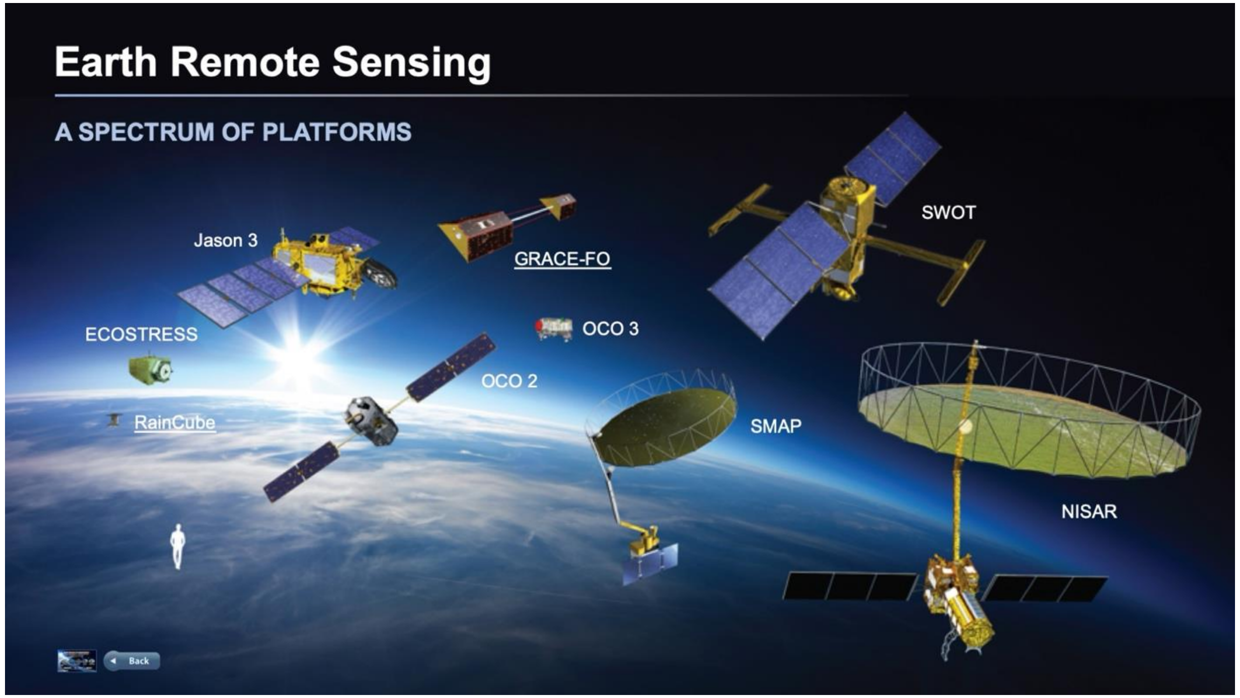
Analyzing and Wrangling Spatiotemporal Remote Sensing Data

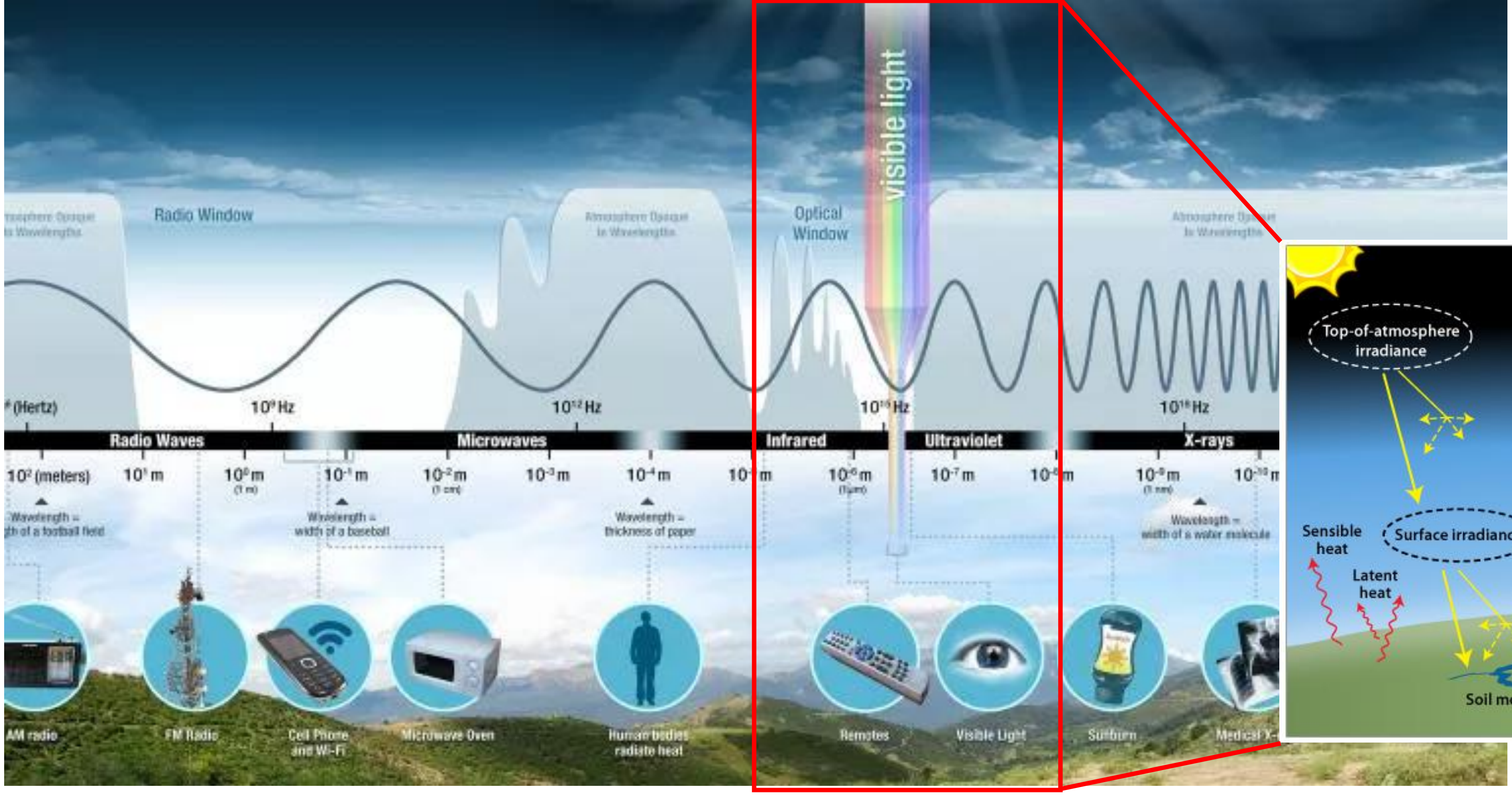
Meredith Franklin, University of Toronto
meredith.franklin@utoronto.ca

