Bayesian inference for aerosol vertical profiles

Shahine Bouabid
Department of Statistics
University of Oxford
Oxford, UK
shahine.bouabid@stats.ox.ac.uk

Duncan Watson-Parris
Atmospheric, Oceanic and Planetary Physics
Department of Physics
University of Oxford
duncan.watson-parris@physics.ox.ac.uk

Dino Sejdinovic
Department of Statistics
University of Oxford
dino.sejdinovic@stats.ox.ac.uk

Abstract

Aerosol-cloud interactions constitute the largest source of uncertainty in assessments of the anthropogenic climate change. This uncertainty arises in part from the difficulty in measuring the vertical distributions of aerosols. We often have to settle for less informative vertically aggregated proxies such as aerosol optical depth (AOD). In this work, we develop a framework to infer vertical aerosol profiles using AOD and readily available vertically resolved meteorological predictors such as temperature or relative humidity. We devise a simple Gaussian process prior over aerosol vertical profiles and update it with AOD observations. We validate our approach using ECHAM-HAM aerosol-climate model data. Our results show that, while simple, our model is able to reconstruct realistic extinction profiles with well-calibrated uncertainty. In particular, the model demonstrates a faithful reconstruction of extinction patterns arising from aerosol water uptake in the boundary layer.

1 Introduction

Aerosols are microscopic particles suspended in the atmosphere such as dust, sea salt or black carbon. They influence the Earth's energy budget with a negative radiative forcing that counteracts the global warming from anthropogenic greenhouse gases emissions. A large fraction of this forcing is due to their modulation of radiative properties of clouds. By acting as cloud condensation nuclei (CCN), aerosols can drive up the cloud droplet number while driving down the mean cloud droplet size. The resulting clouds are brighter, larger and last longer [1, 20]. They hence reflect more solar radiation and cool the Earth.

However, the magnitude of this forcing is difficult to estimate [9] in part because the physical processes underpinning aerosol-cloud interactions are not yet fully understood. To better estimate present day forcing we require accurate, global measurements of CCN concentrations to assess radiative properties of clouds [1, 20]. Unfortunately, measuring CCN concentrations can only be achieved in-situ, and while field campaigns have already undertaken to collect detailed CCN observations, these measurements are spatio-temporally sparse and provide insufficient constraint on global distribution of aerosols [2, 15].

For lack of better observations, the Aerosol Optical Depth (AOD) has been widely adopted as a first order proxy of CCN concentration in an atmospheric column [2, 4, 12]. The AOD is a measure of the extinction of solar radiations through an atmospheric column. It is denoted by τ and defined at

Tackling Climate Change with Machine Learning workshop at NeurIPS 2022.

a given wavelength, time, latitude and longitude by $\tau = \int_0^H b_{\text{ext}}(h)\,\mathrm{d}h$, where b_{ext} is the extinction coefficient[1] and the integral is taken over the height H of an atmospheric column. The AOD is appealing because it is routinely observed on a global scale by satellite products [14] which, as opposed to in situ observations, offer long term global records.

However, the AOD is a column-integrated quantity and does not provide information on the vertical distribution of aerosols. This is limiting since their vertical distribution strongly influences both the magnitude and the sign of the forcing. Stier [18] highlights the importance of determining aerosol vertical distributions to provide stronger constraints on CCN at specific altitudes. In particular, the AOD fails to describe near-surface properties such as the concentration of aerosols in the lowest part of the atmosphere, called the *boundary layer*.

In this work, we propose to probe whether AOD observations can be used to constrain a global prior over aerosol vertical distributions. Formally, given an AOD observation τ, we want to reconstruct the corresponding extinction coefficient profile b_{ext}. Using Gaussian processes (GPs) [13], we design a Bayesian model that maps vertically-resolved meteorological variables that are readily available (e.g. temperature, relative humidity) to a probabilistic estimate of b_{ext}. This probabilistic estimates then integrates into the AOD. The model formulation is simple and makes assumptions explicit, hence granting control and interpretability over predictions while offering built-in uncertainty quantification.

We use ECHAM-HAM global aerosol-climate model simulation data [16, 17, 23] to validate our approach. While our prime motivation is to reconstruct b_{ext} from satellite observations of AOD, the intricacies of combining measurements from different instruments makes it challenging to validate any proposed methodology. On the other hand, ECHAM-HAM is a self-consistent climate model that offers readily available aerosol vertical profiles, and is better suited for model development.

