Reconstructing Aerosol Vertical Profiles with Aggregate Output Learning

Sofija Stefanović [1] **Shahine Bouabid** [2] **Philip Stier** [1] **Athanasios Nenes** [3] **Dino Sejdinović** [2]

Abstract

Aerosol-cloud interactions constitute the largest source of uncertainty in assessments of the anthropogenic climate change. This uncertainty arises in part from the inability to observe aerosol amounts at the cloud formation levels, and, more broadly, the vertical distribution of aerosols. Hence, we often have to settle for less informative two-dimensional proxies, i.e. vertically aggregated data. In this work, we formulate the problem of disaggregation of vertical profiles of aerosols. We propose some initial solutions for such aggregate output regression problem and demonstrate their potential on climate model data.

1. Introduction

Aerosols are atmospheric particles that influence the Earth's energy budget both by scattering radiation directly (direct effect) [8], and by acting as cloud condensation nuclei (CCN) that modulate cloud droplet number and radiative properties (indirect effect). Aerosol-cloud interactions (ACI) contribute the largest uncertainty to the projection of anthropogenic climate change, in part due to the poor estimates of the abundance and vertical distribution of aerosols in the atmosphere [4].

While field measurement campaigns provide detailed aerosol data, these are spatio-temporally sparse [1, 13] and provide insufficient constraints on aerosol global distribution. In contrast, satellite observations offer long term global records, but they are typically limited to measurements of aerosol optical properties [12].

A widespread proxy for aerosol concentration or CCN is Aerosol Optical Depth (AOD), defined at a given wavelength λ as:

$$\text{AOD}(\lambda) = \int_0^H b_{\text{ext}}(\lambda, h) \, dh \qquad (1)$$

where b_{ext} is the extinction coefficient[1] and the integral is taken over the height H of the atmospheric column.

While AOD is useful as a measure of total aerosol load within the column, it does not provide information on the vertical distribution of aerosols, which strongly influences both the magnitudes and even the sign of the aerosol direct and indirect effects. For example, both modelling [14] and observational studies [10] find AOD to be inadequate for assessing aerosol-cloud interactions over vast subtropical ocean areas, which play a key role in determining the radiation balance of the earth. However, in both cases, vertically resolved aerosol extinction coefficient b_{ext} shows significantly higher correlations with CCN or its proxies.

We investigate the reconstruction of aerosol vertical profiles using as inputs meteorology and chemistry covariates. While our prime motivation is to reconstruct b_{ext} from satellite measurements of AOD [12, 18], the intricacies of combining measurements from different instruments makes it challenging to validate any proposed methodology. On the other hand, climate models have readily available aerosol vertical profiles and are self-consistent in the sense that all data is jointly observed. Hence, we propose to use data from NASA's Goddard Earth Observing System, version 5 (GEOS-5) Nature Run (G5NR) simulation [17] — a high-resolution global circulation model [3] for model development. We choose sulfates (SO_4) as a case study — these are a major contributor to anthropogenic aerosol pollution and arise primarily through oxidation of sulfur dioxide (SO_2) emitted from burning fossil fuel. G5NR provides matched samples of the vertically resolved sulfate mass concentrations $[SO_4]$ and the mass column density defined as:

$$\sigma_{SO_4} = \int_0^H [SO_4](h) \, dh. \qquad (2)$$

To mirror our motivating application, i.e. predicting vertically resolved b_{ext} from AOD observations only, we propose in this work to probe the vertical reconstruction of $[SO_4]$ given σ_{SO_4}.

---

[1]Department of Atmospheric, Oceanic and Planetary Physics, University of Oxford, UK [2]Department of Statistics, University of Oxford, UK [3]Laboratory of Atmospheric Processes and their Impacts, École Polytechnique Fédérale de Lausanne, Switzerland. Correspondence to: Sofija Stefanović <sofija.stefanovic@physics.ox.ac.uk>, Shahine Bouabid <shahine.bouabid@stats.ox.ac.uk>.

---

[1]The sum of contributions from particle-light scattering plus absorption of light by particles.

