Beginners Handbook on

Assembling Air Pollution Models

Sarath Guttikunda
Sai Krishna Dammalapati
SIM-air working paper series # 52-2024

About UrbanEmissions.Info

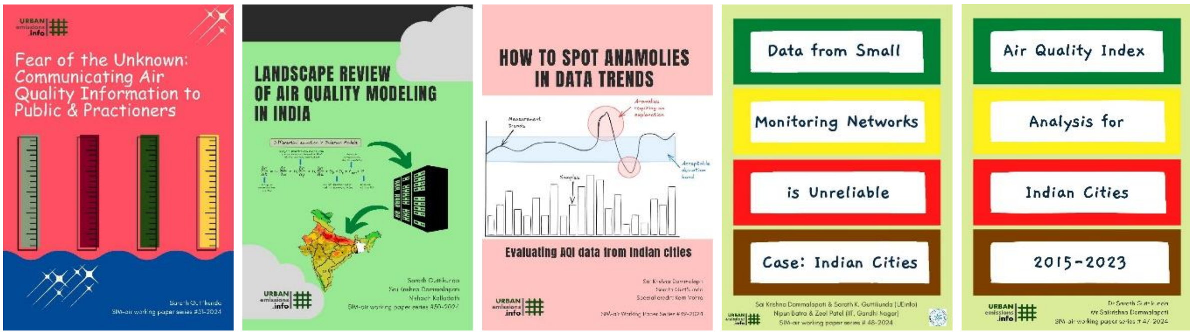
UrbanEmissions.info (UEinfo) was founded in 2007 with the vision to be a repository of information, research, and analysis related to air pollution and focusing on four key objectives:

- Sharing knowledge on air pollution
- Providing science-based air quality analysis
- Promoting advocacy and raising awareness on air quality management
- Building partnerships among local, national, and international airheads

About SIM-air Working Papers

The working papers describe case studies where we applied the SIM-air family of tools, document general notes on various databases available for emissions and pollution modeling and present our reviews on various topics related to air pollution analysis. All the materials are @ https://urbanemissions.info/publications

Last 5 working papers



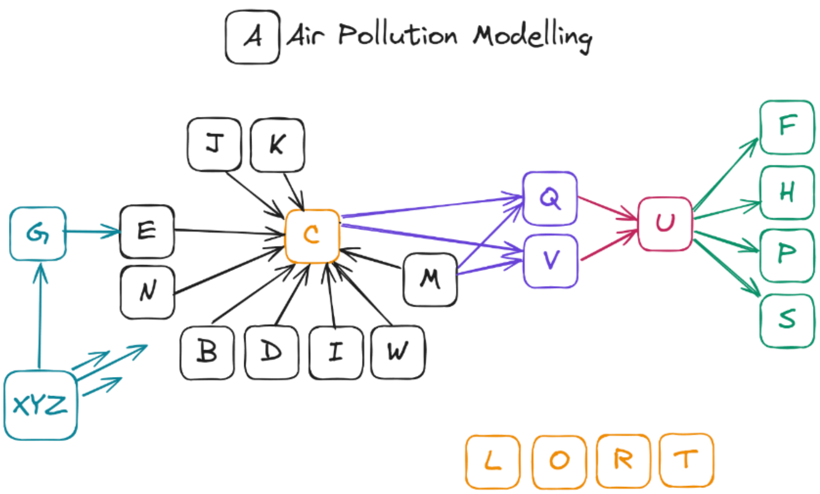
Suggested Citation

Guttikunda, S.K., and Dammalapati, S. (2024) "Beginners Handbook on Assembling Air Pollution Models". SIM-air working paper series #52-2024. Published by UrbanEmissions.Info, New Delhi, India

Disclaimer

Cover design is by Saachi.
Character sketches inside the document are by Saachi and Siddhant.
Doodles were designed using the open software https://excalidraw.com

Air Pollution Modeling Components



A Note to the Reader

This handbook is designed for beginners, as well as those looking to refresh their knowledge, providing a fundamental understanding of the key themes associated with air pollution modeling.

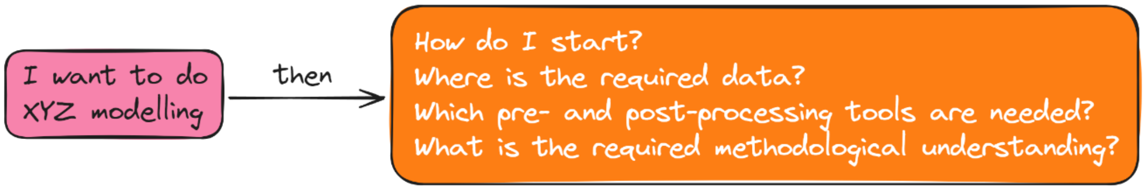
For a beginner starting to learn a specialized skill like air pollution modeling, understanding the **operational basics** is essential for building a strong foundation and becoming comfortable with the procedures.

There's no magic involved—just tools, thumb rules, and repeat applications.

The new and emerging air pollution models, whether simple or advanced, are modular in nature. This means that, with adequate computational resources and a good compiler, you can download these models, compile them, obtain standard inputs, and run the model to generate results. However, the knowledge and skills a beginner needs go beyond simply operating the model.

Important **operational questions to consider** include:

1. How do I interpret the results?
2. How do I validate these results?
3. How can I improve the model inputs?
4. How can I adjust the model parameters to adapt to my specific inputs?
5. Is this the right model for my needs?



Start by familiarizing yourself with the core concepts, terminology, and principles relevant to the topic.

Take small, deliberate steps, allowing yourself time to absorb the underlying ideas behind common questions and to identify where the most essential data resources are located.

In the air pollution modeling community, patience and persistence are crucial as you work through the learning curve, steadily building both your knowledge and confidence.

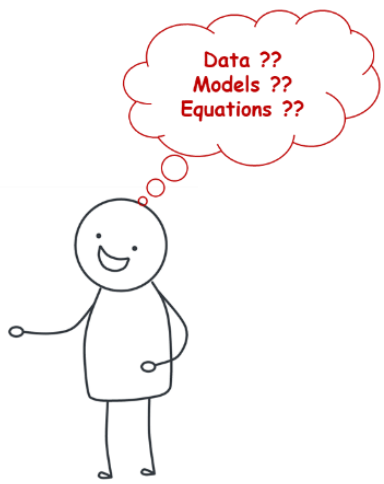
Four mantras for every beginner:

1. **The key is to start the process**. Simply begin, taking it step by step—working with one dataset at a time and focusing on one module at a time. Progress may seem slow at first, but each step builds on the previous one, gradually increasing your understanding and proficiency.
2. **Modeling is never truly complete**. Whether its emissions modeling using activity data or pollution modeling based on emission inventories, there is always something missing and always room for improvement. It's an iterative process, with continual refinements as new data becomes available and methodologies evolve.
3. "**All the models are wrong, but some are useful**" as mathematician George Box famously said. Since modeling is never complete or perfectly accurate, there is always some level of uncertainty—sometimes high, sometimes more manageable. It's important to recognize these limitations and use the modeling results appropriately, depending on the specific needs of your analysis.
4. **There is no perfect dataset**, so don't wait for one to begin your modeling exercise. As you start modeling, you'll discover datasets that fit your needs. If not, adapt and improvise with the data you already have—it's better to start and refine along the way than to wait for ideal conditions.

In this handbook, themes are organized alphabetically to ensure a smooth transition between topics, making it easier for readers to navigate and reference specific sections. This structure enables beginners to grasp each concept at their own pace without feeling overwhelmed.

In this handbook, we did not discuss everything related to air pollution modeling. We discussed only one key topic per alphabet, but enough

- to make you comfortable to start the process
- to make you excited to start compiling and running the models
- to make you understand the basic operations of modeling and
- to answer some frequently asked questions.



In this handbook, the most frequently referenced models are WRF for meteorological simulations and CAMx for chemical transport modeling. We discussed these more, only because we are comfortable using them and they are sufficient for our modeling needs. There are other models that follow a similar logical flow for pre- and post-data processing, offering additional options for exploration, some beyond air pollution modeling. We encourage you to delve into these alternatives to broaden your understanding and find the best fit for your specific needs.

The best resource for resolving doubts or errors related to meteorological and chemical transport models is their respective community forums. The state-of-the-art models (such as WRF, CAMx, and CMAQ) have been in use for decades, and if you encounter a compiling error or issue running a module, it's likely that someone else has faced the same problem. Search the forums with the right keywords, and you'll find a solution. If not, post your error along with the relevant details, and someone in the community will guide you forward.

WRF model forum - https://forum.mmm.ucar.edu
CAMx model FAQs - https://www.camx.com/about/faq

Abbreviations & Resource Links

All the weblinks were last accessed on October 3rd, 2024.

General abbreviations and resources

AOD	Aerosol optical depth
CAAQMS	Continuous ambient air quality monitoring system
CO	Carbon monoxide
ESRI-ArcGIS	Geospatial platform https://www.esri.com/en-us/arcgis/geospatial-platform/overview
GBD/IHME	Global burden of disease program by Institute of health metrics and evaluation https://www.healthdata.org/research-analysis/gbd
GBD-MAPS	Global burden of disease – mapping of air pollution sources program https://www.healtheffects.org/publication/gbd-air-pollution-india
GDAL	GIS information translator https://gdal.org/en/latest Tutorial package - https://spatialthoughts.com/
GEE	Google earth engine https://earthengine.google.com/
Google Earth	https://earth.google.com/web
HCHO	Formaldehyde
HEI/SoGA	Health effects institute/State of the global air https://www.stateofglobalair.org
NO_2	Nitrogen dioxide
NO_x	Nitrogen oxides
OSM	Open Street Maps https://www.openstreetmap.org Open data repository https://download.geofabrik.de/
PM_{10}	Particulate matter with aerodynamic diameter < 10 µm
$PM_{2.5}$	Particulate matter with aerodynamic diameter < 2.5 µm
pyLUR	Python language library for landuse regressions https://link.springer.com/article/10.1007/s11783-020-1221-5
Python	Common environment https://www.anaconda.com/download
QGIS	Open-source Geospatial platform https://qgis.org/download/ Tutorial package - https://spatialthoughts.com/
R-library	Documentation and learning https://www.datacamp.com/category/r
Reanal-WUSTL	Global reanalysis fields from Washington University (St Louis)'s Atmospheric Composition Analysis Group https://sites.wustl.edu/acag/datasets/surface-pm2-5/
Reanal-CAMS	Global reanalysis fields from Copernicus Atmopheric Monitoring System https://www.ecmwf.int/en/research/climate-reanalysis/cams-reanalysis
RGM	Reference/Regulatory grade monitoring
Satellite Data ESA Repository	https://developers.google.com/earth-engine/datasets/tags/tropomi
Satellite Fire Data FIRMS	Fire Information for Resource Management System https://firms.modaps.eosdis.nasa.gov/

Satellite Data NASA Repository	https://giovanni.gsfc.nasa.gov/giovanni/
SO_2	Sulfur Dioxide
SOA	Secondary organic aerosols
VOC	Volatile organic compounds
World View Earth Data	https://worldview.earthdata.nasa.gov/

EI - Emission Inventories

EI-CAMS	A composite of multiple inventories below by Copernicus Atmospheric Monitoring Service https://atmosphere.copernicus.eu/anthropogenic-and-natural-emissions
EI-CEDS	A Community Emissions Data System for Historical Emissions https://www.pnnl.gov/projects/ceds
EI-DICE-Africa	Diffuse and Inefficient Combustion Emissions in Africa https://www2.acom.ucar.edu/modeling/dice-africa
EI-ECLIPSE EI-GAINS	IIASA's Greenhouse Gas and Air Pollution INteractions and Synergies model https://iiasa.ac.at/models-tools-data/global-emission-fields-of-air-pollutants-and-ghgs
EI-EDGAR	Emissions Database for Global Atmospheric Research https://edgar.jrc.ec.europa.eu/
EI-HTAP	By Task Force on Hemispheric Transport of Air Pollution https://htap.org/
EI-FINN	Fire INventory from NCAR https://www2.acom.ucar.edu/modeling/finn-fire-inventory-ncar
EI-GFED	Global Fire Emissions Database https://www.globalfiredata.org/
EI-MEGAN	Model of Emissions of Gases and Aerosols from Nature https://bai.ess.uci.edu/megan
EI-MEIC	Developed under the framework of the Model Inter-Comparison Study for Asia (MICS-Asia) http://meicmodel.org.cn
EI-REAS	Regional Emission inventory in ASia (REAS) https://www.nies.go.jp/REAS/
EI-SMOG-India	Speciated Multipollutant Generator https://sites.google.com/view/smogindia
EI-Global Repository of emissions	Global Emissions Initiative (GEIA) - Emissions of atmospheric Compounds and Compilation of Ancillary Data (ECCAD) for anthropogenic and nature sources https://eccad.aeris-data.fr/

M- Models

M-ADMS	Also, Multi-Model Air Quality System (MAQS) https://www.cerc.co.uk/environmental-software/ADMS-model.html
M-AERMOD M-CALPUFF M-ISC3	Atmospheric dispersion modeling system https://www.weblakes.com/software/air-dispersion/
M-ATMoS	Atmospheric transport modeling system https://urbanemissions.info/tools/atmos/

M-CAM-chem	Community Atmosphere Model with Chemistry https://www2.acom.ucar.edu/gcm/cam-chem
M-CAMS	Copernicus atmospheric monitoring system by ECMWF https://atmosphere.copernicus.eu/
M-CAMx	Comprehensive air quality model with extensions https://www.camx.com/download/
M-CHIMERE	A multi-scale chemistry-transport model for atmospheric composition analysis and forecast https://www.lmd.polytechnique.fr/chimere/
M-CMAQ	Community modeling and analysis system https://www.cmascenter.org/cmaq/
M-DALM	DAnish Lagrangian Model www.au.dk/DALM
M-FARM	Flexible Air quality Regional Model https://www.aria-net.it/it/prodotti/farm/
M-FLEXPART	FLEXible PARTicle dispersion model https://www.flexpart.eu/
M-GEOS-CF	By NASA - GEOS Composition Forecasts https://gmao.gsfc.nasa.gov/weather_prediction/GEOS-CF/
M-GEOS-chem	Global chemical transport mode https://geoschem.github.io/index.html
M-HYSPLIT	Hybrid Single-Particle Lagrangian Integrated Trajectory https://www.ready.noaa.gov/HYSPLIT.php
M-MOZART	Model for OZone and Related chemical Tracers https://www2.acom.ucar.edu/gcm/mozart
M-PAQMS M-TAPM	By CSIRO - Photochemical air quality modelling system and The air pollution model https://www.cmar.csiro.au/airquality/ctm.html
M-SILAM	System for Integrated modeLling of Atmospheric composition https://silam.fmi.fi/
M-UBM	Urban background model www.au.dk/UBM
M-WACCM	Whole Atmosphere Community Climate Model https://www2.acom.ucar.edu/gcm/waccm
M-WRF M-WRF-chem	Weather research forecasting model https://github.com/wrf-model/WRF

A

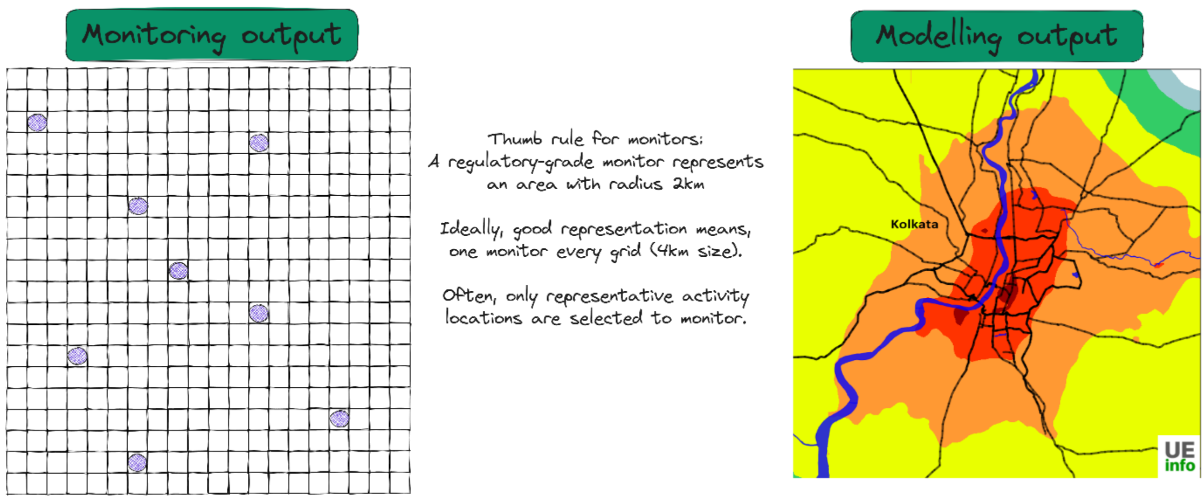
Air Pollution Modeling

Other terms for this exercise include air quality modeling, dispersion modeling, chemical transport modeling, and pollution modeling, among others. While the term "dispersion modeling" is the most used term, this exercise encompasses much more than simply mathematically dispersing emissions.

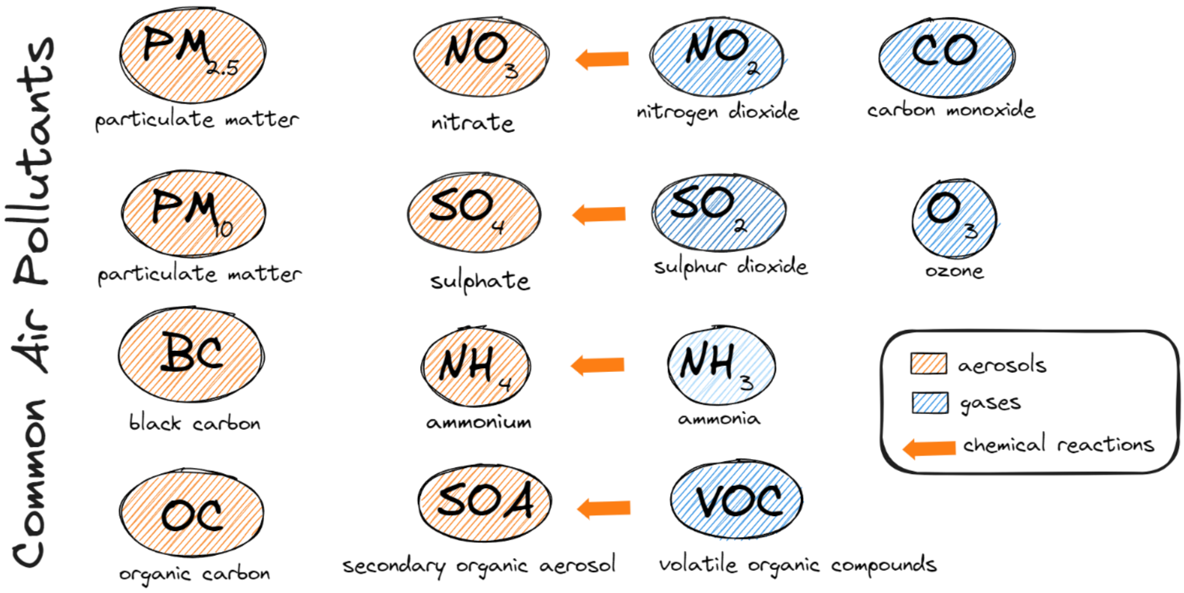
FAQ: Why model air pollution when it can be measured with lesser uncertainty?

For regulatory and policy planning purposes, monitoring data is preferred over modeling data to understand the spatial and temporal trends of air quality. This preference stems from the belief that measurements accurately reflect local activities and are real, reliable, and reproducible using physical instruments. However, monitoring can be technically challenging, financially burdensome, and personnel-intensive, often resulting in a limited number of operational monitors. Consequently, this creates a knowledge gap in understanding the spatial trends of air quality within a city or across a region.

This gap can be minimized through modeling studies, which can complement monitoring efforts by validating the inputs to and outputs from a model. While modeling plays a crucial role, monitoring remains a key component of the air pollution modeling process, as it provides the foundation for validating results and enhancing confidence in the final modeling outputs (Read M and V).



All four main components of air pollution modeling—emissions inventory, meteorological data, chemical transport model inputs, and monitoring data—are data-intensive and require substantial computational resources. The specific requirements for these resources can vary significantly depending on the granularity of the designated airshed, which is influenced by factors such as the number of grids and the size of each grid (Read XYZ). For example, a finer grid resolution may provide more detailed spatial information but will also demand more data and greater computational power. Conversely, coarser grids may simplify the modeling process but could lead to a loss of important local variations. Understanding these dynamics is essential for effectively managing resources and ensuring the accuracy of modeling results.

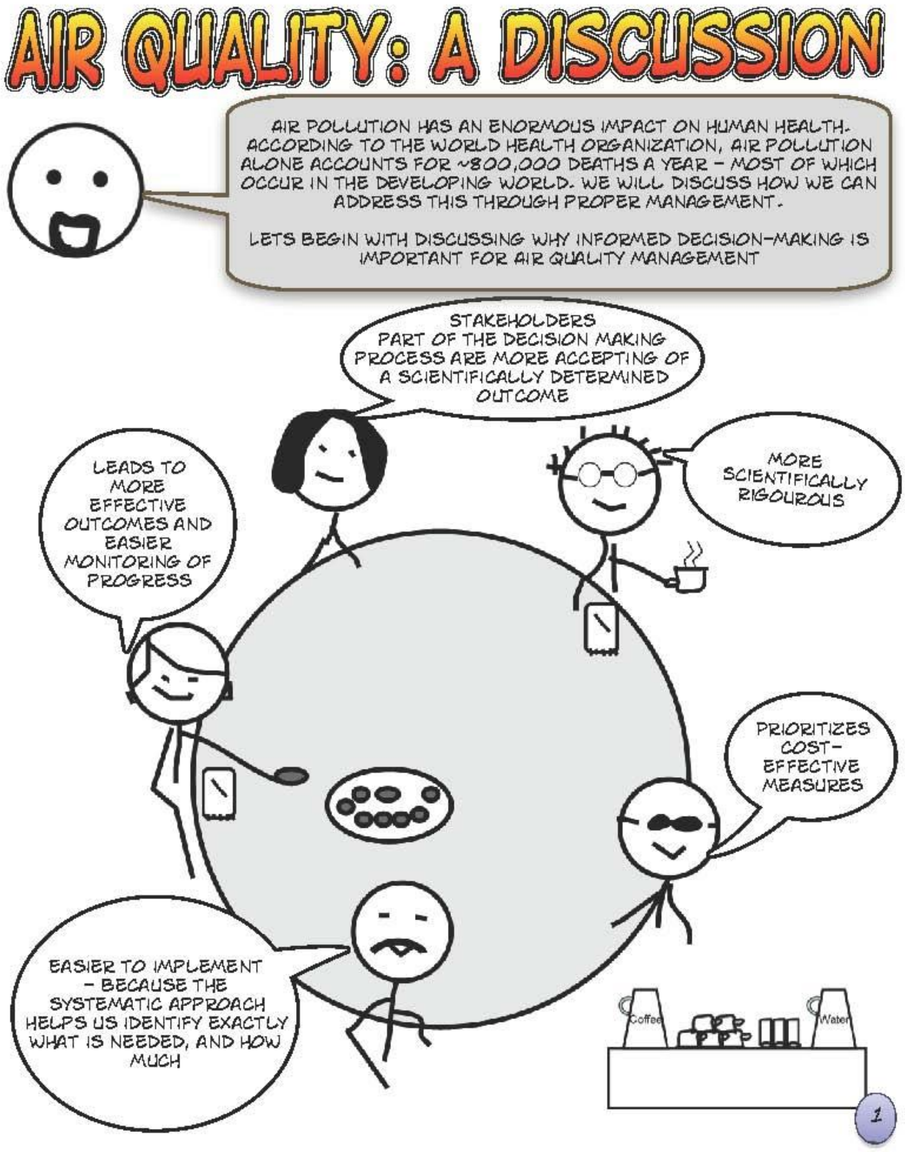


All the chemical transport models include established chemical mechanisms, which cover a range of pollutants from key regulatory pollutants like $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO, and ozone to a range of supporting chemical species likes VOCs, radicals, and other intermediates, which play a key role in apportioning of secondary contributions to PM and formation/destruction of pollutants like ozone, and formaldehyde (Read JK). When conducted correctly, the results from these modeling exercise offer a wide range of applications that surpass the insights gained from data collected through monitoring campaigns alone. For example:

1. **Long-term modeling (past and future)** can support compliance assessments, exposure analyses, and scenario evaluations by providing emissions and concentration data at the grid level. This enables policymakers to make informed decisions about regulations and public health interventions.
2. **Modeling studies can aggregate data** from airshed, urban, or regional areas to develop communication materials such as air quality indices, pollution alerts, and health advisories. This information can help raise public awareness and guide individuals in making informed choices to protect their health.
3. **Modeling results can evaluate source contributions** at varying spatial and temporal scales, allowing for a more nuanced understanding of pollution sources. This capability is crucial for identifying key contributors to air quality issues and targeting mitigation efforts effectively across the airshed or at select hotspots identified from the model outputs.

4. **Models operating in forecast mode** can provide short-term pollution alerts and health warnings, generating timely information for the public and authorities. These forecasts can cover periods ranging from one day to ten days, enabling proactive measures to safeguard health during episodes of poor air quality (proactive = taking actions based on the forecasts to avoid the anticipated episodes).

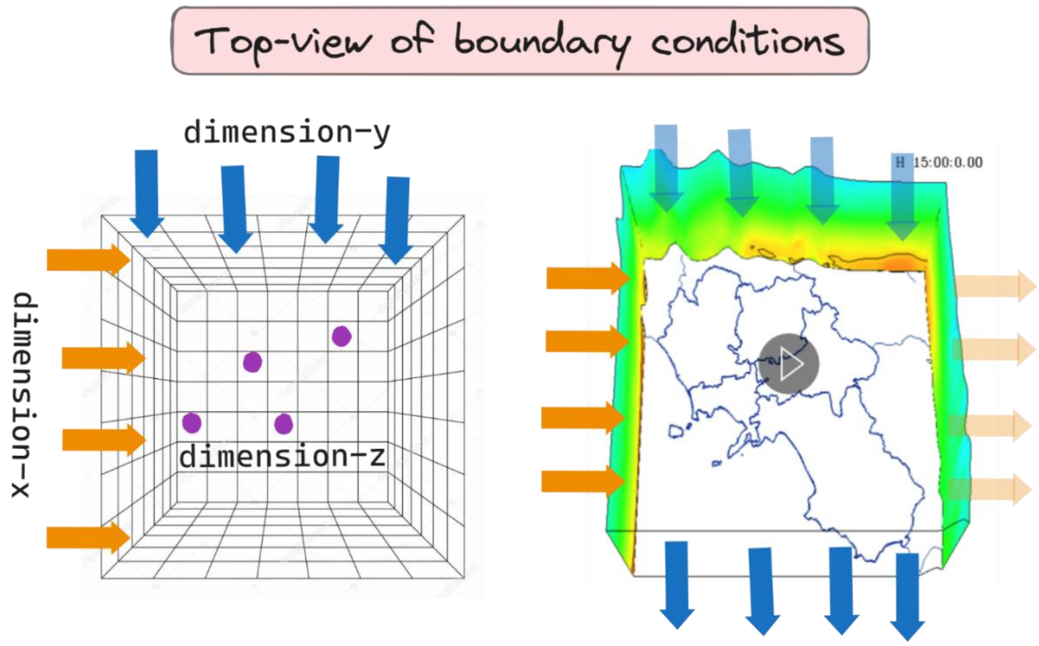
By leveraging these diverse applications, modeling exercises can significantly enhance our understanding of air pollution dynamics and inform effective management strategies.



B

Boundary and Initial Conditions

Every air pollution manager and practitioner must understand where pollution is coming from—specifically, which sectors and regions are contributing to the problem. This process is known as source apportionment (Read S), where pollution is tagged by source type and geographic area or zone. It is crucial that this analysis includes not only contributions from within the designated airshed but also from external sources, which are referred to as **boundary or long-range transport contributions**. Understanding these external influences is vital to creating a comprehensive air pollution management strategy and avoiding blame being misdirected when outside regions contribute significantly to the air quality issues in an airshed.

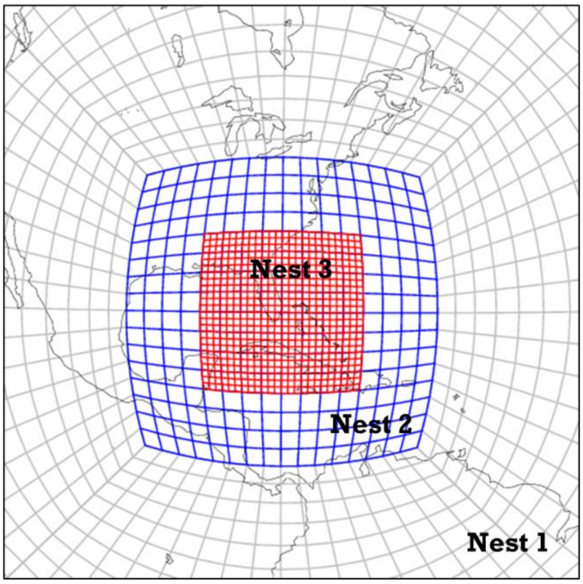


In the chemical transport modeling step (Read C), these contributions are accounted for as 3-dimensional boundary conditions (horizontal x-y and vertical z). If there are any emission sources outside the airshed, their contribution is reflected at the boundary and rest of the work is done by the differential equations of the chemical transport models.

Assigning the boundary conditions, is like "nesting" of grids (Read XYZ). The urban airshed is a nest of the regional simulation. For example, for an urban airshed covering an area of 100 km * 100 km, boundary conditions can be extracted from a regional simulation which covers a much wider area, say 4000 km * 4000 km. The data from the regional (or global) simulations are spatially and temporally interpreted to the grid resolution of the urban airshed.

Most of the pollutants have known regional contributions in the horizontal (x-y) dimensions and vertically spread under the tropopause (Read I). The top-most layer (z-boundary) of the domain is often reserved for ozone concentrations, to account for stratospheric intrusions. Other air pollutants also have non-zero numbers and have limited presence in the z-boundary (at the altitudes greater than 10km).

Before starting a simulation with any established chemical transport model, initial conditions are necessary to jumpstart the calculations. This step is essential to ensure that the chemical mechanisms and mathematical solvers operate without errors, particularly in the context of short-term forecasting systems (see F). For long-term simulations, such as those spanning one year, the initial 10 to 15 days can be utilized to spin up the model, allowing it to accurately reflect the spatial profiles in the emission inventory. This initialization can begin with a non-zero setup based on local measurements or data extracted from a regional or global model.



FAQ: What is a common data-source to build boundary and initial conditions?

