REACH Documentation

Medinat Akindele

2025-03-17

1. REACH Model Description

REACH is a reduced-complexity model with a Gaussian-plume framework and can be used to examine air pollution policies. The model is designed to be location-agnostic by allowing for adjustment of input data and geography information. Every location in the model (sources and receptors) can be designated by their population-weighted/geometric centroids or any coordinates of interest.

There are three main outputs of the REACH model:

1.1 Ambient PM2.5 concentrations

The model predicts ambient PM2.5 concentrations from emissions of primary PM2.5, SO2, NOx, NH3 and VOCs.

1.2 Source-receptor matrix

A source is a point location of emissions and a receptor is a location upwind of an emission source. A source-receptor vector is the change in PM2.5 concentration downwind from a source location from a 1 tonne change in emissions of a PM2.5 precursor. The source-receptor matrices represent the following relationships:

change in emissions –> change in concentrations (ug/m3/tonne)

- primary PM2.5 –> primary PM2.5
- SO2 –> sulfate
- NOx –> total nitrate (gas + particle)
- NH3 –> total ammonium (gas + particle)
- VOCs –> SOA

1.3 Marginal social costs

Source location-specific aggregate PM2.5 health impacts from emitting a tonne of a PM2.5 precursor (mortalities/tonne or $/tonne).

For detailed model description, please see the main paper: Akindele, Medinat, Peter J. Adams, Nicholas Z. Muller. "REACH: an accessible and globally portable air quality model for policy and health impact assessment." In preparation.

Please contact Medinat Akindele at makindel@alumni.cmu.edu for any questions.

The following sections outline description of the main folders, files, and scripts and provides instructions to run the model.

Link to Zenodo for additional files: https://zenodo.org/records/15659310

2. Description of folders

RCM_workshop

This directory contains valuable resources and practice problems for those interested in learning about Reduced-Complexity Models (RCMs). It includes a presentation on RCMs with an accompanying audio recording by Dr. Peter Adams at Carnegie Mellon University. The folder also features background materials and application papers related to RCMs. Additionally, five practice problems are provided to help reinforce key concepts.

REACH_MAIN

This directory contains the model scripts to generate the source-receptor matrices. Version A and Version B scripts are similar, except that Version A is suitable for emission inventories where stack information for point sources is available. For less detailed emission inventories, Version B is more suitable.

In Version_A and Version_B are the following folders:

- model:
  - domain_inputs.R: Edit to specify inputs for the domain (list of emission sources and receptor locations, sub divisions (grids within each source location), and the name specified for the location identifier)
  - input_info.R: Input variables and functions that are used for the model_script
  - model_script.R: Script to generate source-receptor matrix
- source_receptor: source-receptor scripts for each pollutant and height that link to the scripts in the model folder
- dispersion: contains the Gaussian plume equation and relevant loss processes (deposition and/or chemistry) for each pollutant

REACH_SA

Directory for REACH-Southern Africa (SA) with input data, post-processing scripts, and results. For REACH-SA, the Version B scripts in REACH_MAIN were used.

List of folders:

- inputs
  - emissions: emission files for each pollutant and height
  - met_data: meteorology files for each source location. For all files, download met_data_SA.zip at the Zenodo link.
- results
  - concentrations: baseline PM2.5 concentrations
  - matrices: source-receptor matrices for each pollutant and height. Each row represents a source location and each column represents a receptor location. The units are ug/m3/tonne.
  - social_costs: marginal social costs (mortalities per tonne) for each source location
- scripts
  - concentrations: post-processing scripts to get baseline PM2.5 concentrations
  - GBD_mortality: folder with scripts to estimate marginal social costs. Please read GBD_method_MSC.docx prior to preparing input data and running scripts.

REACH_US

Directory for REACH-United States (US) with input data, post-processing scripts, and results. For REACH-US, the Version A scripts in REACH_MAIN were used.

List of folders:

- inputs
  - Sub-folders:
    - emissions: emission files for each pollutant and height
    - met_data: meteorology files for each source location. For all files, download met_data_US.zip at the Zenodo link.
  - Files:
    - input_info.txt: detailed information about each file in this folder
- results
  - Files:
    - output_info.txt: detailed information about each file in this folder (2017 National Emissions Inventory (NEI) results)
    - concentrations: baseline PM2.5 concentrations
    - matrices: source-receptor matrices of each pollutant and height. Each row represents a source location and each column represents a receptor location. The units are ug/m3/tonne. For all files, download matrices_US_2017.zip at the Zenodo link and add to this folder.
    - social_costs: marginal social costs ($/tonne) for each source location in the column named 'marginal_damage'
- additional_outputs: same sub-folders as the results folder, except for the 2005 NEI year. The S-R matrices are located at the Zenodo link (see matrices_US_2005.zip).
- scripts
  - concentrations: post-processing scripts to get baseline PM2.5 concentrations
  - social_costs: scripts to get marginal social costs
- tracts
  - The folder containing the tract source-receptor data can be made available upon reasonable request.

