Project log book

Project:	**Master thesis (Spatialtemporal modeling of NO_2 across Switzerland)**	
Update:	**2021-10-10**	
Working folder:	/Masterarbeit/analysis	(own MacBook)
GitHub repository:	https://github.com/phineeeeeeeeeeeeeeees/Masterarbeit	
Safe folder:	J:/EEH/EEM/Masters/TzeLi/Masterarbeit/analysis	(SwissTPH Drive)

input data	actions	output data	comments
/	initialize-renv.R	- renv/… - rent.lock	- initialize a new project-local environment with a private R library - Call `renv::snapshot()` to save the state of the project library to the lockfile (called renv.lock), - or call `renv::restore()` to revert to the previous state as encoded in the lockfile if your attempts to update packages introduced some new problems
Data collection			
https://biogeo.ucdavis.edu/data/diva/adm/CHE_adm.zip	`get-CH-shp.R`	`1_data/raw/Switzerland_shapefile/` - `CHE_adm0.shp` - `CHE_adm1.shp` - `CHE_adm2.shp` - `CHE_adm3.shp` - `AOI_4326.shp` - `AOI_21781.shp`	- download the Switzerland shapefile from DIVA-GIS - download, unzip - create the AOI • 50km buffer from Swiss boundary • stored as AOI_projection.shp, with two versions with different projections (4326 and 21781)
`1_data/raw/OMI-NO2/subset_OMI_MINDS_NO2d_1_20210419_074014.txt`	`get-OMI-NO2.R`	`1_data/raw/OMI-NO2/OMI-Aura_L3-OMI_MINDS_NO2d_{YYYYMMDD}_AOI.tif`	- download the OMI-NO2 data from NASA GES DISC - the .txt file with the download links must be acquired from https://disc.gsfc.nasa.gov/datasets/OMI_MINDS_NO2d_1/summary , by giving the searching criteria (date) - login required; API key provided • through `2_scripts/login-authentication/login_NASA-earthdata_OMI.R` - clipped to AOI

input data	actions	output data	comments
/	get-TROPOMI-NO2.R	1_data/raw/TROPOMI-NO2/ - query_results/records_filtered.RData - preprocessed/ S5P_OFFL_L2__NO2____{YYYYMMDD}_AOI.tif - preprocessed_resampled/ S5P_OFFL_L2__NO2____{YYYYMMDD}_AOI_rs.tif	- download the TROPOMI NO2 data from Sentinel-5 data hub using R package getSpatialData - login required; through 2_scripts/login-authentication/ login_CopHub.R - slow and unstable connection - clipped to AOI - cloud screened: QA_value > 0.5 - unit converted from mol/m^2 to molec/cm^2 - output: 2-band raster images (NO_2, QA_value) • band 1: NO2_tropospheric_column_QA_screened • band 2: QA_value - the spatial dimension of each image varies slightly (because of data availability) --> resample to the same X,Y dimensions • resample to a standard template (using the first image: x=44 y=32) • saved in a separate folder "**preprocessed_resampled**"
https://land.copernicus.eu/land-files/97824c12f357f50638d665b5a58707cd82857d57.zip	get-EU-DEM.R	1_data/raw/EU-DEM-1.1/ eu_dem_v11_E40N20_AOI.tif	- download the DEM data from the Copernicus Land Monitoring Service website (login required): https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1?tab=download - Switzerland lies in the sub-region E40N20 - clipped to AOI
/	get-ECMWF-CAMS-NO2.py	1_data/raw/ECMWF-CAMS-NO2/ CAMS-NO2.nc	- download CAMS NO2 data from Copernicus Atmosphere Data Store, clipped to AOI at query - Python API: cdsapi - the .grib file is shown to be corrupted; .netCDF is used instead - API key required (ADS): /Users/liutzuli/.cdsapirc • hidden files can be displayed/ edited by command + shift + .
~~/~~	~~get-ECMWF-CAMS-nrt-NO2.py~~	~~1_data/raw/ECMWF-CAMS-NO2/~~ ~~- CAMS-near-real-time-NO2_an.nc (step=0; analysis)~~ ~~- CAMS-near-real-time-NO2_fc.nc (step=0,3,6,9; forecast~~	~~- download CAMS near-real-time NO2 data from ECMWF WebAPI, clipped to AOI at query~~ ~~• use CAMS NRT instead of CAMS Reanalysis because of higher spatial resolution~~ ~~- Python API: ecmwfapi~~ ~~- the .grib file is shown to be corrupted; .netCDF is used instead~~ ~~- API key required (ADS): /Users/liutzuli/.ecmwfapirc~~ ~~• hidden files can be displayed/ edited by command + shift + .~~