Motivated by the study of cloud vertical structures, the task of reconstructing three-dimensional (3D) profiles corresponding to two-dimensional observations (2D) has been framed in the past as fully supervised learning [6]. Collecting high quality observational data of aerosol vertical profiles at large scale is however unfeasible, thus making fully supervised approaches unsuitable. In fact, while in previous work Nair and Yu [9] have addressed the task of CCN number prediction from atmospheric measurements, they resorted to using model data in order to apply fully supervised learning models.

Spatial disaggregation or statistical downscaling is the task of inferring subgrid details given a low resolution observation process. Postulating an underlying fine grained spatial field that aggregates into coarse observations, this problem can be framed as weakly supervised learning [21] with aggregate outputs. Existing works [2, 7, 15, 16, 19, 20] have only considered applications to 2D fields, yet this rationale can be extended to 3D fields that aggregate with respect to height into a 2D field. Furthermore, Kipling et al. [5] show that aerosols' behavior is driven by a relatively small number of atmospheric processes, suggesting that having access to 3D observations of these processes could help vertical profile reconstruction. We thus propose to frame the reconstruction of [SO_4] vertical profile as the vertical disaggregation of σ_{SO_4} using 3D atmospheric covariates.

Our contributions are threefold: **(i)** We propose the novel problem of aerosols vertical profiles disaggregation given vertically aggregated targets and 3D atmospheric covariates; **(ii)** We introduce a dataset of 2D+3D meteorological and chemistry covariates from NASA's G5NR simulation, with focus on sulfates aerosols; **(iii)** We describe baseline vertical disaggregation models and demonstrate them on the sulfate mass column density disaggregation problem.

2. Dataset

NASA's GEOS-5 Nature Run (G5NR) is a 2-year (June 2005-May 2007) global, non-hydrostatic[2] mesoscale[3] simulation with a 7 km horizontal resolution and 72 vertical levels (up to 85km). This simulation includes standard meteorological parameters, as well as aerosol tracers (dust, sea salt, sulfate, black and organic carbon), O_3, CO and CO_2. The simulation is driven by prescribed sea-surface temperature and sea-ice, daily volcanic and biomass burning emissions, and high-resolution inventories of anthropogenic sources.

We used the instantaneous (30min) products relevant for

[2]Where the hydrostatic approximation is not made, i.e. the vertical momentum equation is solved.

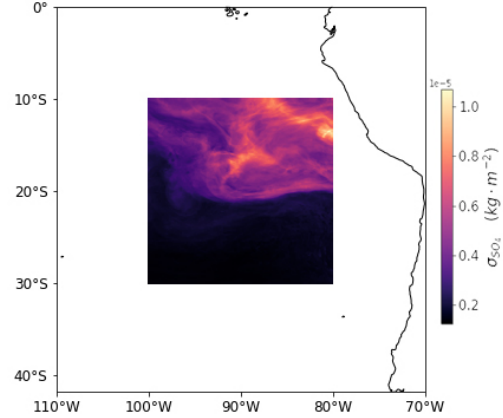[3]Phenomena occurring on scales of tens to hundreds of kilometers.

Figure 1. SO_4 column density field from dataset plotted for one time step; our objective is to obtain a vertically resolved version of this field.

sulfate formation with 7km horizontal resolution from the 15th of January 2007 as the basis for our dataset. Sulfates are a major contributor to the total AOD in our dataset (see Figure 3 in Appendix A). As depicted in Figure 1, the dataset is located in the Pacific Ocean (10-30 °S, 80-100 °W).

Table 1 outlines the variables used in the dataset and the notation that will be used to refer to these in the following sections.

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

The groundtruth 3D SO_4 mass concentration is calculated by multiplying the SO_4 mass mixing ratio with the (moist) air density in the column as:

$$[SO_4] = \rho \cdot r_{SO_4} \tag{3}$$

We verify numerically that this field aggregates with respect to height into the 2D field σ_{SO_4} provided by the dataset.

