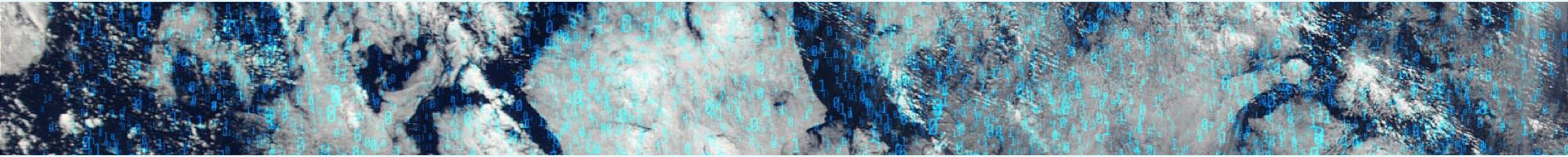


Reconstructing Aerosol Vertical Profiles with Aggregate Output Learning

Sofija Stefanović, Shahine Bouabid, Philip Stier, Athanasios Nenes, Dino Sejdinović

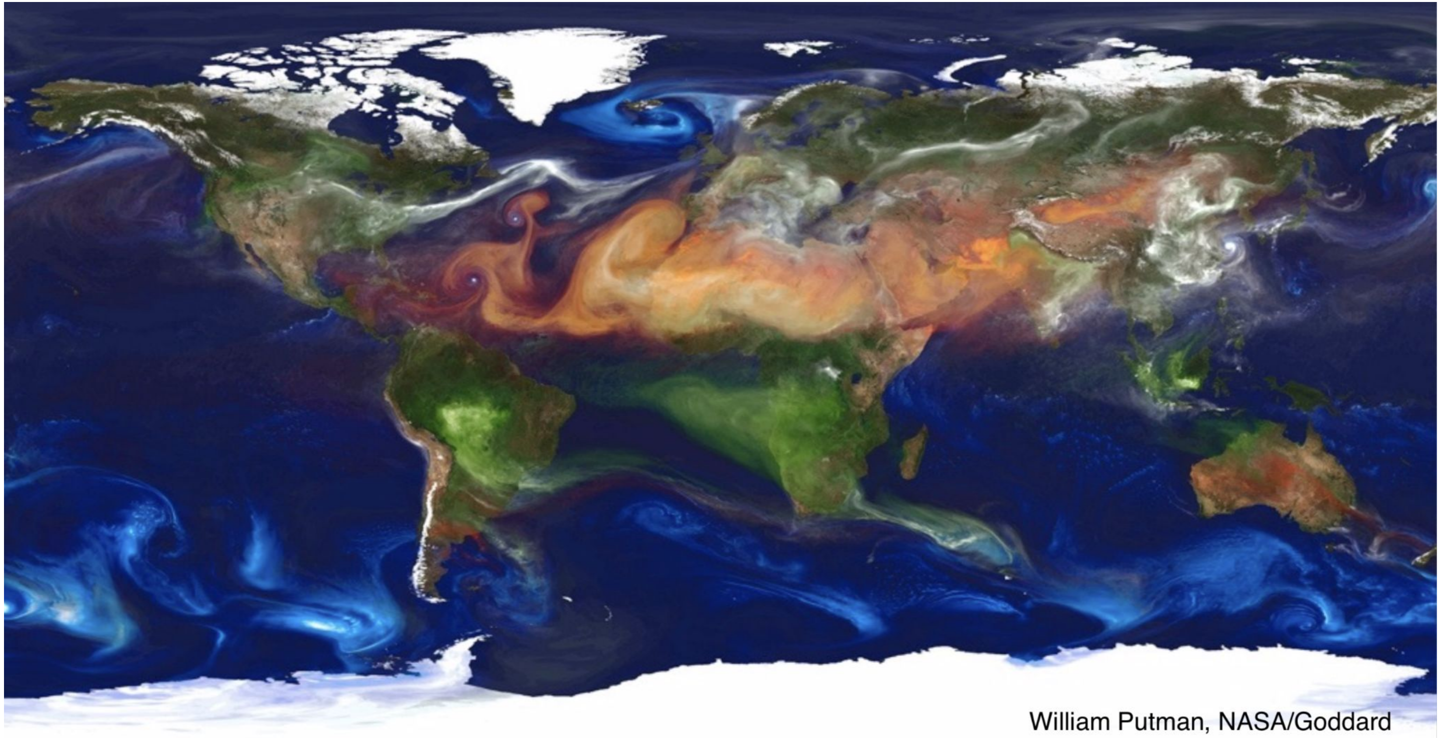
UNIVERSITY OF OXFORD

iMIRACLI

www.imiracli.web.ox.ac.uk