input data	actions	output data	comments
/	get-ERA5-meteoro.py	1_data/raw/ECMWF-ERA5/ERA5-meteorology.nc	- download ERA-5 data from Copernicus Climate Data Store, clipped to AOI at query - Python API: cdsapi - API key required (CDS): /Users/liutzuli/.cdsapirc • Note: although using the same API, the keys used for CDS and ADS are different!! - .netCDF files are easier to work with in stars, therefore .netCDF is used instead of .grib
1_data/raw/MODIS-vegetation/query_result.txt	get-MODIS-vegetation_1.R	1_data/raw/MODIS-vegetation/raw-HDF4/MYD13Q1….hdf	- download MODIS Vegetation Indices (MYD13Q1) Version 6 data from NASA EarthData - the .txt file with the download links must be acquired from this link , by giving the searching criteria (date, area…) - login required; API key provided • through 2_scripts/login-authentication/login_NASA-earthdata_MODIS.R
1_data/raw/MODIS-vegetation/raw-HDF4/MYD13Q1….hdf	get-MODIS-vegetation_2.R	1_data/raw/MODIS-vegetation/NDVI-GTiff/MYD13Q1_NDVI_{YYYYMMDD}_AOI.tif	- *only applicable on Windows* (GDAL support issue for HDF4 on MacOS) - subset NDVI from the MODIS vegetation product - clipped to AOI - convert to GeoTIFF; original .hdf files deleted
https://eogdata.mines.edu/nighttime_light/annual/v20/2019/VNL_v2_npp_2019_global_vcmslcfg_c202101211500.average.tif.gz	get-VIIRS.R	1_data/raw/VIIRS-nighttime-light/VNL_v2_npp_2019_global_vcmslcfg_c202101211500.average_AOI.tif	- download annual global VIIRS nighttime lights from Earth Observation Group (EOG): https://eogdata.mines.edu/products/vnl/ - login required; retrieve access token set in R code (2_scripts/login-authentication/login_EOG_VIIRS.R) - clipped to AOI
http://cidportal.jrc.ec.europa.eu/ftp/jrc-opendata/GHSL/GHS_POP_EUROSTAT_EUROPE_R2016A/GHS_POP_SOURCE_EUROPE_R2016A_3035_100/V1-0/GHS_POP_SOURCE_EUROPE_R2016A_3035_100_v1_0.zip	get-GHS-population.R	1_data/raw/GHS-population/GHS_POP_SOURCE_EUROPE_R2016A_3035_100_v1_0_AOI.tif	- download the Global Human Settlement (GHS) population grid data from https://data.europa.eu/euodp/en/data/dataset/jrc-ghsl-ghs_pop_eurostat_europe_r2016a - login not required - clipped to AOI

input data	actions	output data	comments
https://land.copernicus.eu/land-files/83684d24c50f069b613e0dc8e12529b893dc172f.zip	get-CLC.R	1_data/raw/CORINE-land-cover/ - U2018_CLC2018_V2020_20u1_AOI.tif - metadata……	- download the CORINE 2018 land cover data from the Copernicus Land Monitoring Service website (login required): https://land.copernicus.eu/pan-european/corine-land-cover/clc2018?tab=download (format: 100m GeoTiff) - clipped to AOI
1_data/raw/CORINE-land-cover/ - U2018_CLC2018_V2020_20u1_AOI.tif - Legend/clc_legend_qgis_raster.qml (metadata)	<QGIS:> import .tif as raster layer; import .qml as symbology; Property> Symbology> … > Export color map to file	1_data/raw/CORINE-land-cover/Legend/clc_legend.txt	Get the metadata of the land cover codes (encoded in the .tif raster file) and save it as .txt file. (This file only came as .qml when downloading the data)
Preliminary data exploration			
- 1_data/raw/OMI-NO2/OMI-Aura_L3-OMI_MINDS_NO2d_{YYYYMMDD}_AOI.tif - 1_data/raw/TROPOMI-NO2/preprocessed_resampled/S5P_OFFL_L2__NO2____{YYYYMMDD}_AOI_rs.tif - 1_data/raw/Switzerland_shapefile/CHE_adm0.shp - 1_data/raw/EU-DEM-1.1/eu_dem_v11_E40N20_AOI.tif	compare-OMI-TROPOMI.R	- (graphs and tables) - 3_results/Markdown/compare-OMI-TROPOMI.html	- visualize to compare the OMI and TROPOMI data availability on selected dates - visualize to compare the TROPOMI NO2 product when using the default QA=0.5 and QA=0.75 - pixel-to-pixel comparison between OMI and TROPOMI NO2 products • simple linear regression of pixel-wise values • cross-table of missing values • disagreement of missing values • spatial distribution of missing values (per pixel) • temporal distribution of missing values (per season)
Data preparation			

input data	actions	output data	comments
1_data/raw/ - OMI-NO2/ OMI-Aura_L3- OMI_MINDS_NO2d_{YYYYMMDD }_AOI.tif - ECMWF-CAMS-NO2/ CAMS-NO2.nc - EU-DEM-1.1/ eu_dem_v11_E40N20_AOI.ti f - ECMWF-ERA5/ERA5- meteorology.nc - Switzerland_shapefile/ CHE_adm0.shp - Switzerland_shapefile/ AOI_4326.shp	impute-OMI.R	- (graphs and tables) - 1_data/processed/OMI_imputed/ OMI-Aura_L3- OMI_MINDS_NO2d_2019_daily_imputed_AOI.nc	- resample predictor datasets to OMI grids • CAMS (downscaling; using nearest neighbor resampling) • elevation (upscaling; using bilinear interpolation) • ERA5 (downscaling; using nearest neighbor resampling) - convert u,v wind components to wd and ws - combining spatial and spatialtemporal predictor sets - prepare a data.frame for model development - spatially-blocked 10-fold cross validation - random forest model - optimize predictor variable set • select hour for CAMS and ERA5 data • grid search over every combination • grid search for final model (fixed hour, different variable combinations) - final model: OMI_NO2 ~ CAMS_NO2_15 + CAMS_NO_15 + DEM + DOY + y + x • num.trees = 1000, mtry = 5 - evaluation of the final model: OOB-R2, slope, intercept, CV-R2 - project the model prediction - **output imputation result:** ifelse(is.na(OMI_NO2) , OMI_NO2_pred , OMI_NO2) - visualization of the model development results and the comparison of candidate models