We demonstrate our model is able to reconstruct natural patterns that arise in aerosol vertical distribution, in particular in the boundary layer. We show that very simple and readily available meteorological predictors suffice to obtain a good estimation of the extinction coefficient.

2 Model Design

2.1 Design of the prior

In passive satellite sensors, AOD retrieval algorithms assume a form for vertical profiles, which in the simplest case takes the exponential form $b_{\text{ext}}(h) \propto e^{-h/L}$ [8, 22]. L is a fixed height scale parameter that is typically taken as the top altitude of the boundary layer ($\approx$ 2km). These profiles capture a key element of aerosol vertical distribution: most CCN lies at low altitude in the boundary layer.

Drawing from this, we propose to model the extinction coefficient b_{ext} by weighting an idealized exponential vertical profile with a positive weight function $w : \mathbb{R}^d \to (0, +\infty)$. Namely, let $x|h$ denotes a d-dimensional vector resulting from the concatenation of spatiotemporal and meteorological variables — such as temperature, pressure, humidity — for a given altitude h. The prior for the extinction coefficient profile is denoted φ and takes the simple form $\varphi(x|h) = w(x|h)e^{-h/L}$.

This weight function is meant to capture finer details of variability in the extinction coefficient profile, putting more mass in regions where meteorological predictors suggest aerosol loading is likely to be higher.

To account for the non-linearity and the epistemic uncertainty over the relationship between $x|h$ and the extinction coefficient $b_{\text{ext}}(h)$, we propose a Bayesian formulation of the weighting function. We place a GP prior over the weight function and wrap it with an exponential transform to ensure the weights are strictly positive. Formally, we model the weight function as $w(x|h) = e^{f(x|h)}$ where $f \sim \mathrm{GP}(m, k)$.

2.2 Choice of the observation model

Since the AOD is strictly positive and highly-skewed toward small values (≈ 0.14), the log-normal distribution has been reported to provide a good fit, e.g. for studies focusing on locations in North

[1] the sum of contribution from particle-light scattering plus absorption of light by particles

Model formulation for the i^{th} atmospheric column

Observation Model:

$$\tau_i|\eta_i \sim \mathcal{LN}\left(\log \eta_i - \frac{\sigma^2}{2}, \sigma\right)$$

$$\eta_i = \int_0^H \varphi(x_i|h)\,\mathrm{d}h$$

Prior:

$$\varphi(x_i|h) = e^{f(x_i|h) - h/L}$$

$$f \sim \mathrm{GP}(m, k)$$

τ_i	Observed AOD
$\mathcal{LN}$	Log-normal distribution
η_i, σ	Log-normal mean and scale parameters
φ	Prior for b_{ext}
$x_i\|h$	Input covariates at altitude h
H	Atmospheric column height
L	Idealized profile heightscale parameter
f	GP prior with mean m and kernel k

Figure 1: Observation model and prior formulation for the i^{th} atmospheric column.

America and Europe [11]. We thus propose a log-normal observation model for AOD observations, which we denote $\tau|\mu \sim \mathcal{LN}(\mu, \sigma)$.

We connect the vertical extinction profile φ to the AOD τ through its mean as $\eta := \mathbb{E}[\tau] = \int_0^H \varphi(x|h)\,\mathrm{d}h$. Using the fact that $\eta = e^{\mu+\sigma^2/2} \Leftrightarrow \mu = \log \eta - \sigma^2/2$, we can reparametrise the model to express it in terms of the log-normal mean η. The complete model description for the i^{th} atmospheric column is summarized in Figure 1.

2.3 Posterior inference

Updating our prior with AOD observations means computing the posterior distribution of $\varphi|\tau$. Since φ results from a transformation of the GP, we will focus on the on the posterior distribution of f for convenience, which is given by $p(\mathrm{f}|\tau) = \frac{p(\tau|\mathrm{f})p(\mathrm{f})}{\int_{\mathbb{R}} p(\tau|\mathrm{f})p(\mathrm{f})\,\mathrm{df}}$. Unfortunately, because of the log-normal observation model, the integral denominator is not available in closed-form, making the posterior intractable.

We thus propose to use a variational approximation scheme [7, 10, 19] to substitute this intractable inference problem with a tractable optimization problem. In addition, we make the variational approximation sparse [19] such that the model can scale to large amounts of data. A detailed derivation of the variational approximation is provided in Appendix A.