All chemical transport models come with established pre-processors designed to build the boundary and initial conditions required for simulations, using existing global models, along with instructions to also put together this data from local simulations.

In urban case studies, a regional simulation (typically characterized by more grids and coarser grid resolution) is often necessary to construct this component effectively. If regional simulations are not available, global model results can also be employed.

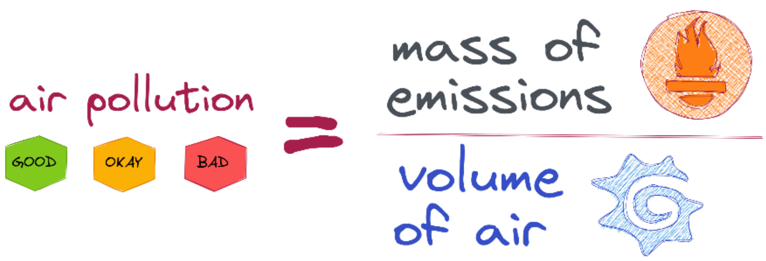
For regional simulations, global model outputs are frequently utilized. The most widely used global databases for constructing boundary, top, and initial conditions include MOZART/CAM-chem, GEOS-Chem, and CAMS.

The chemical kinetics/mechanisms (Read JK) are often different between these models and the model selected for urban/regional simulations. In such cases, matching of the pollutant names and categories is necessary. Most of the pre-processors (for example, of CAMx and CMAQ) already account for this transition.

C

Chemical Transport Models

Chemical transport models are mathematical representations, implemented as computer code, that simulate the physics and chemistry of the atmosphere. They analyze how pollutants disperse, the chemical interactions between various pollutants, and the processes by which these pollutants are removed from the atmosphere. By interpreting the transport and transformation of pollutants, these models provide crucial insights into air quality dynamics and help predict the impact of different pollution sources on local and regional air quality (Read A).



These models vary widely in complexity, allowing users to choose one that best fits their specific needs. Some models focus on fundamental processes, while others incorporate advanced features for treating advection, chemical mechanisms, and pollutant removal. For instance, simpler models may use basic algorithms to simulate pollutant transport, whereas more complex models employ sophisticated numerical techniques and detailed chemical mechanisms to capture the intricacies of atmospheric reactions and interactions.

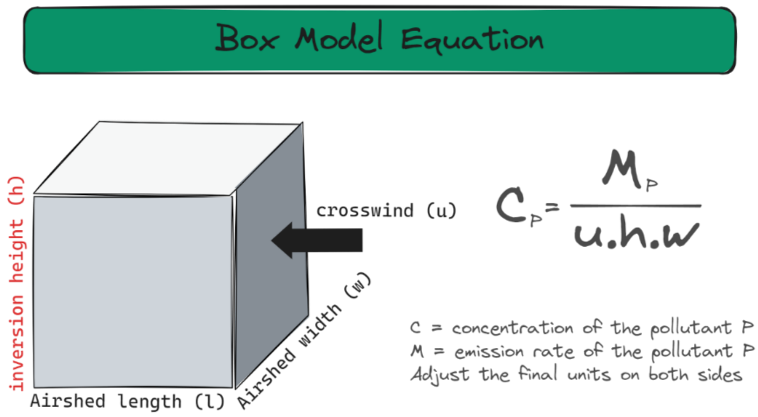
The choice of which model to use depends on factors such as the scale of the study, the specific pollutants of interest, and the required level of detail. There are numerous publications available that users can reference and replicate for each of these methods, offering valuable guidance on selecting the most appropriate model for their specific needs and available resources.

Qualitative Comparison of Chemical Transport Models

	Complexity	Ease of Operations	Data Requirement	Computational Requirement	Pollutant Chemistry	Dispersion (Advection)	Level of Details
Box Models	Easy	Easy	Low	Low	Maybe	None	Low
Plume Models	Easy	Medium	Medium	Low	Maybe	Minimum	Low
3D Eulerian Models	High	Complex	High	High	Full	Full	High

Personnel expertise requirement is a must for all the models

Box models are simplified mathematical representations which treats the atmosphere as a box, neglecting advection, focusing instead on the chemical processes and removal mechanisms occurring within the defined volume of air. These models typically assume uniform conditions within the box, allowing for easy calculations of pollutant concentrations and chemical interactions.



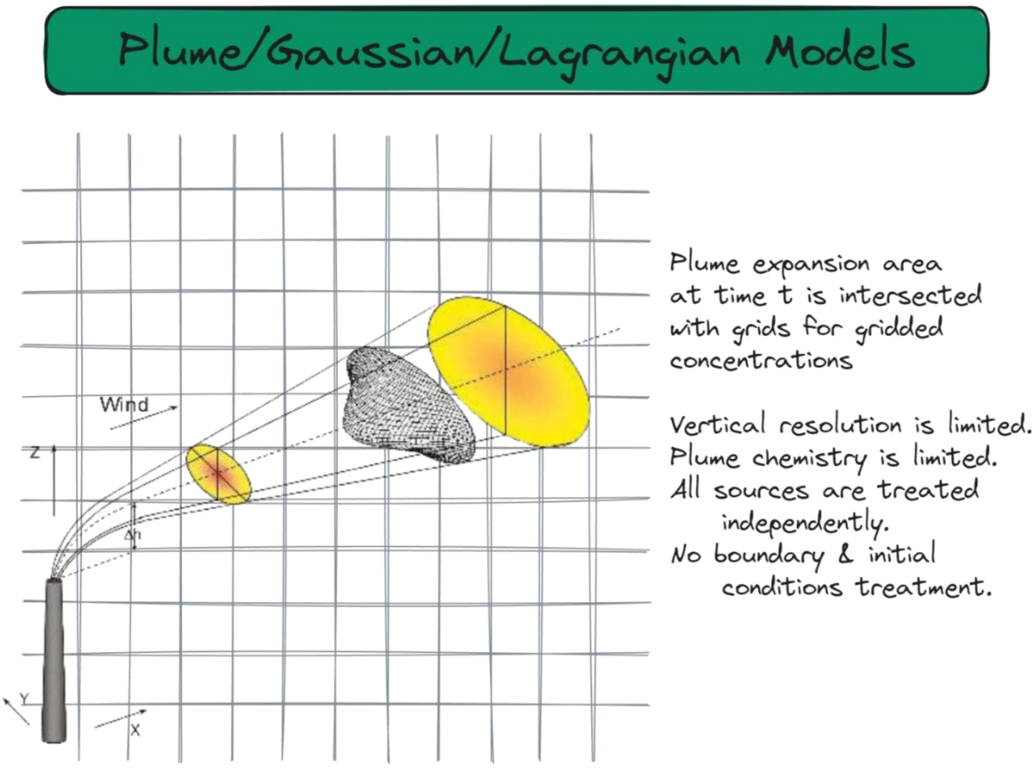
Some common uses of box models are:

1. **Quick Assessments of Pollution Loads**: Box models are particularly useful for conducting rapid and broad calculations of pollution loads within a designated area, as estimates of accumulation and dispersion of pollutants. This makes them ideal for preliminary assessments and to identify key variables and processes for detailed simulations in advanced models.
2. **Impact Studies of Chemical Reactions**: These models are commonly used to study the effects of both existing and new chemical reactions on overall pollutant concentrations. By simulating various scenarios, researchers can analyze how changes in chemical processes affect air quality, helping to identify key reactions that contribute to pollution levels.
3. **Educational Purposes**: Box models serve as valuable tools in educational settings, allowing students and researchers to grasp fundamental concepts of atmospheric chemistry and pollution dynamics without the complexities of more advanced modeling approaches. An example box model setup is included @ https://urbanemissions.info/tools including variations to also inverse calculate concentrations into emissions.
4. **Scenario Analysis**: Box models can facilitate scenario analysis by allowing users to manipulate parameters such as emission rates, reaction rates, and removal processes. This flexibility enables researchers to explore the potential impacts of different pollution control strategies and regulatory measures.

The simplicity of box models does come with limitations, caution is warranted. However, they still provide a practical and efficient means of gaining insights into air quality dynamics and understanding the effects of various chemical processes on pollutant concentrations.

Plume, Gaussian, and Lagrangian (PGL) models advance beyond treating pollution within a static box by modeling pollution over an area using simplified advection functions. These models represent emissions as plumes or puffs, tracking their dispersion through the atmosphere. These models are computationally efficient, offering quick calculations with relatively low resource demands. They simulate the movement of pollutants plume by plume, focusing on individual trajectories, but lacking interactions between plumes. This limits the ability to fully utilize chemical mechanisms, though some parameterizations can account for secondary aerosols (such as sulfate, nitrate, and organic aerosols) and secondary pollutants like ozone.

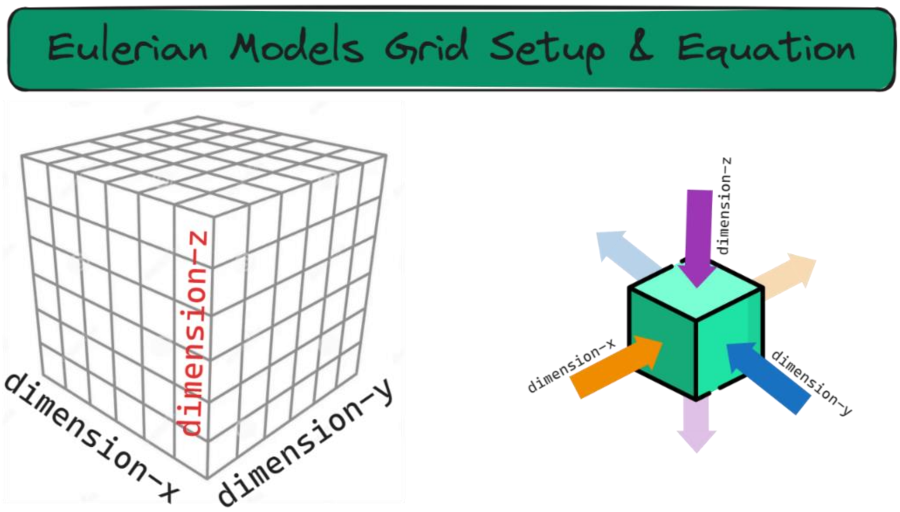
The simplicity in how advection is handled means they are not ideal for studying pollution in environments where topography plays a significant role in pollutant dispersion. Despite these constraints, PGL models are commonly used for rapid pollution assessments, trajectory analysis, and evaluating localized pollutant dispersion from point sources such as industrial facilities or power plants. Their speed and efficiency make them particularly useful for short-term studies and regulatory applications where quick assessments are needed (Read T).



Most used models of this kind are: AERMOD (formerly ISC3, ISCST3), CALPUFF, ADMS, ATMoS/UrBAT, HYSPLIT, FLEXPART, and UBM.

AERMOD, developed by the U.S. Environmental Protection Agency (EPA), is the most widely used dispersion model for regulatory purposes, particularly in the context of compliance with environmental standards for stationary sources such as industrial facilities, power plants, and refineries.

Eulerian models are distinguished by their grid-based approach, where each grid cell is treated independently, offering a highly detailed and state-of-the-art method for simulating atmospheric physics and chemistry. These advanced models can simulate a wide range of atmospheric processes, including turbulent mixing, chemical reactions, emissions, and the dispersion of pollutants. By dividing the modeling domain into a network of interconnected grid cells, Eulerian models allow for precise tracking of pollutants over space and time, providing the highest granularity in capturing spatial, temporal, and chemical variations across different scales (urban, regional, and global).



change in concentrations due to winds in x-y-z dimensions (momentum flux) including boundary contributions

change in concentrations due to chemistry

$$\frac{\delta c}{\delta t} = U_x \frac{\delta c}{\delta x} + U_y \frac{\delta c}{\delta y} + U_z \frac{\delta c}{\delta z} + D_d + D_w + C_{chem} + E$$

change in concentrations over time

dry (d) deposition rates wet (w) scavenging rates

Emissions at time (t)

Each grid cell represents a distinct point in space where atmospheric variables such as temperature, wind speed, chemical concentrations, and humidity are calculated to feed the differential equation shown in the figure. This allows Eulerian models to simulate interactions between adjacent grid cells – pollutants going in and out (after accounting for emission perturbation, chemical transformation, and depositions within each of the grid). The grids cells at the edge of the domain form the boundaries, where the boundary and top conditions are treated as adjacent grid values for furthering the calculations (Read B).

The accuracy and sophistication of Eulerian models come at a cost—they require significant computational resources. The need to perform calculations across potentially thousands or even millions of grid cells make these models highly resource intensive. The finer the grid resolution, the more detailed the simulation, and exponentially more computationally demanding. As a result, Eulerian models are typically run only on high-performance computing systems.

Due to their ability to capture complex interactions and provide detailed insights, Eulerian models are the most desired and widely adopted by the scientific research community. Their capability to integrate various physical and chemical processes makes them the gold standard for academic studies and policy-relevant research.

A variety of applications which these models are already supporting (a) urban, regional, and global modeling (b) source apportionment (c) health impact analysis (d) scenario analysis (e) climate and aerosol science (f) atmospheric chemistry (g) air quality forecasting (h) policy support and (i) emergency response.

Most used models of this kind are:

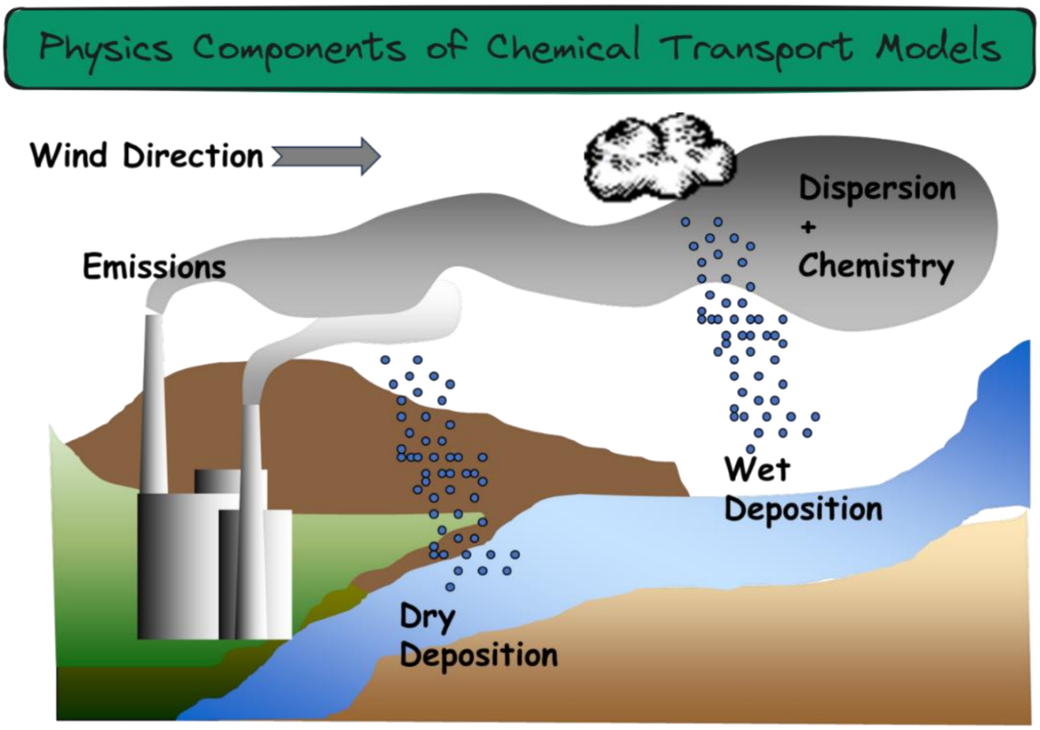
1. CMAQ, CAMx, DEHM (urban and regional applications)
2. WRF-chem, CHIMERE, SILAM (urban, regional, and global applications)
3. GEOS-chem, MOZART/CAM-chem, CAMS (global applications)

D

Deposition Rates

Once emissions are released into the atmosphere, they undergo several key processes: they either get dispersed over long distances following the wind direction, chemically transformed into other pollutants, or removed from the atmosphere through various deposition mechanisms. These removal processes, including **dry deposition** (where pollutants settle onto surfaces like vegetation, soil, or buildings) and **wet scavenging** (where pollutants are washed out of the atmosphere by precipitation), play a crucial role in regulating pollutant concentrations in the air.

The effectiveness of these removal mechanisms must be accurately represented and activated in all chemical transport models—including simpler box models—so that the models do not overestimate the pollutant concentrations. Without proper accounting for dry deposition and wet scavenging, the models can predict unrealistically concentrations, leading to incorrect assessments of air quality and potential health impacts.



There are well-established modules for both dry deposition and wet scavenging, and a wealth of studies that discuss and quantify deposition rates. These modules are integral to chemical transport models, ensuring accurate representation of pollutant removal processes. The deposition rates vary based on factors such as surface type, atmospheric conditions, and pollutant properties, and have been extensively studied to refine the accuracy of air quality modeling.

Dry deposition is calculated in the surface layer only, which is the closest to the ground. Only the pollutant concentration in this layer is subjected to removal calculations, while concentrations in all layers continue to change due to processes like dispersion and chemical transformation. Some simpler models apply a constant deposition rate, regardless of land use or meteorological conditions. However, most state-of-the-art models calculate dry deposition rates that are both time- and grid-dependent, accounting for variables such as land use, surface characteristics, and meteorological conditions, ensuring more accurate simulations by pollutant types.

Deposition Calculations - First Order

$$C = C_o * e^{-k*dt}$$

C = New concentration (at time t)
C = old concentration (at time t-1)
k° = decay rate (1/sec)
dt = calculation timestep in seconds

$$k\ (1/sec) = \frac{\text{deposition rate (m/sec)}}{\text{surface layer height (m)}}$$

Some key considerations are as follows:

1. **Pollutant type and size**: Larger particles tend to have higher deposition rates, as their greater mass causes them to settle more quickly compared to smaller particles. Gaseous pollutants also vary in their deposition rates based on their reactivity and affinity for surfaces (like trees, building, and water).
2. **Land use**: Deposition rates are typically higher over land due to rougher surfaces, such as vegetation and terrain, which enhance the capture of particles. In contrast, deposition rates are lower over water bodies, where the smooth surface and lack of vegetation limit particle trapping. Among land types, vegetated areas have higher deposition rates compared to built-up areas. Vegetation's surface roughness, along with its ability to trap particles through the texture of leaves and stems, significantly enhances dry deposition. Additionally, leaves "breathing" exercise further increasing the efficiency of deposition. In contrast, built-up areas are less rough than textured leaves reducing the overall deposition rate.
3. **Meteorological conditions**: Moderate wind speeds increase deposition by mixing the air mass between vertical layers and bringing more pollutants closer to the surface. However, under extreme wind conditions, much of the air mass is blown away, which can reduce deposition rates. Atmospheric stability also influences deposition: under stable conditions (calm winds), mixing is limited, leading to lower deposition rates. Conversely, under turbulent conditions, increased mixing brings more pollutants closer to the surface, increasing deposition rates.

Wet scavenging is calculated for all the atmospheric layers beneath the cloud base, a parameter provided by 3D meteorological models (see W). Both gaseous and aerosol pollutants are subjected to wet scavenging during precipitation events. Typically, more pollutant mass is removed through this step compared to dry deposition. This is clearly seen in the blue skies that often follow rain events and a sudden improvement in the air quality and air quality index values.

The wet scavenging rate depends on several factors, with the precipitation rate being a key determinant. There is no specific threshold below which scavenging does not occur; even light precipitation can lead to pollutant removal. The amount of wet deposition is directly proportional to both the intensity and duration of the rain. Longer periods of rain, such as 2-3 hours, even in the form of light drizzles, can effectively scavenge a significant portion of the pollutant mass in the atmosphere. Heavier rainfall tends to remove pollutants more rapidly due to the larger number and size of raindrops.

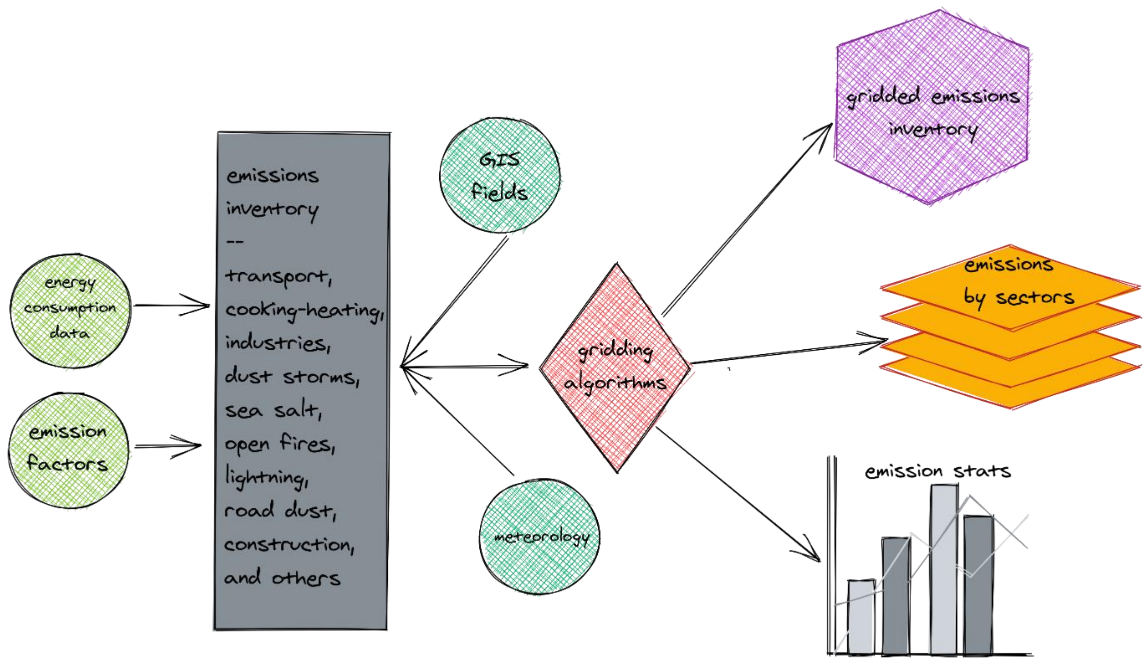
E

Emission Inventories

Emission inventories are the foundation of air pollution modeling, serving as a critical component in understanding air quality dynamics.

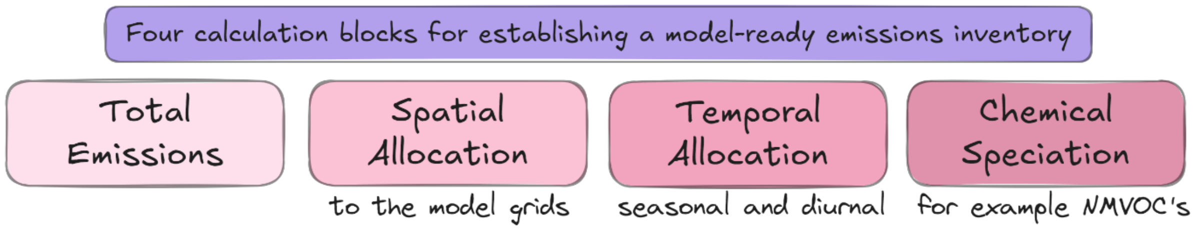
These inventories provide detailed information on the magnitude and sources of pollutants released into the atmosphere, from anthropogenic (vehicles, industries, cooking, heating, waste burning), agricultural, and natural (dust, lightning, and biogenic) sources. By offering this essential baseline data, emission inventories allow researchers and policymakers to assess current air quality, track changes and trends over time, and evaluate the effectiveness of air pollution control measures.

They also play a key role in identifying priority areas for regulatory action and ensuring that models reflect real-world conditions. Without comprehensive emission inventories, it would be nearly impossible to develop effective strategies for reducing pollution and protecting public health.

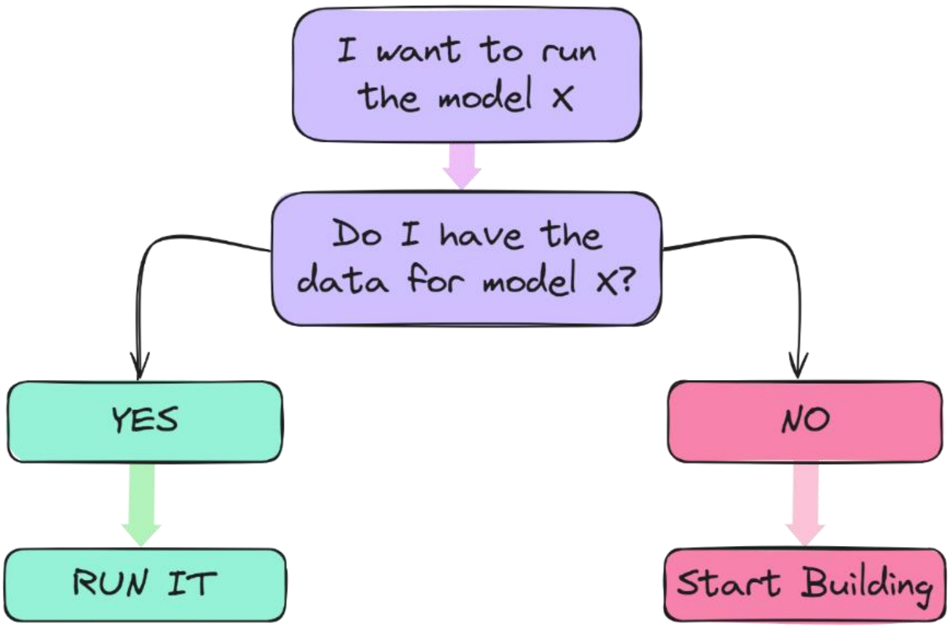


Main pollutants covered in all the inventories are particulate matter (PM as 10µm and 2.5µm fractions); black carbon (BC); organic carbon (OC); sulfur dioxide (SO_2); nitrogen oxides (NO, NO_2, HNO_3, N_2O_5); carbon monoxide (CO); and non-methane volatile organic compounds (NMVOCs). NMVOCs are further speciated into many species, needed to support various chemical mechanisms.

Creating an emissions inventory is a complex and detailed process that requires the collection and analysis of various types of data. Key inputs include emissions factors, which represent the average emission rate of a pollutant for a given source, and activity levels, which describe how often and to what extent the emission-producing activities occur. The process also requires precise information on the spatial and temporal distribution of emissions—where the emissions are released geographically and when they occur—so that the inventory can reflect real-world conditions.



Developing an emissions inventory demands patience, as it often involves overcoming challenges such as incomplete data and uncertainties in emissions estimates. Rigorous validation and quality assurance processes (see Q) are essential to ensure the reliability and accuracy of the inventory. As the saying goes, "garbage in, garbage out"—if the inventories are weak and fail to reflect ground realities, the results of the modeling exercise will be similarly uncertain. However, it's important not to wait for the perfect dataset. Start the process with what is available and refine the inventory as more data becomes accessible over time.



Over time, emissions inventories can be improved by integrating with real-time monitoring data, conducting field studies for better emissions factors, and continuous surveys with localized, sector-specific activity data. Common challenges include lack of reliable source-specific data, uncertainty in self-reported industry figures, and discrepancies between modeled estimates and actual measurements from monitoring stations, which can be addressed over iterations of emission inventories and what-if scenarios.

Establishing an operational emissions inventory is an iterative process that begins with customizing the fundamental equation to fit the available data on activity and emission factors. As more resources become available, the data collection process can become more detailed and complex, allowing for greater accuracy and specificity.

Fundamental equation

Emissions = Activity * Emission Factor

Emissions = fuel consumption (PJ) * emission factor (kg/PJ)
Emissions = fuel consumption (tons) * emissions factor (kg/tons)
Emissions = #vehicles * vehicle usage (km) * emission factor (gm/km)
Emissions = waste burnt (kg) * emission factor (gm/kg)
Emissions = diesel burnt (lit) * emission factor (gm/lit)

A series of MS-excel based calculators to estimate emissions and required inputs are available @ https://urbanemissions.info/tools for transport and non-transport sources. For example:

1. VAPIS 2.1 (Vehicular Air Pollution Information System) - A vehicular exhaust emissions calculator to estimate and compare total emissions by vehicle-age and run scenarios (requires user to activate macros).
2. Demonstration of 4 approaches to estimate total fleet average vehicle exhaust emissions using information on (a) VKT (b) fuel sales (c) modal shares and (d) meteorology. The 4th approach is an extension of box-model concept.
3. Demonstration of fundamental equation in building fleet average emissions, with and without age-mix information.
4. Demonstration of a method to convert fleet average speeds into vehicle km travelled (VKT).
5. Demonstration of a method to calculate how many additional buses are required to support odd-even or an equivalent scheme (with and without fuel mix exemptions).
6. Demonstration of a method to calculate total fuel wasted from idling in the city.
7. Demonstration of a method to calculate benefits of shifting % of 2W and 4W trips to buses and non-motorized transport.
8. Demonstration of a method to estimate vehicle exhaust emission factors using emission standards and deterioration rates (version 1).
9. Example set of survival rates based on vehicle age for 10 vehicle categories (to convert yearly registered vehicle numbers into in-use vehicle numbers).
10. Demonstration of a method to grid the total vehicle exhaust emissions using multiple grid-level proxies as weights (requires user to activate macros).
11. Demonstration of a method to estimate emissions from a coal thermal power plant.
12. Demonstration of a method to estimate open waste burning emissions (and run some scenarios) based on waste generation rate, waste collection efficiency, and waste burnt rate.