input data	actions	output data	comments
1_data/raw/ - TROPOMI-NO2/ preprocessed_resampled/ S5P_OFFL_L2__NO2____{YYYYMMDD}_AOI_rs.tif - ECMWF-CAMS-NO2/ CAMS-NO2.nc - EU-DEM-1.1/ eu_dem_v11_E40N20_AOI.tif - ECMWF-ERA5/ERA5-meteorology.nc - Switzerland_shapefile/ CHE_adm0.shp - Switzerland_shapefile/ AOI_4326.shp	impute-TROPOMI.R	- (graphs and tables) - 1_data/processed/TROPOMI_imputed/ S5P_OFFL_L2__NO2____2019_daily_imputed_AOI.nc	- filter NO2 pixels using QA_value > **0.75** - resample predictor datasets to TROPOMI grids • CAMS (downscaling; using nearest neighbor resampling) • elevation (upscaling; using bilinear interpolation) • ERA5 (downscaling; using nearest neighbor resampling) - convert u,v wind components to wd and ws - combining spatial and spatialtemporal predictor sets - prepare a data.frame for model development - spatially-blocked 10-fold cross validation - random forest model - optimize predictor variable set • select hour for CAMS and ERA5 data • grid search over every combination • grid search for final model (fixed hour, different variable combinations) - final model: TROPOMI_NO2 ~ DEM + sp_12 + CAMS_NO_12 + CAMS_NO2_12 + y + x + DOY + t2m_12 + blh_12 + ws_12 + wd_12 + tcc_12 • num.trees = 500, mtry = 5 - evaluation of the final model: OOB-R2, slope, intercept, CV-R2 - project the model prediction - **output imputation result:** ifelse(is.na(TROPOMI_NO2) , TROPOMI_NO2_pred , TROPOMI_NO2) - visualization of the model development results and the comparison of candidate models

input data	actions	output data	comments
1_data/raw/ - Switzerland_shapefile/ CHE_adm0.shp - EU-DEM-1.1/ eu_dem_v11_E40N20_AOI.tif - GHS-population/ GHS_POP_SOURCE_EUROPE_R2016A_3035_100_v1_0_AOI.tif - VIIRS-nighttime-light/ VNL_v2_npp_2019_global_vcmslcfg_c202101211500.average_AOI.tif - CORINE-land-cover/ U2018_CLC2018_V2020_20u1_AOI.tif - Traffic/traffint - Intersections/int100m.tif - Roads/mjdens.tif - Roads/rdens.tif - Roads/nearmjrd.tif - NDVI_30m/ndvi_100m.tif	preparation_1-resample_sp.R	1_data/processed/cleaned/ spatial/ - elevation_100m.tif - population_100m.tif - nighttime-light_100m.tif - emissions-nox_100m.tif (5 bands for 5 sources) - traffic-intensity_100m.tif - major-road-density_100m.tif - all-road-density_100m.tif - n-intersections_100m.tif - dist-near-major-road_100m.tif - landcover-code_100m.tif (original coding) - landcover-group_binary_100m.tif (6 bands; binary land cover coding for 6 groups) - NDVI_sp_100m.tif	- coordinate system to work with: EPSG: 2056 (CH1903+/ LV95) - AOI: Swiss boundary + 10km - elevation: • bilinear-interpolation resampled (upscaling) to 100m - population • nearest-neighbor resampled (reprojected) - nighttime light: • nearest-neighbor downscaled to 100m - emissions (5 sources) • nearest-neighbor downscaled to 100m - traffic intensity • nearest-neighbor resampled (reprojected) - major road density • nearest-neighbor resampled (reprojected) - all road density • nearest-neighbor resampled (reprojected) - number of intersections • nearest-neighbor resampled (reprojected) - distance to nearest major road • nearest-neighbor resampled (reprojected) - land cover: • nearest-neighbor resampled (reprojected) • grouped into 6 groups (RES, IND, URBGN, BUILT, AGR, NAT) ; binary land cover coding for 6 groups - NDVI (spatial) • nearest-neighbor resampled (reprojected)