3 Experiments

3.1 Dataset and experimental setup

The aerosol-climate model ECHAM-HAM is a self-consistent global climate model of aerosol radiative properties and CCN which demonstrates excellent agreement with AOD measurements from ground-based sun-photometers and satellite retrievals [16]. The simulation used includes aerosols optical properties and meteorological variables at a $1.8° \times 1.8°$ horizontal resolution and over 31 levels of vertical resolution and for 8 regularly spaced time steps over a day (06/06/2008). The resulting dataset counts 147,456 atmospheric columns used as training points. Its detailed dimensions are reported in Appendix B

We use the AOD at 550nm as the response variable to vertically disaggregate and the extinction coefficient at 533nm as the groundtruth variable to evaluate our predictions against. To select the vertically resolved variables used as predictors, we limit ourselves to standard meteorological variables that could be easily obtained from reanalysis data: temperature T, pressure P, relative humidity RH and updraft ω. Indeed, while aerosol satellite imagery do not provide any vertically resolved measurements of meteorological variables, the latter can be reliably extracted on different pressure-levels from reanalysis data or from atmospheric sounders. The input variable writes $x = (t, \mathrm{lat}, \mathrm{lon}, T, P, \mathrm{RH}, \omega)$. The data and code are openly available [3].

Table 1: Comparison of our method to an idealized exponential baseline for deterministic (RMSE, Bias, Bias98) and probabilistic (ELBO, ICI) metrics; "Entire column" means scores are computed and averaged for every altitude levels; "Boundary layer" means scores are computed and averaged for altitude levels of the boundary level only (<2 km); ↓ : closer to 0 means better; ↑ : higher means better; runs with our method are averaged over 5 seeds; we report 1 standard deviation.

Region	Method	RMSE (10^{-5})↓	Bias (10^{-6})↓	Bias98 (10^{-5})↓	ELBO↑	ICI (10^{-2})↓
Entire column	Our method	**3.29**±0.02	**-0.167**±0.105	**-0.646**±0.151	**13.1**±0.1	5.29±0.59
	Idealized	4.10	-2.40	-4.08	13.1	**5.05**
Boundary layer	Our method	**6.06**±0.03	**-1.25**±0.45	**-4.64**±0.32	**10.6**±0.1	**8.27**±0.29
	Idealized	7.55	-12.9	-11.7	10.2	19.1

3.2 Results

We compare the predicted extinction coefficient profile to the extinction coefficient from ECHAM-HAM simulations. We use as a comparative baseline an idealized exponential profile. We report results in Table 1 in terms of deterministic metrics — root mean square error (RMSE), mean bias (Bias), 98$^{\text{th}}$ quantile bias (Bias98) — and probabilistic metrics — evidence lower bound (ELBO), integrated calibration index (ICI). Details on the metrics are provided in Appendix C

The posterior mean profile arising from the proposed method outperforms the idealized exponential baseline for all deterministic metrics. Evaluating over the entire column consistently yields better scores than over the boundary layer only. This is to be expected as the extinction coefficient outside the boundary layer tends to vanish to zero and most of the variability happens within the boundary layer.

The probabilistic metrics are comparable for both methods when computed over the entire column. For the boundary layer however, our method outperforms the baseline, with a significant difference for the ICI. This suggests that, in the boundary layer, the predicted posterior probability distribution with our method is a sounder fit to the ECHAM-HAM extinction coefficient profiles.

In Appendix D, we display vertical slices at fixed latitude of the ECHAM-HAM extinction coefficient and the corresponding predicted extinction coefficient. For comparison, we also display a prediction with the idealized exponential baseline. We observe that our predicted mean profile is able to reconstruct extinction patterns that are visually very similar the the groundtruth extinction coefficient. In comparison with the idealized exponential baseline, the extinction profiles predicted with our method look much more realistic. This is encouraging given the only aerosol optical property used is the AOD. In particular, since aerosols water uptake is related to relative humidity, we observe a good capacity to recover patterns corresponding to extinction due to aerosol water uptake in the boundary layer. We also observe that tail extinction coefficient values are consistently captured within the 95% confidence region of the posterior distribution.

4 Conclusion

We introduce a simple GP-based methodology to vertically disaggregate the AOD using simple vertically resolved meteorological predictors such as temperature, pressure or relative humidity and demonstrate its successful application to the reconstruction of ECHAM-HAM simulated extinction profiles. Our model outperforms an idealized baseline and displays capacity to recover realistic extinction patterns, in particular for extinction patterns arising from aerosol water uptake in the boundary layer.