This project receives funding from the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement No 860100.

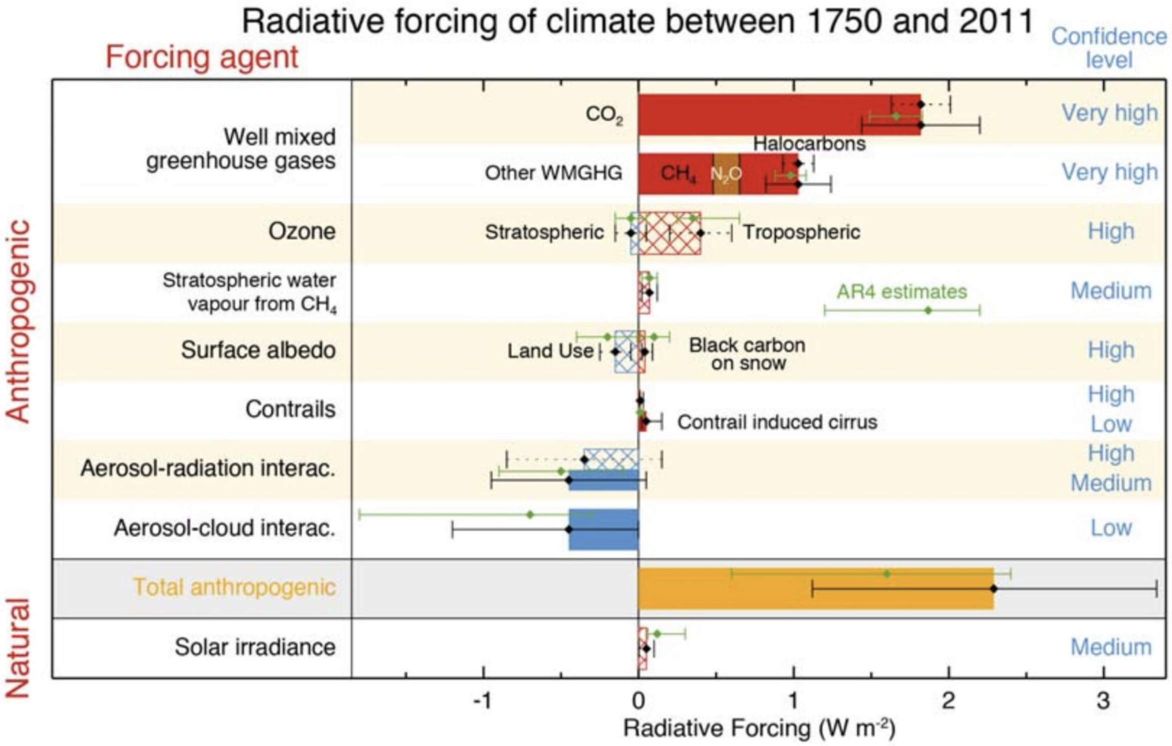
Motivation



GEOS-5 10km resolution
Red: Dust Blue: Sea Salt Green: Smoke White: Sulfate

This project receives funding from the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement No 860100.