input data	actions	output data	comments

1_data/processed/
- Switzerland_shapefile/CHE_adm0.shp
- OMI_imputed/OMI-Aura_L3-OMI_MINDS_NO2d_2019_daily_imputed_AOI.nc
- TROPOMI_imputed/S5P_OFFL_L2__NO2____2019_daily_imputed_AOI.nc
- ECMWF-ERA5/ERA5-meteorology.nc
- MODIS-vegetation/NDVI-GTiff/MYD13Q1_NDVI_{YYYYMMDD}_AOI.tif
	preparation_1-resample_st.R	1_data/processed/cleaned/spatialtemporal/
		- OMI_daily_1000m.tif (365 bands)
		- TROPOMI_daily_1000m.tif (365 bands)
		- meteorological variables (each w/ 365 bands)
		  - blh_12H_daily_1000m.tif
		  - sp_12H_daily_1000m.tif
		  - t2m_12H_daily_1000m.tif
		  - tcc_12H_daily_1000m.tif
		  - tp_12H_daily_1000m.tif
		  - wd_12H_daily_1000m.tif
		  - ws_12H_daily_1000m.tif
		  - blh_15H_daily_1000m.tif
		  - sp_15H_daily_1000m.tif
		  - t2m_15H_daily_1000m.tif
		  - tcc_15H_daily_1000m.tif
		  - tp_15H_daily_1000m.tif
		  - wd_15H_daily_1000m.tif
		  - ws_15H_daily_1000m.tif
		- NDVI_daily_250m.tif (365 bands)
		  - NDVI_daily_EPSG9001.tif (without reprojection)
		- AOI_spatial-lookup-table_1000to100m.csv
		- AOI_spatial-lookup-table_250to100m.csv
			- coordinate system to work with: EPSG: 2056 (CH1903+/ LV95)
			- AOI: Swiss boundary + 10km
			- OMI-NO2, TROPOMI-NO2 (daily):
			  • bilinear-interpolation resampled to 1*1km
			- meteorological variables (7 variables, 3-hourly)
			  • only use 12H and 15H
			  • bilinear-interpolation resampled to 1*1km
			  • calculate ws and wd from u- and v-components
			- NDVI (16-day spatialtemporal):
			  • temporal interpolation to daily
			  • re-project (resample) to EPSG: 2056 at 250*250m (re-projection without resampling results in curvilinear grid)

input data	actions	output data	comments
1_data/processed/cleaned/spatialtemporal/ - blh_12H_daily_1000m.tif - blh_15H_daily_1000m.tif - NDVI_daily_250m.tif - OMI_daily_1000m.tif - sp_12H_daily_1000m.tif - sp_15H_daily_1000m.tif - t2m_12H_daily_1000m.tif - t2m_15H_daily_1000m.tif - tcc_12H_daily_1000m.tif - tcc_15H_daily_1000m.tif - tp_12H_daily_1000m.tif - tp_15H_daily_1000m.tif - TROPOMI_daily_1000m.tif - wd_12H_daily_1000m.tif - wd_15H_daily_1000m.tif - ws_12H_daily_1000m.tif - ws_15H_daily_1000m.tif	preparation_2-aggregate_st.R	1_data/processed/cleaned/spatialtemporal/ - blh_12H_annual_1000m.tif - blh_12H_monthly_1000m.tif - blh_15H_annual_1000m.tif - blh_15H_monthly_1000m.tif - NDVI_annual_250m.tif - NDVI_monthly_250m.tif - OMI_annual_1000m.tif - OMI_monthly_1000m.tif - sp_12H_annual_1000m.tif - sp_12H_monthly_1000m.tif - sp_15H_annual_1000m.tif - sp_15H_monthly_1000m.tif - t2m_12H_annual_1000m.tif - t2m_12H_monthly_1000m.tif - t2m_15H_annual_1000m.tif - t2m_15H_monthly_1000m.tif - tcc_12H_annual_1000m.tif - tcc_12H_monthly_1000m.tif - tcc_15H_annual_1000m.tif - tcc_15H_monthly_1000m.tif - tp_12H_annual_1000m.tif - tp_12H_monthly_1000m.tif - tp_15H_annual_1000m.tif - tp_15H_monthly_1000m.tif - TROPOMI_annual_1000m.tif - TROPOMI_monthly_1000m.tif - wd_12H_annual_1000m.tif - wd_12H_monthly_1000m.tif - wd_15H_annual_1000m.tif - wd_15H_monthly_1000m.tif - ws_12H_annual_1000m.tif - ws_12H_monthly_1000m.tif - ws_15H_annual_1000m.tif - ws_15H_monthly_1000m.tif	- pixel-wise temporal aggregation of the spatial-temporal variables

input data	actions	output data	comments
1_data/processed/cleaned/spatial - all-road-density_100m.tif - elevation_100m.tif - emissions-nox_100m.tif - landcover-group_binary_100m.tif - major-road-density_100m.tif - n-intersections_100m.tif - nighttime-light_100m.tif - population_100m.tif - traffic-intensity_100m.tif	preparation_2-aggregate_sp.R	1_data/processed/cleaned/spatial - all-road-density_radius{}_100m_.tif - elevation_radius{}_100m.tif - emissions-nox_radius{}_100m.tif - landcover-group_binary_radius{}_100m.tif - major-road-density_radius{}_100m.tif - n-intersections_radius{}_100m.tif - nighttime-light_radius{}_100m.tif - population_radius{}_100m.tif - traffic-intensity_radius{}_100m.tif	- Calculate focal ("moving window") values for the neighborhood of focal cells using a matrix of weights - various window sizes (radius=100, 200, 500, 1000, 2000, 5000, 10000m)

1_data/raw/
- Switzerland_shapefile/CHE_adm1.shp
- NO2-monitoring/
  - NO2_2019_idb_daily.csv
  - metadaten_idbluft.csv
1_data/processed/cleaned/spatialtemporal/
- (everything)
1_data/processed/cleaned/spatial
- (everything)
	preparation_3-extract.R	1_data/processed/cleaned/extracted/
		- annual_scaled.csv
		- monthly_scaled.csv
		- daily_scaled.csv
			- the coordinates of the measurement sites are in “metadaten_idbluft.csv”
			- clean the measured NO2 data
			  • wide table to long table
			  • clean Station_name
			  • annual and monthly mean of the measured NO2 are also calculated
			- log-transformation for: enox, light, population, traffint, mjdens, rdens, intersection
			- scale (standardize) the predictor variables
			- pixel values of the predictor variables (spatial and spatial-temporal) are extracted using the coordinates
			- summarize (join) the measured NO2 values and the extracted pixel values of the predictor variables
			  • annual, monthly, daily
			- the meteorological variables come with 12H and 15H: use 15H for the OMI model and 12H for the TROPOMI (because of the CAMS and ERA tilmestep used in the imputation models)