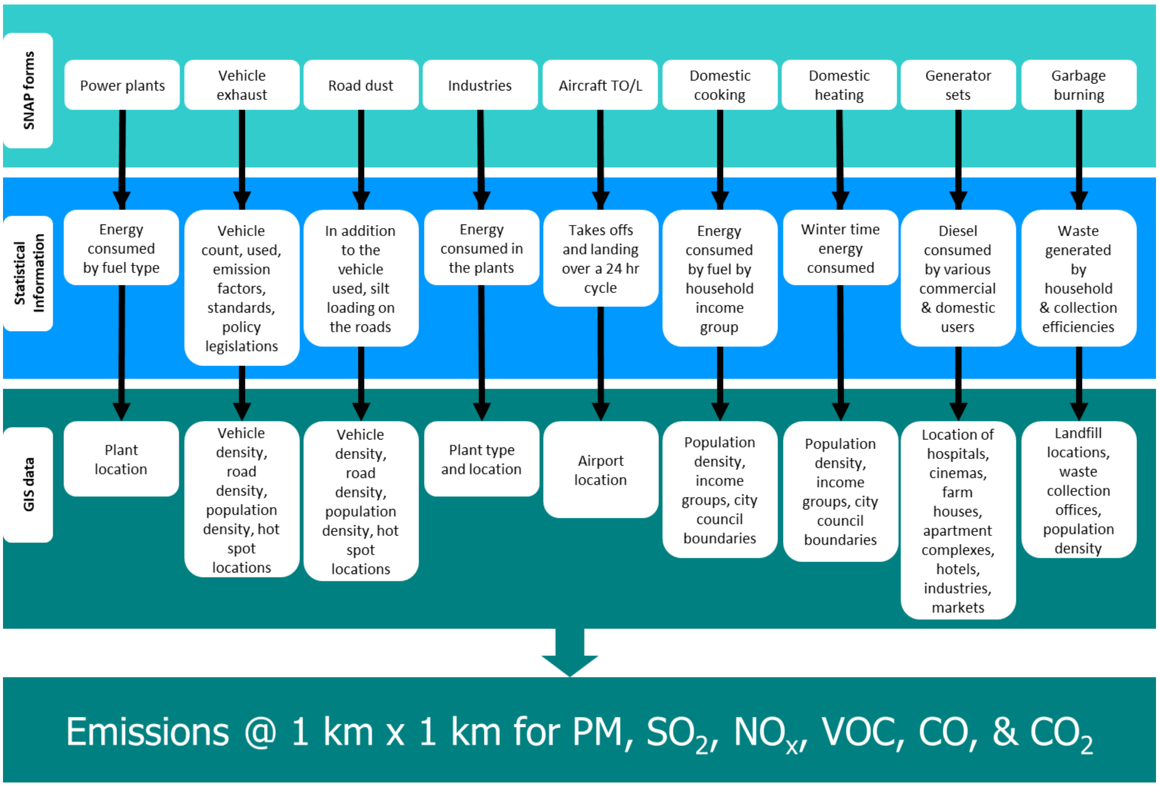
An **emission factor** is a critical component of emissions inventory exercise, representing the amount of a pollutant released into the atmosphere per unit of activity, such as the amount of fuel burned or the level of industrial activity. Essentially, it provides a quantitative measure of pollution generated by a specific source. For instance, an emission factor for a car would measure how much CO_2, PM, or NO_x is emitted per kilometer driven. Still, it is not a simple number as it can vary significantly depending on several variables:

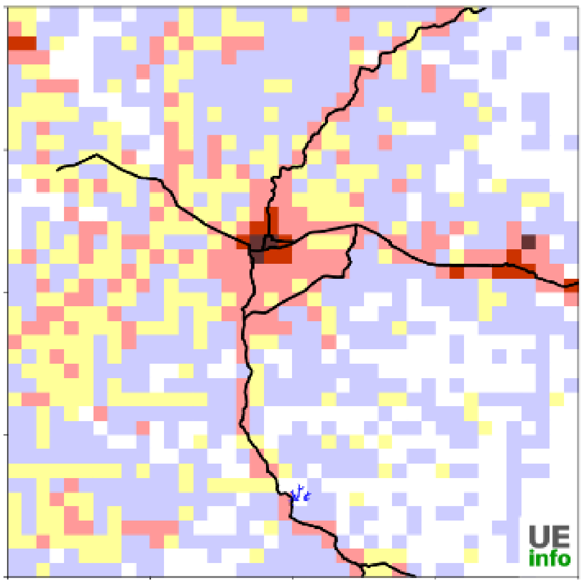
1. **Technology**: The type of combustion technology in use, such as older engines versus newer and more efficient industrial boilers. Newer technologies often have lower emission factors due to improvements in design, efficiency, and pollution control devices like catalytic converters in vehicles.
2. **Fuel Type**: The type and the amount of fuel being burnt is critical information. For example, burning coal will result in higher SO_2 emissions compared to burning natural gas. Similarly, diesel vehicle engines tend to emit more PM than petrol or gas engines.
3. **Environmental Conditions**: Emission factors can also vary with external conditions, such as temperature and humidity. For instance, cold temperatures can lead to incomplete combustion in vehicles, resulting in higher emissions of CO and VOCs, while high temperatures can increase NO_x emissions with potential to alter ozone formation rates.
4. **Behavioral Patterns**: Human behaviors, such as driving habits and road conditions, play a significant role in altering emission factors, especially for vehicles. For instance, aggressive driving, stop-and-go traffic, and poor road conditions can increase fuel consumption and emissions, while smooth, efficient driving on well-maintained roads typically results in lower emissions.

Due to the wide range of influencing factors, it is not easy to generalize emission factors across different technologies, fuels, and conditions. The most accurate approach to determining emission factors is through direct measurement—conducting emission tests at the source. This involves sampling emissions from a wide range of sources in a city or region to obtain representative numbers. However, conducting such tests requires significant resources, technical expertise, and time.

In the absence of local data, emission factors can be borrowed from established libraries or databases that provide standardized factors based on studies from similar regions or industries. These libraries, like the one available at (https://urbanemissions.info/tools) provide pre-determined emission factors for various sources, which can be used to estimate total emissions, when localized data is unavailable. While these representative factors are not as precise as locally measured data, they offer a reliable starting point and can be refined over time as more specific and local data becomes available.

For the **spatial allocation of emissions to grids**, most procedures rely on Geographic Information System (GIS) tools and require additional information about the airshed. This process can range from simple tasks, such as identifying the exact locations of large point sources like power plants or factories, to more complex tasks, such as assigning appropriate weights to grid vehicle exhaust emissions from different modes of transportation (for example, heavy trucks on highways or motorcycles in residential and commercial areas).





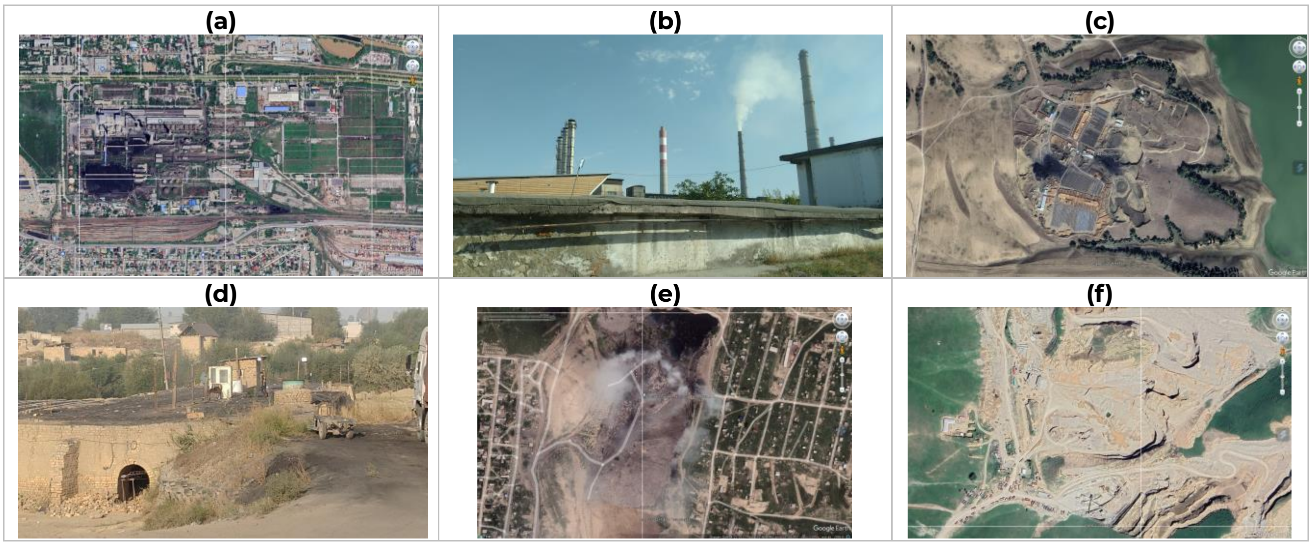
Although spatial allocation is a data-intensive process, GIS data has become more readily available in recent years. Detailed geographic datasets, including road networks, population density, land use, and industrial locations, can now be easily accessed from government databases and open-source platforms. This wealth of spatial data enables more representative emission allocation by reflecting real-world activity patterns.

An example methodology as a MS-Excel file to allocate total emissions to grids using a variety of GIS layers is available @ https://urbanemissions.info/tools

Satellite feeds are also increasingly used to enhance the spatial allocation of emissions. Satellite-derived data, such as land cover, vegetation, and even NO_2 or CO_2 concentrations (for example), can be incorporated into GIS models to nudge and refine spatial patterns. This is particularly useful for large areas or regions where ground-based monitoring or local information on the sources is sparse.

Where specific information on industries, points of interest, or activity hubs is lacking, satellite imagery can be an invaluable tool for manually scanning through grids to identify potential emission sources. High-resolution satellite images can allow to visually inspect areas for industrial facilities, construction sites, transportation hubs, stone quarries, brick kilns, and other unknown activities that may contribute to emissions but are not well-documented in existing inventories.

This method is suitable for urban airsheds where the total area to scan is of reasonable size, as shown in the figure below for the city of Bishkek, Kyrgyzstan.



Sample screenshots from Google Earth satellite imagery scan over Bishkek airshed (a) screenshot of the central heating plant (b) Photo of the central heating plant stack (c) screenshot of a cluster of brick kilns to the north of the city (d) Photo of an operational brick kiln in the airshed (e) screenshot of the landfill (f) screenshot of an operational quarry. Photos by the authors.

Temporal allocation of emissions is also a critical step establishing an operational emissions inventory, since emissions from any of the known sectors is emitted at the same intensity at all the time. A good allocation algorithm helps to accurately represent how emissions vary over time.

One common approach is to use monitoring data from air quality stations to develop diurnal profiles. When all stations are considered collectively, an overall temporal pattern can be established. However, if stations are categorized by specific sources—such as transport or industrial sectors—more tailored diurnal profiles can be built. For instance, transport stations might show distinct rush-hour peaks, while industrial stations could highlight emissions patterns linked to factory operating hours. These profiles allow for the allocation of emissions more precisely throughout the day, reflecting operational conditions.

Additionally, meteorological data can be used for seasonal emissions allocation. For example, surface temperature data can be used to adjust emissions from heating sources, with higher heating-related emissions in colder months. Similarly, power demand, which can be influenced by both air conditioning in summer and heating in winter, serves as a useful proxy for emissions related to electricity generation.

Temporal allocation can also benefit from vehicle speed data sourced from platforms like Google Maps. By analyzing speed trends, it is possible to differentiate between busy and non-busy hours, and even identify distinct weekday versus weekend traffic patterns. These insights help to allocate vehicular emissions representatively across different times of day, accounting for both traffic volume and congestion, leading to a more refined temporal distribution of emissions.

Depending on the chemical mechanism (see M) selected for use in the chemical transport models, the **emission inventory must also provide speciated information for NMVOC's**. One useful tool to accomplish this is SPECIATE from the US EPA. Speciation of VOC emissions is integral requirement of air quality simulations, particularly those focusing on the formation and destruction of ozone, the behavior of photochemical regimes, and the production of secondary organic aerosols (SOA). Different VOCs have varying reactivities and roles in photochemical and aerosol processes, meaning that accurately speciating VOC emissions—identifying the specific compounds and their quantities—is essential. Typical list of species included in an emissions inventory via CAMx system are:

Typical list of emission species for CB6 mechanism in CAMx

FPRM = Fine PM
CPRM = Coarse PM
PEC = Elemental carbon
POC = Organic carbon
SO2 = Sulphur Dioxide
NO = Nitric Oxide
NO2 = Nitrogen Dioxide
CO = Carbon Monoxide
NH3 = Ammonia
VOC species ==>

More sub-categories for PM are defined in the model. See model manual.

ACET = Acetone
ALD2 = Acetaldehyde
ALDX = Propionaldehyde and higher aldehydes
BENZ = Benzene
ETH = Ethene
ETHA = Ethane
ETHY = Ethyne
ETOH = Ethanol
FORM = Formaldehyde
IOLE = Internal olefin carbon bond (R-C=C-R)
ISOP = Isoprene
KET = Ketone carbon bond (C=O)
MEOH = Methanol
OLE = Terminal olefin carbon bond (R-C=C)
PAR = Paraffin carbon bond (C-C)
PRPA = Propane
TERP = Monoterpenes
TOL = Toluene and other monoalkyl aromatics
XYL = Xylene and other polyalkyl aromatics
IVOC = Intermediate VOCs

When building an emission inventory (both totals and gridding), the most rewarding moment comes from using entirely local data—ranging from activity data and surveys to emission factors and spatial and temporal gridding algorithms. However, for those just starting an air pollution modeling exercise, it's practical to **begin by using existing urban, regional, or global emission inventories**. This approach allows for immediate insights before dedicating more time and resources to develop customized, locally tailored inventories. Most used global and regional emission inventories in air pollution modeling are:

1. GEIA (Global and regional repository)
2. EDGAR, CEDS, CAMS, HTAP (Global – gridded - anthropogenic)
3. SMOG-India (India – gridded - anthropogenic)
4. MIX, REAS (Asia – gridded – anthropogenic)
5. DICE, DACCIWA (Africa – gridded – anthropogenic)
6. MEGAN (Global – gridded – biogenics)
7. FINN, GFED (Global – gridded – open fires)
8. GAINS (Regional, Global – anthropogenic)

The latest versions of chemical transport models, such as WRF-Chem, come equipped with built-in emission pre-processors that streamline the integration of global anthropogenic emission inventories like EDGAR or HTAP. Additionally, WRF-Chem includes modules that calculate natural emission sources online, such as biogenic emissions from vegetation, sea salt aerosols from oceans, and lightning-generated NO_x (Read N), making the modeling process more dynamic and user friendly.

In contrast, models like CAMx have independent pre-processors that offer more flexibility and customization options. These pre-processors allow users to tailor emission inventories to specific needs, whether regional or local, and to incorporate additional data sources. CAMx provides detailed instructions for users to customize emission inputs and calculate natural emissions, such as dust storms, sea salt, biogenics, and lightning, outside of the core model. This modularity enables researchers to modify and refine emission inventories to suit specific study areas, ensuring that both anthropogenic and natural emissions are represented.

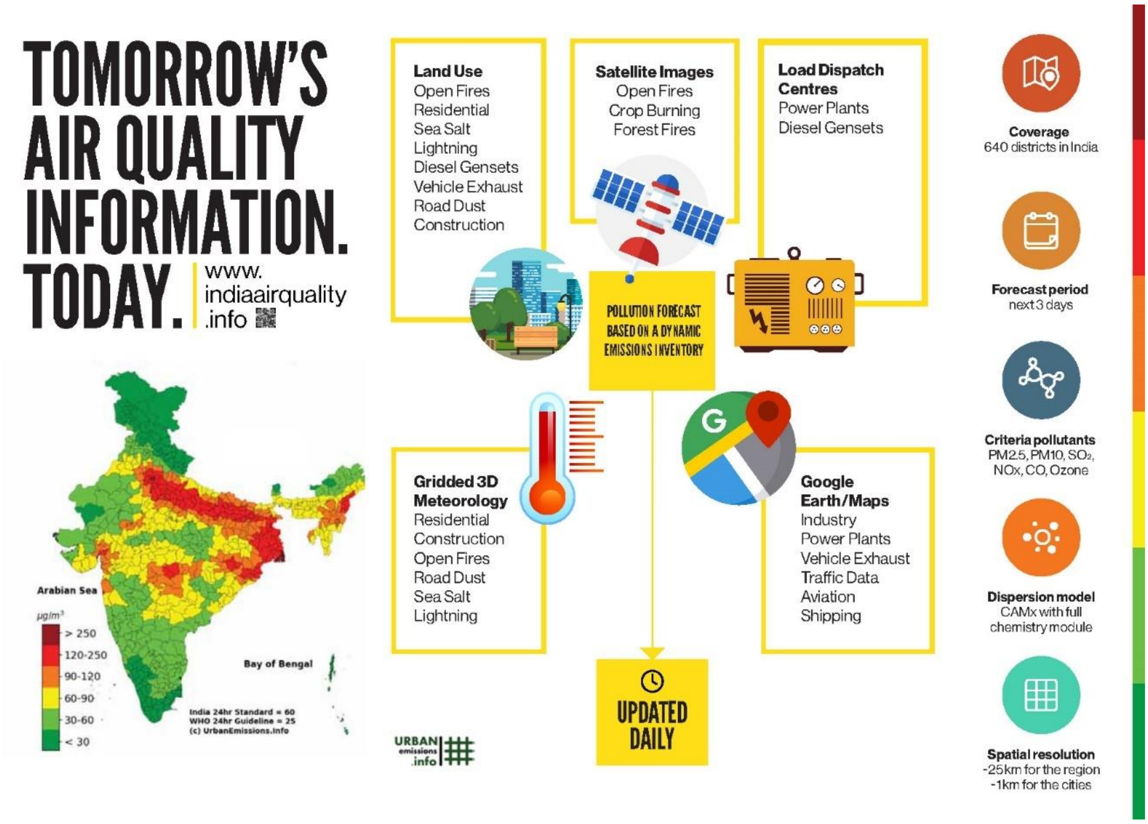
By supporting both default global databases and highly customizable options, these models ensure that anthropogenic and natural emissions data is precisely integrated into air quality simulations. This versatility makes them indispensable tools for both global-scale atmospheric research and localized air pollution studies.

A library of example tools to calculate total emissions including average emission factors and layers of GIS information which can be used as proxies for allocating these emissions to grids are included @ https://urbanemissions.info/tools

F

Forecasting Air Quality

Air pollution modeling for short-term periods, such as the next day, 3 days, or up to 10 days, is referred to as "**air quality forecasting**". The methodology for air quality forecasting largely mirrors that of long-term modeling, using the same meteorological and chemical transport modeling setups (Read C and W). This involves employing weather forecasts for the same period and running the chemical transport model to predict how pollutants will disperse, transform, and settle under various atmospheric conditions. These forecasts are invaluable for predicting pollution events like high ozone levels or particulate matter spikes, allowing authorities to issue alerts or advisories, to reduce the anticipated exposure levels.



However, one key distinction between short-term forecasting and long-term modeling lies in the need for highly specific and up-to-date emission inventories. For air quality forecasting, the emissions data must be vetted and adjusted to reflect the anticipated activities for the upcoming days. This could involve accounting for changes in traffic patterns, industrial operations, or even special events that might alter regular emission patterns.

This is in addition to anticipated meteorological conditions, such as temperature, wind speed, and precipitation, modeled from systems integrating WRF and global forecasting systems (GFS).

Some examples of dynamic corrections implemented in the air quality forecasting systems at UrbanEmissions include:

- **Reduction in daytime traffic-related emissions** in response to prior notice of a holiday, leading to school or office closures.
- **Integration of surface temperature profiles** from WRF forecast simulations (by grid and by hour) to adjust space heating demand—modifying both the duration and intensity of emissions based on heating needs.
- **Reduction in road-dust resuspension emissions** using precipitation data, as rain reduces the likelihood of dust being kicked up.
- **Adjustments to traffic-related emissions** based on precipitation fields, such as eliminating motorcycle emissions in grids and hours with more than 1 mm of rain and reducing car emissions under more rainy conditions.
- **Reduction in emissions from open waste burning and construction activities** during rainy periods, as precipitation dampens these activities.
- **Increase in road-dust resuspension emissions** during windy conditions in non-precipitation grids/hours, as stronger winds stir up more dust.
- **Modulation of congestion-related emissions** based on grid-average vehicle speeds, reflecting real-time traffic conditions.
- **Integration of daily open-fire detection data** into emission processors to adjust for fire-related pollution.
- **Incorporation of daily power plant electricity generation and fuel usage data** into emission processors to reflect real-time changes in energy demand and fuel consumption.

These dynamic corrections help refine the accuracy of short-term air quality forecasts by accounting for real-time changes in human activities, weather conditions, and other factors that influence emissions. The algorithms, while sound logical, needed some back and forth, before settling on an operational function to reflect the changes with meteorology. All these corrections are also useful for long-term applications but considered crucial for short-term, where the accuracy in the alert system is vital. Day to day changes in the emission intensities are archived and used for long-term assessments in hindcast mode.

A list of known short-term forecasting platforms:

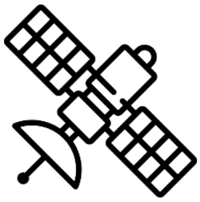
1. CAMS+ECMWF (from the EU)
2. GEOS-CF (from NASA, USA)
3. MOZART/CAM-chem (from NCAR, USA)
4. SILAM (from Finnish Meteorological Institute)

A list of open visualization tools (last accessed September 2024):

1. www.earth.nullSchool.net
2. www.windy.com
3. www.waqi.info

G

Geospatial Information Systems (GIS)



Maps and lines help us visualize locations and make connections between different places. On a map, lines can represent roads, rivers, or boundaries, making it easier to understand how things are laid out in an airshed and how people or goods move within and across regions.

GIS data is essential for mapping various elements in a city or region, such as infrastructure, land use, and environmental features. By providing detailed, location-based information, GIS helps identify spatial trends and patterns that are critical for urban planning and environmental management. For example, GIS can show how population density is distributed across different neighborhoods, where industrial zones are concentrated, and how traffic flows through major corridors—factors that influence air pollution, water pollution, noise pollution, and the urban heat island effect. GIS enables decision-makers to identify these problem areas (hotspots) and optimize geographic options to implement more targeted interventions for public health.

In emissions and pollution modeling exercises, GIS plays a central role by providing spatial context and detailed geographic data. Some key applications include:

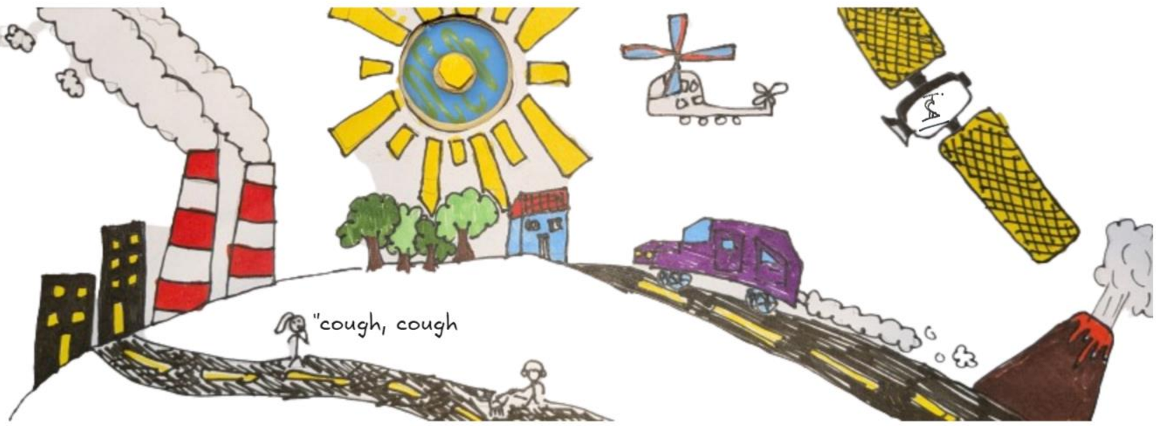
1. **Spatial allocation of total emissions**: (Read E) GIS is used to allocate total emissions to specific airshed grids using proxy weights such as road density, population density, commercial activity density, the location of point sources (e.g., factories, brick kilns), and information on urban-rural classification.
2. **Monitoring station placement**: (Read M) GIS helps in determining the best representative locations for air quality monitoring stations by using information like topography, land use, and meteorological data like wind patterns and temperature. This ensures that monitoring stations are optimally placed to capture locations that are representative of the activities in the entire airshed.

3. **Land use regression**: (Read L) GIS supports land use regression models by linking spatial data, such as traffic volume, industrial activity, and green spaces, with pollution levels. Through this pollution exposure (at high resolutions) can be estimated across different areas, even where monitoring data is sparse.

4. **Reduced complexity models** (RCMs): (Read R) These models simplify the representation of complex atmospheric processes and use GIS data to manage and visualize inputs and outputs, making it easier to assess pollution impacts on a wide spatial scale while reducing computational effort.
5. **Trajectory analysis**: (Read T) Trajectory analysis helps to identify the source regions of pollution and GIS is used to further these hotspots by integrating meteorological data and pollutant dispersion models.

These GIS-based tools and techniques are fundamental for **spatially accurate pollution modeling**, helping decision-makers better understand emissions distribution, pollution source hotspots, and exposure risks, ultimately leading to more informed and targeted interventions. In addition, these databases play a critical role in environmental justice studies by revealing how pollution disproportionately affects marginalized and vulnerable communities. By **integrating spatial data with socio-economic indicators**, these tools help highlight disparities in environmental health, enabling policymakers to address inequities and ensure that interventions benefit all populations, particularly those most at risk.

Everything in the image is a geospatial data



A repository of resources useful for emissions and energy modeling are documented @ https://urbanemissions.info (click on resources).

One of the largest GIS data links is available @ https://freegisdata.rtwilson.com – "*This page contains a categorised list of links to over 500 sites providing freely available geographic datasets - all ready for loading into a Geographic Information System*".

Some commonly used resources are listed below:

1. GADM is a repository of global administrative boundaries of countries at various levels (0,1,2,3,4) (https://gadm.org/download_country.html)
2. The Global Human Settlements (GHS) program by the European Commission (https://ghsl.jrc.ec.europa.eu/datasets.php) is a valuable initiative that provides comprehensive data on urban and rural classification, population density, and future projections, based on satellite imagery and a library of geospatial data worldwide. This enables informed

decision-making in fields such as urban planning and disaster risk management, making it a crucial tool for long-term planning.

3. The LANDSCAN program (https://landscan.ornl.gov) developed by Oak Ridge National Laboratory, provides high-resolution global gridded population data by combining data from various sources, including satellite imagery, census data, land use and land cover, and transportation networks, to create a detailed representation of where people are located at different times of the day (a proxy for activity levels). This gridded data (at spatial resolution of 1 km) is updated annually, ensuring that the estimates use a fixed set of methods and resources. Other population databases include: Socioeconomic Data and Applications Center (sedac) - https://sedac.ciesin.columbia.edu/data/collection/gpw-v4; WorldPop - https://www.worldpop.org;
4. Open Street Maps (OSM) database (https://www.openstreetmap.org) is an open public repository that provides free, editable maps of the world, created and maintained by a community of volunteers. The platform contains various types of data, including roads, buildings, land use, natural features, and points of interest such as schools, hotels, hospitals, shops, banks, traffic junctions, petrol stations, worship houses, and other businesses. As an example of citizen science, OSM allows everyday users to contribute their local knowledge, helping to improve the accuracy and detail of global maps.
5. Like OSM, Google also provides a variety of mapping databases via Google Maps, which are widely used for navigation and geospatial tagging. While Google's resources are open for public viewing and small-scale use, they are not free for extensive applications or large-scale commercial projects, often requiring paid licenses for full access to their data and services.
6. Google Earth (https://earth.google.com) is a widely accessible, open-source application available both as an app and in a browser, allowing users to explore satellite imagery and scan for points of interest. It offers an easy-to-use interface with simple navigation and zoom features, enabling users to quickly scan areas to identify roads, landmarks, and various activities, including those not always mapped, such as quarries. When official maps for industrial and large-scale commercial activities are unavailable, manual scans on Google Earth can be used to create useful GIS layers for these activities, including roads.
7. Google Earth Engine (https://developers.google.com/earth-engine/datasets) is a powerful platform that provides access to a vast library of satellite data and cloud-based processing tools for large-scale environmental analysis. It enables users to analyze geospatial data efficiently by leveraging Google's infrastructure for high-performance computing, making it a valuable resource for researchers, scientists, and public. For instance, the platform allows access to ultraviolet aerosol index (UVAI) from the TROPOMI satellite and aerosol optical depth (AOD) data from MODIS-MAIAC. These datasets can be used to track and analyze trends in particulate pollution and compare methodologies, enabling the creation of ensemble-style correlations with other factors such as energy consumption or meteorological patterns in a specific region. However, while small-scale applications are free, large-scale applications may require additional cloud storage and computational resources, which fall under paid services. Other examples include:

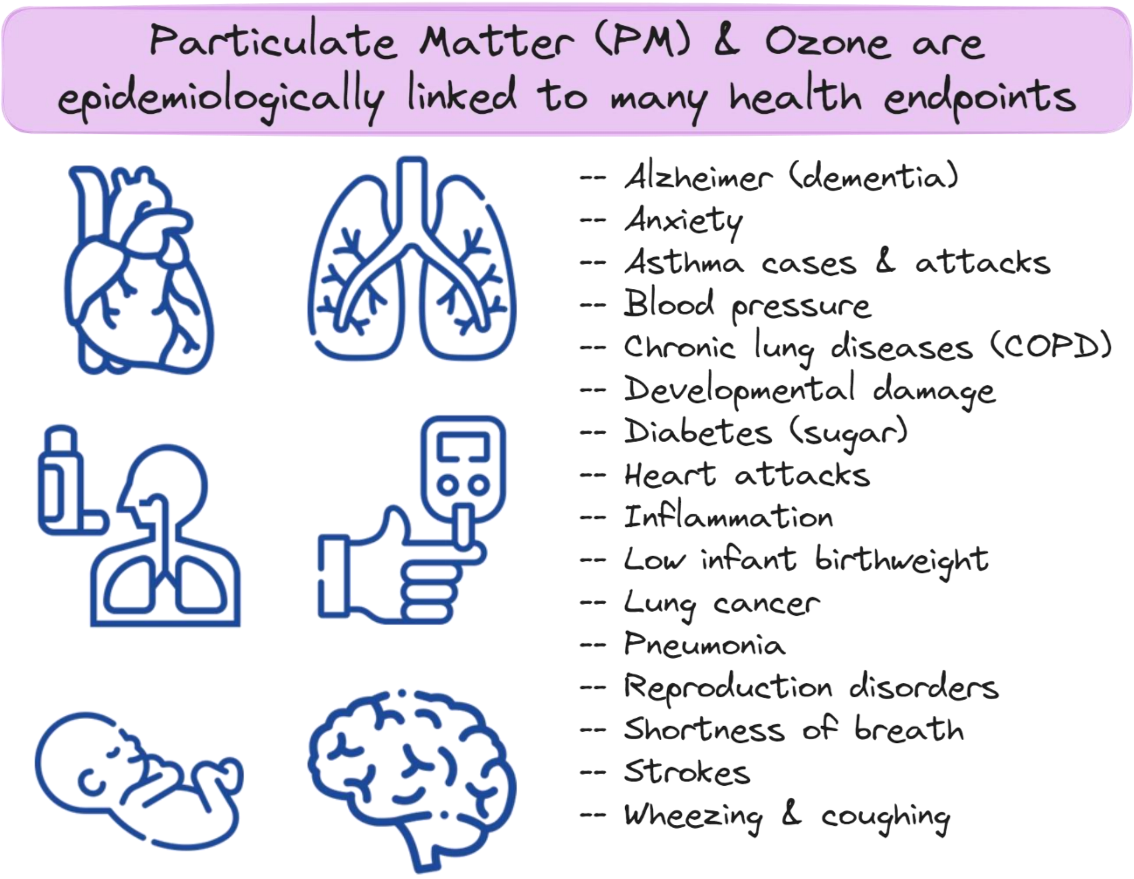
    a. Sentinel-5P OFFL SO2
    b. Sentinel-5P OFFL NO2
    c. Sentinel-5P OFFL HCHO
    d. Sentinel-5P OFFL AER AI
    e. Sentinel-5P OFFL O3 TCL
    f. MERRA-2 M2T1NXAER: Aerosol Diagnostics V5.12.4
    g. MCD19A2.061: Terra & Aqua MAIAC AOD Daily 1km
    h. ECMWF-CAMS Global NRT
    i. ECMWF-ERA5 Reanalysis

8. FlightStats (https://www.flightstats.com) is a comprehensive flight tracking and data service platform that provides real-time and historical information on global flight operations by airlines and airports. In addition to its free consumer-facing services, FlightStats offers APIs and data analytics solutions on historical data as a commercial service, which are highly useful for studying emission trends from aviation activities.
9. The European Space Agency's (ESA) Land Cover Climate Change Initiative (CCI) is a project (https://www.esa-landcover-cci.org/?q=node/164) provides high-resolution global land cover maps from 1992-2022. These are most useful in studying the rate of urbanization and other landuse changes in and around cities across the world.
10. Open Buildings dataset (https://beta.source.coop/repositories/vida/google-microsoft-open-buildings/description) is a large-scale, merged collection of building footprints from two major sources: Google's V3 Open Buildings and Microsoft's latest Building Footprints dataset. This comprehensive dataset contains a staggering 2,579,035,323 building footprints and is divided into 185 partitions for easier access and processing. Its high-resolution coverage across diverse geographical regions makes it a crucial resource for both researchers and developers in fields of urban planning and environmental sustainability.
11. The Night-Lights Database (https://eogdata.mines.edu/products/vnl/) is a powerful resource that uses satellite imagery to capture artificial lighting at night across the globe. This dataset, which records the intensity and distribution of nighttime lights, provides valuable insights into human activity, urbanization, economic development, and energy consumption. Night-lights data are also used as a proxy for socioeconomic indicators, as brighter areas often correspond to more economically developed regions, while areas with less light may indicate lower levels of development or access to electricity.
12. Web-based Reanalysis Intercomparison Tool: Monthly/Seasonal Time-Series (https://psl.noaa.gov/data/atmoswrit/timeseries) – This portal provides access to global reanalysis fields from multiple systems as maps, tables, and raw data for a number of meteorological parameters, such as wind speeds, wind directions, pressure, temperature, heat content, water content, and precipitation.