3. Baseline Methodologies

In this section, we describe baseline models. While our experiments focuses on SO_4 column mass density disaggre-

gation, we employ broad notations for the sake of generality.

3.1. Problem Statement

Let $\{\{x_j^{(i)}\}_{i=1}^H, y_j, z_j\}_{j=1}^n$ a collection of bagged observations. $x_j^{(i)} \in \mathbb{R}^{d_x}$ is a spatial covariate that admits 3D coordinates following latitude, longitude and height and belongs to j^{th} height column. We denote its tensor concatenation as $\boldsymbol{x}_j := [x_j^{(1)} \ \ldots \ x_j^{(H)}]^\top \in \mathbb{R}^{H \times d_x}$ and denote $\mathbf{X} := [\boldsymbol{x}_1 \ \ldots \ \boldsymbol{x}_n]^\top \in \mathbb{R}^{n \times H \times d_x}$. $y_j \in \mathbb{R}^{d_y}$ is a column-level covariate that admits 2D latitude and longitude coordinates and $z_j \in \mathbb{R}$ is the column covariate we wish to disaggregate along height. Likewise, we use tensor representations $\mathbf{Y} := [y_1 \ \ldots \ y_n]^\top \in \mathbb{R}^{n \times d_y}$ and $\mathbf{z} = [z_1 \ \ldots \ z_n]^\top \in \mathbb{R}^n$.

Let $f : \mathbb{R}^{d_x} \to \mathbb{R}$ be the function of interest we want to recover and consider the linear aggregation operator defined for the jth column as $\text{Agg}_j : f \mapsto \int_{j^{\text{th}}\text{column}} f(x) \, \mathrm{d}h(x)$ where h is a positive measure of height. Then we postulate an aggregate observation model

$$z_j = \text{Agg}_j(f) + \varepsilon_j \tag{4}$$

where $\varepsilon_j \sim \mathcal{N}(0, \sigma^2)$ is observation noise.

In finite data size, we substitute $\text{Agg}_j(f)$ with an approximation $\text{Agg}(f(\boldsymbol{x}_j))$ using trapezoidal rule

$$\text{Agg}(f(\boldsymbol{x}_j)) = \sum_{i=1}^{H-1} \frac{f(x_j^{(i)}) + f(x_j^{(i+1)})}{2} \Delta h_j^{(i)} \tag{5}$$

where $\Delta h_j^{(i)}$ is the height difference between $x_j^{(i+1)}$ and $x_j^{(i)}$. With notation abuse, we will use in what follows notation $\text{Agg}(\cdot)$ for the approximate aggregation over height of any tensor that admits a height dimension. For example, $\text{Agg}(f(\mathbf{X})) = [\text{Agg}(f(\boldsymbol{x}_1)) \ \ldots \ \text{Agg}(f(\boldsymbol{x}_n))]^\top \in \mathbb{R}^n$.

To probe the potential of the vertical disaggregation task, we propose as a first approach to adopt a plain linear model baseline and make hypothesis $f(x) = x^\top \beta + c$ with $\beta \in \mathbb{R}^{d_x}$ and $c \in \mathbb{R}$. Without loss of generality, we assume that intercept is included in the covariates, i.e. $c = 0$.

3.2. Aggregate Ridge Regression

Given observation model (4), it is natural to consider as an objective the regularized quadratic risk between aggregate observations and the aggregated prediction, which writes

$$\min_{\beta \in \mathbb{R}^{d_x}} \|\mathbf{z} - \text{Agg}(\mathbf{X}\beta)\|_2^2 + \lambda \|\beta\|_2^2 \tag{6}$$

with regularization weight $\lambda > 0$.

By linearity of the aggregation operator, the solution to (6) is simply the solution to the ridge regression of 3D covariates

aggregated over height $\text{Agg}(\mathbf{X})$ against aggregate targets $\mathbf{z}$, given by

$$\beta = \left(\text{Agg}(\mathbf{X})^\top \text{Agg}(\mathbf{X}) + \lambda \mathbf{I}_{d_x}\right)^{-1} \text{Agg}(\mathbf{X})^\top \mathbf{z} \tag{7}$$

While straightforward, this method has the advantage of scaling seamlessly to a huge amount of data since it only incurs a $\mathcal{O}(d_x^3)$ matrix inversion computational cost.