3. Steps to prepare the source-receptor matrices

Follow these steps to run the scripts for the source-receptor matrices. The folder and file descriptions have been described above.

- Step 1: Download Rstudio
  - https://posit.co/products/open-source/rstudio/
- Step 2: Create a main directory for the REACH project
- Step 3: Create two sub-directories in the main directory:

  - Folder for the domain: Add an R project folder (.Rproj extension)
  - Copy over the REACH_Main folder: Use VERSION A or VERSION B scripts

- Step 4: Create a file specifying centroid lat, lon coordinates of emission locations (sources) and receptor locations

- Step 5: Prepare emissions and meteorology data files There are scripts available for global datasets: ERA5 meteorology (used for REACH-US and REACH-SA), EDGARv6.1 emissions (used for REACH-SA), and GFED biomass burning emissions (used for REACH-SA). For more information, see section 4. If using other datasets, prepare emission files for each pollutant and source location. Designate annual emissions in units of metric tonnes.

- Step 6: Generate additional input data

  - Sub-grids (05_subgrids.R)

- Step 7: Open the R project folder in Rstudio

  - Open model_script.R and input.info.R in VERSION A/VERSION B
  - Edit the domain_input.R script to designate emission locations and receptor locations for the domain

- Step 8: To run the source-receptor script for each pollutant and height

  - In VERSION_A/source_receptor or VERSION B/source_receptor folder:
  - Run each SR_pollutant_height script

4. Preparation of input data

The 'Inputs' folder contains scripts to prepare emissions and meteorology data for REACH. The scripts were prepared for global EDGAR emissions and ERA5 meteorology datasets as the default option. These datasets can be used when local emissions and meteorology data are unavailable. For the entire directory with input and output files, download 'Inputs.zip' at the Zenodo link. There are also scripts for DACCIWA emissions and can be made available upon reasonable request.

4.1 EDGAR emissions

Dataset: EDGAR v6.1 Global Air Pollutant Emissions https://edgar.jrc.ec.europa.eu/dataset_ap61

Note that a more updated version is available. Adjust the script (01_download_emissions.R) accordingly when downloading data from the new dataset.

The following scripts are used to prepare the EDGAR emissions data for the REACH model. Input information from the Southern Africa domain is applied to illustrate the use of the scripts. Replace the inputs accordingly for the domain of interest.

01_download_emissions.R

This script downloads EDGARv6.1 annual emissions for all sectors from the online website. There is a function to unzip the folders and extract the emissions in nc format.

02_process_emissions.R

This script maps EDGAR sector emissions from 0.1 degrees x 0.1 degrees spatial resolution into administrative geography resolution. The annual emission fluxes are converted into units of tonnes/year.

03_aggregate_edgar.R

This script consolidates the EDGAR sector emissions for each location and pollutant. EDGAR doesn't provide detailed information about the stack heights of emission sources at elevated heights. For that reason, certain sectors were assumed to have all emissions at the ground-level or at an elevated height. These assumptions are documented in a file called edgar_sectors.csv, which can be adjusted based on knowledge of your domain. If using the newer version of EDGAR, please check the pollutant and sector list on the website and edit edgar_sectors.csv if needed.

The EDGAR sector emissions are then aggregated for each location and pollutant according to the assumption on whether the sector comprises of entirely ground-level or elevated sources. For ground-level sources, an effective stack height of 10 m was assumed. For elevated sources in Southern Africa, an effective stack height of 220 m was assumed based on information from South African power plants. In the script, adjust the stack information for your domain accordingly.