STATSTRO

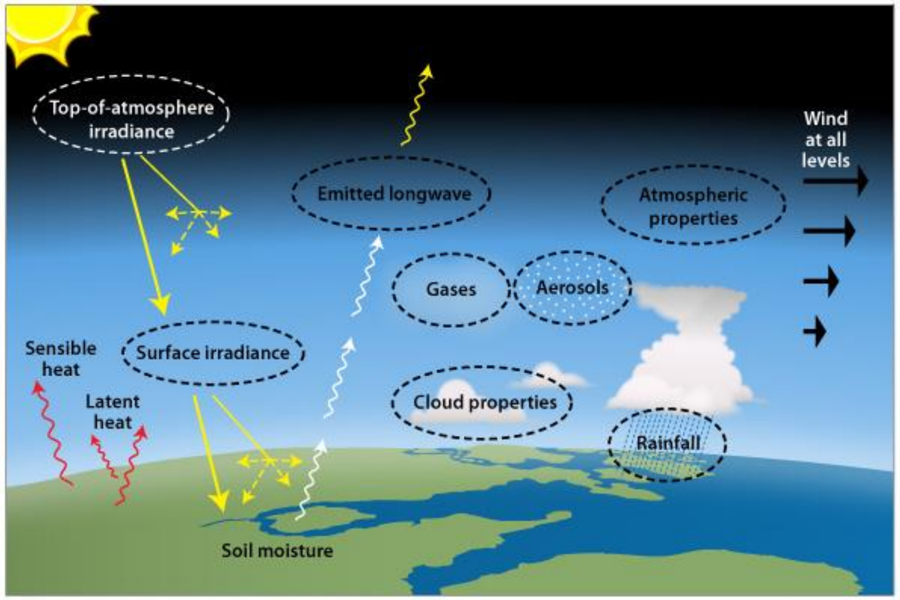
May 13, 2025

Earth Observing Satellites





Passive sensors rely on reflected sunlight or emitted thermal energy



Source: NOAA and Remote Sensing Research Center Australia

In environmental applications we primarily focus on instruments that observe:

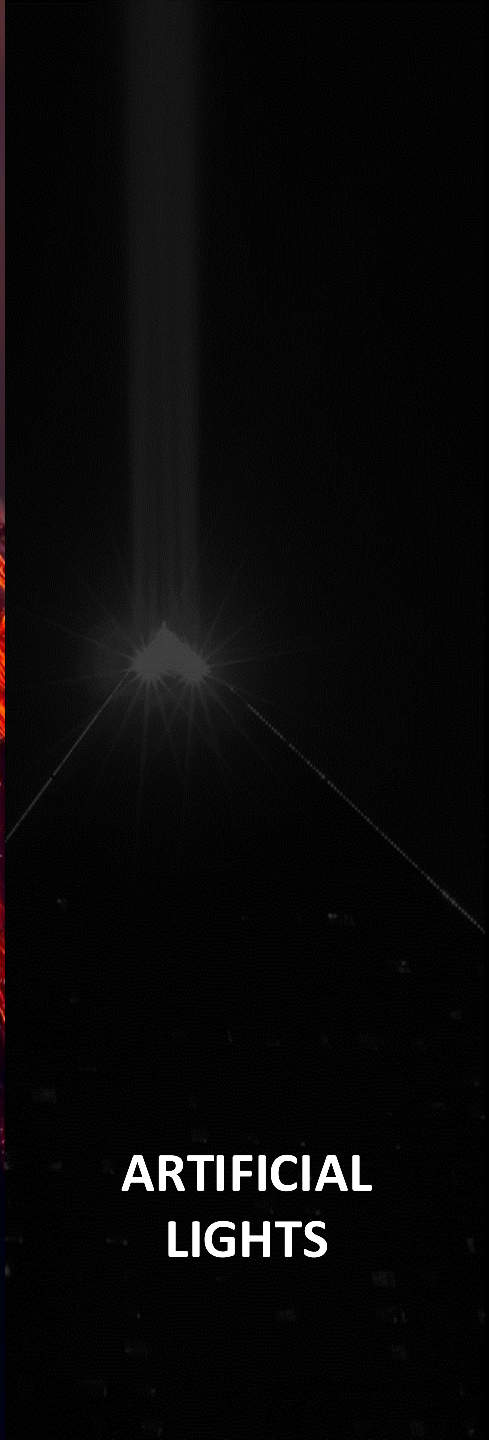
- Aerosols
- Gases
- Wildfires, biomass burning, gas flares
- Land characteristics

Instruments include:

- Moderate Resolution Imaging Spectroradiometer (MODIS)
- Multiangle Imaging Spectroradiometer (MISR)
- Visible Infrared Imaging Radiometer Suite (VIIRS)