3.3. Two-Stage Aggregate Ridge Regression

In the above, it is implicitly assumed we can establish a mapping $j \mapsto \text{Agg}_j(f)$ that associates each column to its aggregated value based on its index only. In doing so, each column is treated and regressed independently from the others. This is unrealistic since we expect continuous fields to correlate across spatial and temporal dimensions. On the other hand, the 2D column-level covariates y_j provide information about columns correlations. This in turn can be used to embed columns dependence information by learning a mapping $y_j \mapsto \text{Agg}_j(f)$.

In a second baseline, we augment the aggregate ridge regression model by a two-stage process that enables leveraging the information conveyed by 2D covariates.

Stage 1 We first regress the 2D covariates $\mathbf{Y}$ against aggregated columns values $\text{Agg}(\mathbf{X}\beta)$. For the sake of simplicity, we assume linear model $g(y) = \gamma^\top y$ and minimize regularized empirical quadratic risk

$$\min_{\gamma \in \mathbb{R}^{d_y}} \|\text{Agg}(\mathbf{X}\beta) - \mathbf{Y}\gamma\|_2^2 + \nu \|\gamma\|_2^2 \tag{8}$$

where $\nu > 0$ is a regularization weight. As per above, this admits closed form solution

$$\gamma = \left(\mathbf{Y}^\top \mathbf{Y} + \nu \mathbf{I}_{d_y}\right)^{-1} \mathbf{Y}^\top \text{Agg}(\mathbf{X}\beta) \tag{9}$$

Stage 2 We now use the prediction provided by $g(\mathbf{Y})$ to regress against aggregate targets $\mathbf{z}$. Let $\boldsymbol{\Upsilon} := \mathbf{Y} \left(\mathbf{Y}^\top \mathbf{Y} + \nu \mathbf{I}_{d_y}\right)^{-1} \mathbf{Y}^\top$. The evaluation of the regressor learnt in Step 1 writes $g(\mathbf{Y}) = \mathbf{Y}\gamma = \boldsymbol{\Upsilon} \text{Agg}(\mathbf{X}\beta)$. Hence, substituting the latter to the aggregated columns values in (6), we obtain empirical risk

$$\min_{\beta \in \mathbb{R}^{d_x}} \|\mathbf{z} - \boldsymbol{\Upsilon} \text{Agg}(\mathbf{X}\beta)\|_2^2 + \lambda \|\beta\|_2^2 \tag{10}$$

which admits closed form minimizer

$$\beta = \left(\text{Agg}(\mathbf{X})^\top \boldsymbol{\Upsilon}^2 \text{Agg}(\mathbf{X}) + \lambda \mathbf{I}_{d_x}\right)^{-1} \text{Agg}(\mathbf{X})^\top \boldsymbol{\Upsilon} \mathbf{z}. \tag{11}$$

Comparing (11) to (7), we can interpret $\boldsymbol{\Upsilon}$ as a regularizing term enforcing functional smoothness across 2D covariates.
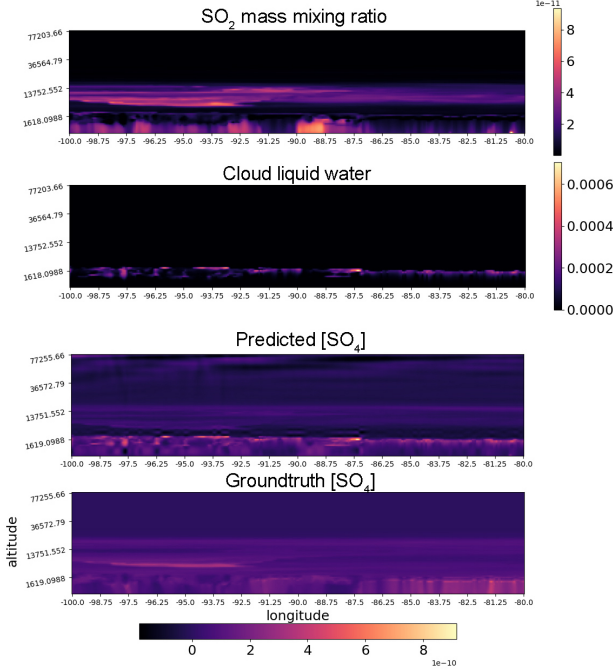
Figure 2. **Top/Middle-Top** vertical profiles of SO_2 mixing ratio and cloud liquid water, used as 3D covariates for prediction; **Middle-Bottom:** predicted $[SO_4]$ vertical profile with two-stage ridge regression baseline; **Bottom:** groundtruth $[SO_4]$ vertical profile.