In future work we intend to apply our model to AOD arising from satellite observations and meteorological predictors from reanalysis data. Since observations of groundtruth extinction coefficient are not available in 2D satellite products, we intend to collocate observations from MODIS 2D AOD product [14] with CALIOP vertical lidar measurements [21] to validate the model.

References

[1] Bruce A. Albrecht. Aerosols, Cloud Microphysics, and Fractional Cloudiness. *Science*, 1989.

[2] M. O. Andreae. Correlation between cloud condensation nuclei concentration and aerosol optical thickness in remote and polluted regions. *Atmospheric Chemistry and Physics*, 2009.

[3] Anonymous Author(s). `https://anonymous.4open.science/r/aodisaggregation-BBBD`, 2022.

[4] Antony Clarke and Vladimir Kapustin. Hemispheric Aerosol Vertical Profiles: Anthropogenic Impacts on Optical Depth and Cloud Nuclei. *Science*, 2010.

[5] Durk P Kingma, Tim Salimans, and Max Welling. Variational Dropout and the Local Reparameterization Trick. In *Advances in Neural Information Processing Systems*, 2015.

[6] Ho Chung Law, Peilin Zhao, Leung Sing Chan, Junzhou Huang, and Dino Sejdinovic. Hyperparameter Learning via Distributional Transfer. In *Advances in Neural Information Processing Systems*, 2019.

[7] Felix Leibfried, Vincent Dutordoir, ST John, and Nicolas Durrande. A tutorial on sparse Gaussian processes and variational inference. *arXiv preprint arXiv:2012.13962*, 2020.

[8] Chong Li, Jing Li, Oleg Dubovik, Zhao-Cheng Zeng, and Yuk L. Yung. Impact of Aerosol Vertical Distribution on Aerosol Optical Depth Retrieval from Passive Satellite Sensors. *Remote Sensing*, 2020.

[9] Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.). *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2021.

[10] Alexander G. de G. Matthews, James Hensman, Richard Turner, and Zoubin Ghahramani. On Sparse Variational Methods and the Kullback-Leibler Divergence between Stochastic Processes. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, 2016.

[11] O'Neill N. T., A. Ignatov, B. N. Holben, and T. F. Eck. The lognormal distribution as a reference for reporting aerosol optical depth statistics; Empirical tests using multi-year, multi-site AERONET Sunphotometer data. *Geophysical Research Letters*, 2000.

[12] Teruyuki Nakajima, Akiko Higurashi, Kazuaki Kawamoto, and Joyce E. Penner. A possible correlation between satellite-derived cloud and aerosol microphysical parameters. *Geophysical Research Letters*, 2001.

[13] C Rasmussen and C Williams. Gaussian process for machine learning, 2005.

[14] 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.

[15] 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*, 2011.

[16] P. Stier, J. Feichter, S. Kinne, S. Kloster, E. Vignati, J. Wilson, L. Ganzeveld, I. Tegen, M. Werner, Y. Balkanski, M. Schulz, O. Boucher, A. Minikin, and A. Petzold. The aerosol-climate model ECHAM5-HAM. *Atmospheric Chemistry and Physics*, 2005.

[17] P. Stier, J. H. Seinfeld, S. Kinne, and O. Boucher. Aerosol absorption and radiative forcing. *Atmospheric Chemistry and Physics*, 2007.

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

[19] Michalis K. Titsias. Variational Learning of Inducing Variables in Sparse Gaussian Processes. In *AISTATS*, 2009.

[20] S. Twomey. The Influence of Pollution on the Shortwave Albedo of Clouds. *Journal of Atmospheric Sciences*, 1977.

[21] 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*, 2013.

[22] Yerong Wu, Martin de Graaf, and Massimo Menenti. The impact of aerosol vertical distribution on aerosol optical depth retrieval using CALIPSO and MODIS data: Case study over dust and smoke regions. *Journal of Geophysical Research: Atmospheres*, 2017.

[23] K. Zhang, D. O'Donnell, J. Kazil, P. Stier, S. Kinne, U. Lohmann, S. Ferrachat, B. Croft, J. Quaas, H. Wan, S. Rast, and J. Feichter. The global aerosol-climate model ECHAM-HAM, version 2: sensitivity to improvements in process representations. *Atmospheric Chemistry and Physics*, 2012.