Motivation



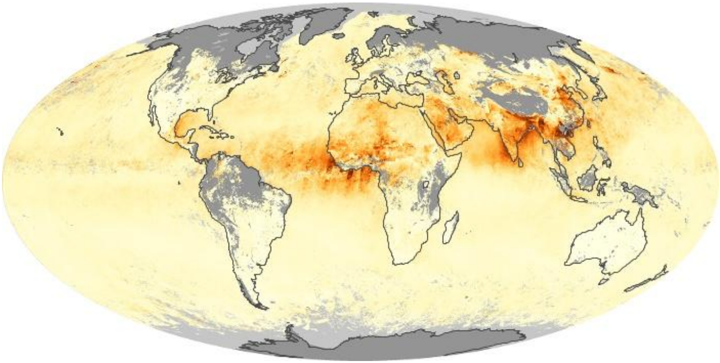
IPCC 2013

This project receives funding from the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement No 860100.

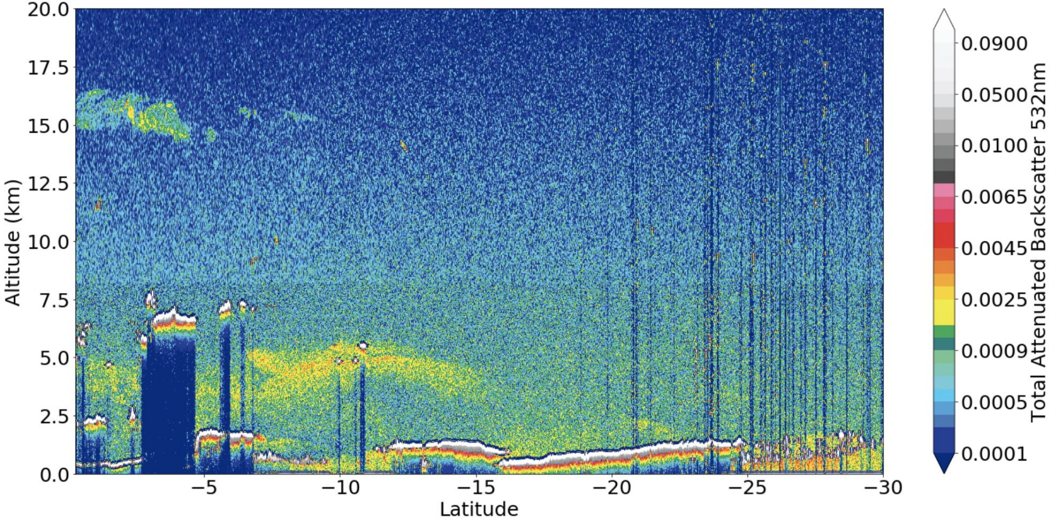
Motivation

2D proxies (vertically aggregated data) often insufficient to understand aerosol distribution

e.g. aerosol optical depth from satellites $\mathrm{AOD} = \int_0^H b_{\mathrm{ext}}(h)\mathrm{d}h$



MODIS AOD



CALIOP lidar vertical profile (‘scene’)

This project receives funding from the European Union’s Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement No 860100.

Problem Statement

General setup:

- Collection of bagged observations: $\{\{x_j^{(i)}\}_{i=1}^{H}, y_j, z_j\}_{j=1}^{n}$
- Function to disaggregate: $f : \mathbb{R}^{d_x} \to \mathbb{R}$
- Aggregation operator over column height: $\mathrm{Agg}_j : f \mapsto \int_{j^{\mathrm{th}}\mathrm{column}} f(x)\,\mathrm{d}h(x)$
- Aggregate observation model:

$$z_j = \mathrm{Agg}_j(f) + \varepsilon_j$$

Specific problem to develop a proof of concept for the methodology:

Reconstruct vertical profiles of sulfate concentrations from aggregated column mass density + chemical and meterological covariates

$$\sigma_{\mathrm{SO}_4} = \int_0^H [\mathrm{SO}_4](h)\mathrm{d}h$$

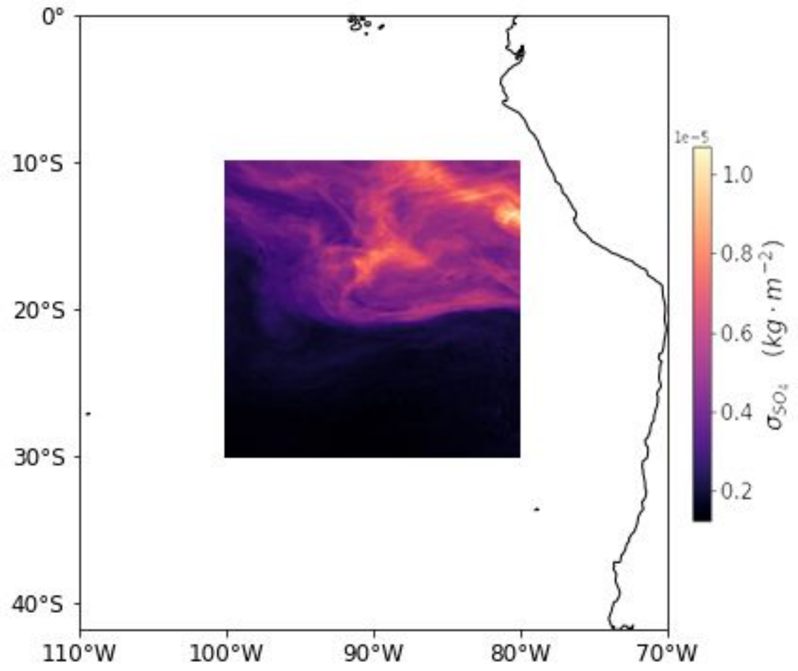
This project receives funding from the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement No 860100.

Dataset

NASA's GEOS-5 Nature Run output used as dataset basis.

	Name	Notation	Units
2D	SO_4 column density	σ_{SO_4}	$kg \cdot m^{-2}$
	Liquid water path	LWP	$kg \cdot m^{-2}$
3D	SO_4 mass mixing ratio	r_{SO_4}	$kg \cdot kg^{-1}$
	SO_2 mass mixing ratio	r_{SO_2}	$kg \cdot kg^{-1}$
	Relative Humidity	RH	1
	Air temperature	T	K
	Vertical velocity	w	$m \cdot s^{-1}$
	Cloud liquid water	q	$kg \cdot kg^{-1}$
	Moist air density	ρ	$kg \cdot m^{-3}$

Table 1. Dataset variables, "2D" corresponds to variables indexed by time, latitude and longitude while "3D" corresponds to variables that also have a height dimension.



This project receives funding from the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement No 860100.

Initial Solutions - Baseline 1

Input 3D covariates : $x = (\text{latitude}, \text{longitude}, \text{altitude}, r_{\text{SO}_2}, \text{RH}, T, w, q)$

Objective : $$\min_f \sum_{i=1}^{n} \left(\sigma_{\text{SO}_4\, i} - \int_{\text{i}^{\text{th}}\ \text{column}} f(x) \mathrm{d}h(x) \right)^2$$

Hypothesis : $f(x) = \beta^\top x$

Solution : Closed form ridge regressor of column-aggregate inputs against AOD

UNIVERSITY OF OXFORD

iMIRACLI

This project receives funding from the European Union’s Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement No 860100.