4. Experiments

We demonstrate and evaluate the baseline models at the vertical disaggregation of sulfate mass column density using dataset introduced in Section 2. The ridge regression baseline is referred to as RIDGE and the two-stage ridge approach as TWO-STAGE. We report evaluation against the unobserved groundtruth 3D sulfate mass concentration profiles $[SO_4]$ and also compare the vertically aggregated prediction against the 2D sulfate mass column density σ_{SO_4} used for training. Scores are reported in root mean square error (RMSE), mean absolute error (MAE) and Pearson correlation (Corr.). Experiments are implemented in *PyTorch* [11] and code and dataset are made publicly available [4].

Model Setup: Since RIDGE only uses 3D covariates, we use as input 3D variable $x = (\text{latitude}, \text{longitude}, \text{altitude}, r_{SO_2}, \text{RH}, T, w, q)$ where covariates notations are defined in Table 1. For TWO-STAGE, we can additionally leverage column-level 2D covariates in the first stage. We use 2D covariate $y = (\text{latitude}, \text{longitude}, \sigma_{SO_4}, \text{LWP})$ to fit the first step, and then use the same covariates as RIDGE for the second step. We emphasize that while σ_{SO_4} is also our aggregate target z, it can nonetheless be used as a column-level covariate. All input variables are standardized.

___

[4] https://github.com/shahineb/
aerosols-vertical-profiles

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

Results: As depicted in Figure 2, we observe the model is able to resolve vertically distributed details that correlate with the input covariates. Table 2 suggests that the column-level knowledge conveyed by 2D covariates is reflected in better performance of the two-staged ridge regression model on the reconstructed 3D profiles.

The scene plotted in Figure 2 shows the patterns of the cloud liquid water reproduced in the prediction — while the cloud layer can be identified in the groundtruth there are no clusters of high concentration within the cloud layer as seen in the prediction. Cloud liquid water is taken as a proxy for where oxidation of SO_2 would occur — having explicit oxidant fields could help reduce the bias due to the cloud field. Both prediction and groundtruth feature a layer of SO_4 that extends across all longitudes and beyond 13km in altitude and that are consistent with the SO_2 mass mixing ratio. Although the prediction in this scene fails to reproduce the thin layer of higher concentration (around 90°W longitudes) consistent with high SO_2, some predictions are strongly influences by the SO_2 mass mixing ratio and reproduce its pattern in full (see Appendix B).

5. Discussion

Motivated by the prediction of better vertically resolved aerosol proxies, we introduced the new task of vertical disaggregation from aggregated 2D observations. We provide a dataset of G5NR model data including diverse meteorological and chemical covariates, propose baseline vertical disaggregation models, and demonstrate their performance at sulfates mass column density disaggregation.

In future work, we intend to apply baseline models to collocated observations from MODIS 2D AOD product [12], CALIOP vertical lidar measurements and more widespread measurements of atmospheric states and compositions. A major benefit from doing that would be increasing the spatiotemporal resolution of CALIOP. Simultenously, we aim to define evaluation metrics that sensibly penalise vertical incoherence, and are hence naturally suited to this problem. Finally, while we limit ourselves to demonstrating simple linear models operating on aggregate output in this

work, further directions will include both non-linear (e.g. kernel-based) and Bayesian methods. This will enable a finer treatment of input covariates along with uncertainty quantification.