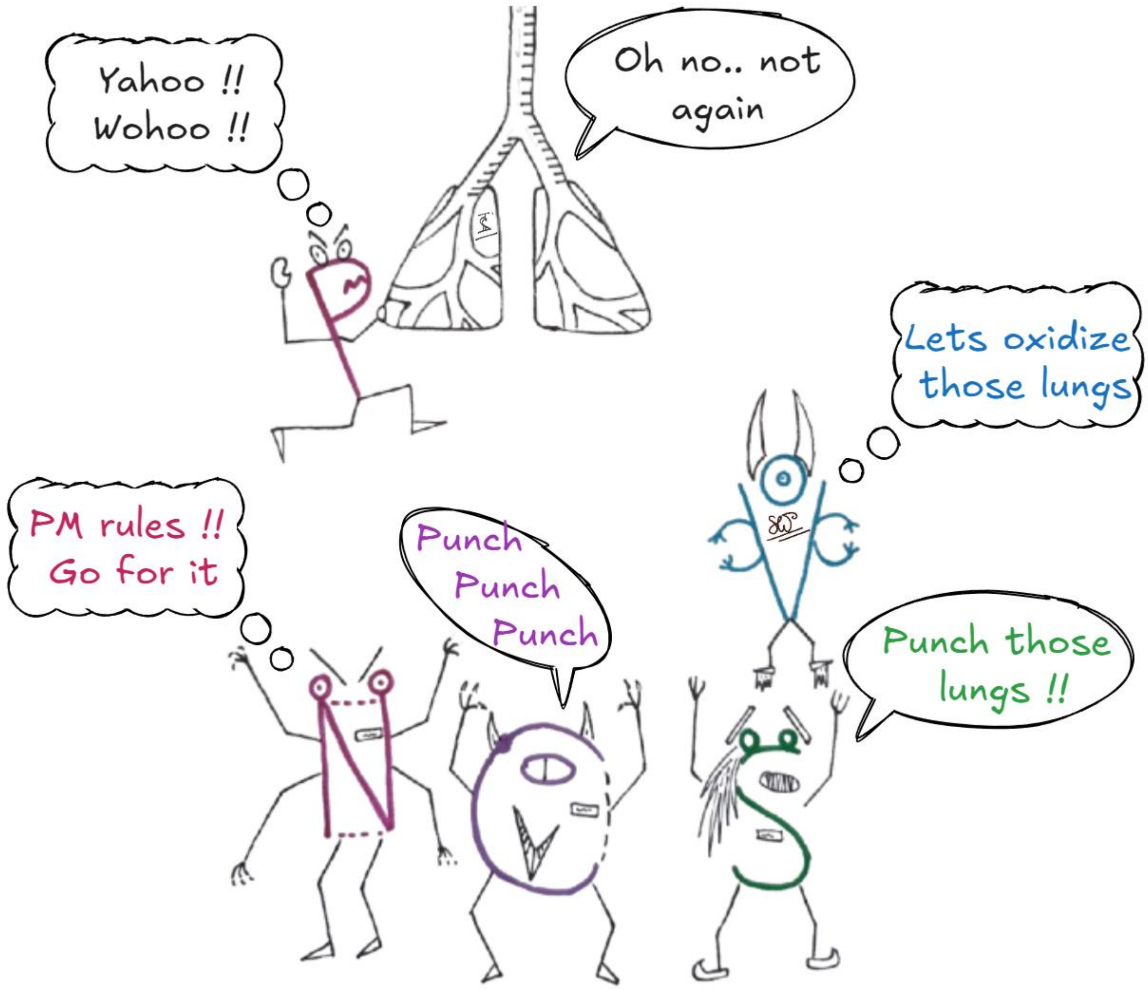
H

Health (Exposure) Impacts Modeling

One of the most studied and widely discussed endpoints of modeling is "**health impacts of air pollution**", which are immediate and global in reach. Health impacts serve as an ideal metric for raising public awareness, linking policy measures, and personalizing the often-abstract concepts of air quality and pollution. These impacts occur at multiple scales: **chronic effects** from long-term exposure, such as respiratory and cardiovascular diseases; **acute effects** from short-term exposure, like asthma attacks and irritation; **occupational impacts**, where workers in certain industries face heightened risks; and **personal effects**, which vary depending on individual vulnerability and proximity to pollution sources.



Air pollution affects nearly every part of the human body, with some pollutants contributing to serious health issues. Across the globe, numerous epidemiological studies have been conducted and are ongoing to better understand the linkages between air pollution and a wide range of health impacts, including the incidence of premature death in many cases. These studies are not simple exercises; they require significant time, effort, and collaboration among experts in various fields, such as medicine, data science, statistics, atmospheric science, and the contributions of many volunteers. While this chapter does not delve into the methodologies for conducting such studies, it builds on the knowledge gained from them to explore how we can model these health impacts and use the results for cost-benefit analysis, ultimately helping to inform policy and decision-making.



All pollutants have an impact on our health, though the severity varies by pollutant and exposure level. Among them, $PM_{2.5}$ and ozone concentrations are most used to assess the health impacts of chronic exposure. The health effects of $PM_{2.5}$ range from respiratory issues and cardiovascular diseases to premature mortality, as well as exacerbating conditions like asthma and bronchitis. Chemically, $PM_{2.5}$ carries signatures of various gaseous pollutants, as secondary PM in the form of sulfate aerosols from SO_2, nitrate aerosols from NO_x, organic aerosol from VOCs. Beyond $PM_{2.5}$, gaseous pollutants also affect human health in different ways: SO_2 can cause respiratory problems and aggravate lung diseases, NO_2 contributes to respiratory inflammation and worsens asthma, CO impairs oxygen delivery to the body, and ozone causes respiratory distress and long-term lung damage. Together, these pollutants have widespread and serious health implications.

Over the past few decades, methodologies for evaluating these health impacts have evolved significantly. The most recent advancement is the development of **Integrated Exposure Response** (IER) functions, linking various levels of air pollution exposure to specific health outcomes across different populations. This library of work was created as part of the Global Burden of Disease (GBD) study led by the Institute for Health Metrics and Evaluation (IHME), in collaboration with the Health Effects Institute (HEI) and a consortium of leading research institutions.

IER Health Impacts Calculator

$$HI_i = Y \times AF \times PoP_i$$

$$AF = 1 - \frac{1}{RR}$$

$$RR(z) = 1 \quad \text{for } z < z_{cf}$$

$$RR(z) = 1 + \alpha\left[1 - e^{-\gamma (z - z_{cf})^{\delta}}\right] \quad \text{for } z > z_{cf}$$

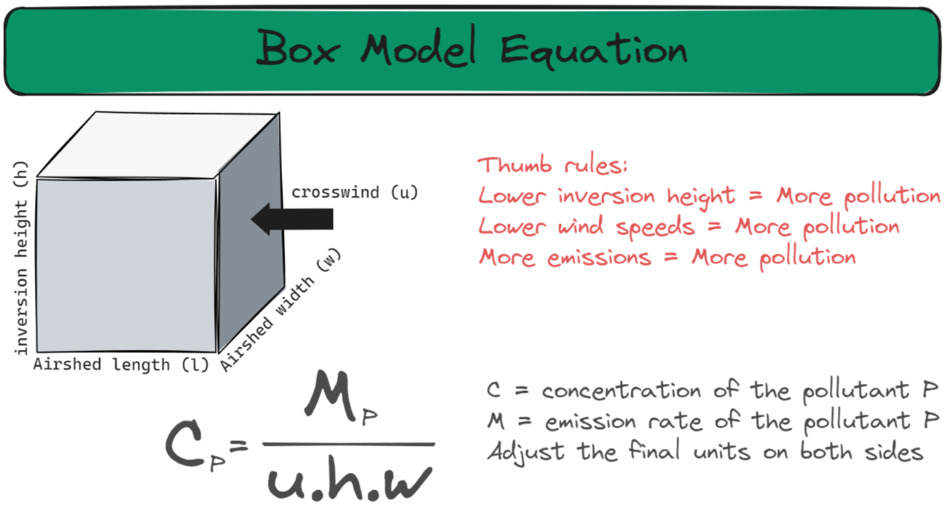
HI = estimated health endpoint impacts in zone i
POP = population exposed in zone i
Y_o = incidence/prevalence rate of health endpoint
AF = attributable fraction
RR = relative risk of the health endpoint
z = PM2.5 concentration
z_{cf} = counterfactual concentration
α δ γ = IER parameters for calculating RR

These IER functions quantify the relationship between exposure to certain health risk and the subsequent health outcomes at the population level. The end metric is as both premature morality and disability adjusted life years (DALYs). These functions incorporate epidemiological evidence and dose-response relationships from across the world and information on national disease incidence rates and health management systems. All the inputs to run the IER function for regional and global assessments are available from **the State of the Global Air program**.

An example tool to estimate health impacts using this methodology is included @ https://urbanemissions.info/tools.

Inversion (Mixing) Layer

Mixing height is a key meteorological parameter that significantly impacts pollution levels by determining how well pollutants disperse in the atmosphere (refer to box model description in C). Without doing any modeling, just considering this parameter, we can determine whether local air quality falls into the "poor," "very poor," or even "severe" air quality alert zones, as lower heights contribute to higher pollution loads in an urban environment with history of high emission intensity.

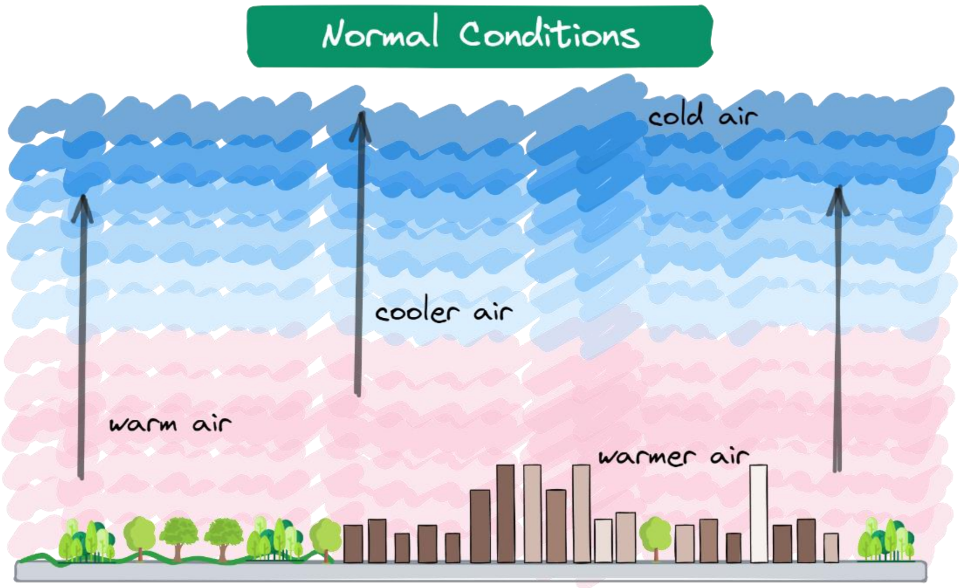


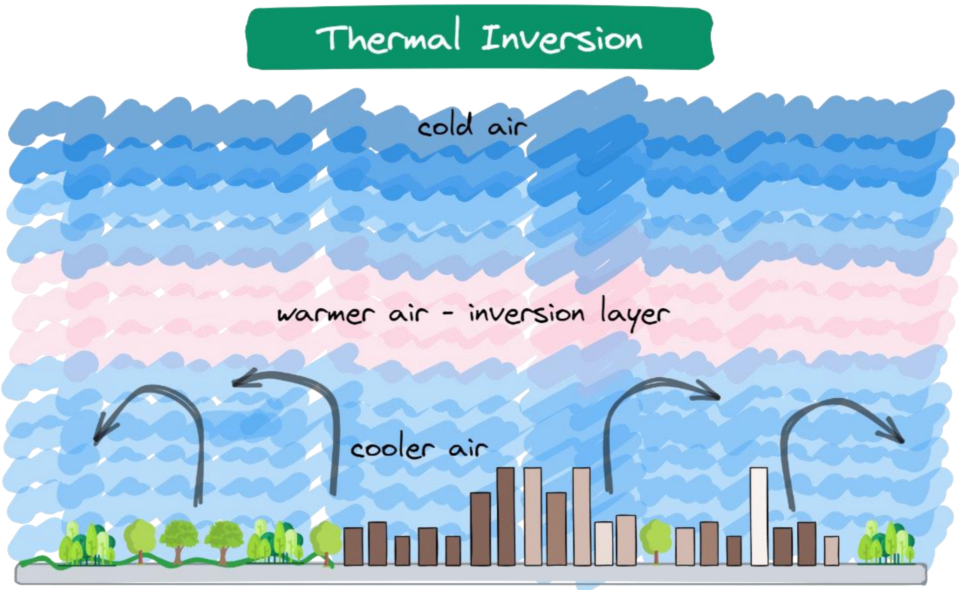
Under normal conditions, the air closer to the surface is warmer due to heat absorbed by the Earth, and as we move higher, the air cools. This temperature gradient allows air to rise from the surface to higher altitudes, carrying emissions with it. As a result, the mixing of air is more uniform, which generally helps to disperse pollutants and lowers ambient pollution levels.

In contrast, during thermal inversion conditions, a layer of warm air becomes trapped between layers of colder air, acting like an invisible barrier that prevents the vertical mixing of air. This inversion layer inhibits the upward movement of emissions, keeping pollutants closer to the surface. During colder weather, this inversion layer often forms near the ground, significantly increasing ambient pollution levels by trapping pollutants in the lower atmosphere. This difference is evident between daytime and nighttime periods and between summer and winter months. The most favorable conditions for thermal inversion are:

- **Clear skies at night without cloud cover**: Without cloud cover, the Earth's surface loses heat rapidly after sunset, causing the air near the ground to cool quickly. This leads to cooler air getting trapped under a warmer layer.
- **Calm winds**: Light or calm winds prevent the mixing of air layers, allowing the cooler air to remain close to the ground and creating a stable inversion layer.

- **Long nights during colder months**: Longer nights, especially during the winter, allow for extended periods of radiational cooling, which enhances the formation of inversion layers.
- **Topography**: Valleys and low-lying areas are more prone to thermal inversions because cool air tends to sink and accumulate in these regions, with warmer air layers forming above.





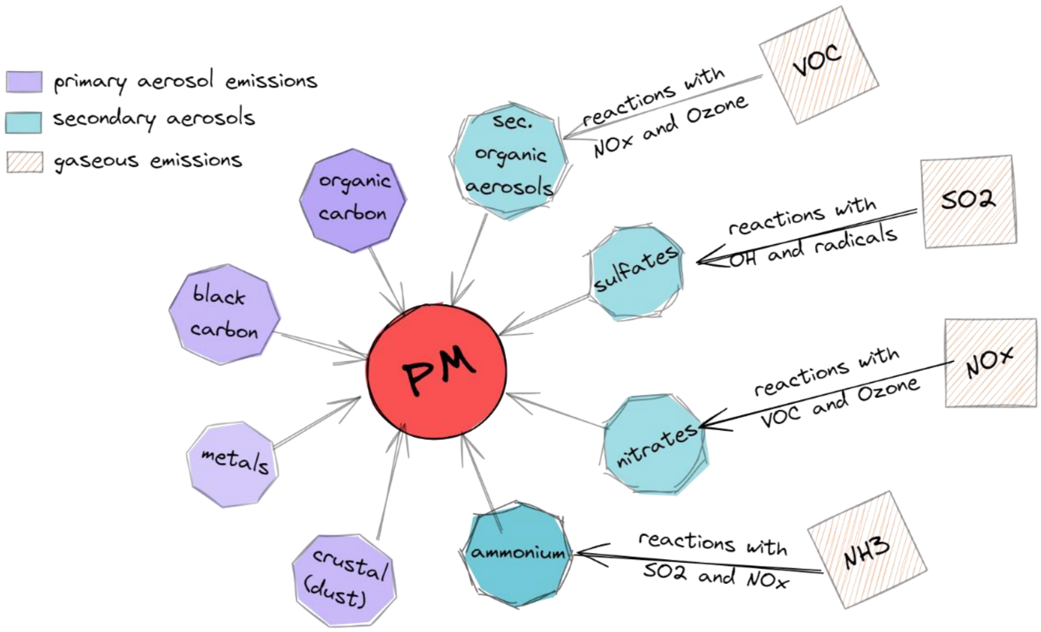
In general, the terms "temperature inversion" and "inversion layer" are synonymous, and the terms like "potential boundary layer" or "planetary boundary layer" (PBL) describe a broader atmospheric region where the inversion layer may influence but isn't the same. PBL is often discussed in the chemical transport model results and often all the layers under PBL are used to report changes in advection of pollutants and chemical mechanisms. This is the lowest part of the atmosphere where most mixing occurs due to friction with the Earth’s surface. The inversion layer can cap the boundary layer, preventing vertical air movement, but the PBL typically refers to the entire region where surface-atmosphere interactions take place.

JK

J-rates & K-rates (Chemical Mechanisms)

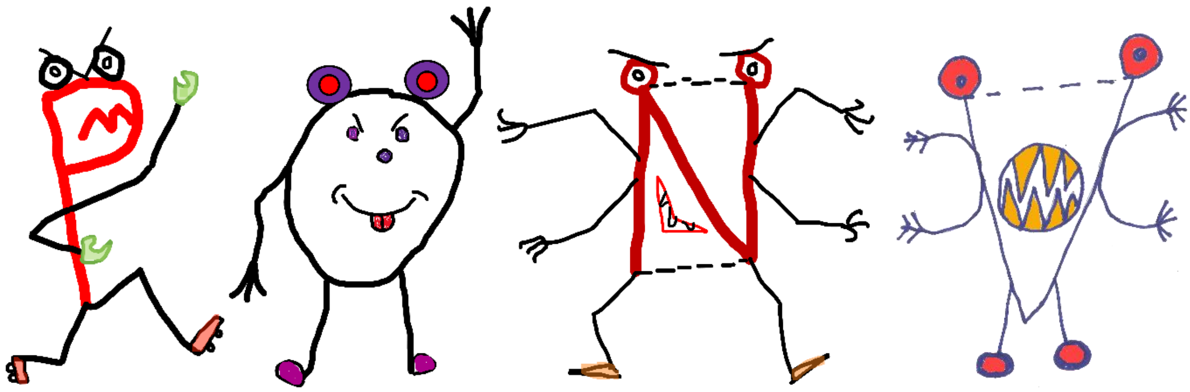
Air pollution is a complex mixture of pollutants, including long-lived, short-lived, and extremely short-lived species such as radicals. The lifespan of these pollutants varies, with some capable of traveling long distances and undergoing chemical transformations as they move across different regions. Their behavior can also change depending on environmental conditions; some pollutants react differently in the presence of sunlight, while others behave distinctly at night. This variability, combined with the uncertainty in their proportions, makes the mixture even more toxic. Understanding the chemical nature of these pollutants—their concentrations, interactions, and transformations—is crucial for assessing their full impact on health and the environment.

Here, **photolysis** is the process by which chemical compounds are broken down in the presence of sunlight and triggering various chemical reactions in the atmosphere and **chemical mechanisms** describe the sequence of chemical reactions and processes that transform pollutants in the atmosphere, detailing how different substances react, form intermediates, and result in new compounds under specific meteorological conditions. The rates of reactions during photolysis are typically referred to as J-rates, while the rates of reactions between gaseous and aerosol species are called K-rates.



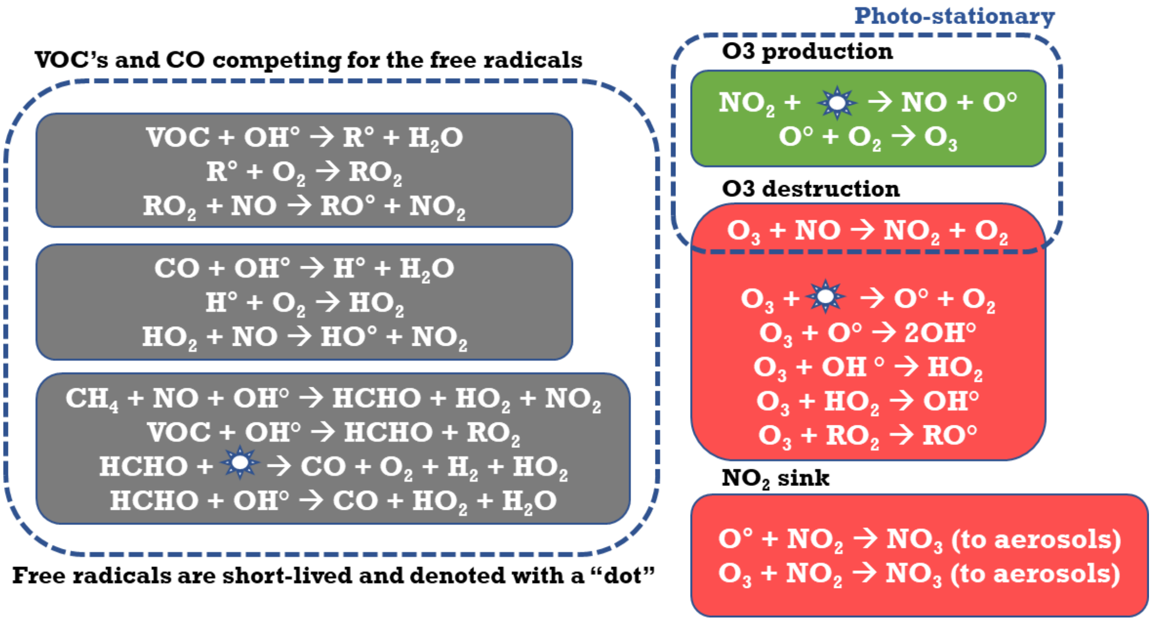
This chapter is not intended to cover everything about chemical mechanisms and reaction rates in exhaustive detail. Rather, it aims to give you a taste of the key elements that are important to understand in these areas. Established chemical transport models (Read C) include detailed descriptions of various mechanisms, ranging from simpler models with around 100 reactions to more advanced ones featuring over 500 reactions. These models account for numerous intermediates and radicals, capturing the complexity of atmospheric chemistry. Understanding even the basics of these mechanisms is crucial, as they form the backbone of how pollutants are transformed and transported in the atmosphere.

Take $PM_{2.5}$, for instance—the most talked about and studied pollutant globally. $PM_{2.5}$ is not a single entity, but a complex mixture formed from both direct emissions and chemical transformations over time and space. It incorporates contributions from various criteria gases, including sulfates from SO_2, nitrates from NO_x, ammonium from NH_3, and secondary organic aerosols (SOA) from VOCs (Read H). These chemical transformations are not the result of a single reaction but are dependent on a range of factors, including the ratios of pollutants and meteorological conditions. For example, SOA forms through reactions between NO_x, VOCs, and ozone, which are more active in the presence of sunlight, and their reaction rates vary based on the relative concentrations of each component. Similarly, distinct chemical regimes govern non-organic aerosol components like sulfates, nitrates, and ammonium, each with its own intricate dependencies.



Ozone, on the other hand, undergoes even more complex chemical transformations—both in the presence of sunlight and in reactions with other pollutants. Ozone photolysis occurs when ozone molecules absorb ultraviolet light, causing them to break apart into oxygen molecules and oxygen atoms. The free oxygen atom can then react with NO_x or VOCs, contributing to the formation of secondary pollutants like photochemical smog or ground-level ozone. While ozone is beneficial when it forms in the stratosphere, its formation near the surface is harmful.

In urban environments, O_3-NO_x-VOC chemistry is very nonlinear, with two key regimes affecting ozone production. In a NO_x-limited regime, reducing NO_x emissions can effectively control ozone levels while VOC concentrations remain unchanged. On the other hand, in a VOC-limited regime, reducing VOC emissions lowers ozone production. Under conditions of high NO_x, known as NO_x-saturated conditions (which is often the case in the cities), reducing NO_x emissions decreases ozone titration, which can maintain or even increase ozone levels. Typically, at regional scales and in rural areas, ozone production is primarily NO_x-limited, whereas in urban areas, ozone production is VOC-limited. Some researchers use the formaldehyde (HCHO) to nitrogen dioxide (NO_2) ratio (FNR) as a proxy to explain this dependency.



The chemistry represented in these reactions is broad and covers several critical processes in atmospheric chemistry. These chemical processes include sulfur chemistry, NO_x chemistry, ozone chemistry, halogen chemistry, aerosol chemistry, radical chemistry (OH, HO_2, and RO_2), VOC oxidation, methane oxidation, long-range transport chemistry, urban chemistry, tropospheric chemistry, and stratospheric chemistry. There is published literature on each of these topics to explore in detail. **Common chemical mechanisms used in atmospheric modeling** include the following:

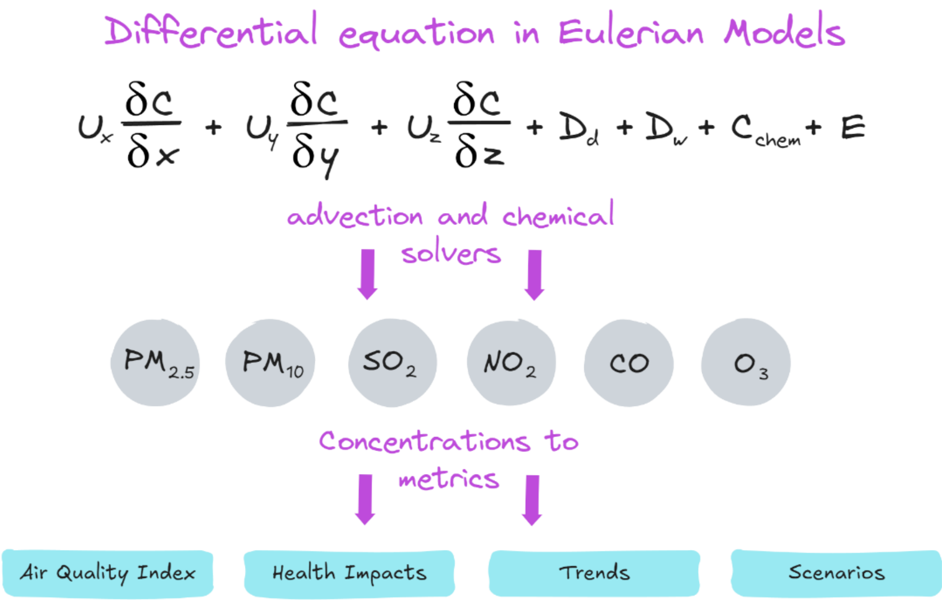
- **Carbon Bond Mechanism (CBM)**: CBM simplifies atmospheric chemistry by lumping organic compounds into groups based on the types of bonds they contain (e.g., single, double, or triple carbon bonds). Variants of this mechanism, such as CBM-IV, CB05, and CB6, are commonly used in air quality models for simulating urban air pollution, particularly for ozone and secondary organic aerosol formation.
- **SAPRC (Statewide Air Pollution Research Center) Mechanism**: Developed by the University of California, Riverside, the SAPRC mechanism is one of the most detailed chemical schemes used for modeling ozone, particulate matter, and other pollutants. It includes hundreds of reactions involving VOCs, NO_x, and other species. Versions like SAPRC-99, SAPRC-07 and SAPRC-11 are used in various regional air quality models.
- **Master Chemical Mechanism (MCM)**: Developed by the University of York, MCM is a complex chemical mechanism, with thousands of reactions and covering more than 140 VOC species alone. It is used for detailed simulations of the chemical degradation of VOCs in the atmosphere and for understanding the complex pathways of ozone and secondary pollutant formation.
- **GEOS-Chem Mechanism**: GEOS-Chem is a global chemical transport model with applications in regional and global air quality and tropospheric and stratospheric chemistry analysis. It maintains an internal detailed chemical mechanism that covers a wide range of atmospheric processes,

including oxidation of VOCs, nitrogen oxides, sulfur compounds, halogens, and aerosols (in gas and aqueous phase). It is widely used for studying global air pollution, long-range transport of pollutants, and climate interactions.

- **MOZART (Model for Ozone and Related Chemical Tracers) Chemical Mechanism**: This mechanism is comprehensive, covering a wide range of species and reactions that are critical to understanding global atmospheric chemistry. The mechanism's inclusion of long-range transport, radical chemistry, and stratosphere-troposphere exchange enables it to capture the complex processes that govern the transformation and distribution of pollutants on a global scale. Its detailed and tested representation of ozone, NO_x, VOCs, and aerosols makes it a valuable tool for simulating global air quality and the interactions between atmospheric chemistry and climate. The new versions of the model are called CAM-chem and WACCM.