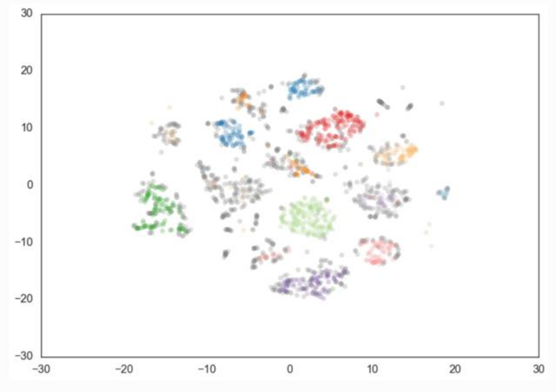

GAS FLARES

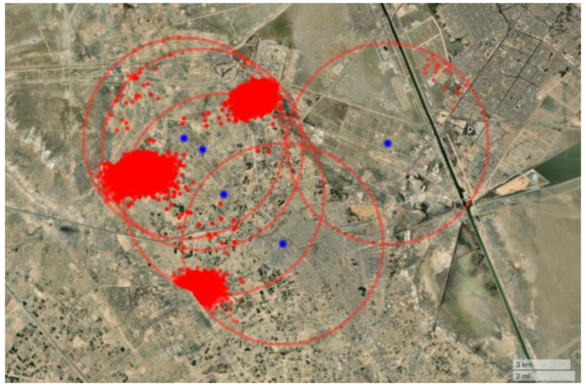

VOLCANOS

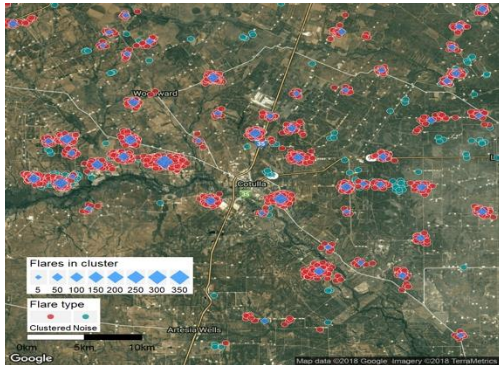

ARTIFICIAL
LIGHTS

Thermal Sources

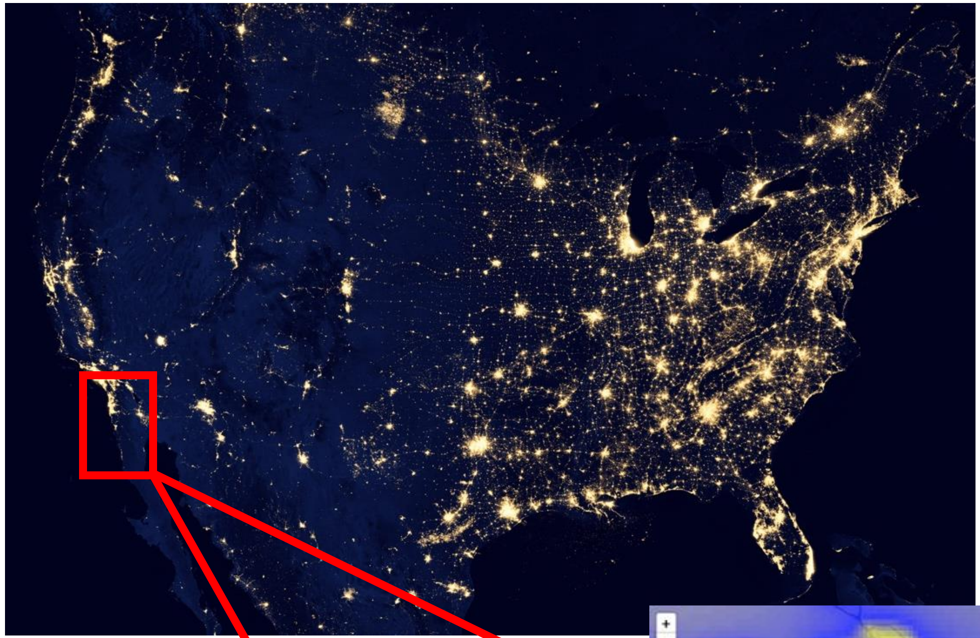
- **Wildfires, biomass burning and gas flares** are observed by satellites (MODIS and VIIRS) as thermal hotspots (points).
- Spatial hierarchical density-based clustering uses a hierarchy of clusters and iteratively merges smaller clusters into larger ones based on their density connectivity.
- Identify clusters of varying densities, which is a challenge for many clustering algorithms.