Acknowledgements

References

[1] M. O. Andreae. Correlation between cloud condensation nuclei concentration and aerosol optical thickness in remote and polluted regions. *Atmospheric Chemistry and Physics*, 9(2):543–556, 2009.

[2] Siu Lun Chau, Shahine Bouabid, and Dino Sejdinovic. Deconditional Downscaling with Gaussian Processes. *arXiv preprint arXiv:2105.12909*, 2021.

[3] Gregory M. Flato. Earth system models: an overview. *WIREs Climate Change*, 2011.

[4] IPCC. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

[5] Z. Kipling, P. Stier, C. E. Johnson, G. W. Mann, N. Bellouin, S. E. Bauer, T. Bergman, M. Chin, T. Diehl, S. J. Ghan, T. Iversen, A. Kirkevåg, H. Kokkola, X. Liu, G. Luo, T. van Noije, K. J. Pringle, K. von Salzen, M. Schulz, Ø. Seland, R. B. Skeie, T. Takemura, K. Tsigaridis, and K. Zhang. What controls the vertical distribution of aerosol? relationships between process sensitivity in hadgem3–ukca and inter-model variation from aerocom phase ii. *Atmospheric Chemistry and Physics*, 16(4):2221–2241, 2016.

[6] Jussi Leinonen, Alexandre Guillaume, and Tianle Yuan. Reconstruction of cloud vertical structure with a generative adversarial network. *Geophysical Research Letters*, 2019.

[7] Ho Chung Leon Law, Dino Sejdinovic, Ewan Cameron, Tim C.D. Lucas, Seth Flaxman, Katherine Battle, and Kenji Fukumizu. Variational learning on aggregate outputs with Gaussian processes. In *Advances in Neural Information Processing Systems*, 2018.

[8] Robert A. McCormick and John H. Ludwig. Climate Modification by Atmospheric Aerosols. *Science*, 156 (3780):1358–1359, June 1967.

[9] A. A. Nair and F. Yu. Using machine learning to derive cloud condensation nuclei number concentrations from commonly available measurements. *Atmospheric Chemistry and Physics*, 20(21):12853–12869, 2020.

[10] D. Painemal, F.-L. Chang, R. Ferrare, S. Burton, Z. Li, W. L. Smith Jr, P. Minnis, Y. Feng, and M. Clayton. Reducing uncertainties in satellite estimates of aerosol–cloud interactions over the subtropical ocean by integrating vertically resolved aerosol observations. *Atmospheric chemistry and physics*.

[11] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*. 2019.

[12] L. A. Remer, Y. J. Kaufman, D. Tanré, S. Mattoo, D. A. Chu, J. V. Martins, R.-R. Li, C. Ichoku, R. C. Levy, R. G. Kleidman, T. F. Eck, E. Vermote, and B. N. Holben. The MODIS Aerosol Algorithm, Products, and Validation. *Journal of the Atmospheric Sciences*, 2005.

[13] D. V. Spracklen, K. S. Carslaw, U. Pöschl, A. Rap, and P. M. Forster. Global cloud condensation nuclei influenced by carbonaceous combustion aerosol. *Atmospheric Chemistry and Physics*, 11(17):9067–9087, 2011.

[14] Philip Stier. Limitations of passive remote sensing to constrain global cloud condensation nuclei. *Atmospheric Chemistry and Physics*, 2016.

[15] Yusuke Tanaka, Toshiyuki Tanaka, Tomoharu Iwata, Takeshi Kurashima, Maya Okawa, Yasunori Akagi, and Hiroyuki Toda. Spatially aggregated Gaussian processes with multivariate areal outputs. *Advances in Neural Information Processing Systems*, 2019.

[16] Arto Klam Ville Tanskanen , Krista Longi. Non-Linearities in Gaussian Processes with Integral Observations. *IEEE international Workshop on Machine Learning for Signal*, 2020.