4.2 ERA5 meteorology

Dataset: ERA5 hourly data on single levels from 1940 to present https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=download

1. Download the ERA5 reanalysis data

   Step 1: Select the following variables:

   - 10m u-component of wind
   - 10m v-component of wind
   - 2m dewpoint temperature
   - 2m temperature
   - Surface solar radiation downwards
   - Total cloud cover
   - Surface pressure

   Step 2: Select year of interest, all days and times

   Step 3: Select geographical region of interest

   Step 4: Download ERA5 nc file

2. Prepare meteorology files for REACH

   Run 04_meteorology.R

4.3 GFED biomass burning emissions

- Scripts located in biomass folder
- See GFED_preparation.docx for instructions

5. Post-processing scripts

5.1 Baseline PM2.5 concentrations

- Open the REACH_US.Rproj folder in Rstudio
- Edit and run the 01_concentration.R script
- Edit and run the 02_partition.R script

Adjust inputs as neccessary for the domain of interest

5.2 Marginal social costs

There are two methods used for estimating marginal social costs. If using the Global burden of disease method for PM2.5-attributed mortalities, use scripts in the REACH_SA/GBD_mortality folder. Otherwise if using all-cause mortality rates and relative risks for the United States, use scripts in the REACH_US/social_costs folder.

6. Developing domain-specific calibration coefficients

This section outlines the steps to determine calibration coefficients for primary PM2.5, sulfate, nitrate, ammonium, and secondary organic carbon.

Inorganic PM2.5 calibration

Follow this order: calibrate sulfate first, then nitrate, and lastly ammonium

1. Assume an initial guess for the sulfate calibration and multiply the model predictions by this value. Calculate the mean bias between the model predictions and measured sulfate. Adjust the sulfate calibration until the mean bias is close to 0.
2. After finalizing the sulfate calibration, determine the nitrate calibration. Assume an initial guess for the total nitrate calibration and and multiply the model predictions by this value. Estimate particulate nitrate with the regression in the 01_parition.R script. Calculate the mean bias between the model predictions and measured particulate nitrate. Adjust the nitrate calibration until the mean bias is close to 0.
3. Lastly, determine the ammonium calibration. Assume an initial guess for the total ammonium calibration and multiply the model predicitons of total ammonium by this value. Estimate particulate ammonium with the equations in 01_partition.R script. Calculate the mean bias between the model predictions and measured particulate ammonium. Adjust the ammonium calibration until the mean bias is close enough to 0. Because of the non-linear interactions between sulfate, nitrate and ammonium, and the assumptions in estimating particulate ammonium, the mean bias will likely be larger than the sulfate and nitrate mean biases against observations.

Primary PM2.5 calibration

If there are quality measurements of primary PM2.5 or sub-components OC and BC, then compare the model predictions of primary PM2.5 to observations of primary PM2.5 and minimize the mean bias to be close to 0.

Otherwise if observations are unreliable or unavailable, then follow the same process as REACH-US. For REACH-US, the national emission-weighted marginal social cost for primary PM2.5 was calculated. This was compared to the average primary PM2.5 marginal social costs for APEEP, InMAP and EASIUR reduced-complexity models. The REACH national average primary PM2.5 marginal social cost was then scaled to fall within the range of the other models. Similar comparisons for other regions can be done given marginal social costs for other models are available. If data to conduct either of the above is unavailable, then the REACH-US primary PM2.5 calibration can be assumed for another region.

Secondary organic carbon (OC) calibration

This section outlines the process to estimate the VOC/secondary organic carbon calibration. For the Unites States, measurements of total (primary and secondary) OC are available and not available for secondary

OC. Because of this reason, total OC in the model is calibrated to measured total OC and the resulting calibration coefficient is used to tune secondary OC.

The process is summarized in the following steps:

Step 1: Apply PM2.5 calibration to OC primary

```
OC_primary_cal = OC_primary * PM25_cal
```

Step 2: Set an initial guess for the OC calibration

Step 3: Multiply the OC calibration in step 2 by secondary OC

```
OC_secondary_cal = OC_secondary * OC_cal
```

Step 4: Add total organic carbon

```
OC_tot_cal = OC_primary_cal + OC_secondary_cal
```

Step 5: Compare modelled total OC (OC_tot_cal) to measurements of total OC and calculate mean bias

Step 6: If the mean bias is not close to 0, adjust initial guess for OC calibration and repeat steps 2-6

NOTE: If speciated PM2.5 observations are not available to develop region-specific calibrations, use REACH-US calibrations for your domain and note assumption. The REACH-US calibrations have been shown to work reasonably well for REACH-SA.

7. Evaluation of total PM2.5

Lastly, after calibrating each PM2.5 species, aggregate the concentrations of each calibrated PM2.5 species to determine total PM2.5 (01_partition.R). Compare to measurements of total PM2.5 and ensure that the mean bias is relatively low due to the calibration process.

8. License