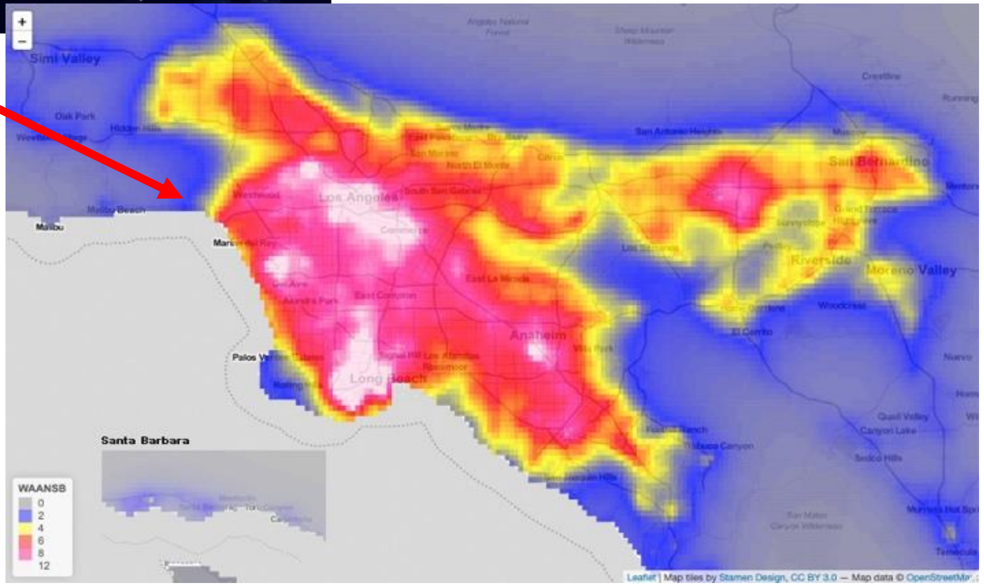


Artificial Light at Night



NIR and Visible Bands

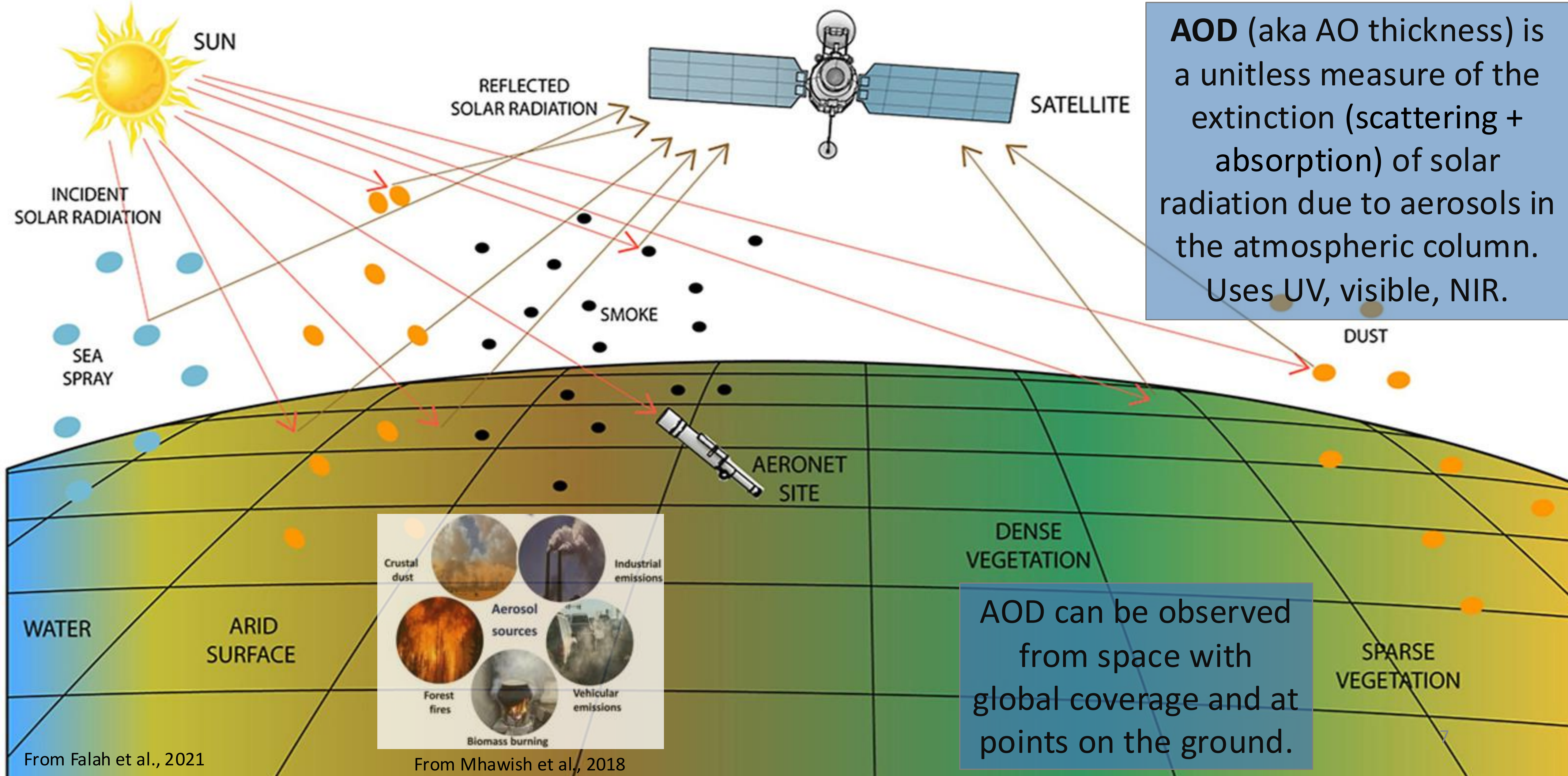
- **Artificial light at night** is a marker of urbanization, is one of the most pervasive environmental pollutants in urban settings.
- VIIRS satellite observations linked with handheld sky quality meters to generate "skyglow" or illuminance.



From Franklin et al 2020

Satellite Aerosol Optical Depth (AOD)

AOD (aka AO thickness) is a unitless measure of the extinction (scattering + absorption) of solar radiation due to aerosols in the atmospheric column. Uses UV, visible, NIR.