[17] L.E. Ott W. Putman, A.M. da Silva and A. Darmenov. Model configuration for the 7-km geos-5nature

run, ganymed release (non-hydrostatic 7 km global mesoscale simulation). gmao office note no.5 (version 1.0), 2014. URL http://gmao.gsfc.nasa.gov/pubs/office_notes.

[18] D. M. Winker, J. L. Tackett, B. J. Getzewich, Z. Liu, M. A. Vaughan, and R. R. Rogers. The global 3-d distribution of tropospheric aerosols as characterized by caliop. *Atmospheric Chemistry and Physics*, 13(6): 3345–3361, 2013.

[19] Fariba Yousefi, Michael Thomas Smith, and Mauricio A. Álvarez. Multi-task learning for aggregated data using Gaussian processes. In *Advances in Neural Information Processing Systems*, 2019.

[20] Yivan Zhang, Nontawat Charoenphakdee, Zhenguo Wu, and Masashi Sugiyama. Learning from aggregate observations. In *Advances in Neural Information Processing Systems*, 2020.

[21] Zhi-Hua Zhou. A brief introduction to weakly supervised learning. *National Science Review*, 2017.
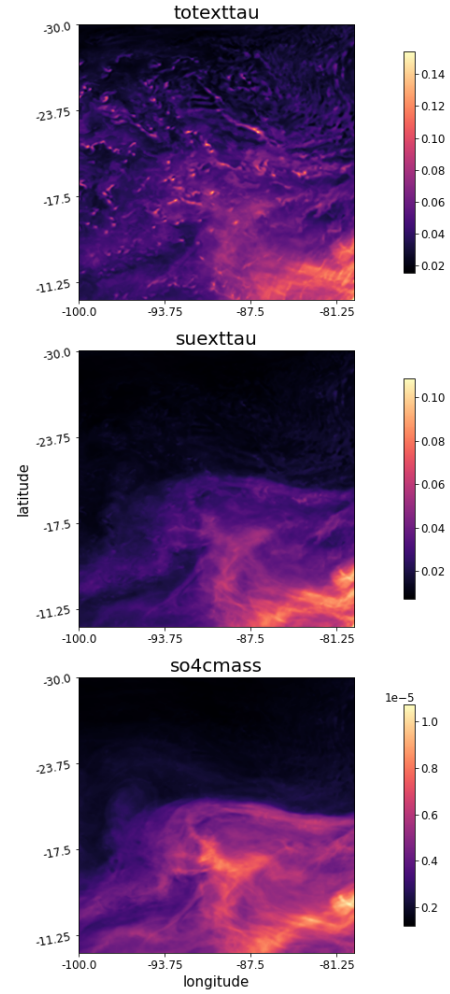
A. Dataset Specifications