1_data/processed/cleaned/spatial - (everything)	preparation_4-spatialDF.R	1_data/processed/cleaned/data-frame/spatial_df.csv	- reshape the spatial predictor variables into a data.frame (8361297*148) for later modeling prediction - coordinate system: EPSG2056
1_data/processed/cleaned/spatialtemporal - (everything)	preparation_4-spatialtemporalDF.R	1_data/processed/cleaned/data-frame/ - spatialtemporal_annual_df.csv - spatialtemporal_NDVI_annual_df.csv - spatialtemporal_monthly_df.csv - spatialtemporal_NDVI_monthly_df.csv - spatialtemporal_daily_df.csv - spatialtemporal_NDVI_daily_df.csv	- reshape the spatial-temporal predictor variables into several data.frame for later modeling prediction - (daily NDVI not yet succeeded)
Modeling			

input data	actions	output data	comments

1_data/processed/cleaned/extracted/annual_scaled.csv
1_data/raw/Switzerland_shapefile/CHE_adm0.shp
1_data/raw/NO2-monitoring/metadaten_idbluft_supplemented.csv
	spatial-CV.R	1_data/processed/cleaned/extracted/NO2-sites_5-fold-CV.shp	- cross validation design: k-fold random and spatially-blocked cross validation
			- 5-fold
			- output shapefile can be used in later modeling for partitioning cross validation
			- columns:
			  - Station_name: 100 monitoring sites
			  - CV: random-split
			  - spatial_CV: spatially-blocked

(model prediction data.frame)	utils_model-eval.R	/	- source(utils_model-eval.R) in the modeling scripts to call the functions
			- summary:
			  - eval_performance_indices(): summarize the model performance indices as a data.frame
			  - Training and cross validation R2 and RMSE, slope, intercept
			  - eval_resid_moran(): Moran's I test for spatial autocorrelation of the residuals
			- visualization:
			  - plot_obs_pred(): plot the predicted values against the observed values
			  - plot_resid(): plot the residual diagnostics plots
			  - observed—residuals; histogram and QQ-plot of residuals
			  - plot_resid_month(): plot the distribution of residuals by month
			  - plot_resid_map(): plot the annual-average residual map

input data	actions	output data	comments

1_data/processed/cleaned/
extracted/
- annual_scaled.csv
- NO2-sites_5-fold-CV.shp
	annual_SLR.R
	
	source("2_scripts/
	utils_model-eval.R")
		3_results/output-data/model_annual/
		- SLR_summary/
		  - OMI.csv
		  - TROPOMI.csv
		  - spatial.csv
		- observed-predicted/
		  - SLR_OMI.csv
		  - SLR_TROPOMI.csv
		  - SLR_spatial.csv
		- indices/
		  - SLR_OMI.csv
		  - SLR_TROPOMI.csv
		  - SLR_spatial.csv
		- Moran/
		  - SLR_OMI.csv
		  - SLR_TROPOMI.csv
		  - SLR_spatial.csv
		3_results/output-model/model_annual/
		- SLR_OMI.rds
		- SLR_TROPOMI.rds
		- SLR_spatial.rds
			- Supervised stepwise linear regression
			  • define the expected direction of effect
			  • initial univariate regression
			  • iterations of stepwise selection
			  - increase at least 0.01 R^2
			  • final check
			  - remove the variables whose p-value > 0.10 (iteratively)
			  - remove the variables with VIF>3 to avoid multi-collinearity
			- cross validation
			  • conventional random-split, spatially-blocked
			- Evaluation
			  • linear regression summary table
			  • performance indices
			  • diagnostic plots:
			  - predicted versus observed
			  - residual diagnostic plots
			  - map of annual average residuals

1_data/processed/cleaned/ extracted/ - annual_scaled.csv - NO2-sites_5-fold-CV.shp 1_data/raw/ Switzerland_shapefile/ CHE_adm0.shp 3_results/output-data/ model_annual/SLR_summary/ - spatial.csv - OMI.csv - TROPOMI.csv	annual_GWR.R source("2_scripts/ utils_model-eval.R")	3_results/output-data/model_annual/ - observed-predicted/ - indices/ - Moran/ - GWR_coef (tables and figures) 3_results/output-model/model_annual - GWR_spatial.rds - GWR_OMI.rds - GWR_TROPOMI.rds	- model development - cross validation - model evaluation - spatially varying regression coefficients
1_data/processed/cleaned/ extracted/ - annual_scaled.csv - NO2-sites_5-fold-CV.shp	annual_RF_gridsearch.R	3_results/output-data/model_annual/ RF_grid-search/hyper_evaluation.csv	grid search of random forest hyperparameters (with TROPOMI)