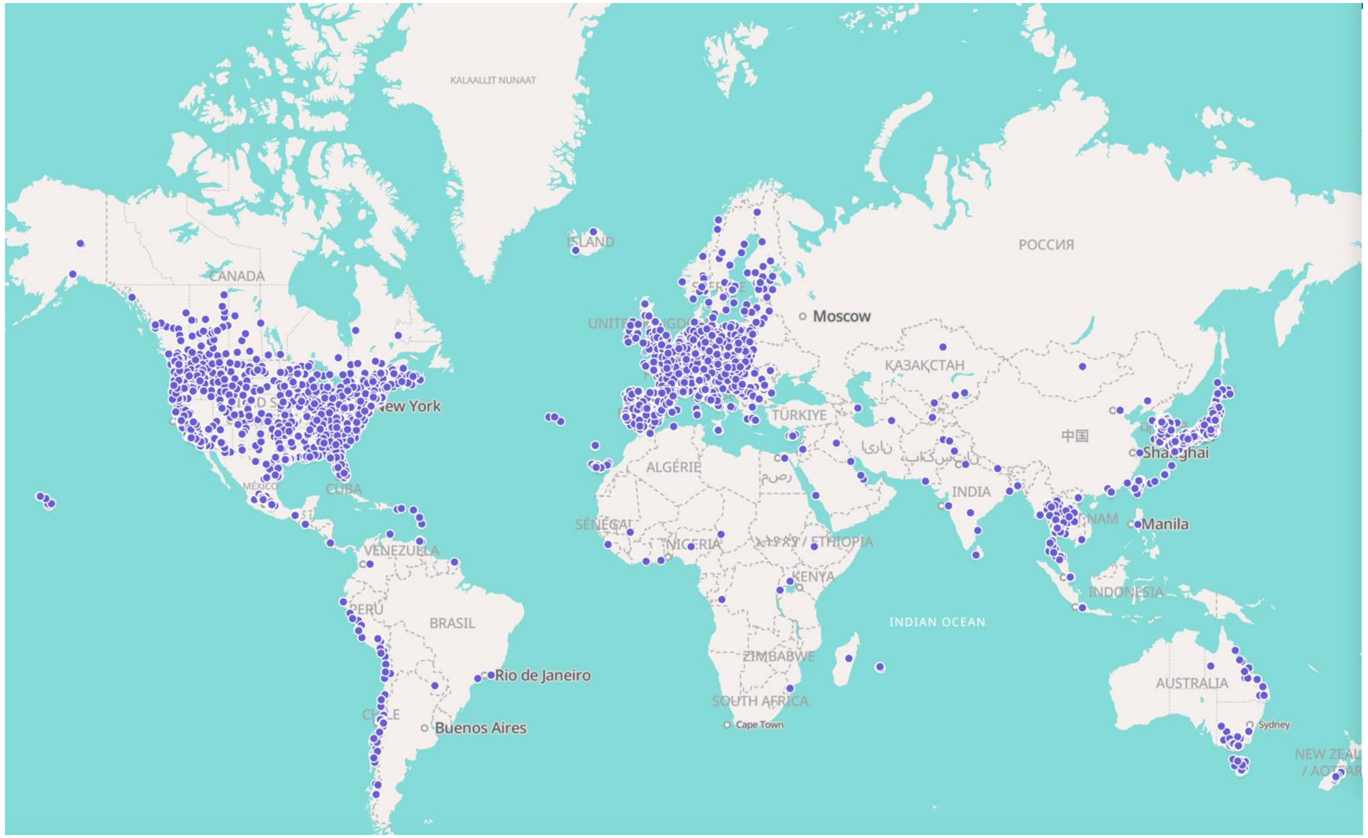


From Falah et al., 2021

From Mhawish et al., 2018

AOD can be observed from space with global coverage and at points on the ground.

Satellite Data for Air Quality Studies

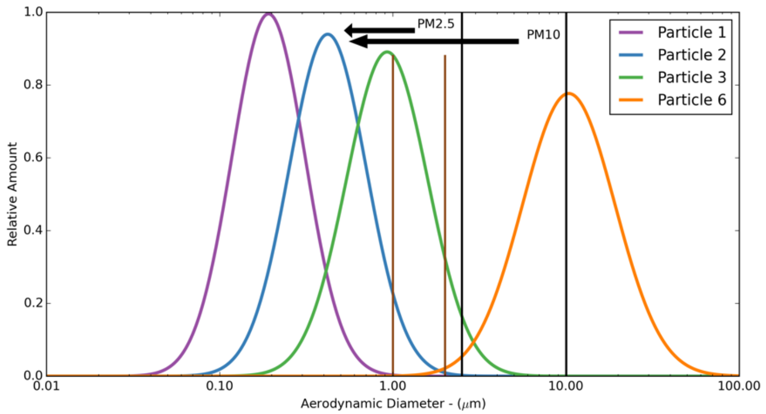


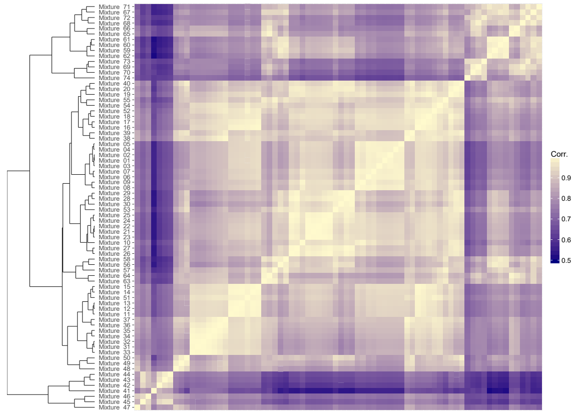
Global surface $PM_{2.5}$ network openaq.org

- Air quality studies are limited by the spatial sparsity of ground-level monitoring networks.
- This sparsity is an issue in many parts of the world where infrastructure is limited (and pollution is generally higher).
- With global coverage, satellite observations fill in these gaps.
- Nevertheless, several steps are required to take satellite observations <u>to surface level mass concentrations</u>.

AOD and Particulate Matter Air Quality

- Combine multi-source data through spatial and temporal alignment.
- Develop models with AOD and other independent variables to predict $PM_{2.5}$