Initial Solutions - Baseline 2

Input 3D covariates : $x = (\text{latitude}, \text{longitude}, \text{altitude}, r_{\mathrm{SO}_2}, \mathrm{RH}, T, w, q)$

Input 2D covariates : $y = (\text{latitude}, \text{longitude}, \sigma_{\mathrm{SO}_4}, \mathrm{LWP})$

Step 1 : Fit $g : y_i \mapsto \int_{\mathrm{i^{th}\ column}} f(x) \mathrm{d}h(x)$

Step 2 : $\min_f \sum_{i=1}^{n} (\sigma_{\mathrm{SO}_4\ i} - g(y_i))^2$

Hypothesis : $f(x) = \beta^\top x$ $\quad$ $g(y) = \gamma^\top y$

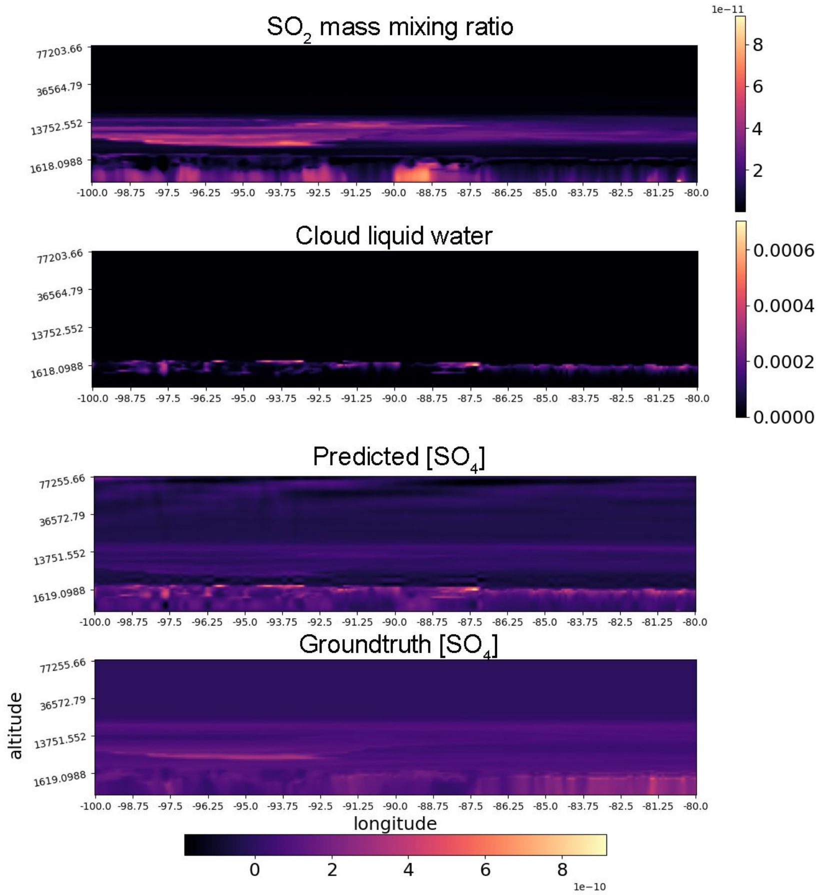
Solution : Closed form two-stage ridge regressor

This project receives funding from the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement No 860100.

Experiments

		RIDGE	TWO-STAGE
	RMSE (10^{-6})	3.47	3.52
2D	MAE (10^{-6})	3.39	3.39
	Corr. (%)	93.5	87.5
	RMSE (10^{-10})	2.71	2.50
3D	MAE (10^{-10})	1.07	1.10
	Corr. (%)	62.5	63.9

Table 2. Evaluation scores on vertical profile reconstruction; "2D" refers to evaluation against aggregate σ_{SO_4} targets used for training; "3D" refers to evaluation against vertical groundtruth



This project receives funding from the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement No 860100.

Discussion & future work

- Two-stage regression shows a slight increase in performance over simple kernel ridge regression with metrics used
- Unclear why the influence of SO2 profiles (important sulfate precursor gas) on predictions varies in experiments
- Metrics more suited to the problem should be developed
- Next step: use specialised aerosol models with lidar simulator and develop kernel-based model to tackle the AOD disaggregation problem:

Reconstruct vertical profiles $b_{\text{ext}}(h)$ from aggregated observations of the AOD and chemistry + meterological covariates

Code and data available at:
https://github.com/shahineb/aerosols-vertical-profiles

This project receives funding from the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement No 860100.