Figure 3. Dataset 2D aerosol fields for one timestep; **Top:** Total extinction AOD; **Middle:** SO_4 extinction AOD; **Bottom:** SO_4 column mass density
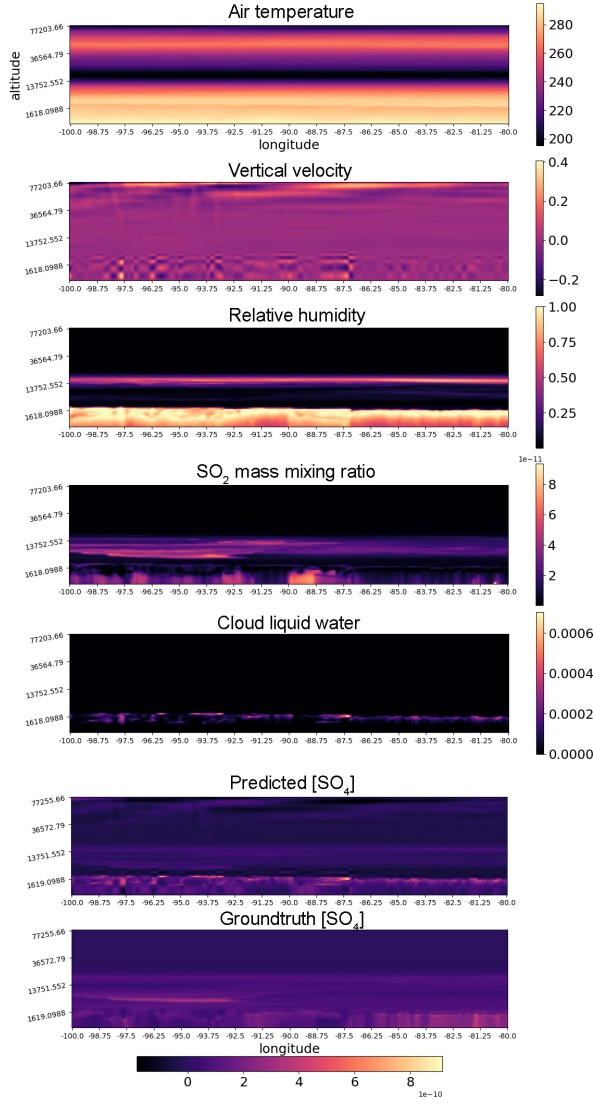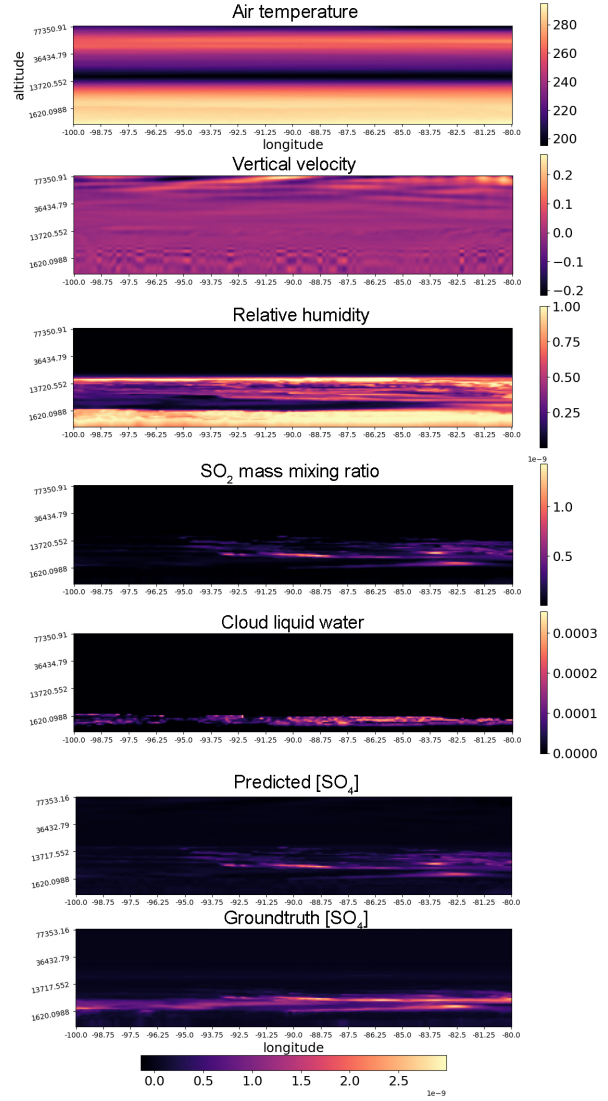
B. Additional Experimental Results

Figure 4. Vertical slices at latitude 23.75°S; **Rows 1-5:** vertical profiles of 3D input covariates; **Row 6** predicted vertical profile of [SO$_4$] with two-stage aggregate ridge regression model; **Row 7** groundtruth [SO$_4$].

Figure 5. Vertical slices at latitude 12.5°S; **Rows 1-5:** vertical profiles of 3D input covariates; **Row 6** predicted vertical profile of [SO$_4$] with two-stage aggregate ridge regression model; **Row 7** groundtruth [SO$_4$].