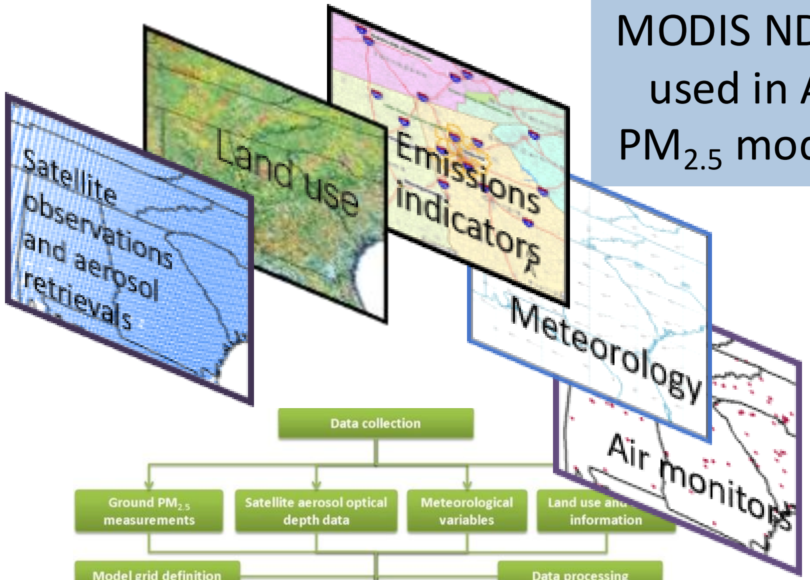




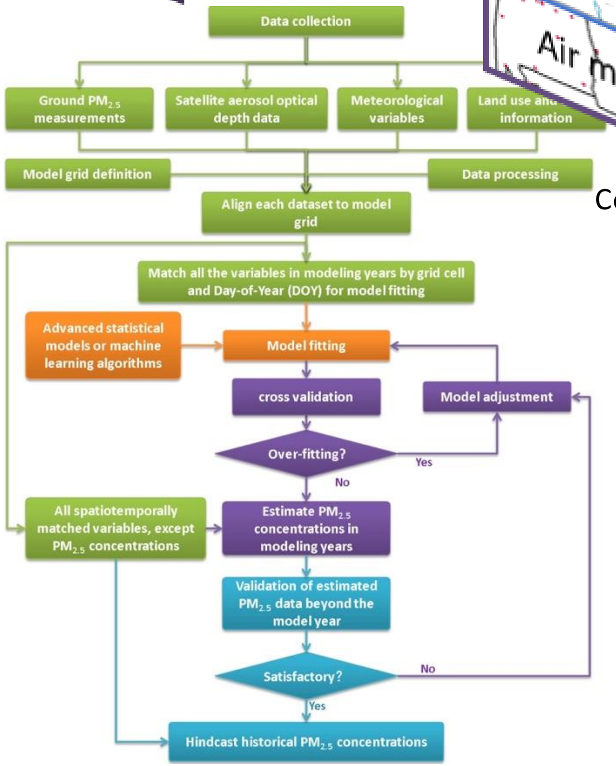
From Franklin et al., 2017 & 2018

- MISR has the additional feature of AOD "fractionated" by size (small, medium, large), spherical, non-spherical, absorbing, non-absorbing.
- AOD types have been further broken down into 74 "mixtures", which have been used to estimate PM chemical speciation (sulfate, nitrate, EC, OC, dust) and PM_{10} and $PM_{2.5}$.

MODIS Fire 3.9um and 11um bands MODIS NDVI are used in AOD-$PM_{2.5}$ modeling!