input data	actions	output data	comments
1_data/processed/cleaned/ extracted/ - annual_scaled.csv - NO2-sites_5-fold-CV.shp	annual_RF.R source("2_scripts/ utils_model-eval.R")	3_results/output-data/model_annual/ - observed-predicted/ - indices/ - Moran/ (tables and figures) 3_results/output-model/model_annual - RF_OMI.rds - RF_TROPOMI.rds	- model development - model evaluation
1_data/processed/cleaned/ extracted/ - annual_scaled.csv - NO2-sites_5-fold-CV.shp	annual_XGB_gridsearch.R	3_results/output-data/model_annual/ GBM_grid-search/hyper_evaluation.csv	grid search of xgboost hyper parameters (eta, max_depth, min_child_weight, subsample, colsample_bytree)
1_data/processed/cleaned/ extracted/ - annual_scaled.csv - NO2-sites_5-fold-CV.shp 3_results/output-data/ model_annual/GBM_grid-search/ hyper_evaluation.csv	annual_XGB.R source("2_scripts/ utils_model-eval.R")	3_results/output-data/model_annual/ - observed-predicted/ - indices/ - Moran/ (tables and figures) 3_results/output-model/model_annual - XGB_OMI.rds - XGB_TROPOMI.rds	- model development - model evaluation
1_data/processed/cleaned/ extracted/ - annual_scaled.csv - NO2-sites_5-fold-CV.shp	annual_LGB_gridsearch.R	3_results/output-data/model_annual/ LGB_grid-search/hyper_evaluation.csv	grid search of LightGBM hyper parameters
1_data/processed/cleaned/ extracted/ - annual_scaled.csv - NO2-sites_5-fold-CV.shp 3_results/output-data/ model_annual/LGB_grid-search/ hyper_evaluation.csv	annual_LGB.R source("2_scripts/ utils_model-eval.R")	3_results/output-data/model_annual/ - observed-predicted/ - indices/ - Moran/ (tables and figures) 3_results/output-model/model_annual - LGB_OMI.txt - LGB_TROPOMI.txt	- model development - model evaluation
/	keras_configure.R	/	- configure the keras and TensorFlow in R • create a new environment "r-reticulate" • install scipy

input data	actions	output data	comments

hyperparm_vector: a named vector with the hyperparameters	utils_define-NN.R
	
	source("2_scripts/utils_model-eval.R")
		/	- defining the Keras neural network models by providing the hyperparameters
			  • layers, neurons, epochs, batch.size , regularization, regularization_factor, dropout_rate
			  • n_var
			- loss: mean absolute error
			- optimizer: adam

1_data/processed/cleaned/extracted/ - annual_scaled.csv - NO2-sites_5-fold-CV.shp	annual_NN_gridsearch.R	3_results/output-data/model_annual/NN_grid-search/hyper_evaluation.csv output-graph/model_annual/NN/…(graphs)	grid search of DNN hyper parameters (number of hidden layers, number of neurons in each hidden layer, epochs, batch sizes, regularization, garson's variable selection)
1_data/processed/cleaned/extracted/ - annual_scaled.csv - NO2-sites_5-fold-CV.shp 3_results/output-data/model_annual/NN_grid-search/hyper_evaluation.csv	annual_NN.R source("2_scripts/utils_model-eval.R")	3_results/output-data/model_annual/ - observed-predicted/ - indices/ - Moran/ (tables and figures) 3_results/output-model/model_annual - NN_OMI.rds - NN_TROPOMI.rds	- model development - model evaluation
1_data/processed/cleaned/extracted/ - monthly_scaled.csv - NO2-sites_5-fold-CV.shp	monthly_SLR.R source("2_scripts/utils_model-eval.R")	3_results/output-data/model_monthly/ - observed-predicted/ - indices/ - Moran/ (tables and figures) 3_results/output-model/model_monthly - SLR_OMI.rds - SLR_TROPOMI.rds	- transformation of month (fit cosine wave) - variable selection: supervised stepwise linear regression algorithm - cross validation - model evaluation
1_data/processed/cleaned/extracted/ - monthly_scaled.csv - NO2-sites_5-fold-CV.shp	monthly_lmer.R source("2_scripts/utils_model-eval.R")	3_results/output-data/model_monthly/ - observed-predicted/ - indices/ - Moran/ (tables and figures) 3_results/output-model/model_monthly - SLMER_OMI.rds - SLMER_TROPOMI.rds	- variable selection: supervised stepwise linear regression algorithm - cross validation - model evaluation

input data	actions	output data	comments
1_data/processed/cleaned/ extracted/ - monthly_scaled.csv - NO2-sites_5-fold-CV.shp	monthly_RF_gridsearch.R	3_results/output-data/model_monthly/ RF_grid-search/hyper_evaluation.csv	grid search of random forest hyperparameters (with TROPOMI)

1_data/processed/cleaned/
extracted/
- monthly_scaled.csv
- NO2-sites_5-fold-CV.shp
 	
3_results/output-data/
model_monthly/RF_grid-
search/
hyper_evaluation.csv
	monthly_RF.R
	
	source("2_scripts/
	utils_model-eval.R")
		3_results/output-data/model_monthly/
		- observed-predicted/
		- indices/
		- Moran/
		(tables and figures)
		
		3_results/output-model/model_monthly
		- RF_OMI.rds
		- RF_TROPOMI.rds
			- model development
			  - screening of important predictor variables
			  - cross validation
			- model evaluation

1_data/processed/cleaned/ extracted/ - monthly_scaled.csv - NO2-sites_5-fold-CV.shp	monthly_XGB_gridsearch.R	3_results/output-data/model_monthly/ XGB_grid-search/hyper_evaluation.csv	grid search of xgboost hyperparameters (with TROPOMI)