Chemical solvers in the chemical transport models

Gone are the days when differential equations of chemical mechanisms were manually solved using logarithmic tables. Today, all chemical mechanisms have coded solver modules, allowing for efficient, automated solutions of complex atmospheric chemistry. These solvers are designed to be user-friendly, operating in a "select and play" mode where users can choose their preferred mechanism, and the model will handle the numerical computations.



However, one critical challenge remains: the names and lists of species differ between chemical mechanisms. This means that while the underlying chemistry may be similar, the way species like VOCs, NO_x, or ozone precursors are represented can vary from one mechanism to another. For example, in CBM, VOCs are grouped by carbon bonds, whereas in the SAPRC mechanism, VOCs are represented in more detail based on individual compounds. Because of this, emission inventories must be carefully tailored to match the chosen chemical mechanism. **Failure to match the inventory with the mechanism can lead to significant errors in model outputs**.

KPP (Kinetic PreProcessor) from the University of Iowa, is a widely used tool for generating efficient and customizable chemical solvers in atmospheric modeling. It automates the process of converting chemical mechanisms into numerical solvers, making it easier for researchers and modelers to simulate complex chemical reactions without manually solving each equation. This interoperability makes KPP a valuable tool for researchers studying air quality, climate change, and atmospheric chemistry and this module is integrated into most of the established regional and global chemical transport models.

Photolysis rates and kinetic reaction rates

Not all chemical reactions can be mathematically represented in the same way due to the varying complexity and nature of atmospheric processes. Different reactions may follow different kinetics, and their representation in models must account for factors like temperature, pressure, and the presence of other chemicals. For instance, reaction rate equations the CAMx model employs to simulate atmospheric chemistry are illustrated below.

Table 3-3a. Rate constant expression types supported in CAMx and order of expression parameters for the chemistry parameters file.

Expression Type	Description	Expression
1	Constant	$k = k_{298}$
2	UAM (Arrhenius expression)	$k = k_{298} \exp\left[E_a\left(\frac{1}{298} - \frac{1}{T}\right)\right]$
3	General temperature dependence	$k = A\left(\frac{T}{T_R}\right)^B \exp\left(-\frac{E_a}{T}\right)$
4	Troe-type temperature and pressure dependence	$k = \left(\frac{k^0[M]}{1+k^0[M]/k^\infty}\right)F^G$ $k^0 = A\left(\frac{T}{T_R}\right)^B \exp\left(-\frac{E_a}{T}\right)$ $k^\infty = A'\left(\frac{T}{T'_R}\right)^{B'} \exp\left(-\frac{E'_a}{T}\right)$ $G = \left[1+\left(\frac{\log\left(k^0[M]/k^\infty\right)}{n}\right)^2\right]^{-1}$
5	Equilibrium with a previously defined reaction (k_{ref})	$k = k_{ref}\left[A\left(\frac{T}{T_R}\right)^B \exp\left(-\frac{E_a}{T}\right)\right]^{-1}$
6	Lindemann - Hinshelwood as used for OH + HNO3	$k = k^0 + \frac{k_3[M]}{1+k_3[M]/k_2}$
7	Simple pressure dependence used for OH + CO	$k = k_1 + k_2[M]$

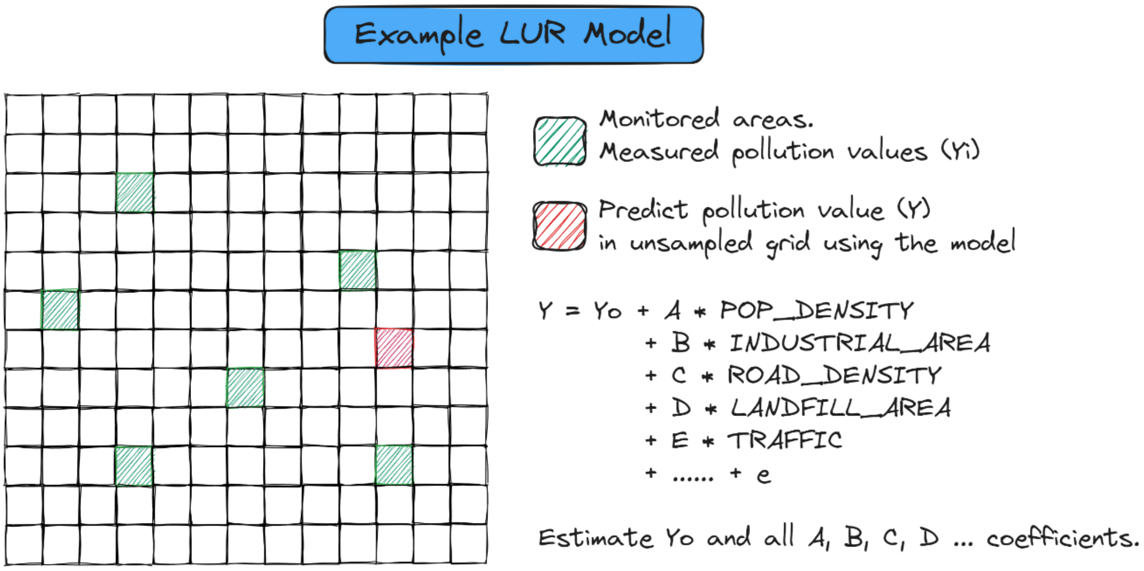
Modern solvers have streamlined the computational process in chemical transport modeling but ensuring that species in emission inventories align with those in chemical mechanisms remains the user's responsibility.

L

Land Use Regression Models

Land Use Regression (LUR) models offer an alternative to computationally demanding chemical transport models (Read C). These models can predict pollution levels at high resolutions, even in areas with limited monitoring data or just some modeled values, along with spatial predictor variables, primarily obtained through GIS (Read G). Common predictors include population density, road density, industrial land use, total traffic load, vegetation, and commercial activity density. LUR models can be used to downscale regional-scale model results, which typically have resolutions of 10 to 25 kilometers, to more detailed urban-scale resolutions of 1 kilometer and in some cases can also generate pollution heatmaps for an entire city at spatial resolutions as fine as 100 meters.

LUR models are highly localized because the relationships between predictor variables and pollution levels are specific to the unique characteristics of the city or region being studied. Factors like local traffic patterns, industrial activity, meteorological conditions, and geographical features influence how pollution disperses and accumulates in different areas. As a result, the regression functions developed for one city or region may not be directly applicable to another without recalibrating the model to account for local conditions. This specificity is both a strength—providing high accuracy in local predictions—and a limitation, as LUR models need to be tailored to each new location.



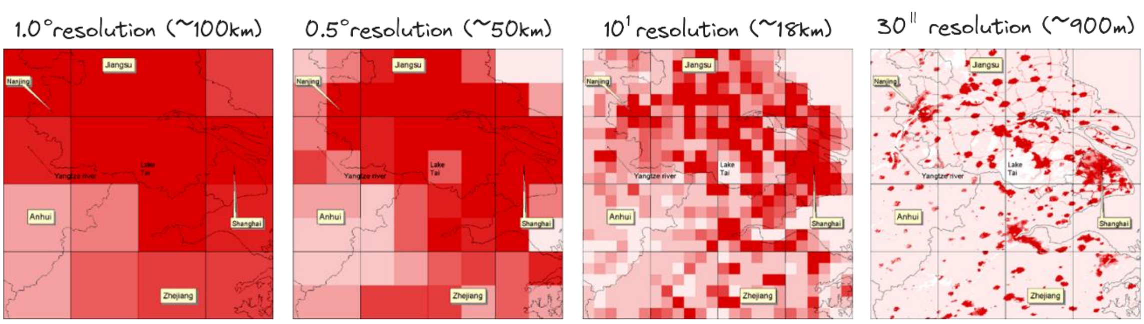
All the challenges that apply to a regression model, also apply to a LUR model. For example, if we miss out an important predictor variable that can explain air pollution in the city (say, construction activity), then the model will underestimate pollution in the grids that have high construction activity. This is called "omitted variable bias". At the same time, if we use a lot of predictor variables to get a perfect fit on the input and output pollution values, it causes "overfitting" and the LUR model has limited use for long-term assessments.

Creating reliable LUR models requires a significant amount of data collection, including reliable monitoring data for long-term overlapping with meteorological data at similar resolutions for the same period, geographic information, and predictor variables. However, once the necessary datasets are available, the methods for building LUR models are replicable. As long as similar data sources—such as GIS-based predictors and local air quality measurements—are accessible, the same modeling techniques can be applied to different regions to develop location-specific pollution predictions.

As data science and technology advance, researchers are increasingly looking to new data sources to enhance LUR models. Satellite imagery provides high-resolution data on land cover, vegetation, and urban development, while street view statistics offer granular details about local infrastructure, road conditions, and even vehicle types in specific areas. These emerging tools allow for the deduction of pollution levels with more confidence, especially in areas where ground-level monitoring is sparse or unavailable. By incorporating these advanced data streams, LUR models are paving the way for innovative and data-driven air quality modeling approaches.

Examples of re-allocation of monitored and modelled pollution values using these empirical methods are:

1. There are only 6 monitors in an airshed covering an area of 10 x 10 grids with a grid resolution of 4km (see the figure in the previous page). Using a LUR model with pre-defined coefficients linking various activity and landuse parameters of the airshed, concentrations can be regressed for the grids without any monitoring stations. This is not a dynamic model and does not account for the influences of meteorological parameters. Data from more monitoring stations can produce a better fit with the parameters. Emerging models also include use of Google streetview images to account for vehicle density for hyperlocal regressions using hyperlocal monitoring data.
2. Downscaling emissions or concentrations from a low-resolution to a high-resolution using local activity and landuse parameters. This methodology is beneficial in cases where (a) an emission inventory is only available at coarser resolutions and there is a need to run the chemical transport model at a high-resolution (b) a pollution heatmap is only available at a coarser resolution and computational costs are too high to run a high-resolution chemical transport model.



An integrated GIS package for conducting LUR is pyLUR – an open-source python language library.

It's important to emphasize that LUR is a statistical method based on identifying relationships (regressions) between various predictor variables—such as population density, road networks, industrial activity—and observed or modeled pollution levels. While LUR models can offer valuable insights, especially in areas with limited monitoring infrastructure, they are not a substitute for direct ground monitoring of pollution, or the more comprehensive analysis provided by classical chemical transport models.

LUR should therefore be viewed as an alternative tool, particularly useful for generating pollution estimates in areas without extensive monitoring or where computational resources for classical modeling are unavailable. It offers a practical solution for developing high-resolution pollution maps in urban environments and filling gaps in monitoring data. However, the method benefits from and complements classical models—using more data, including ground monitoring, making them more reliable.

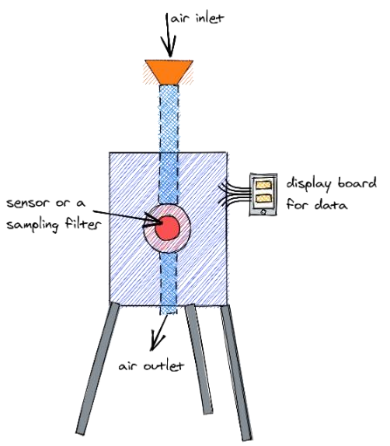
Ultimately, all approaches, when used together, contribute to a better understanding of air quality and pollution dynamics.

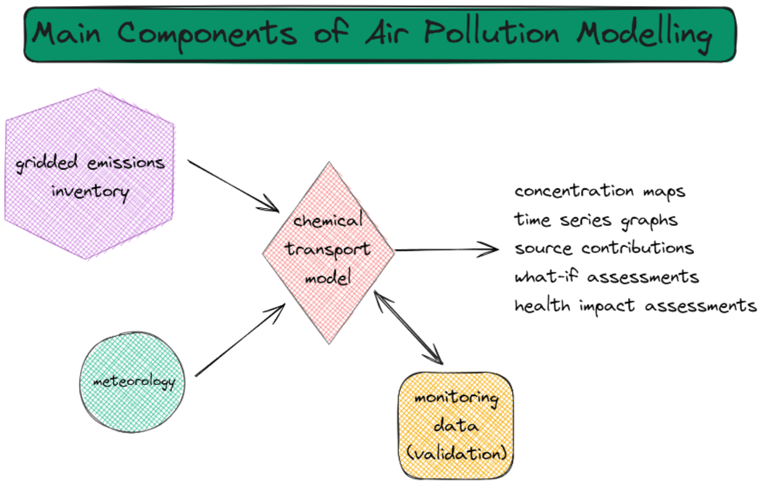
M

Monitoring Data

Seeing a number directly from a physical instrument often carries more credibility than one generated by a model. On any given day, monitoring data from a well-established network, sufficiently large to represent the airshed both spatially and temporally, is ideal for telling the full story of pollution levels. This is because monitoring stations provide direct, real-time measurements that capture the actual conditions of air quality across different locations and times.

In most countries, monitoring data is trusted more, as it reflects the ground reality and is often used for immediate air quality assessments and public health advisories. Additionally, when comparing air quality on a global scale—whether evaluating which country or city has good or bad air quality—these assessments are typically based on monitoring data. Monitoring provides a consistent, empirical basis for global rankings and comparisons, allowing for more reliable assessments of pollution levels.





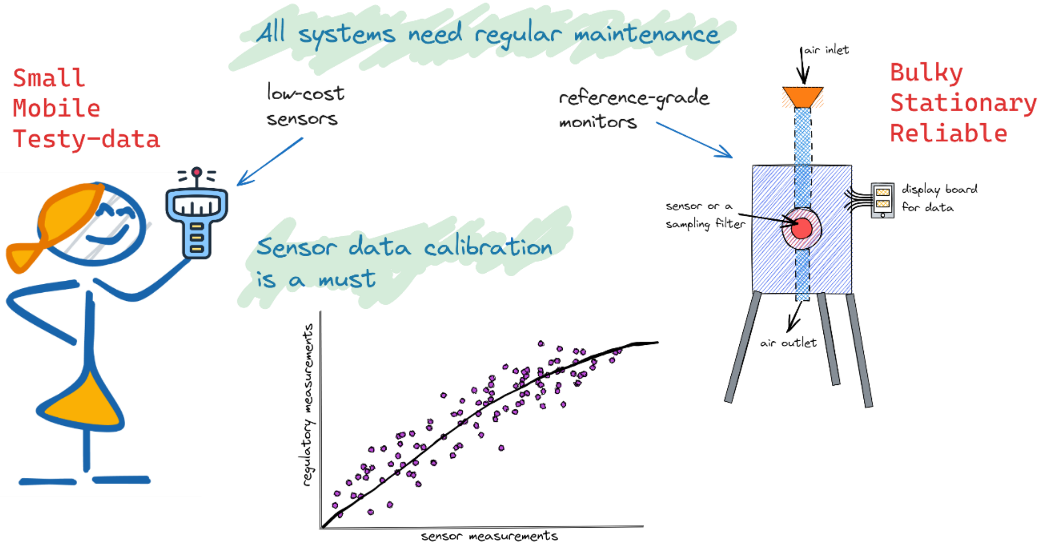
Air pollution modeling is a highly data-intensive and resource-demanding process. No matter which air pollution modeling approach is used—whether a simpler box model or a complex Eulerian chemical transport model—validation is crucial (Read V). Validation involves comparing the modeled outputs against real-world data, typically from ground-based air quality monitoring stations, to ensure that the model can replicate qualitatively and quantitatively the pollution levels and atmospheric dynamics. Validation also helps identify discrepancies, allowing for adjustments to improve the model's accuracy.

The success of the validation process hinges on the availability of a reliable and comprehensive air quality monitoring database. There is no upper limit to how much monitoring data is sufficient—more data is always beneficial. A larger pool of monitoring data provides more opportunities for comparison, improves the efficiency of model validation, and allows for better understanding of pollution patterns over time.

In addition to validation, monitoring data (including satellite data) is used to nudge boundary and initial conditions (Read B) and in the reanalysis of historical databases, such as global reanalysis fields, which combine results from various models and monitoring sources. Reanalysis fields are highly valuable for examining long-term trends in pollution levels alongside the energy consumption patterns of a city or country. They provide insight into whether pollution levels are rising in tandem with increased energy consumption, indicating that more energy use is leading to higher emissions. Alternatively, they can help assess if pollution levels are decreasing despite growing energy demand, which could suggest that the country is successfully implementing emission-reduction programs at the source, such as cleaner energy initiatives, improved industrial practices, or stricter environmental regulations.

Air quality monitoring data can be collected through various means, with some methods being more preferred than others.

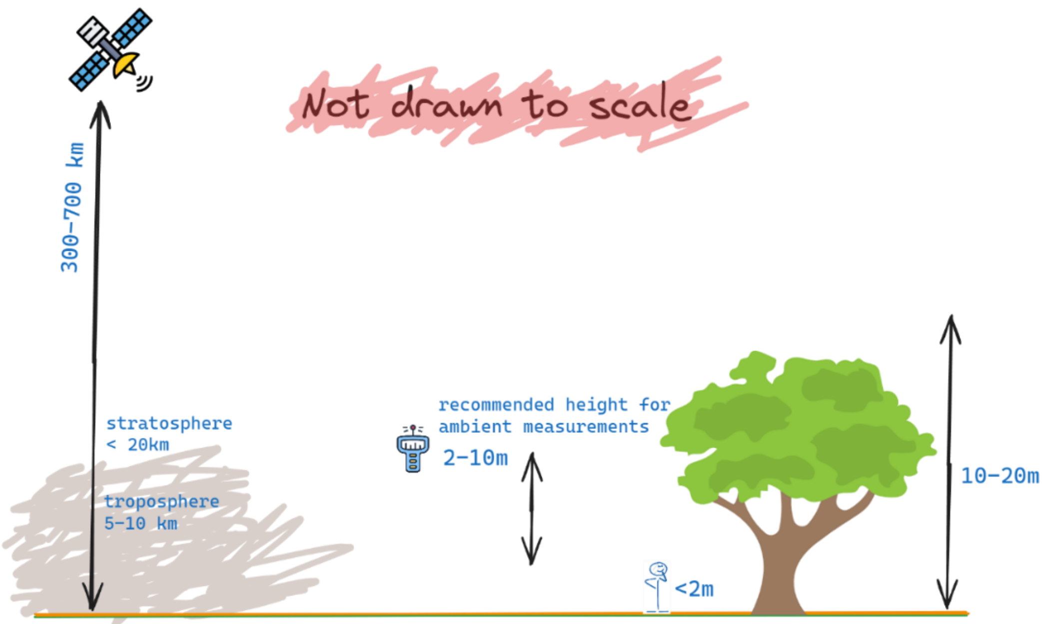
Reference-grade monitoring (RGM): Also known as regulatory-grade monitoring, this method follows strict protocols regarding the type of equipment used, operational guidelines for quality assurance, data processing for quality control, data storage, and dissemination. For all official purposes, RGMs are preferred due to their adherence to standardized procedures across the world. Because these systems operate continuously, high-resolution data can be used to study short-term pollution episodes and long-term air quality trends and the data from a network of these stations can be used for scenario evaluations (Read P). Since RGMs measure a wide array of pollutants, including meteorological parameters, this contributes to their high capital and operational costs. In addition to providing information on the amount of pollution, the continuous $PM_{2.5}$ filters from these stations can be used for chemical analysis and source apportionment (Read S).



Low-cost sensor monitoring: RGMs are expensive due to the technical complexity, personnel requirements, and strict protocols for quality assurance and control (QA/QC – Read Q) they must follow. Low-cost sensors represent an emerging technology in air quality monitoring, gaining popularity among various research groups and organizations due to their affordability and ease of deployment. As more groups adopt these sensors, their use is expanding, providing valuable data in areas where RGMs are limited or unavailable. Some governments are beginning to recognize the potential of low-cost sensors and are starting to include them in official air quality monitoring networks, supplementing the existing infrastructure.

Additionally, these sensors have played a significant role in **encouraging citizen science**, allowing the public to actively engage in environmental monitoring and helping bridge knowledge gaps by making air quality data more accessible and widespread. This democratization of data collection empowers communities to take an active role in understanding and addressing local air quality issues. While low-cost sensors offer the potential to fill the spatial and temporal gaps associated with RGMs, they require regular calibration to ensure replicability and usability. Although these sensors are more affordable to deploy, the ongoing maintenance costs are often comparable to those of RGMs.

Satellite retrievals represent an emerging big-data technique in air quality monitoring, offering the potential to provide large-scale, high-resolution atmospheric data across regions. However, effective methods must be developed to accurately interpret these satellite retrievals, as the data is often localized and influenced by specific atmospheric conditions, requiring careful calibration and validation against ground-based observations. Despite these challenges, the future potential for using satellite feeds as a primary source of air quality information is immense, with the possibility of these technologies surpassing traditional ground monitoring networks in terms of coverage, frequency, and accessibility.



The new generation satellite feeds can cover the regions with limited or no monitoring for the key pollutants, such as aerosols (as aerosol optical depth, AOD), SO_2, NO_2, CO, Ozone, HCHO, Ammonia, and CO_2. Here is a list of most referred feeds: (a) MODIS, OMI, MOPITT, CALYPSO, GHGsat (with geostationary setups over the continental US) (b) TROPOMI (with geostationary setups over continental Europe) and (c) GEMS (geostationary over Korea and Japan).

Over an operational period of 24 to 72 hours to 1-week, all these satellite feeds can provide a product covering the entire globe, which can be used as a proxy for understanding the spatial spread of these pollutants. It is important to keep note of the following when using the satellite retrievals:

a. Ground monitoring (using RGM's and low-cost sensors) is always preferred for validating the chemical transport models.
b. Geostationary satellite feeds (as columnar densities) are used for validating the chemical transport models, where data for a region is available for longer periods in a consistent format (barring cloud cover).
c. Satellite data must go through vertical interpretation (Read V) before it can be used for direct comparison with on-ground measurements. These interpretations depend on the availability of local emission inventories and use of chemical transport models.

There is a primer on "air quality monitoring" @ https://urbanemissions.info (in English and 10 other languages) which provides detailed information on

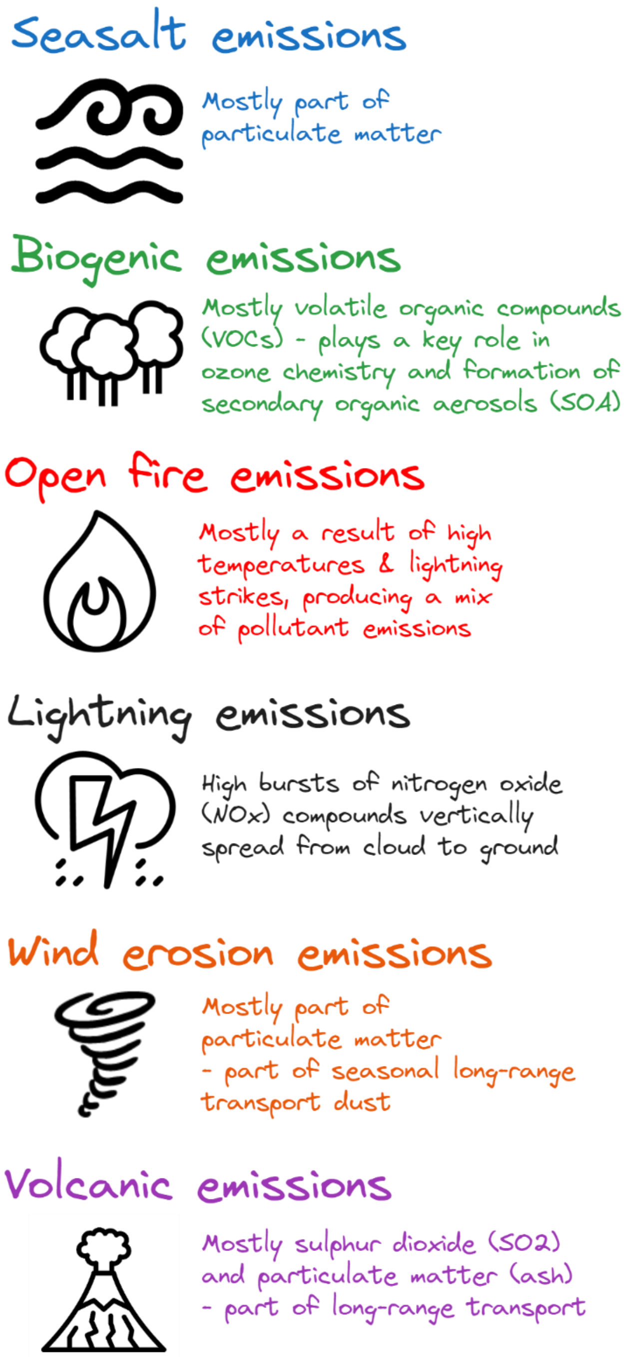
1. What is ambient air quality monitoring and what are the uses of the data?
2. What are the different types of monitoring?
3. What are the thumb rules for minimum number of locations?
4. What is the difference between emissions and pollution monitoring?

More monitoring data
= More understanding of the spatial & trends
= More data to validate the modelling results
= More confidence in the modelling results
= More usable baseline information for policy dialogues.

N

Natural Emission Inventories

Natural emission inventories cover all emissions that are not directly tied to fossil fuel combustion or human-made activities.



These emissions originate from various natural processes, such as wildfires, volcanic activity, sea salt spray from oceans, dust from arid regions, and the release of biogenic VOCs from vegetation. These natural emissions play a significant role in the overall air quality and atmospheric composition, contributing to the formation of pollutants like ozone and particulate matter.

While they occur independently of human activities, their impact on local and regional air quality can be substantial, especially in areas prone to events like wildfires or dust storms.

Good representation of natural emission inventories is critical for air quality modeling, as these sources can vary greatly depending on geographical, seasonal, and meteorological factors.

Except for the open-fire and volcanic emissions, other sources are calculated using the meteorological fields and pre-defined parametrization. All the established chemical transport models have pre-processors to estimate these emissions in forecast and hindcast mode.

Open fire emissions are linked to satellite observations of daily fire instances (for example: NASA FIRMS). These observations are converted to area burnt and respective emissions using pre-defined parametrization like landuse and its associated emission factor by season (for example: GFED and FINN).

Volcanic emissions are primarily estimated using historical observations and global inventories are regularly updated to document the frequency, intensity, and composition of emissions from volcanic eruptions. These inventories provide crucial data on gases like SO_2, CO_2, and PM, which play a significant role in atmospheric processes over their long-range transport. Volcanic emissions are not only important for understanding local pollution but also for their impact on climate change. Large eruptions can inject massive amounts of aerosols and SO_2 into the stratosphere, leading to global cooling by reflecting sunlight away from the Earth. Historical events, such as the eruption of Mount Pinatubo in 1991, are well-documented instances where volcanic activity temporarily cooled the planet by reducing solar radiation. These global inventories help researchers assess both the short-term environmental effects and the long-term implications for climate and atmospheric chemistry.

A popular biogenic emissions package is MEGAN (Model of Emissions of Gases and Aerosols from Nature), which is integrated into most global and regional chemical transport models. This inventory is available as pre-calculated global emission fields, provided as monthly and annual totals. Alternatively, it can be dynamically calculated in forecast or hindcast mode using localized meteorological fields, current leaf area index (often derived from MODIS satellite retrievals), and pre-defined emission factors. Key pollutants from this inventory include isoprene and terpenes, which play a significant role in atmospheric chemistry, particularly in the formation of ozone and secondary organic aerosols.

When performing chemical transport modeling with full atmospheric chemistry, we must include all the natural emissions alongside anthropogenic emissions. These emissions are especially important in regional air pollution modeling, where they can have a substantial impact on air quality, particularly in rural or non-urban areas. Excluding natural emissions can lead to incomplete or misleading results, especially in regional studies where their presence is felt the most (proportionally, compared to the cities).

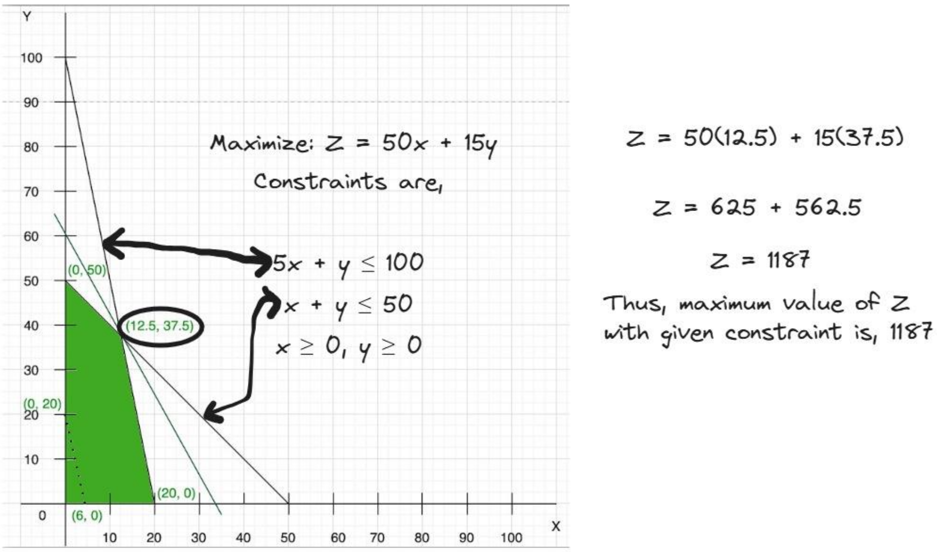
O

Optimization of Action Plans

Cities have numerous options to control emissions, including transitioning to cleaner fuel for road transport, promoting public transportation, promoting walking and cycling, enforcing vehicle maintenance, promoting the use of electric vehicles, increasing green spaces, transitioning to cleaner fuels for cooking and heating, stricter regulations on open waste burning, paving of roads to reduce dust loading, implementing industrial emission controls, and encouraging energy efficiency at industries.

When cities around the world are required to design an air quality management plan, final decisions are often based on the costs of interventions listed alongside potential benefits—whether health-related, social, or political. These decisions are frequently made without any or limited mathematical input, but this approach can be improved by incorporating quantitative methods, such as optimization and cost-benefit analysis, to make informed and effective choices.

Optimization is a mathematical process used to deduce the best decision among various options by adjusting variables to achieve the most favorable outcome, often targeting the "low-hanging fruit"—measures that are easy to implement and offer significant (and sometimes immediate) benefits with minimal cost or effort. Mathematically, in linear programming, we try to maximize (or minimize) an objective function with given several constraints, as illustrated in a simple case study below.