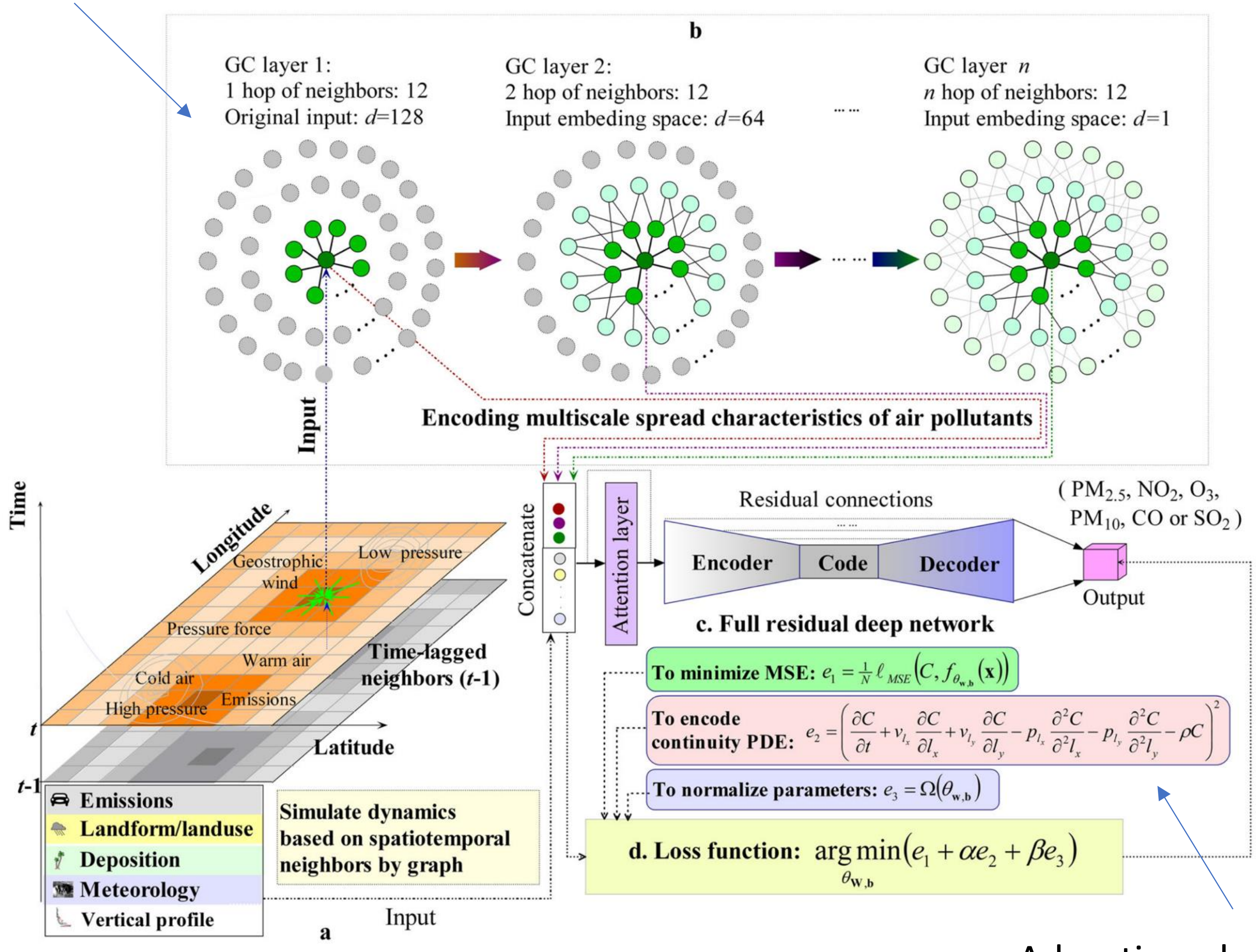
Courtesy of Dave Diner



From Ma et al., 2022

Spatiotemporal GCNs

Graph layers simulate spatial spread



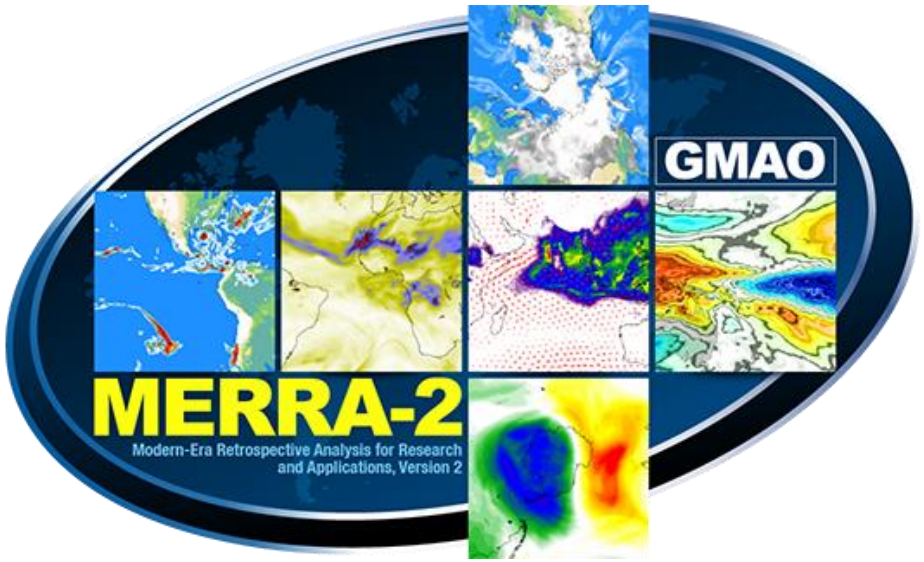
From Li et al 2023

Advection plus diffusion minus deposition

- Graph Convolutional Networks
- Incorporate spread into the graph convolution layers to simulate the spatiotemporal dynamics (spread) of pollutants.
- Recurrent layers for temporal dynamics.
- Incorporate physical law into the regularization term in the loss function of the GCN, ensures the solution by the network does not violate the laws of physics.
- We also include spatial statistical methods in this framework to account for missing data.

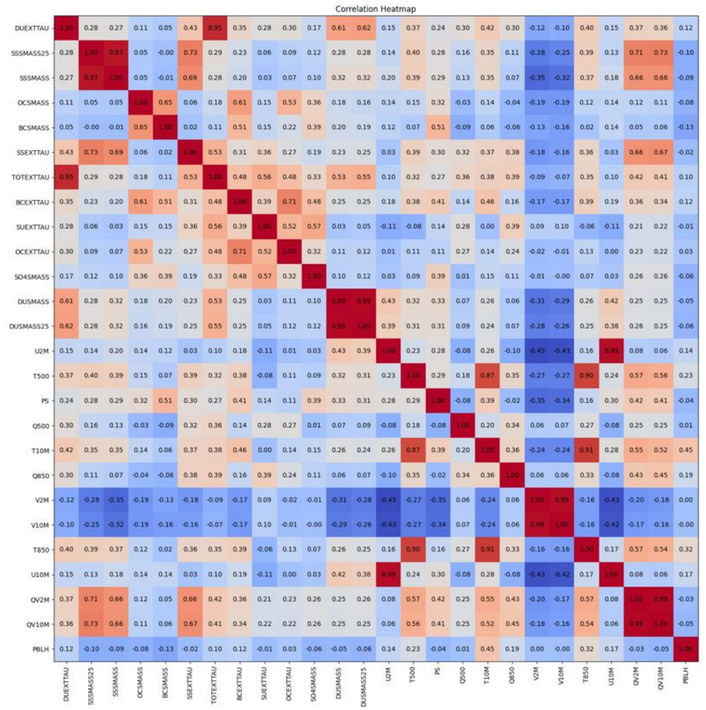
Satellite ‘Image’ Calibration and Downscaling

- The Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) is a reanalysis produced by NASA’s Global Modeling and Assimilation Office.
- MERRA-2 is available back to 1980, but at low spatial resolution (50 km).
- It is coupled with a numerical model GOCART (Goddard Chemistry Aerosol Radiation and Transport) to provide simulations of **dust, sulfate, black carbon, organic carbon, sea salt**. It does **not provide $PM_{2.5}$ directly**, it is a function of the sum of the components.

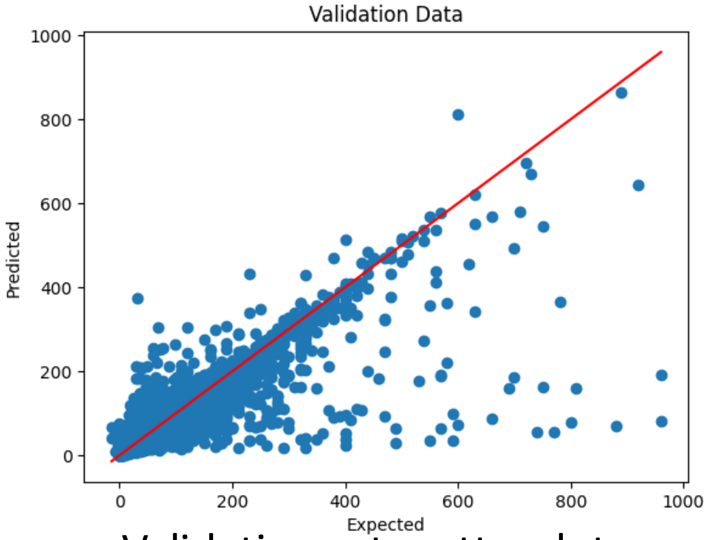


MERRA-2 Calibration

- MERRA-2 components are not well calibrated to $PM_{2.5}$ mass. This was also found in other studies (Sayeed et al 2022).
- We developed a calibration model using deep learning.
  - Matched ground-level OpenAQ and MERRA-2 data by location and time
  - Tried different ML approaches and neural networks, training hyperparameters, and feature engineering (separate arid and non-arid).
  - Used MERRA-2 $PM_{2.5}$ components and meteorology to predict $PM_{2.5}$ mass.
  - Deep-learning validation $R^2 = 0.79$, root-mean-square-error 24 ug/m^3



