) - Sp. res.: 100 * 100m

Category	Name	Source	Format	N. files	Total size	Original projection	Metadata, units, acquisition parameter

Spatialtemporal predictors	ERA-5 meteorological variables	ERA5 hourly data on single levels from 1979 to present (link), ECMWF	netCDF	1 (but more than one time points)	19.4 MB	? (Guessing: EPSG: 4326 Geographic Latitude/Longitude)	- Acquisition parameter:
							  - Variables (and units):
							  - 2m temperature (K)
							  - 10m u-component of wind (m/s)
							  - 10m v-component of wind (m/s)
							  - Surface pressure (Pa)
							  - Total precipitation (m)
							  - Total cloud cover (0-1)
							  - Boundary layer height (m)
							  - Year: 2019
							  - Month: select all
							  - Day: select all
							  - Time: 00, 03, 06, 09, 12, 15, 18, 21
							  - Area: 3.50(W), 11.11(E), 44.52(S), 48.44(N)
							  - Format: netCDF
							- Sp. res.: 0.25° * 0.25°
							- Temp. res.: 3 hour

	MODIS vegetation MYD13Q1: NDVI	The Aqua Moderate Resolution Imaging Spectroradiometer (MODIS) Vegetation Indices (MYD13Q1) Version 6 data NASA EarthData (link)	(original) .hdf	24	(original) ~ 180 MB * 24	+proj=sinu +lon_0=0 +x+0=0 +y_0=0 +R-6371007.181 +units=m +no_defs (more info: https://proj.org/operations/projections/sinu.html)	- Acquisition parameter:
							  - start: 2019/01/01
							  - end: 2019/12/31
							- Sp.. res: ~250*250m

			GeoTIFF		~ 9 MB * 24 = 236 MB		

Summary of the cleaned variables

Category	Abbreviation	Variable	Grid resolution	Buffer size (in meter)	Direction	Properties
ground-[NO_2] monitoring	NO2	NO_2 concentration measured at monitoring stations	/	/	/	/
spatial-temporal	OMI_NO2	OMI Tropospheric column NO_2 (imputed)	1000*1000m	/	+	- standardized to mean=0 and sd=1
	TROPOMI_NO2	TROPOMI Tropospheric column NO_2 (imputed)	1000*1000m	/	+	- standardized to mean=0 and sd=1
	temperature_{}H	air temperature at 2m	1000*1000m	/	~	- standardized to mean=0 and sd=1 - orignial unit: K
	blh_{}H	boundary layer height	1000*1000m	/	-	- standardized to mean=0 and sd=1 - orignial unit: m
	pressure_{}H	air pressure	1000*1000m	/	~	- standardized to mean=0 and sd=1 - orignial unit: Pa
	precipitation_{}H	total precipitation	1000*1000m	/	~	- standardized to mean=0 and sd=1 - orignial unit: m
	tcc_{}H	total cloud cover	1000*1000m	/	~	- not standardized - unit: [0-1]
	ws_{}H	wind speed	1000*1000m	/	-	- standardized to mean=0 and sd=1 - orignial unit: m/s
	wd_{}H	wind direction	1000*1000m	/	~	- standardized to mean=0 and sd=1 - orignial unit: pi
	NDVI	daily NDVI interpolated from 16-day	250*250m	/	-	- standardized to mean=0 and sd=1

Category		Abbreviation	Variable	Grid resolution	Buffer size (in meter)	Direction	Properties
spatial		elevation_{}	elevation	100*100m	100, 200, 500, 1000, 2000, 5000, 10000	-	- standardized to mean=0 and sd=1 - orignial unit: m
	emission	enox_agfo_{}	NOx emission from agriculture	100*100m	100, 200, 500, 1000, 2000, 5000, 10000	+	- log-transformation -> standardization
		enox_hoco_{}	NOx emission from household	100*100m	100, 200, 500, 1000, 2000, 5000, 10000	+	- log-transformation -> standardization
		enox_indu_{}	NOx emission from industry	100*100m	100, 200, 500, 1000, 2000, 5000, 10000	+	- log-transformation -> standardization
		enox_traf_{}	NOx emission from traffic	100*100m	100, 200, 500, 1000, 2000, 5000, 10000	+	- log-transformation -> standardization
		enox_wood_{}	NOx emission from wood smoke	100*100m	100, 200, 500, 1000, 2000, 5000, 10000	+	- log-transformation -> standardization
	land cover	lc_RES_percent_{}	percentage of residential area	100*100m	100, 200, 500, 1000, 2000, 5000, 10000	+	- percent (0-1), not standardized
		lc_IND_percent_{}	percentage of industrial and commercial area	100*100m	100, 200, 500, 1000, 2000, 5000, 10000	+	- percent (0-1), not standardized
		lc_URBGR_percent_{}	percentage of urban green area	100*100m	100, 200, 500, 1000, 2000, 5000, 10000	-	- percent (0-1), not standardized
		lc_BUILT_percent_{}	percentage of total built up area	100*100m	100, 200, 500, 1000, 2000, 5000, 10000	+	- percent (0-1), not standardized
		lc_AGR_percent_{}	percentage of agricultural area	100*100m	100, 200, 500, 1000, 2000, 5000, 10000	~	- percent (0-1), not standardized
		lc_NAT_percent_{}	percentage of natural area	100*100m	100, 200, 500, 1000, 2000, 5000, 10000	-	- percent (0-1), not standardized
		light_{}	nighttime light	100*100m	100, 200, 500, 1000, 2000, 5000, 10000	+	- log-transformation -> standardization

Category		Abbreviation	Variable	Grid resolution	Buffer size (in meter)	Direction	Properties
		population_{}	population	100*100m	100, 200, 500, 1000, 2000, 5000, 10000	+	- log-transformation -> standardization
		traffint_{}	traffic intensity (daily traffic volume times road length)	100*100m	100, 200, 500, 1000, 2000, 5000, 10000	+	- log-transformation -> standardization
		mjdens_{}	major road density	100*100m	100, 200, 500, 1000, 2000, 5000, 10000	+	- log-transformation -> standardization
		rdens_{}	all road density	100*100m	100, 200, 500, 1000, 2000, 5000, 10000	+	- log-transformation -> standardization
		intersection_{}	number of intersections	100*100m	100, 200, 500, 1000, 2000, 5000, 10000	+	- count - log-transformation -> standardization
		nearmjrd	distance to nearest major road	100*100m	/	-	- standardization
		NDVI_sp	NDVI (aggregated from 30m sp. res.)	100*100m	/	-	- standardized to mean=0 and sd=1

- standardization: $x' = \frac{x - \bar{x}}{\sigma_x}$
- log-transformation: $x' = \ln(x + 1)$