1_data/processed/cleaned/
extracted/
- monthly_scaled.csv
- NO2-sites_5-fold-CV.shp
-
3_results/output-data/
model_monthly/XGB_grid-
search/
hyper_evaluation.csv
	monthly_XGB.R
	
	source(“2_scripts/
	utils_model-eval.R")
		3_results/output-data/model_monthly/
		- observed-predicted/
		- indices/
		- Moran/
		(tables and figures)
		
		3_results/output-model/model_monthly
		- XGB_OMI.rds
		- XGB_TROPOMI.rds
			- data preparation for xgboost
			- model development
			  - screening of important predictor variables
			  - cross validation
			- model evaluation

1_data/processed/cleaned/ extracted/ - monthly_scaled.csv - NO2-sites_5-fold-CV.shp	monthly_LGB_gridsearch.R	3_results/output-data/model_monthly/ LGB_grid-search/hyper_evaluation.csv	grid search of LightGBM hyperparameters (with TROPOMI)

input data	actions	output data	comments

1_data/processed/cleaned/extracted/
- monthly_scaled.csv
- NO2-sites_5-fold-CV.shp
 	
3_results/output-data/model_monthly/LGB_grid-search/hyper_evaluation.csv
	monthly_LGB.R
	
	source("2_scripts/utils_model-eval.R")
		3_results/output-data/model_monthly/
		- observed-predicted/
		- indices/
		- Moran/
		(tables and figures)
		
		3_results/output-model/model_monthly
		- LGB_OMI.txt
		- LGB_TROPOMI.txt
			- data preparation for LightGBM
			- model development
			  - screening of important predictor variables
			  - cross validation
			- model evaluation

1_data/processed/cleaned/extracted/ - monthly_scaled.csv - NO2-sites_5-fold-CV.shp	monthly_NN_gridsearch.R	3_results/output-data/model_monthly/NN_grid-search/hyper_evaluation.csv	grid search of NN hyperparameters (with TROPOMI)
1_data/processed/cleaned/extracted/ - monthly_scaled.csv - NO2-sites_5-fold-CV.shp 3_results/output-data/model_monthly/NN_grid-search/hyper_evaluation.csv	monthly_NN.R source("2_scripts/utils_model-eval.R") source("2_scripts/utils_define-NN.R")	3_results/output-data/model_monthly/ - observed-predicted/ - indices/ - Moran/ (tables and figures) 3_results/output-model/model_monthly - NN_OMI.hdf - NN_TROPOMI.hdf	model development - feature selection - hyperparameters (selected from the hyperparameters of the monthly model because of high grid search time requirement) - cross validation model evaluation

1_data/processed/cleaned/extracted/
- daily_scaled.csv
- NO2-sites_5-fold-CV.shp
	daily_SLR.R
	
	source("2_scripts/utils_model-eval.R")
		3_results/output-data/model_daily/
		- observed-predicted/
		- indices/
		- Moran/
		(tables and figures)
		
		3_results/output-model/model_daily
		- SLR_OMI.rds
		- SLR_TROPOMI.rds
			- data inspection and preparation
			  - square-root transformation for NO2
			  - fit cosine wave: DOY, weekly cycle
			- variable selection: supervised stepwise linear regression algorithm
			- cross validation
			- model evaluation

input data	actions	output data	comments
1_data/processed/cleaned/ extracted/ - daily_scaled.csv - NO2-sites_5-fold-CV.shp	daily_lmer.R source("2_scripts/ utils_model-eval.R")	3_results/output-data/model_daily/ - observed-predicted/ - indices/ - Moran/ (tables and figures) 3_results/output-model/model_daily - SLMER_OMI.rds - SLMER_TROPOMI.rds	- square-root transformation for NO2 - variable selection: supervised stepwise linear regression algorithm - cross validation - model evaluation
1_data/processed/cleaned/ extracted/ - daily_scaled.csv - NO2-sites_5-fold-CV.shp	daily_RF_gridsearch.R	3_results/output-data/model_daily/ RF_grid-search/hyper_evaluation.csv	grid search of random forest hyperparameters (with TROPOMI)

1_data/processed/cleaned/
extracted/
- daily_scaled.csv
- NO2-sites_5-fold-CV.shp
 	
3_results/output-data/
model_daily/RF_grid-
search/
hyper_evaluation.csv
	daily_RF.R
	
	source("2_scripts/
	utils_model-eval.R")
		3_results/output-data/model_daily/
		- observed-predicted/
		- indices/
		- Moran/
		(tables and figures)
		
		3_results/output-model/model_daily
		- RF_OMI.rds
		- RF_TROPOMI.rds
			- model development
			  - screening of important predictor variables
			  - cross validation
			- model evaluation

1_data/processed/cleaned/ extracted/ - daily_scaled.csv - NO2-sites_5-fold-CV.shp	daily_XGB_gridsearch.R	3_results/output-data/model_daily/ XGB_grid-search/ - hyper_evaluation.csv - hyper_evaluation_full.csv	grid search of xgboost hyperparameters (with TROPOMI)

1_data/processed/cleaned/
extracted/
- daily_scaled.csv
- NO2-sites_5-fold-CV.shp
 	