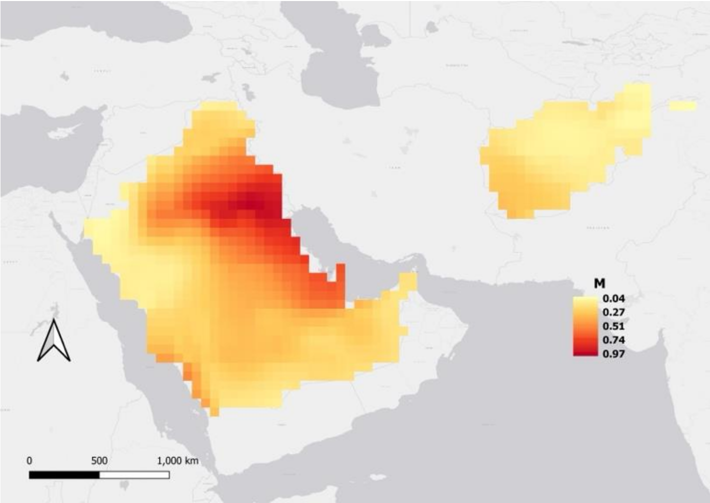
Correlation matrix of the MERRA-2 variables



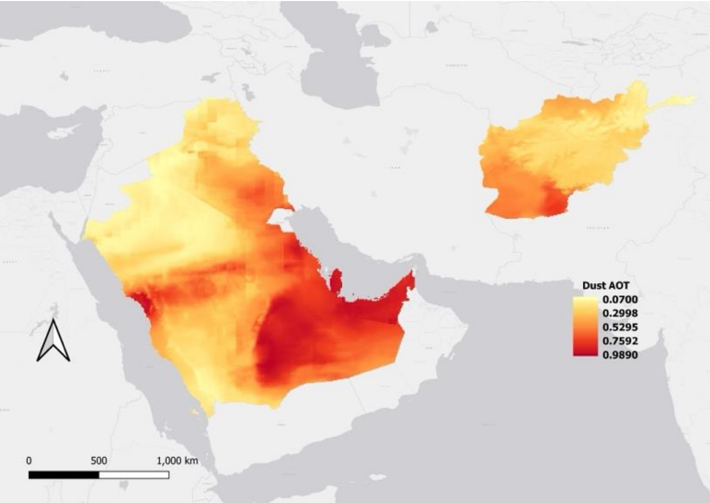
Validation set scatterplot

MERRA-2 Downscaling

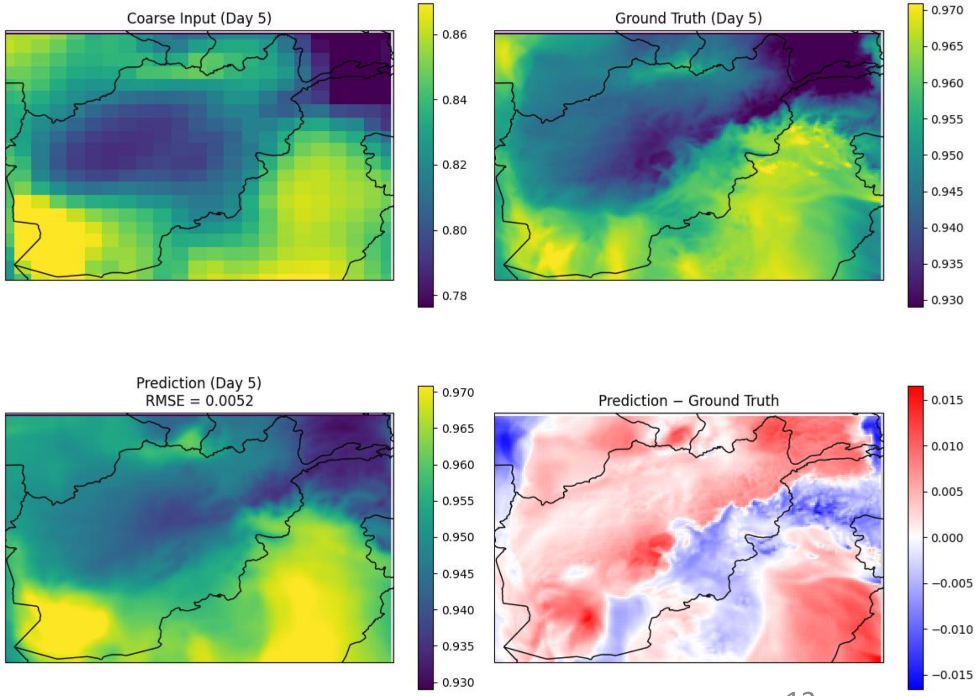
- Using NASA's Goddard Earth Observing System Model, Version 5 (GEOS-5) "Nature Run" (G5NR) of MERRA-2 over a subset of time (2005-2007) we trained a **transfer learning** model to downscale MERRA-2 from 50 km to 7 km.
- Use the **calibrated $PM_{2.5}$** model to rescale the components and apply the downscaling to the components.



MERRA-2 Dust $PM_{2.5}$



Downscaled 7km Dust $PM_{2.5}$



Summary

- Remote sensing can be used to understand a variety of environmental factors on a global scale. Many used for downstream health studies.
- Remote sensing data are spatiotemporal, so methods we develop must capture these trends.
- Some satellite data require more wrangling and modeling than others (e.g. PM air pollution vs light at night).
- Statistical and machine learning approaches are needed to synthesize and generate meaningful quantities from large datasets.
- Importance of uncertainty quantification, which is still lacking.

Thank you!