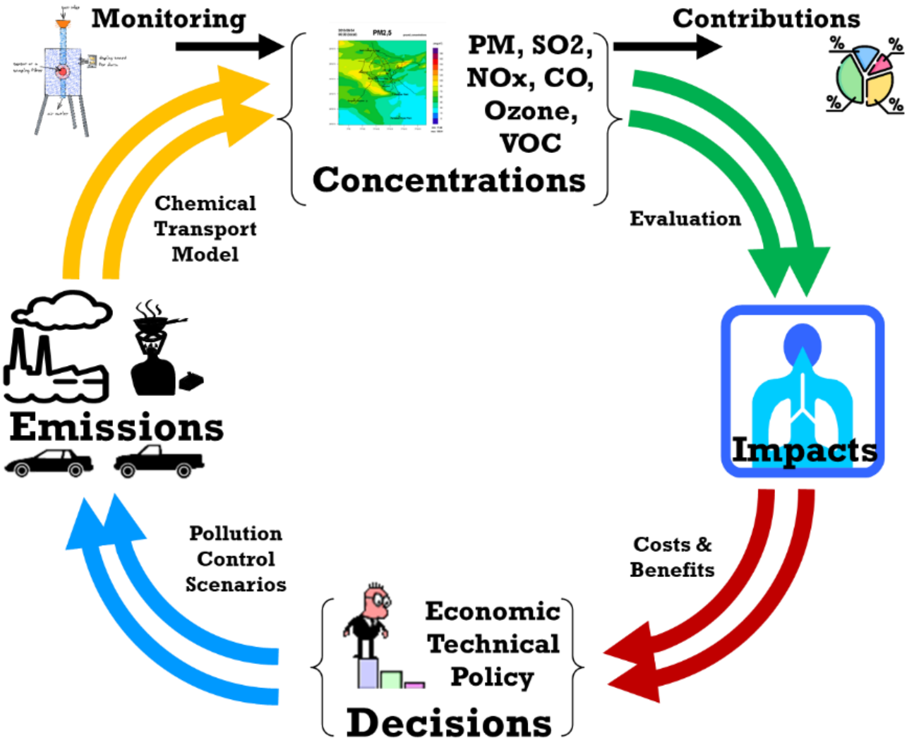
However, managing air pollution is not as simple as it may seem. While every city aims to reduce air pollution, each faces unique constraints that make one-size-fits-all policies ineffective. These constraints must be considered when developing action plans. For example, a city may target a 10% pollution reduction at minimum cost, but it may lack the infrastructure to support clean fuels or the capacity to install pollution control systems in all industries. These limitations define the boundaries within which action plans are developed to achieve the desired pollution reduction at the lowest cost.

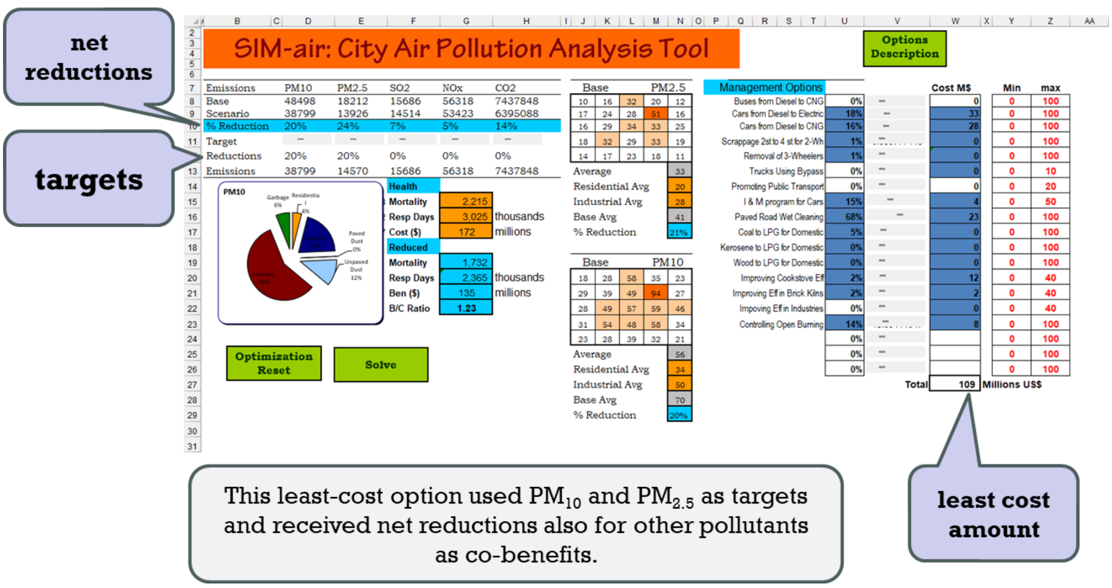
In addition to health-related concerns, we must also consider climate change-related pollutants when optimizing for an effective strategy. This is **the concept of co-benefits**: since the sources of both air pollution and climate change are often the same, addressing one problem benefits the other. While minimizing cost is the goal, the target for reducing either air pollution emissions or climate change emissions can depend on the source of funding. Ultimately, addressing air pollution benefits both public health and the climate.

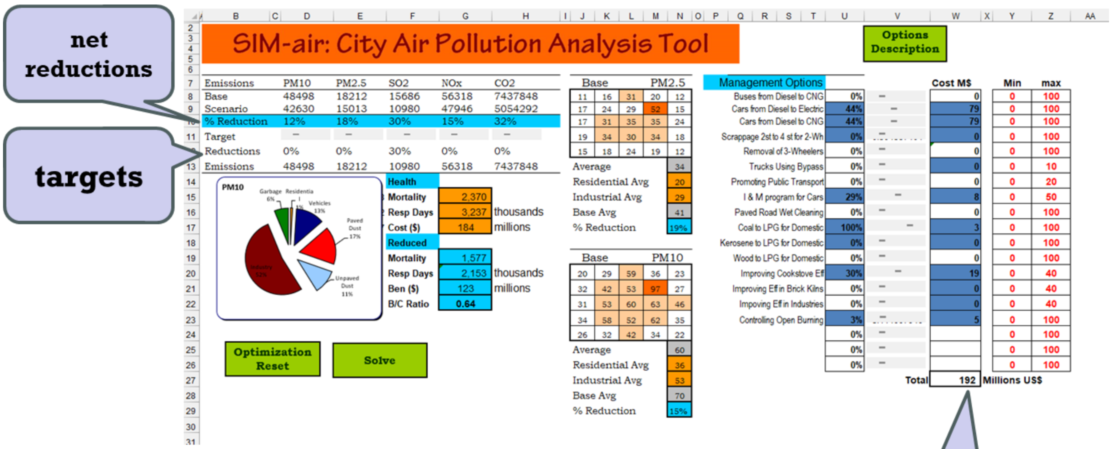
Examples of co-benefit strategies that address both air pollution and climate change include: (a) Shifting from coal and oil to renewable energy sources such as wind, solar, and hydropower reduces greenhouse gas emissions and cuts harmful air pollutants like SO_2 and PM (b) Replacing traditional gasoline and diesel vehicles with electric vehicles reduces CO_2 (when green electricity is used for charging) and eliminates tailpipe emissions of NO_x and PM on the roads (c) Retrofitting buildings with better insulation, efficient appliances, and lighting reduces energy demand, lowering both CO_2 and air pollution emissions from power plants (d) Planting trees and developing green areas in cities not only absorbs CO_2 but also improves air quality by filtering pollutants like PM and ozone.

The **SIM-air model** (as an example from https://urbanemissions.info) serves as a demo of an integrated air quality management system, allowing users to estimate emissions, translate them into pollution levels, and assess impacts for a given scenario, coupled with an optimization feature (a MS-Excel feature) to explore least-cost options for achieving a set of air quality targets in a co-benefits setting.

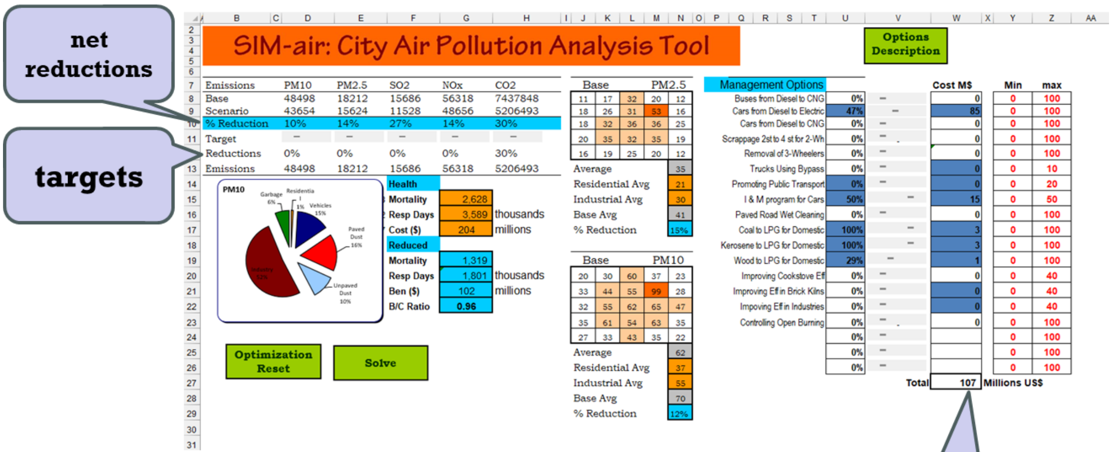
Microsoft Excel provides powerful tools for optimization through its Solver add-in, which allows users to solve linear and non-linear optimization problems. This application is demonstrated through a series of screenshots from the SIM-air model, which allows for multiple constraints—both on the emissions side and the intervention options side—adjusted independently or collectively. The screenshots represent random scenarios, all optimized for least cost. The cost functions used are based on historical literature and are intended for reference only. However, for city-specific applications, it is essential not only to customize the emission activity data and source-receptor matrix functions but also to tailor the cost functions to reflect the economic context of the local area.



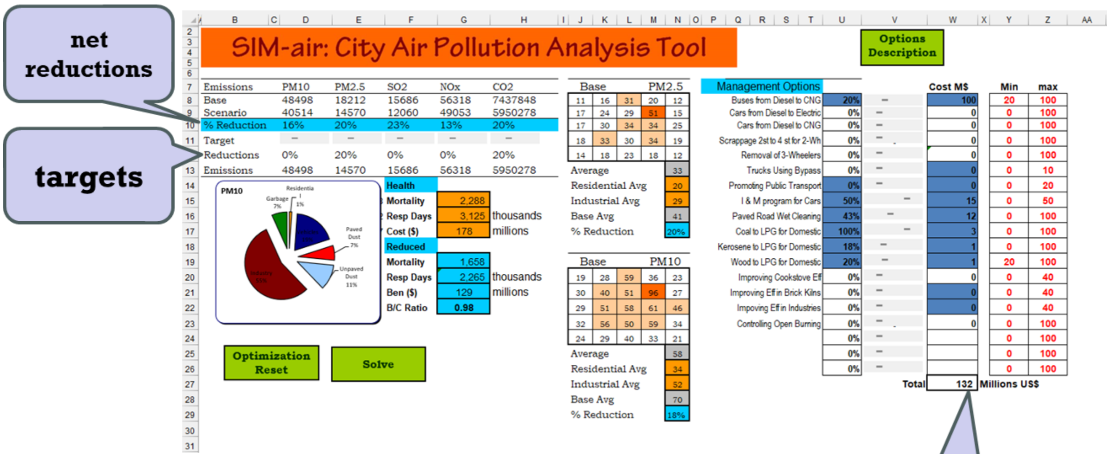




This least-cost option used only SO_2 as target and received net reductions also for other pollutants as co-benefits.



This least-cost option used only CO_2 as target and received net reductions also for other pollutants as co-benefits.



This least-cost option used only PM2.5 and CO_2 as target, with additional constraint on 2 options and also received co-benefits for other pollutants.

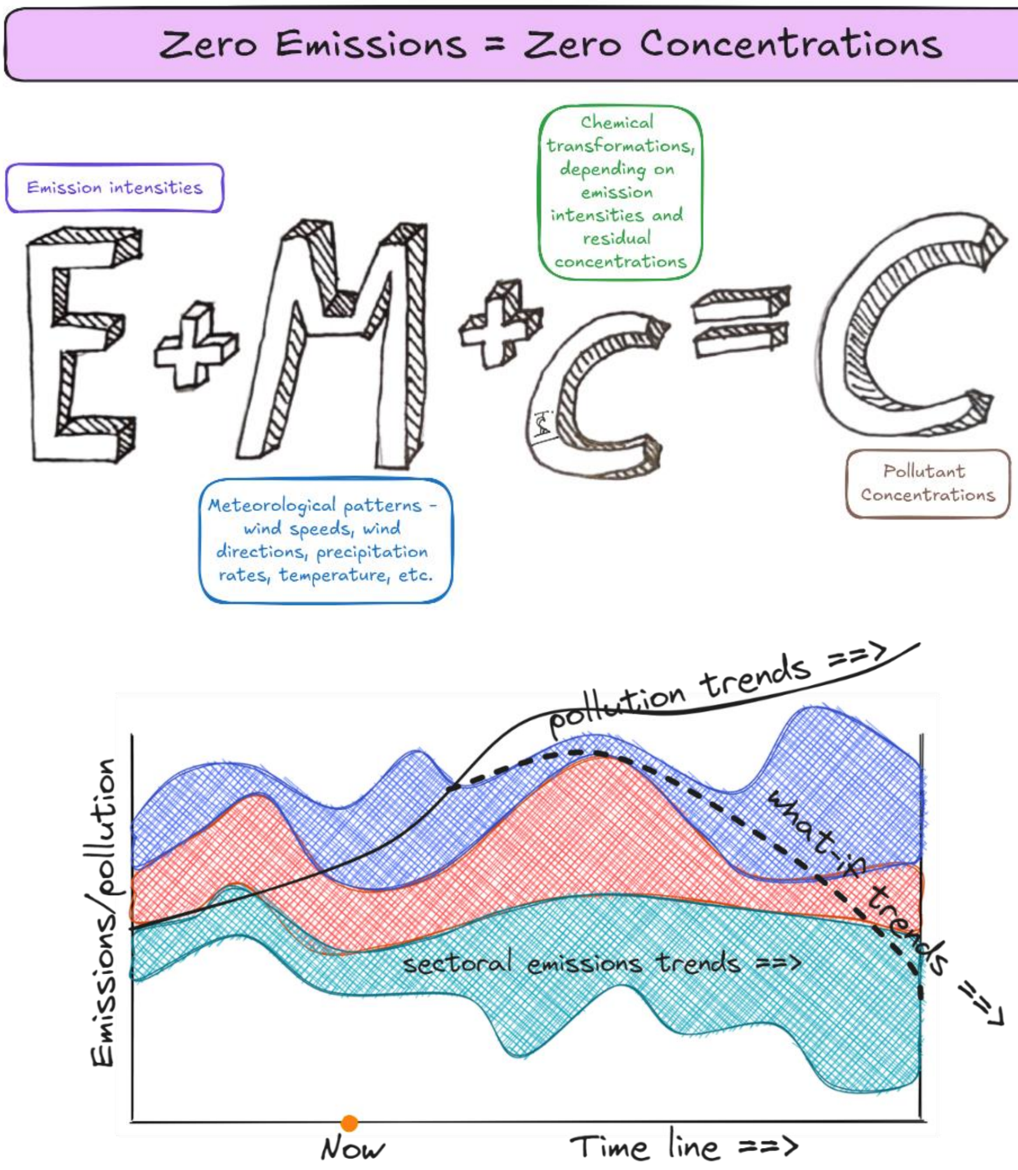
Using the Solver function in MS-Excel is just one way to approach optimization. There are also advanced models, such as linear programming and non-linear programming models, and stochastic optimization techniques, which can handle more complex and large-scale problems. Additionally, specialized software like MATLAB, GAMS, and Python-based libraries (e.g., SciPy, PuLP, and Pyomo) provide advanced tools for optimization, offering greater flexibility and precision in tackling diverse decision-making scenarios. These advanced models are often used for complex systems where multiple variables and constraints interact dynamically.

All these systems are data-intensive and require advance technical training.

P

Projections and Scenario Analysis

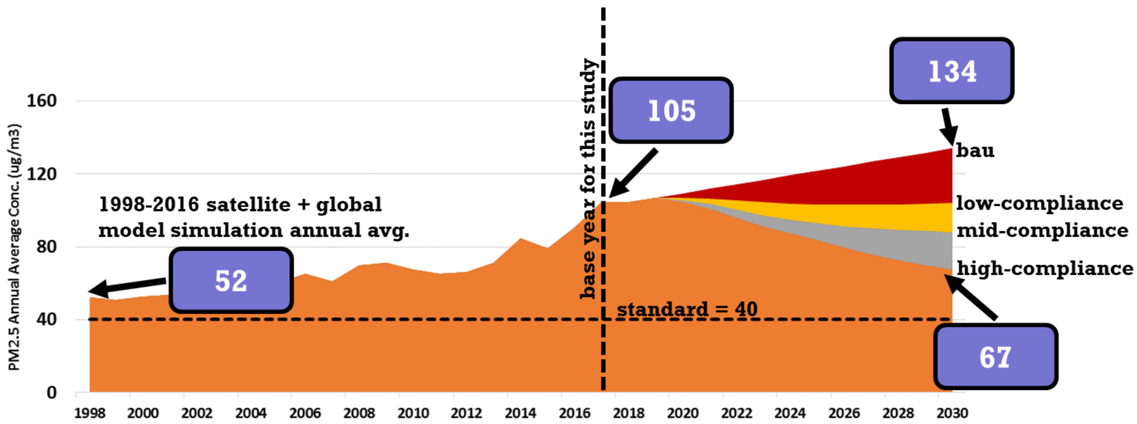
Here's a rendition of Einstein's famous equation of relativity, E = mc², applied to the story of air pollution management. The goal of any air quality management program is to reduce pollution levels (big C, for concentrations). We can't blame meteorology for our pollution problems—wind and rain come and go, moving pollution around, but not solving the root issue. On rainy days, we experience good air quality, while winter nights often bring bad air quality. The main driver of pollution is emissions (big E), and atmospheric chemistry (small C) which is simply a dependent factor, shaped by the emissions and pollution already present.



Best use case of air pollution modeling results is **scenario (what-if) analysis using projected emissions**. This analysis explores potential future outcomes by modeling different emission scenarios based on changes in known sectors, such as industry, transportation, or energy production. By adjusting variables like emission rates, fuel types, or adoption of new technology, chemical transport models simulate the resulting pollution levels under various "what-if" conditions.

What-if analysis allows researchers and policymakers to predict the impacts of potential policy decisions, such as the introduction of cleaner technologies or stricter regulations. These modeled outcomes are often combined with **cost-benefit analysis**, a method that weighs the economic costs of implementing the selected actions against the public health and environmental benefits they would generate. For example, a policy to reduce emissions might have upfront costs, but the long-term benefits of improved air quality, reduced healthcare expenses, and enhanced public health could outweigh those costs. This process is part of a broader **scenario analysis**, where multiple potential futures are simulated to help policymakers make informed decisions. These analyses provide valuable insights for guiding policy dialogues and planning effective air quality management strategies.

This analysis can also be conducted using ambient monitoring data alone when resources for developing emission inventories and projections are limited. In this case, the focus shifts to reducing the health impacts of current pollution levels by comparing the concentrations against national standards or World Health Organization (WHO) guidelines. The goal is to implement strategies that bring pollution levels in line with established health benchmarks.



While scenario analysis is typically conducted through a series of steps, there are MS-Excel-based example simulators (@ https://urbanemissions.info/tools) that utilize data from both monitoring and modeling studies to explore various air quality scenarios and their potential impacts.

A what-if analysis is exactly what the term suggests—it explores different possibilities, offering numerous ways (even thousands) to reach the same answer. This process allows us to map out all potential outcomes and pathways. However, it is important to remember that these simulations are just models, not definitive predictions. Nothing is set in stone, so it's crucial to explore and test the model as many times as possible to understand the full range of possibilities and refine decision-making.

For example:

1. **SIM-air model** serves as a demo of a reduced-complexity integrated air pollution analysis system, allowing users to estimate emissions, translate them into pollution levels, and assess impacts for a given scenario. It also includes an optimization feature to explore options for achieving air quality targets. Using a grid-to-grid source receptor matrix derived from a chemical transport model, the tool converts emissions into ambient pollutant concentrations. This reduced-complexity model (Read R) simplifies the process while still providing valuable insights for decision-making.
2. Simulator 1 - calculates population weighted concentrations based on zonal averages of an airshed and allows to build scenarios against % reduction in zonal pollution.
3. Simulator 2 - based on source apportionment results, allows to build scenarios against % reduction in each source.
4. Simulator 3 - solves for net zonal reductions required to reach an overall target using maximum possible reductions by zone as input.
5. Simulator 4 - solves for net zonal reductions by source required to reach an overall target using zonal source-apportionment results and maximum possible reductions by zone as input.
6. Simulator 5 - solves for net zonal reductions by source required at least cost using zonal source-apportionment results, maximum possible reductions by zone, and cost per unit pollution reduction by source as input.
7. Simulator 6 - demonstration of source-receptor matrix concept linking zonal concentrations to associated grids (15 x 15 grid airshed) and allows to build scenarios against % reduction in each zone.
8. VAPIS - Vehicular Air Pollution Information System (versions 1.01 and 2.1) – A vehicular emissions calculator to estimate and compare total vehicle exhaust emissions by vehicle-age and run scenarios.

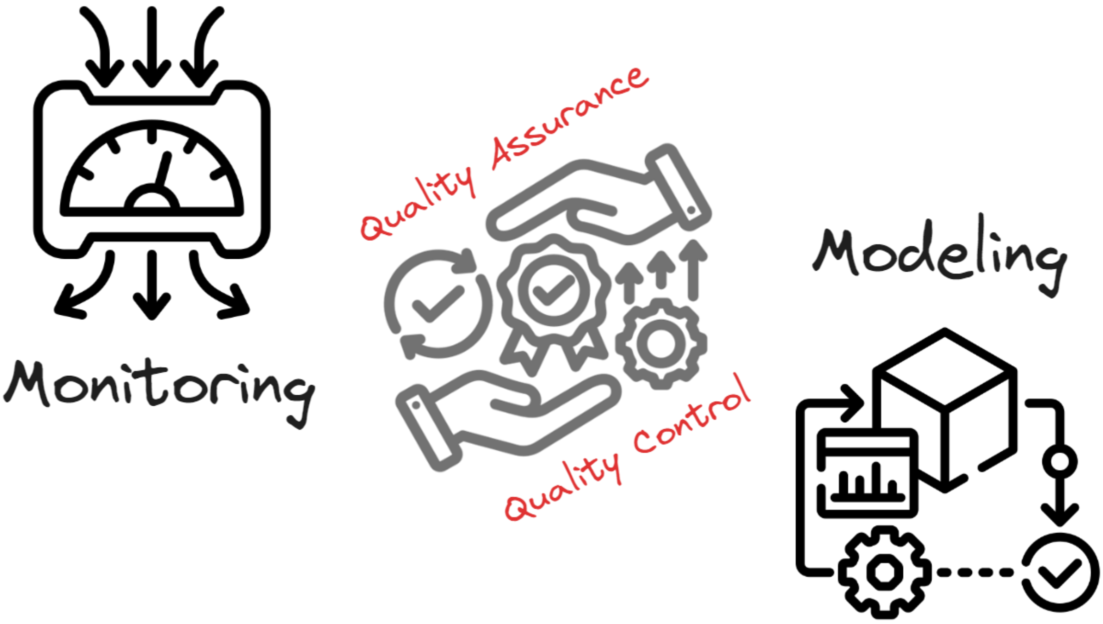
There is a line of thought that suggests we don't need extensive analysis to implement control strategies, as we already know what needs to be controlled. However, the value of what-if and cost-benefit analyses lies in their ability to identify the "low-hanging fruit"—the easiest and most cost-effective options for achieving faster reductions in emissions or concentrations of specific pollutants. These analyses provide insights into which strategies can yield the greatest benefits with the least effort and prioritize actions that lead to quicker results (Read O).

While the idea of skipping detailed analyses might seem appealing to avoid potential “analysis-paralysis”, having solid data and projections in hand can only lead to informed decision-making. This ensures that resources are allocated effectively and that the chosen strategies align with both environmental and economic goals.

Q

Quality Assurance and Quality Control (QA/QC)

Computational models help in decision making, tracking progress, and evaluating impacts to support public policy dialogues. To ensure that these models are useful, a quality assurance and quality control check must be performed.



This generally includes:

1. **Data Validation** (Read V): Modelled results must replicate spatial and temporal trends observed in the airshed, for the model to be used for projections and scenario analysis (Read P) with better confidence. Model validation is a key step of quality assurance for policy makers and researchers. When representative (and enough) monitoring data (Read M) to validate the air pollution models is not available, other aspects of quality assurance (discussed below) will provide us confidence.

2. **Fit for Purpose**: Every air pollution model (Read C) requires specific data inputs in specific formats to perform a set of tasks. These inputs range from an emissions inventory (Read E), meteorology (Read W), chemical mechanisms (Read JK) and model resolution (Read XYZ). Depending on the level of inputs available, a model fit to use them must be selected. The quality of the model results depends on the quality of input data (garbage in = garbage out). So, using dummy data or data with unreasonable assumptions to fill the model requirements will result in lower quality of model results.

3. **Crosschecking**: All the inputs to run a chemical transport model must be crosschecked at all the pre-processing stages, to maintain 100% data quality. Especially, when preparing an emissions inventory or downscaling an emissions inventory from a global inventory, or converting an existing emissions inventory into model ready format, a thorough data cleaning and crosschecks for consistency with units and total mass must be performed before using them as the model inputs.

4. **Consistency**: Where available, meta data for the inputs must be tracked to ensure consistency in the data quality. For example, if low-cost sensor data is used for model validation, then the data must be calibrated and validated against reference grade systems for consistency; during primary surveys, pre-defined protocols must be followed for collection, processing, storage, and usage by all the participants.

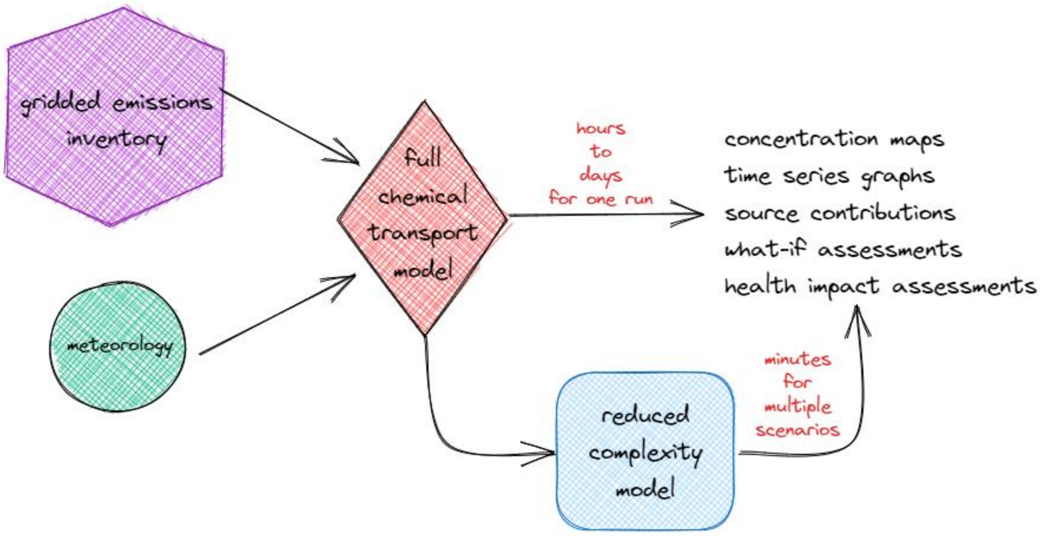
5. **Transparency and Accountability**: The inputs, assumptions, methodologies, data uncertainties, data processing techniques, and results of air pollution models must be transparent to the users and the decision makers alike. This increases model confidence levels and enables accountability among public bodies to make informed decisions and avoid the "model said this" syndrome.

6. **Engagement**: The model's performance must be continuously monitored and evaluated by developing pertinent post-processing metrics (Read U), which will build an environment for constant engagement between users (research bodies and public bodies).

R

Reduced Complexity Models

A common complaint for using established chemical transport models is their high demand for computational resources, both in terms of space and time. This limitation can be addressed using "reduced complexity models" (RCMs) which simulate marginal changes in pollutant concentrations relative to marginal in emission inventories. This approach is also known as a "source-receptor matrix," where the source represents a predefined area, a group of grids, or an emission sector (spanning all grids), and the receptor refers to all predefined areas or grids within the airshed. RCMs allow for quicker assessments of the impact of emission changes on pollutant concentrations, making them ideal for large-scale or real-time applications.