3_results/output-data/
model_daily/XGB_grid-
search/
hyper_evaluation_full.csv
	daily_XGB.R
	
	source("2_scripts/
	utils_model-eval.R")
		3_results/output-data/model_daily/
		- observed-predicted/
		- indices/
		- Moran/
		(tables and figures)
		
		3_results/output-model/model_daily
		- XGB_OMI.rds
		- XGB_TROPOMI.rds
			- data preparation for xgboost
			- model development
			  - screening of important predictor variables
			  - cross validation
			- model evaluation

input data	actions	output data	comments
1_data/processed/cleaned/ extracted/ - daily_scaled.csv - NO2-sites_5-fold-CV.shp	daily_LGB_gridsearch.R	3_results/output-data/model_daily/ LGB_grid-search/hyper_evaluation.csv	grid search of LightGBM hyperparameters (with TROPOMI)

1_data/processed/cleaned/
extracted/
- daily_scaled.csv
- NO2-sites_5-fold-CV.shp
 	
3_results/output-data/
model_daily/LGB_grid-
search/
hyper_evaluation.csv
	daily_LGB.R
	
	source("2_scripts/
	utils_model-eval.R")
		3_results/output-data/model_daily/
		- observed-predicted/
		- indices/
		- Moran/
		(tables and figures)
		
		3_results/output-model/model_daily
		- LGB_OMI.txt
		- LGB_TROPOMI.txt
			- data preparation for LightGBM
			- model development
			  - screening of important predictor variables
			  - cross validation
			- model evaluation

	~~daily_NN_gridsearch.R~~		
1_data/processed/cleaned/ extracted/ - daily_scaled.csv - NO2-sites_5-fold-CV.shp 3_results/output-data/ model_monthly/NN_grid- search/ hyper_evaluation.csv	daily_NN.R source("2_scripts/ utils_model-eval.R") source("2_scripts/ utils_define-NN.R")	3_results/output-data/model_daily/ - observed-predicted/ - indices/ - Moran/ (tables and figures)	model development - feature selection - hyperparameters (selected from the hyperparameters of the monthly model because of high grid search time requirement) - cross validation model evaluation
Projection			
/	projection_Zurich.R	1_data/processed/cleaned/data-frame/ - Zurich_area.shp - Zurich_area.csv	Create a shapefile of the Zurich area for model projection

3_results/output-model/
model_annual
- the models
  (as .rds, .txt, .hdf
  files)
1_data/processed/cleaned/
data-frame/
annual_full_df_Zurich.csv
	annual_projection.R	3_results/output-data/model_annual/
		- projection_annual_Zurich.csv
		- projection_annual_Zurich.tif
			model projection
			(only making model projection at Zurich area)

input data	actions	output data	comments
3_results/output-model/ model_monthly - the models (as .rds, .txt, .hdf files) 1_data/processed/cleaned/ data-frame/ monthly_full_df_Zurich.csv	monthly_projection.R	3_results/output-data/model_monthly/ projection_monthly_Zurich - projection_monthly_Zurich.csv - SLR_OMI.tif - SLR_TROPOMI.tif - SLMER_OMI.tif - SLMER_TROPOMI.tif - RF_OMI.tif - RF_TROPOMI.tif - XGB_OMI.tif - XGB_TROPOMI.tif - NN_OMI.tif - NN_TROPOMI.tif	model projection (only making model projection at Zurich area)
3_results/output-data/ model_annual/ projection_annual_Zurich.t if 3_results/output-data/ model_monthly/ projection_monthly_Zurich	mapping.R	3_results/output-graph/model_annual/ - mapping.png - mapping_zoomed.png 3_results/output-graph/model_monthly - mapping.png - mapping_zoomed.png	Visualization of the model projection maps
Further model evaluation, comparison, discussion			
1_data/processed/cleaned/ extracted//NO2-sites_5- fold-CV.shp 1_data/raw/Roads/ Roads_SonBase_sel.shp	site-road-dist.R	1_data/processed/cleaned/extracted/ site-road-distance.csv	Calculate the distance between the monitoring sites in the training data and the roads
3_results/output-data/ model_annual/ - indices - observed-predicted - moran 1_data/processed/cleaned/ extracted/site-road- distance.csv	annual_model-comparison.R	(graphs and tables)	- tidy table for presentation and paper - spatial CV residual diagnostics - correlations between different model CV residuals

input data	actions	output data	comments
3_results/output-data/ model_monthly/ - indices - observed-predicted - moran 1_data/processed/cleaned/ extracted/site-road- distance.csv	monthly_model-comparison.R	(graphs and tables)	- tidy table for presentation and paper - spatial CV residual diagnostics - temporal CV residual diagnostics - correlations between different model CV residuals
3_results/output-data/ model_daily/ - indices - observed-predicted - moran 1_data/processed/cleaned/ extracted/site-road- distance.csv	daily_model-comparison.R	(graphs and tables)	- tidy table for presentation and paper - spatial CV residual diagnostics - temporal CV residual diagnostics - temporal autocorrelation of model residuals - correlations between different model CV residuals - aggragate daily estimation to annual mean - comparison: variable selection
3_results/output-data/ model_daily/indices 3_results/output-data/ model_monthly/indices 3_results/output-data/ model_annual/indices	model-comparison.R	(graphs) 3_results/output-graph/ - model_comparison_TROPOMI.png - comparison_OMI_TROPOMI.png	Graphs: comparing all models (only TROPOMI) and comparison the two products