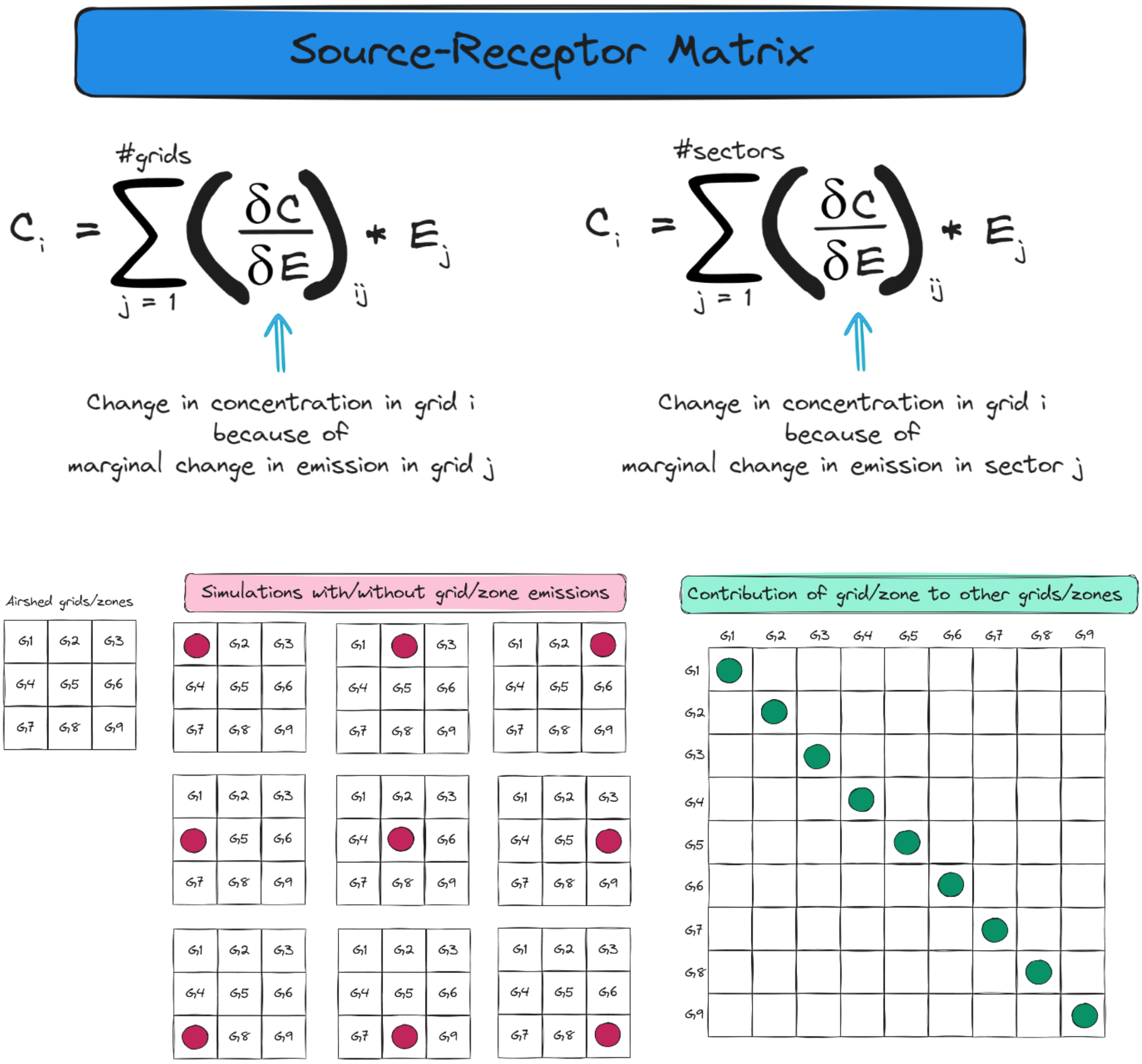


Any chemical transport model can be used to build the source-receptor matrix, which can then simulate multiple scenarios (Read P) at a fraction of the time required to run a full chemical transport model simulation for each scenario. However, there are uncertainties in this approximation approach that must be accounted for and clearly documented for discussion (Read Q). It is important to remember that this method does not entirely bypass traditional modeling. A representative emissions inventory and an operational chemical transport model are still required to build the matrix as a one-time effort.

The matrix can reduce model complexity by representing the airshed in various ways and capturing the corresponding contributions of different sources. The complexity of the matrix itself, however, depends on how the sources and receptors are defined. For example:

1. The matrix can represent the contribution of every designated grid in the airshed to every other grid in the airshed. If the airshed contains 3 rows and 3 columns, then at least 9 simulations are necessary, to build matrix of 9 x 9, to represent the contribution of every grid to every other grid.

2. This can be further simplified, to represent the zones instead of the grids. If the airshed consists of 9 zones, then the matrix can represent the contribution of one zone to every other zone.
3. A combination of the two is also possible, where the perturbation in the zonal emissions can be linked to marginal changes in the airshed grids. In this example, with 9 zones and 9 grids, the matrix size will be 9 x 9 but representing the change in concentrations in all the grids for a change in total emissions in a zone.



The simulations must include all the tracers necessary to substitute a full-scale chemical transport model. For example, when simulating $PM_{2.5}$, separate matrices must be constructed to account for primary $PM_{2.5}$ concentrations (based on the primary $PM_{2.5}$ emissions) and secondary SO_4 (based on the primary SO_2 emissions), NO_3 (based on the primary NOx emissions), and SOA (based on the primary VOC emissions). In short, reduced complexity model for $PM_{2.5}$ will consist of at least 3 or 4 sub-components, which can be further broken down into source categories like a matrix for area sources and point sources, whose advection schemes are significantly different.

One example lagrangian chemical transport model (ATMoS) with instructions to build this matrix as grid to grid and some example simulators for grid/zone combinations and more are included @ https://urbanemissions.info/tools

1. **SIM-air model** serves as a demo of a reduced-complexity integrated air pollution analysis system, allowing users to estimate emissions, translate them into pollution levels, and assess impacts for a given scenario. It also includes an optimization feature to explore options for achieving air quality targets. Using a grid-to-grid source receptor matrix derived from a chemical transport model, the tool converts emissions into ambient pollutant concentrations. This reduced-complexity model (Read R) simplifies the process while still providing valuable insights for decision-making.
2. Simulator 6 - demonstration of source-receptor matrix concept linking zonal concentrations to associated grids (15 x 15 grid airshed) and allows to build scenarios against % reduction in each zone.

RCMs are most effective for rapid scenario analysis, but there are important limitations to consider:

a) These models rely on pre-built relationships between sources and receptors, which are based on a single set of meteorological conditions. This means that while they may be broadly applicable, they cannot account for inter-annual variations in weather patterns.
b) Each pre-built matrix is tailored to a specific city, region, or grid configuration. It cannot be generalized or easily transferred to a different geographical context.
c) These models assume a linear relationship between emissions and concentrations, making them less effective at simulating pollutants like ozone, which are formed through complex, non-linear chemical reactions.

If proper quality control is ensured and the limitations are carefully considered, RCMs can be an asset for both technical and non-technical pollution managers and practitioners, helping to build informed baseline assessments and contributing to public-policy dialogue with robust, defensible data.

S

Source Apportionment

For an effective air quality management plan, it is essential not only to know (1) the amount of pollution, (2) where the pollution occurs, and (3) when it happens, but also (4) the contributions of various sources to the air pollution problem. This includes **sources both within the airshed and beyond its administrative boundaries**, extending to the designated airshed. The process of identifying and quantifying these sources and their contributions is known as "source apportionment".

All cities, whether in developed or developing countries, are required to create a source contribution chart at some point to effectively address air quality challenges. For air quality management, a clear and reliable chart is the most valuable information that any pollution manager or practitioner needs to have.

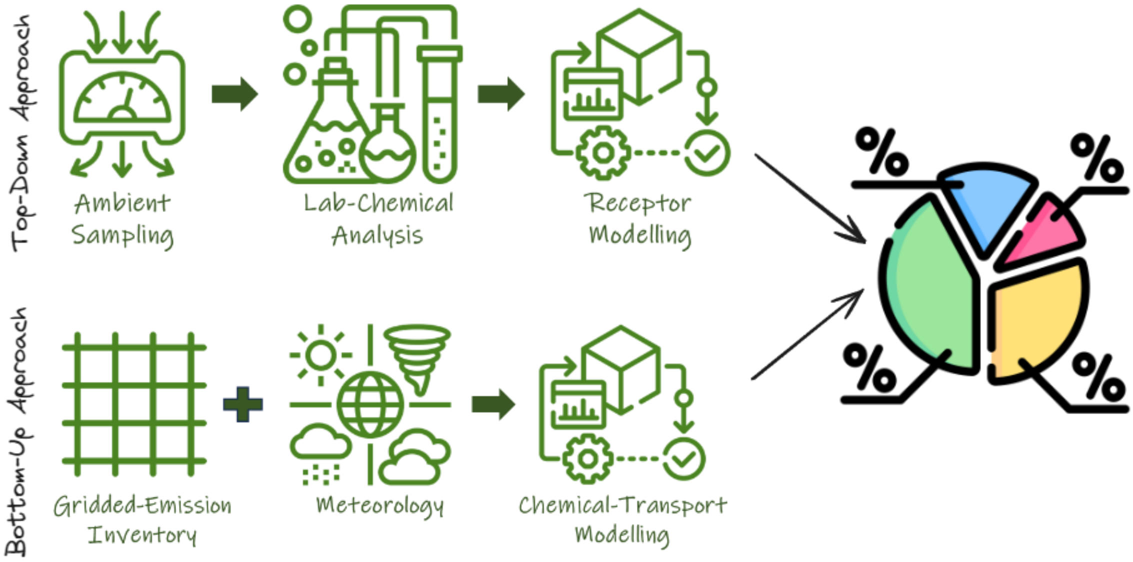
In policy discussions, the term "contribution" often refers specifically to pollution rather than emissions, making it important to distinguish between the two. Pollution refers to the concentrations of harmful substances in the air, while emissions refer to the release of those substances from sources. These terms are sometimes used interchangeably, along with references to energy sources, which can create confusion. Caution is needed, as the sources being discussed in the context of pollution, emissions, or energy may be the same, but the implications and solutions differ.

What's the difference?

Emissions	Concentrations
Definition: Amount of pollutant directly emitted at a source (like a vehicle tail pipe, industrial chimney, or a pile of openly burning garbage).	Definition: Amount of pollutant present in a unit volume of ambient air.
Typical unit: kg/day or kg/kg-fuel	Typical unit: $\mu g/m^3$ or ppm

When discussing the sources contributing to pollution, the geographical context is crucial. The contribution chart for a city will differ from that of the entire country. In cities, vehicle exhaust often dominates pollution issues, while at the national level, key sources may include rural energy use or power-plants. Best practices recommend conducting these studies more frequently and over a broader geographic area to address these differences.

Source apportionment studies are scientifically robust but challenging to conduct. They require skilled personnel for monitoring, data modeling, access to laboratory facilities, and substantial financial support. There are multiple institutions worldwide with the necessary resources and expertise that can assist in conducting these studies. However, resources to conduct these studies are limited in low- and middle-income countries (LMICs), though the global research community is providing support to help bridge this gap.



There are two popular ways:

The **top-down approach, also known as the "sampling way"**, involves collecting pollution samples from specific locations within a city. These locations are chosen to represent different areas and pollution sources (Read M). The results from these points are then chemically analyzed in a lab and processed through a receptor model to create a city-wide pollution profile. While this method provides detailed, location-specific data, it may not fully capture variations in pollution levels across the entire city. Nonetheless, it offers valuable insights into pollution sources and concentrations at specific locations, which can then be extrapolated to inform city-level air quality management strategies.

The **bottom-up approach, also known as the "emissions way"**, involves estimating air pollution levels and source contributions by using a detailed multi-pollutant and multi-sectoral emissions inventory for the city's designated airshed and processed through a chemical transport model. This approach provides spatially granular and comprehensive results of pollution across the entire airshed rather than at specific sampling points. It accounts for complex interactions between pollutants and meteorological conditions. While this method relies heavily on the completeness of the emissions inventory, it offers a more systematic view of pollution across the city and contribution of sources within and outside the airshed.

The methodological details, requirements, and pros-n-cons are discussed in detail in a primer "what is source apportionment" @ https://urbanemissions.info

While both approaches are expected to provide similar results, there are procedural differences that make them complementary. The top-down approach can chemically quantify contributions (e.g., by fuel type), offering insights that can be used to mass balance the contributions estimated using emissions inventories and chemical transport modeling. Conversely, the bottom-up approach can identify pollution hotspots and key representative locations within the airshed, guiding where sampling should be conducted for a better understanding of the sources and their chemical composition. By combining both methods, a more comprehensive picture of air quality can be achieved**. For a more complete assessment, it is always recommended that a city conducts source apportionment using both approaches**.

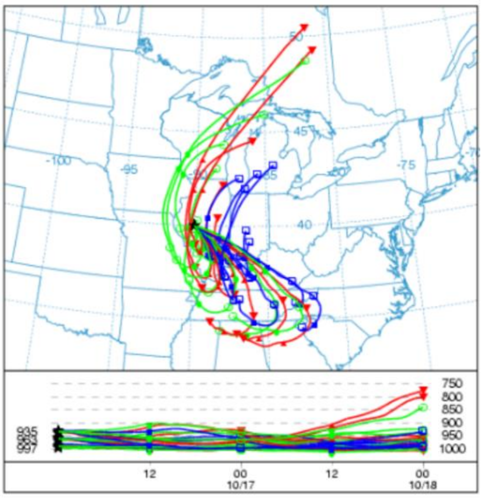
Despite the known uncertainties and differences between the methods, when conducted to its fullest extent—with comprehensive spatial coverage across the entire airshed and temporal coverage spanning all seasons (preferably over multiple years)—the bottom-up approach (emissions-based) provides more policy-relevant results compared to the top-down approach, often at a fraction of the cost. **Among the two approaches, the emissions-based method is generally more advantageous for informed decision-making, offering the flexibility to repeat the analysis as frequently as every month if needed**.

Source apportionment can be conducted for a wide range of pollutants, such as quantifying the contributions of sources to SO_2, NO_2, or ozone pollution. However, most studies around the world, particularly in low- and middle-income countries, focus on quantifying the contributions to $PM_{2.5}$ and PM_{10} pollution. This focus on PM is largely because it is the most harmful pollutant to human health, with strong links to respiratory and cardiovascular diseases and includes contributions from other gases via chemical transformation (Read H).

T

Trajectory Analysis

Most chemical transport models can replicate ambient pollutant concentrations both quantitatively (by providing absolute values and source contributions) and qualitatively (by showing trends and ranges). These models account for the physical and chemical transformations that emissions undergo in the atmosphere. However, in certain situations, we are primarily interested in understanding where an emission parcel is moving to or identifying the specific area source of a pollution parcel, such as the air one is breathing. This is trajectory analysis - **tracking the movement of particles, gases, or tracers through a meteorological field** to determine the likely regional sources of air pollution. These models are largely qualitative in nature, offering insights into the movement of pollutants, though they may include limited quantitative assessments, by incorporating a decay rate (for example, for SO_2).



Forward trajectory example:
Blue lines are carrying pollution to the layers closer to the surface and the red lines are carrying the pollution into the top layers.

Backward trajectory example:
Areas/grids covered by the blue lines are carrying pollution to the receptor grid and the red lines are originating in the top (likely clean) layers. Even though, all the lines are converging at the receptor grid, blue lines are more important.

In air quality management, trajectory analysis is conducted for two main purposes:

1. **Forward Trajectory Analysis**: Assessing the likely contribution of emissions from a source or grid to neighboring zones or grids. This involves tracking the movement of pollutants as they disperse from their origin, helping to predict how emissions affect downwind areas.
2. **Backward Trajectory Analysis**: Determining the likely origins of pollution observed at a specific grid or zone. This method traces the path of air parcels backward in time to identify potential upwind sources contributing to the observed pollution levels.

Both methods are valuable for identifying sources and source regions, playing a critical role in the overall air quality management of an airshed. The model can be run over extended periods to generate heatmaps: forward trajectories create heatmaps of receptor regions for a source grid, while backward trajectories generate heatmaps of source regions for a receptor grid.

In principle, a tracer parcel is released, and its movement is tracked through wind fields in three dimensions (X, Y, and Z). Since these trajectories are three-dimensional, it is crucial to account for whether the parcel is within the surface layer—where it can affect ground-level measurements—or if it is merely passing above the receptor or source grid, without influencing ground-level concentrations.

Common trajectory models in use are:

1. HYSPLIT
2. FLEXPART
3. ATMoS/UrBAT

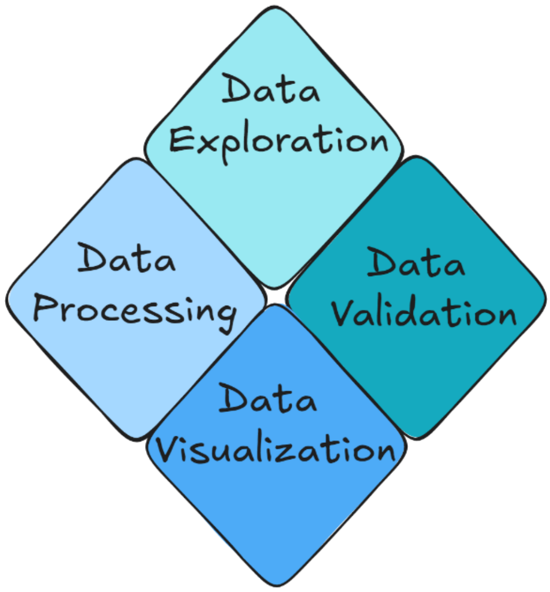
Another benefit of this methodology is its application in creating source-receptor matrices for running reduced complexity models (Read R). The qualitative contributions of tracers, with limited decay rates, can be used to generate this matrix, which represents marginal contributions—showing the change in pollutant concentrations (delta-change) in response to changes in emissions (delta-change). This matrix simplifies the analysis by linking emissions to air quality impacts, facilitating efficient scenario testing (Read P).

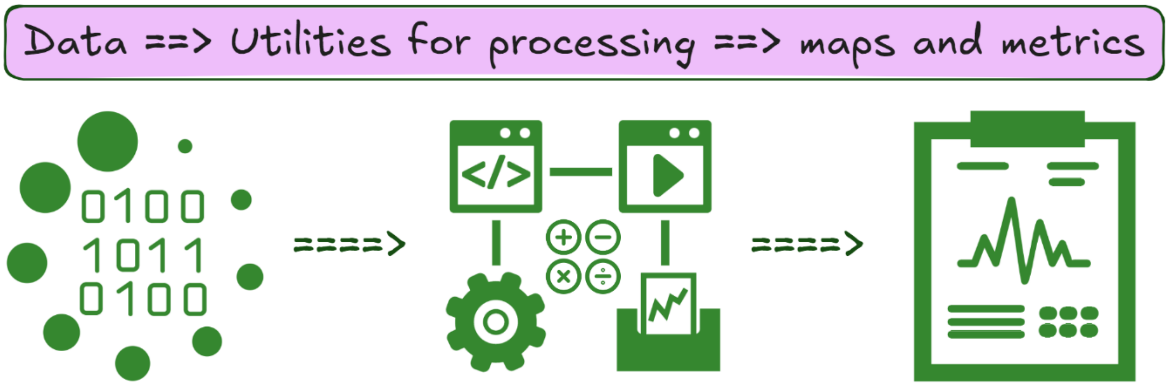
Use this approach only as a general guidance to identify potential sources and source regions. While trajectory models can provide useful insights into pollutant movement and regional contributions, they are not precise enough to serve as a standalone tool for detailed source apportionment. Caution should be exercised, particularly in urban environments, where local emissions can play a significant role. Relying solely on trajectory analysis, especially with low-resolution meteorological data, may lead to misleading conclusions, such as overestimating the role of distant sources while underrepresenting the impact of local emissions. Therefore, trajectory models should be complemented with other approaches, such as emissions inventories or chemical transport models, to ensure a more comprehensive and accurate understanding of pollution sources.

U

Utilities

Data is everything in air pollution modeling, serving as the foundation for understanding pollution sources, impacts, and solutions. However, as the saying goes, "garbage in, garbage out," meaning we must be cautious about the quality of the data we use. Air pollution modeling requires vast amounts of data, but not all of it is useful all the time. The challenge is to explore and identify the meaningful databases that provide accurate insights. Often, the data we collect doesn't come in the format we need, so it requires processing, validation, and transformation into a usable format. Additionally, since not everyone can easily interpret raw data, it's important to find ways to communicate this information effectively, whether through maps, graphs, or concise summaries. In short, while access to data is invaluable, there is significant work involved in preparing and presenting it, and for this, we rely on various utilities and tools to make the data speak clearly.





Data processing is a crucial step in air pollution modeling, involving several important tasks. Cleaning ensures that inaccuracies, inconsistencies, or missing values are corrected for reliable analysis. Transformations convert raw data into usable formats, such as normalizing or aggregating variables. Standardizations make sure that data from different sources follow the same units or scales, allowing for consistent comparisons. Geo-spatial operations process location-based data, enabling the analysis of spatial patterns by mapping coordinates or overlaying geographical layers. Data imputation addresses missing or incomplete data, using statistical methods to fill gaps and ensure the dataset remains viable for analysis.

Data validation is a key part of ensuring data quality. Instead of manually checking data, it's more efficient to use automated tools like Great Expectations, a Python library. Data engineers can set clear rules for the data, such as the type, range, or values expected. For example, you can set a rule that pollutant concentration values should not be negative. If any negative values appear, the system will alert you. This helps maintain data quality without needing to check everything manually.

Data exploration, also known as exploratory data analysis (EDA), involves summarizing data and gathering basic insights to guide further analysis. It helps to identify patterns, trends, and relationships within the data, setting the stage for more advanced questions. In tools like MS Excel or Google Sheets, formulas can be used to calculate important metrics, while Python's Pandas library offers functions like *df.describe()* to inspect datasets quickly. Visual tools such as histograms, scatterplots, and boxplots are essential for understanding the distribution and relationships between data variables. These features are available across most software platforms, making data exploration an accessible and crucial first step in data analysis.

Most simple visualizations, such as bar charts, line charts, and pie charts, can be easily created in Microsoft Excel. For public sharing, online tools like Datawrapper offer accessible options. For more advanced visualizations, Python and R scripting are more efficient, with Python offering various packages like matplotlib, seaborn, and plotly to create complex and geospatial visualizations. Geospatial visualizations can also be created using GIS software like QGIS and ArcGIS. To make visualizations more interactive, Python libraries such as Streamlit and Dash can be used.

Best tutorial library for GDAL and QGIS is https://spatialthoughts.com

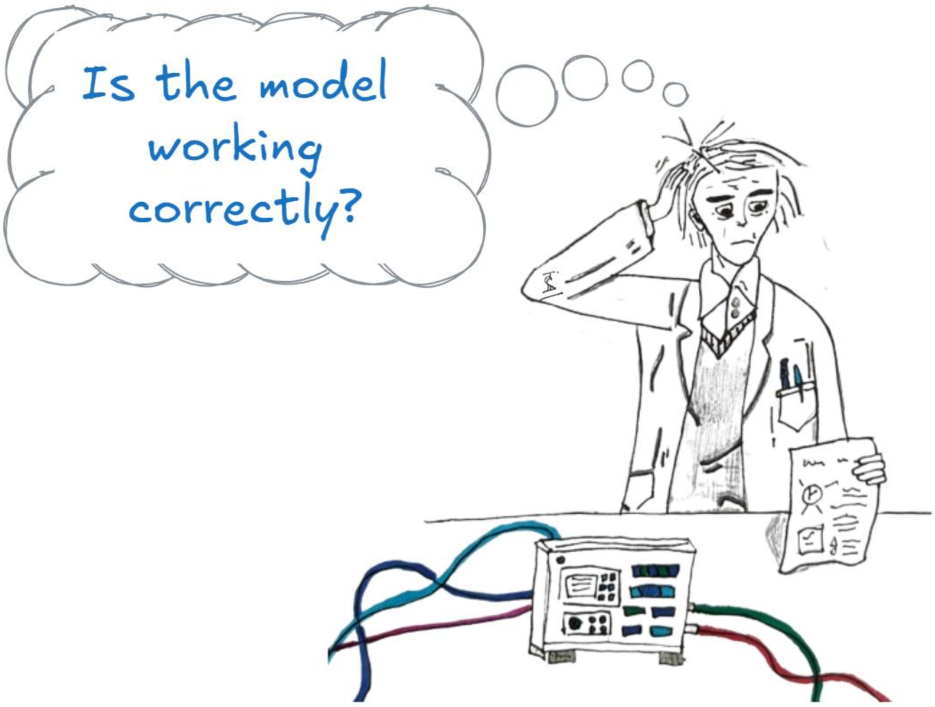
Best tutorial library for Google Earth Engine is https://github.com/opengeos

We have also created a repository of visualization tools and examples that can be made using Python @ https://github.com/urbanemissions-info

V

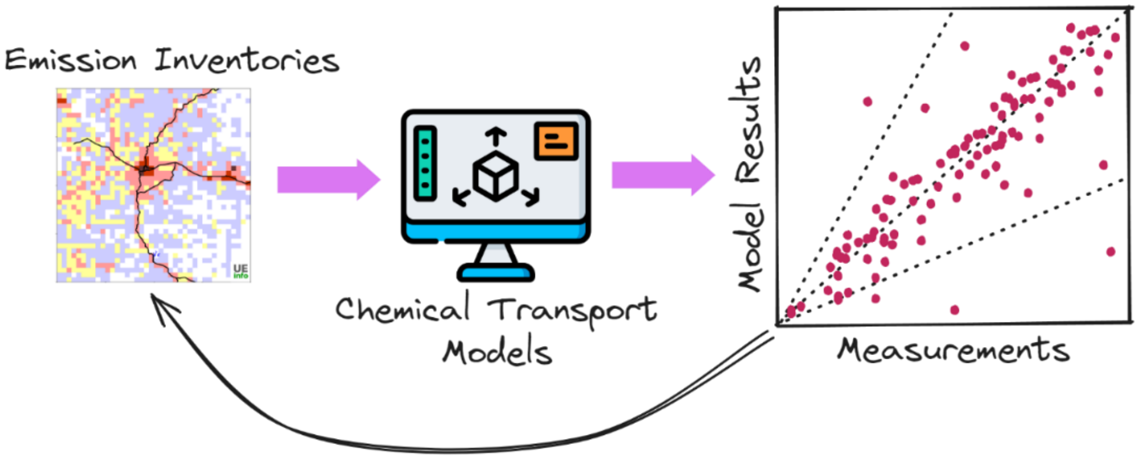
Validation of Model Results

Why do we need models? Simply put, it is impossible to monitor everything, everywhere, all the time. Models fill this gap by providing estimates where only limited or no monitoring is available or feasible. Meteorological and chemical transport models generate outputs that help us understand air quality and pollutant behavior over urban, regional, and global scales at various levels of complexities (Read C and W). However, since these **outputs are estimates, they must be validated to ensure reliability**. This validation process is essential before using estimated inputs, such as emission inventories, **for quality assurance of the models** (Read Q), enhancing their credibility for scenario analysis (Read P), policy formulation, and pollution-health alerts (Read F).



The physical and chemical processes in an urban or regional environment are highly dynamic, making model validation an ongoing process. As new data becomes available, whether from monitoring networks or updated emission estimates, continuous validation and refinement of the model are essential to ensure its accuracy and relevance.

A good fit between modeled results and actual measurements increases our confidence in the estimated emission intensities of various sectors and their spatial and temporal distribution (Read E). Conversely, a poor fit serves as a valuable lesson, encouraging a review of the methodologies used and offering an opportunity to improve the inputs and assumptions, for better modeling in the future.

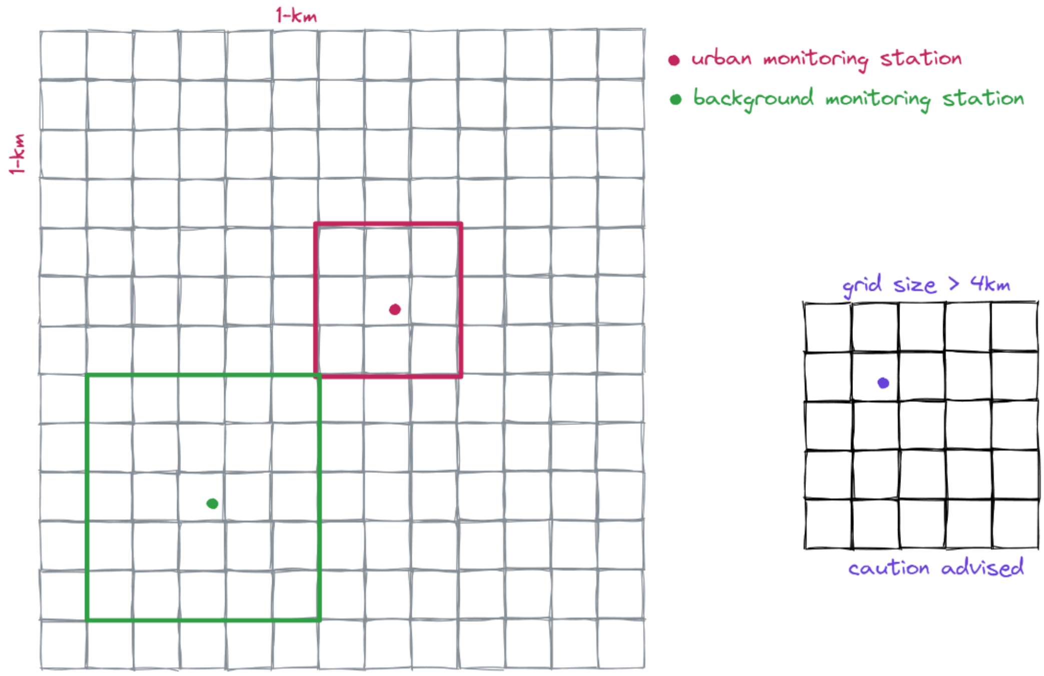


In air pollution modeling, key components that need immediate validation are:

1. **Meteorological parameters**: Validating data such as wind speed, wind direction, temperature profiles, and precipitation rates is critical. These parameters play a central role in chemical transport models by influencing the movement of pollutants (advection) and the chemical transformation processes. For example, predicting more precipitation in the model will result in increased wet scavenging of pollutants (Read D), which can lead to underestimating their final concentrations. Similarly, predicting stronger winds will disperse pollution to farther regions, potentially shifting the hotspots within the airshed. Additionally, variations in temperature will alter chemical reaction rates and atmospheric chemistry (Read JK).
2. **Pollutant concentrations at surface level**: It is crucial to validate pollutant concentrations in the surface layer of the model. This validation assesses how well the emission inventories fed into the chemical transport model represent real-world conditions both spatially and temporally. Comparisons should not be limited to annual averages; they must also be conducted for seasonal, monthly, and diurnal patterns to capture the temporal variability of pollution. This again ensures that the model accurately simulates fluctuations in meteorological data. Additionally, validating data at both monitoring sites and zonal averages builds confidence in the model's spatial allocation algorithms, ensuring that emissions are properly distributed across the airshed. This multi-level comparison, along with a good fit between model outputs and observed data, enhances the reliability of the activity data and emission factors used to build the inventory.
3. **Columnar densities across the airshed**: Validation of model's columnar densities (the vertical integration of pollutants) across the entire airshed with satellite-retrieved data, allows for a broader assessment of pollution levels, especially in regions where ground-level monitoring stations may be sparse or unavailable. Satellite data provides a valuable reference point for verifying that the model's spatial distribution of pollutants aligns with real-world observations across large areas and how the model deals with the vertical distribution of the pollutants.

Bridging the differences between model resolutions and representativeness of the monitoring locations

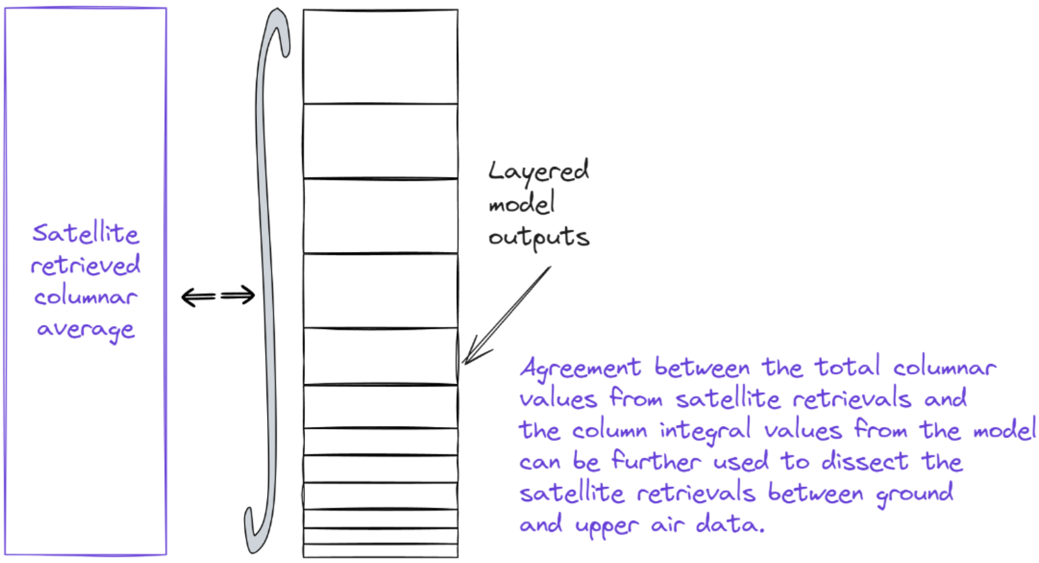
For pollutant concentrations, a general rule of thumb suggests that a reference-grade monitoring station can represent emission activities within a 2-km radius. Therefore, for a model running at a 1-km grid resolution, it is recommended to take the average of a 3 x 3 grid area (a total of 9 grids), with the grid containing the monitoring station at the center. A direct comparison with the single grid containing the monitoring station is not advised due to significant uncertainties in emission inventory estimates and their spatial allocations. For background monitoring stations, where the variability in emission intensities and source contributions is less pronounced, an average of 5 x 5 grids can be used to provide a more reliable representation of the overall area



When the model grid size exceeds 4 km, such as in regional models with grid sizes of 10 or 25 km, caution is advised. It is unrealistic to expect a perfect match between measurements and the modeled grid averages due to the larger scale of the grid, which may not capture local variations in emissions and meteorological conditions accurately. For example, in an urban grid where vehicle exhaust emissions are concentrated along roads, using resolutions of 4 km, 10 km, or larger leads to these emissions being averaged across the entire grid. As a result, the modeled concentrations are also averaged, potentially smoothing out localized pollution hotspots and underrepresenting areas with higher emission densities, such as busy roads. This doesn't mean the model is inaccurate, but caution is advised when expecting a direct match between localized monitoring data and grid-averaged concentrations at these resolutions. The model captures broader trends, while localized variations, such as traffic hotspots, may not align perfectly with the averaged grid outputs.

Bridging the differences between model resolutions and representativeness of the satellite observations

When comparing model outputs with satellite retrieval data, the comparison is typically made using the column-integrated values, which represent the total concentration of pollutants across all vertical layers (z-layers – Read XYZ) in the atmosphere. This approach provides a more holistic view of atmospheric pollution as seen from space.



It is important to note that there is no direct comparison between ground station measurements and satellite retrievals, as they capture different aspects of air quality. Ground stations measure surface-level concentrations, while satellites measure the entire atmospheric column. Therefore, these two data sources complement each other but cannot be directly compared without considering their distinct perspectives on pollutant distribution.

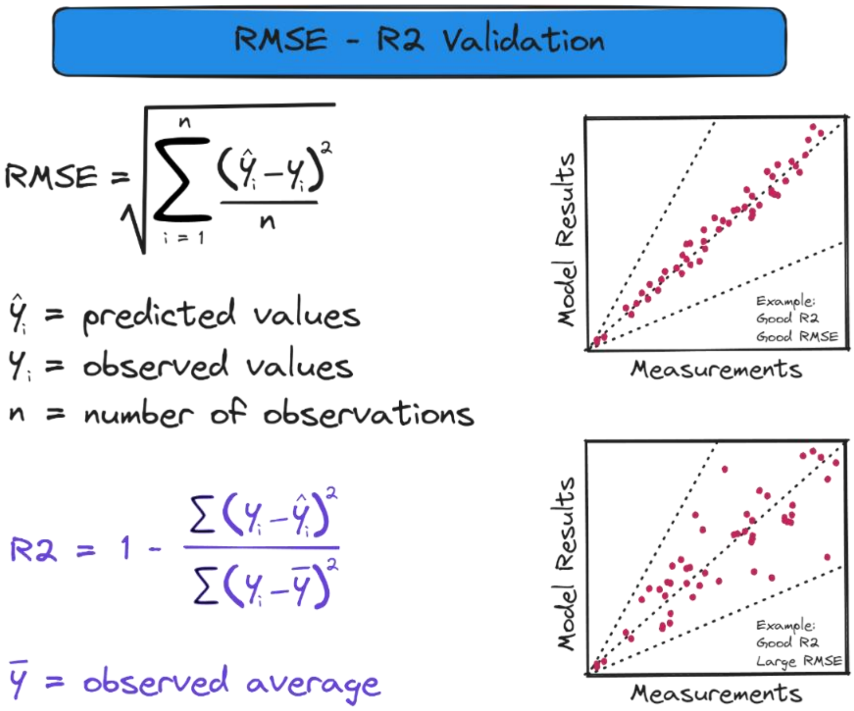
Bridging temporal differences

For long-term modeling results, most comparisons are made between daily, monthly, seasonal, or sometimes even annual average concentrations. For short-term simulations, such as 1- or 3-day forecasts (Read F), validating the model's diurnal profiles is critical for ensuring temporal accuracy. Diurnal profiles represent the variation in pollutant concentrations over a 24-hour period, reflecting the influence of daily activities such as traffic patterns, industrial operations, and changes in day-night meteorological conditions like temperature and wind speed. These profiles capture peak pollution events during rush hours or nighttime stagnation periods. Ensuring that the model can replicate these short-term fluctuations is essential for reliable air quality forecasting and to build confidence in its short-term predictive capabilities.

Statistical validation between model results and (any) measurements

Common metrics used to make this comparison are Root Mean Square Error (RMSE) and Coefficient of Determination (R-squared – R2).

1. RMSE measures the magnitude of error between model predictions and observed data. Low RMSE values indicate that the model closely reflects real-world conditions in the measurements, signifying good representation of pollutant dynamics. Conversely, high RMSE values suggest discrepancies, pointing to areas for model improvement. RMSE is especially useful for evaluating model performance over different time periods or spatial scales, providing insights into how well the model inputs are calibrated and validated.
2. R2 measures the goodness of fit between observed and model-predicted values. It is best visualized through a scatter plot of observed versus predicted values, with a regression line fitted through the points. The R2 value ranges from 0 to 1: an R2 of 0 indicates no correlation or pattern between the data, while an R2 of 1 represents a perfect fit, where the regression line passes exactly through both the model and observed values.



3. However, it is possible for the R-squared value to be close to 1 while the RMSE is large. In such cases, the model may not accurately represent the physical reality as measured by monitors, but it can still be useful after proper calibration. This highlights the importance of interpreting model validation metrics in the context of the model’s specific application. When such discrepancies arise, other methods like Mean Absolute Error (MAE - the average of absolute differences between model predictions and observations, offering insight into the model's overall accuracy) and Mean

Bias Error (MBE - the tendency of the model to overestimate or underestimate values. A bias close to zero indicates balanced performance) become valuable for providing a more comprehensive evaluation of model performance.

Sensitivity Analysis

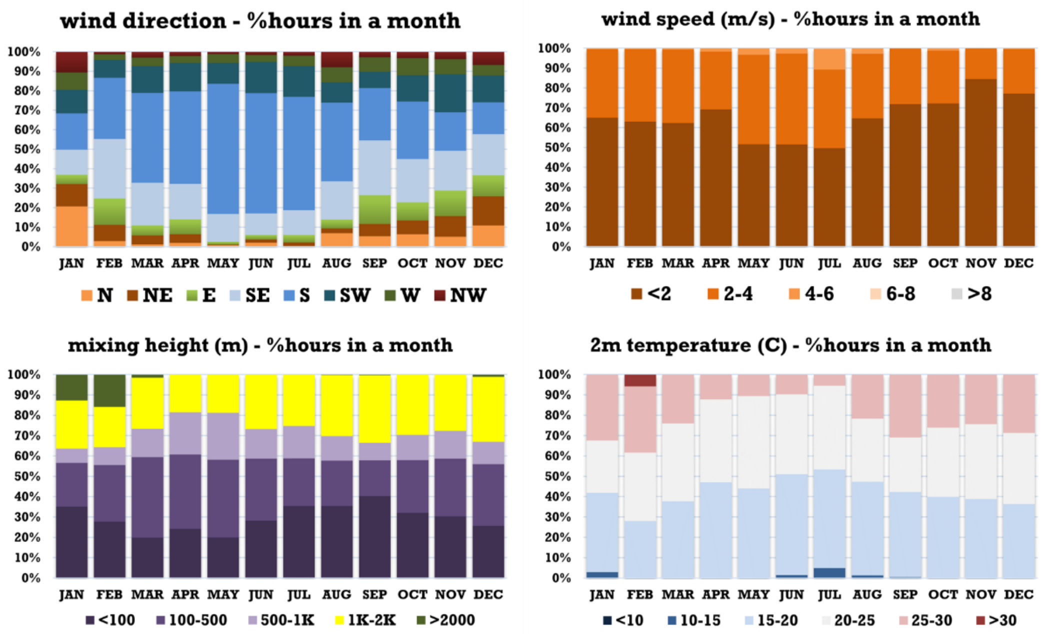
In addition to data-driven validation, model validation also involves sensitivity analysis to assess the relative influence of various input parameters, initial-boundary conditions, and alternative assumptions on the model's output. Sensitivity analysis helps identify which inputs have the greatest impact on the results. If a particular parameter is found to exert undue influence on the simulation output that does not align with observed reality, the input characteristics must be reevaluated and adjusted to improve model accuracy.

One common method for conducting sensitivity analysis is through Monte-Carlo simulations. These simulations involve running the model multiple times, each time with slightly varied input parameters, to observe how changes affect the output. This process provides a statistical range of possible outcomes and helps identify which inputs are most sensitive, offering valuable insights into where refinements are needed for more reliable predictions.

W

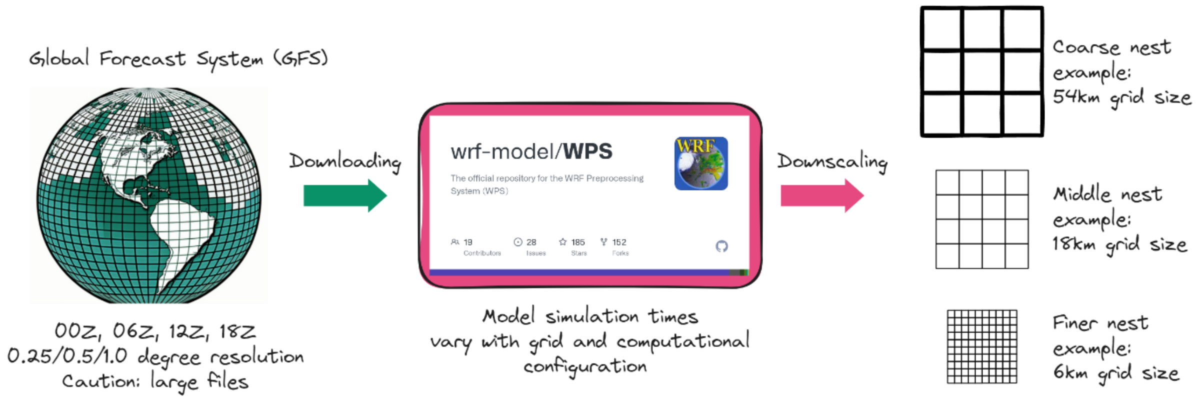
Weather Research Forecasting (WRF) Model

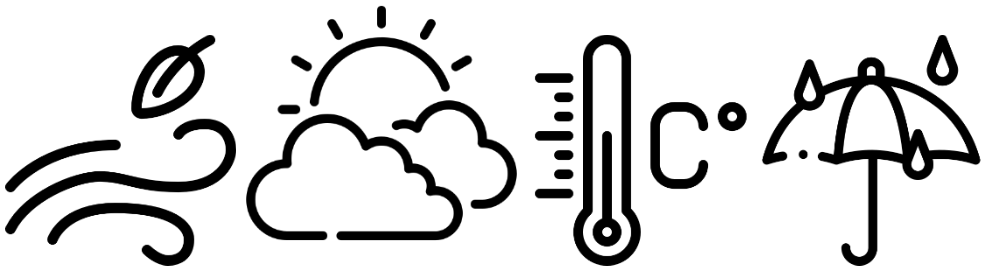
Even before analyzing monitoring or modeled pollution data, much can be inferred about a city's air pollution trends simply by examining its meteorological data, particularly its seasonal and diurnal patterns over the months. This "first guess" information can come from automated weather stations, reanalysis fields, or high-resolution meteorological modeling. Understanding these meteorological patterns can give pollution managers valuable insights into local air quality issues and serve as a strong foundation for more detailed analysis.



For example, the statistics presented in the figure allow for several important conclusions: (a) From April to September, the contribution of sources to the South (SE-S-SW) should be investigated more closely; (b) Sources to the North (NE-N-NW) are particularly important during January and December; (c) Surface temperatures remain pleasant throughout the year, indicating no significant demand for space heating from the residential or industrial sectors; (d) The percentage of hours with mixing heights below 100m is consistent across the months, mostly occurring during the nighttime (not shown in the figure), while a substantial share of hours also experience mixing heights above 1000m, suggesting good vertical mixing and, therefore, greater dispersion and lower pollution concentrations; and (e) Wind patterns remain consistent throughout the year, with no significant zonal stagnation issues, indicating steady air circulation and reduced potential for pollution buildup.

Meteorological modeling has advanced significantly over the past 10 to 20 years, with increased standardization of models and inputs. For generating the necessary meteorological fields for chemical transport models, the **Weather Research and Forecasting (WRF)** model is the most widely used system. WRF has a large user community, supported by forums that address various queries, from downloading and installing (compiling) the model to its operational use. Additionally, WRF includes a chemical module, known as WRF-Chem, which allows users to perform meteorological and chemical transport model calculations simultaneously, providing a comprehensive tool for air quality and atmospheric research.





The WRF system can be run in both hindcast and forecast modes, depending on the application. In **hindcast mode**, the model simulates past atmospheric conditions by using historical meteorological data, allowing for validation against observed data. In **forecast mode** (from 1 day to 10 days), the model predicts future atmospheric conditions based on real-time or near real-time inputs (using a combination of on-ground automated weather stations and satellite observations). To improve accuracy, the model uses various input fields, such as reanalysis data or observational data (like upper sondes), to "nudge" the calculations towards more realistic results. Most of the modules are internalized in the WRF system and can be activated on demand.

An open resource for historical meteorological data from multiple reanalysis fields is available @ https://psl.noaa.gov/data/atmoswrit/timeseries

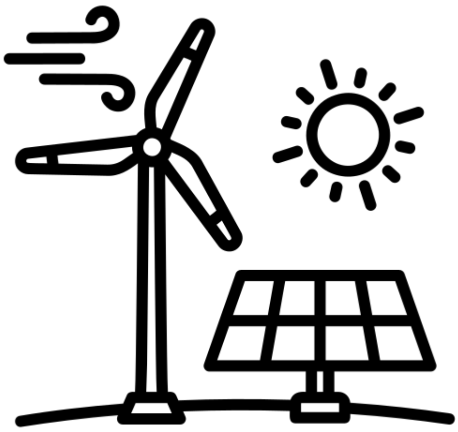
Additionally, WRF can be configured with different nesting settings (Read B and XYZ), which involve running the model at varying spatial resolutions. Finer resolution nests can be embedded within coarser grids to capture detailed local processes while maintaining computational efficiency at larger scales. This flexibility in configuration allows the WRF model to be tailored for a wide range of applications, from global/intercontinental to regional weather forecasting to detailed urban air quality studies.

The model also provides a variety of physics combinations, which should be studied carefully, before settling on a combination that works best for the region. For example: cu_physics (=1, KF), bl_pbl_physics (=0, 3DTKE with km_opt = 5, diff_opt=2), mp_physics (=16, WDM6), is a good combination for high-resolution (2.5 km or 3 km) nests over the Indian Subcontinent. This is based on a series of simulations over multiple urban areas. Benchmarking notes on the WRF model, input fields, and some notes on nesting, layering, and physics parameters from these case studies is documented @ https://urbanemissions.info/tools.

The WRF model is supported by well-populated user forums and extensive support structure combined with comprehensive documentation, making it highly accessible to users at all levels. The model is thoroughly documented in literature with applications across the world, and if adequate computational resources are available, running the model should be straightforward by following the instructions in the manual.

The chemical transport models like CAMx and CMAQ and global models have integrated the WRF outputs and have established pre-processors to convert WRF outputs to their model required formats.

Beyond weather reports



The WRF model has applications that extend far beyond simply supporting daily weather reports or analyzing historical weather events. Its flexible architecture makes it a powerful tool for other practical purposes. One such application is in the energy sector, where high resolution simulations in the WRF model can be used to evaluate renewable energy supply metrics. For example, it can provide detailed insights into weather patterns that influence the potential for wind and solar energy installations. By simulating wind speeds, solar radiation, and other meteorological factors, WRF helps determine the best locations and times for renewable energy projects, thus supporting decision-making in energy planning, optimizing infrastructure, and assessing future energy production potential. This capability makes WRF an invaluable tool for integrating weather data into renewable energy strategies.

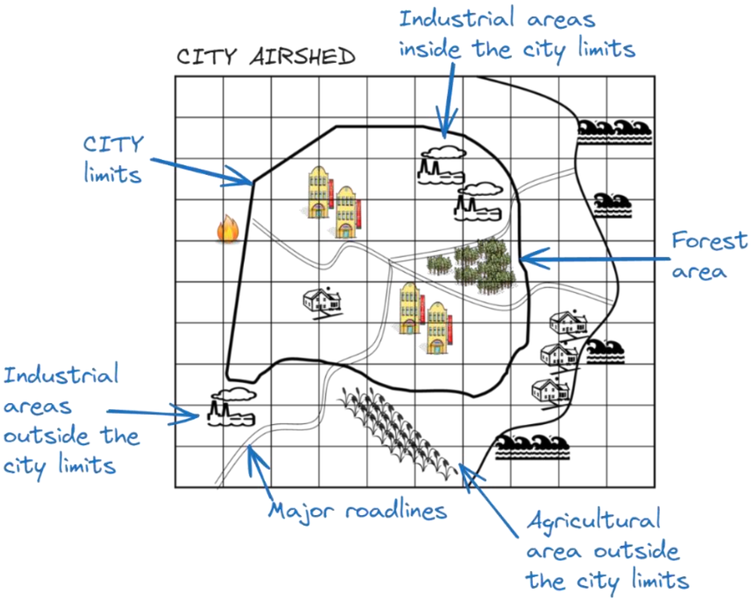
XYZ

XYZ-dimensions (airsheds, nesting, and model-layer heights)

A common phrase used for describing the modeling domain is an “airshed” - the length and breadth of an area to conduct emissions and pollution modeling. How do we define this airshed? How big or small should the airshed be, which can cover the needs of a city or a region in addressing the air pollution problem?

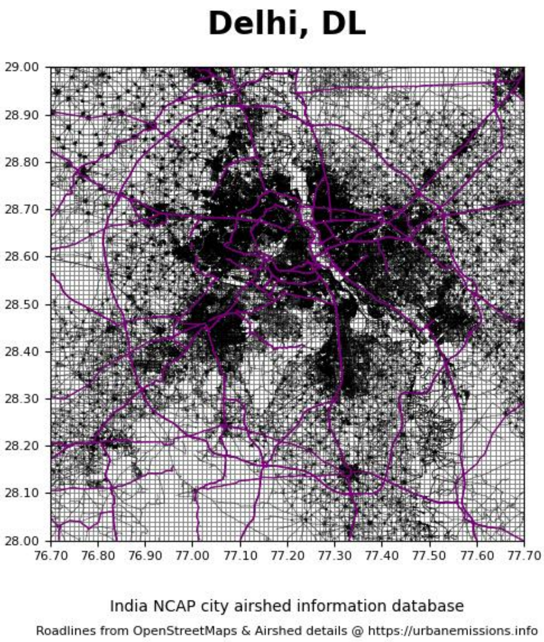
There is no set definition for an airshed size. Designating an airshed size (x-y dimensions) is a subjective assessment and here are some **thumb rules for setting an urban airshed**:

1. The airshed must include any of the surroundings that influence the city's air quality -- large settlements like satellite cities, large point sources like power plants, cement plants, brick kiln clusters, etc. A clear understanding of the city’s geography and the emission strengths (including inventory) of various sources inside and outside the city limits is necessary.

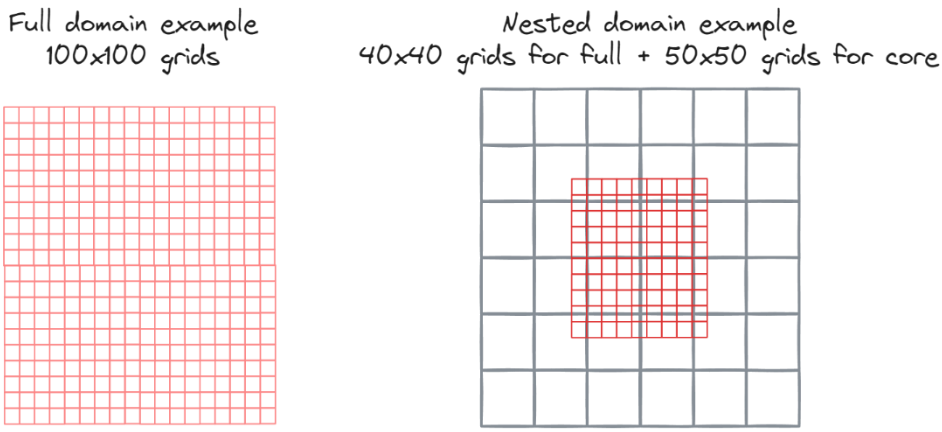


2. Keep the airshed size to a size mathematically solvable and a size manageable by the computational system. For urban-scale studies, typical grid-size is 1-km (approximately 0.01 degrees near the Equator) and typical airshed size is 30 x 30 grids for small cities to 80 x 80 grids for big cities, in north-south and east-west directions. It is a good practice to set rounded number of grids to support parallel processing and other server configurations.
3. In some cases, it may not be possible to compromise on the airshed size or grid resolution. For example, in Delhi, a 100 x 100 grid at 0.01° resolution defines one of the largest urban airsheds under India’s NCAP program. This

airshed not only covers Delhi but also its satellite cities and numerous point sources, such as brick kilns, many of which lie outside the city limits. Due to the densely interconnected geography of the region, a smaller airshed would likely result in an underestimation of local influences on air quality. For reference, the main city of Delhi itself covers only about 900 km², emphasizing the need for a larger airshed to capture the full scope of pollution sources.



4. If computational resources are limited, running meteorological and chemical transport models on a domain larger than 100x100 grids at 0.01° resolution can become challenging. To reduce the computational load, one option is to use a nested configuration, such as a 40x40 grid at 0.01° resolution nested within a 50x50 grid at 0.02° resolution. This reduces the total number of grids from 10,000 to 4,100, significantly easing the computational burden. In this setup, the finer grid draws boundary conditions from the coarser domain. While there will be some compromise in accuracy, this approach can still provide valuable results, though it is important to test the model thoroughly before finalizing the configuration.

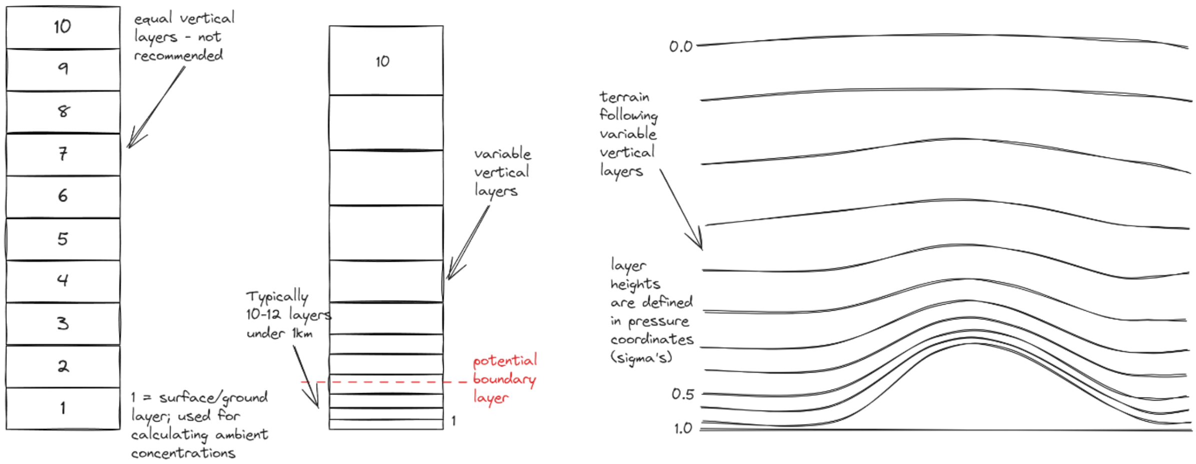


For **regional and sub-regional simulations**, a larger grid resolution of 0.1° to 0.25° (~10 to 25 km) is typically used. This choice is often determined by the availability of regional emission inventories. Since most global emission inventories are available at a 0.1° resolution, many regional modeling systems are run at this resolution. While this is sufficient to capture most urban signatures, if high-resolution local inventories are available, specific areas (airsheds) can be embedded as nests within the larger domain for higher-resolution modeling. This approach allows for more detailed analysis in key areas while maintaining computational efficiency for the broader region.

Thumb rules for assigning z-layers

The vertical configuration (z-dimension) of models is critically important, especially in regions with varying terrain elevations between grids. Proper vertical layering ensures that the model represents atmospheric processes at different altitudes and properly captures the air flow, pollutant dispersion, and chemical transformations.

Contrary to common belief, the vertical layers in these models are not equidistant. This uneven layering allows the model to focus higher resolution near the surface, where pollutant concentrations and meteorological changes are most impactful for air quality and human exposure and where intermixing of pollutants between layers is more important to study.



In the WRF model, vertical layers are often pressure-dependent (following sigma levels), meaning that the layers are spaced according to pressure levels rather than a fixed altitude. This design reflects the natural thinning of the atmosphere at higher altitudes, with finer resolution in the lower atmosphere (where weather and pollution processes are more active) and coarser resolution higher up. This pressure-based layering improves the model's ability to simulate realistic atmospheric conditions across different altitudes.

Additionally, WRF uses a terrain-following coordinate system, meaning the model's vertical layers adjust to follow the contours of the terrain. In regions with mountains or valleys, the vertical grid "bends" to align with the land elevation. This terrain-following feature enables the model to more accurately simulate the effects of topography on air flow and pollutant dispersion, capturing how terrain influences local weather patterns and pollution transport. A common question is **whether the model accounts for topography—this feature is, in fact, built-in and enabled by default in the WRF model**.

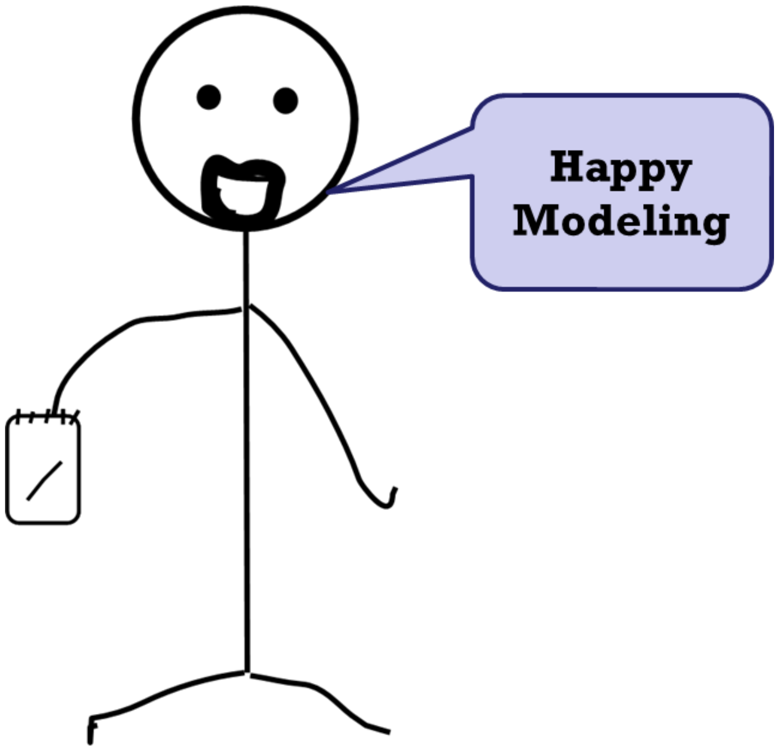
An example setup of pressure level sigmas with 41 layers (for use in the WRF namelist file and subsequently in chemical transport models) is as follows:

> 1.0000,0.9980,0.9955,0.9930,0.9900,0.9869,0.9833,0.9778,0.9705,0.9603,0.9472,0.9307,0.9110,0.8859,0.8559,0.8212,0.7825,0.7396,0.6939,0.6479,0.6043,0.5629,0.5213,0.4798,0.4381,0.3975,0.3583,0.3192,0.2809,0.2436,0.2080,0.1743,0.1430,0.1148,0.0903,0.0691,0.0508,0.0350,0.0215,0.0098,0.0000

This setup has 8 layers under 500 m and the first 12 layers under 1 km. For the airsheds with high mountains, it is necessary to increase the number lower layers under 1 km, for mathematical stability of the model and for reasonable vertical mixing during the application of the advection schemes in the chemical transport models.

Airshed definition

The size of an airshed, both horizontally and vertically, is often subjective and can vary depending on the specific study or region. While there are certain thumb rules that must be followed to ensure the mathematical and physical stability of the models, the final arrangement of grids and vertical layers is frequently determined by the user's computational capacity. However, it is crucial that a region is defined, models are set up, and initial calculations are made to begin the air pollution modeling process. This step is essential for gaining insights and improving air quality management strategies.



Send your comments to simair@urbanemissions.info

"Everything you need to know to apply for a job working at the crossroad of policy and research" -- Zbigniew Klimont, IIASA, Austria

"It covers so many key topics in modelling and explains them all really well. It is a great resource for my teaching on atmospheric modeling" -- Rebecca Garland, University of Pretoria, South Africa

"A great resource for those starting in air quality modeling, providing clear explanations of key atmospheric science concepts as well. Ideal for PhD students and early-career researchers looking to build their modeling skills. Highly recommended" -- Shahzad Gani, Indian Institute of Technology Delhi, India

"A truly priceless resource for scholars and practitioners working on urban air pollution across the world" -- Madhav Badami, McGill University, Canada

"Starts with a question to get readers focused. I will surely recommend to all my students" -- Jun Wang, University of Iowa, USA

www.urbanemissions.info