A Variational approximation of the posterior

A.1 Finite-sample problem formulation

Assume we observe the AOD for n independent columns, which we stack into the vector $\boldsymbol{\tau} = \begin{bmatrix}\tau_1 & \dots & \tau_n\end{bmatrix}^\top \in \mathbb{R}^n$. For the i^{th} column, we also observe m_i vertically resolved meteorological covariates $x_i^{(1)}, \dots, x_i^{(m_i)}$ and their respective altitudes $h_i^{(1)} < \dots < h_i^{(m_i)}$, such that $x_i^{(j)} \sim p(x|h_i^{(j)})$. We concatenate these observations into a dataset $\mathcal{D} = \left\{\left(x_i^{(j)}, h_i^{(j)}\right)_{j=1}^{m_i}, \tau_i\right\}_{i=1}^{n}$ and denote $M = \sum_{i=1}^n m_i$ the total number of vertically resolved samples.

If f denotes a realization of f at any input $x|h$, the posterior distribution of $f(x|h)$ given observations is denoted $p(\mathrm{f}|\boldsymbol{\tau})$. Since $\varphi(x|h) = e^{f(x|h)}$, the prior $\varphi(x|h)$ is a log-normal random variable given by

$$\varphi(x|h) = e^{f(x|h)-h/L} \sim \mathcal{LN}\left(m(x|h) - h/L, k(x|h,x|h)^{1/2}\right). \tag{1}$$

A direct consequence is that we can obtain analytical expressions of its mean and variance, now given by

$$\mathbb{E}[\varphi(x|h)] = e^{m(x|h) - \frac{h}{L} + \frac{1}{2}k(x|h,x|h)}. \tag{2}$$

$$\mathrm{Var}(\varphi(x|h)) = \left(e^{k(x|h,x|h)} - 1\right) e^{2m(x|h) - 2\frac{h}{L} + k(x|h,x|h)}. \tag{3}$$

Hence, this also applies to the posterior and having access to the posterior $p(\mathrm{f}|\boldsymbol{\tau})$ allows to compute the predictive means and variance of $\varphi(x|h)|\boldsymbol{\tau}$ using the posterior mean and variance of $p(\mathrm{f}|\boldsymbol{\tau})$.

A.2 A sparse variational approximation of the posterior

The predictive posterior distribution of interest is given by

$$p(\mathrm{f}|\boldsymbol{\tau}) = \frac{p(\boldsymbol{\tau}|\mathrm{f})p(\mathrm{f})}{\int_{\mathbb{R}} p(\boldsymbol{\tau}|\mathrm{f})p(\mathrm{f})\,\mathrm{df}}. \tag{4}$$

The integral denominator is not available in closed-form, making the posterior intractable. We propose to use a variational approximation scheme [7, 10, 19] to substitute this intractable inference problem with a tractable optimization problem. In addition, we make the variational approximation sparse [19] such that the model can scale to large amounts of data.

Let $\mathbf{w} = \begin{bmatrix}w_1 & \dots & w_p\end{bmatrix}^\top \in \mathbb{R}^p$ be a set of $p \ll M$ inducing locations over the space of inputs. Their evaluation by the GP follows a multivariate normal distribution $f(\mathbf{w}) \sim \mathcal{N}(0, \mathbf{K_{ww}})$, where $\mathbf{K_{ww}} = k(\mathbf{w}, \mathbf{w})$. We denote $\mathbf{u} = f(\mathbf{w}) \in \mathbb{R}^p$ and refer to this vector as *inducing variables*.

A p-dimensional parametric distribution is then set over these inducing variables. We choose this distribution as a multivariate normal defined by $q(\mathbf{u}) := \mathcal{N}(\mathbf{u}|\boldsymbol{\mu}_\mathbf{w}, \boldsymbol{\Sigma}_\mathbf{w})$. $\boldsymbol{\mu}_\mathbf{w}, \boldsymbol{\Sigma}_\mathbf{w}$ are called the *variational parameters* and need to be tuned such that $q(\mathbf{u})$ best approximates the true posterior $p(\mathbf{u}|\boldsymbol{\tau})$.

Once this is achieved, we take as an approximation to $p(\mathrm{f}|\boldsymbol{\tau})$ the variational posterior defined by $q(\mathrm{f}) := \int_{\mathbb{R}^p} p(\mathrm{f}|\mathbf{u})p(\mathbf{u})\,\mathrm{d}\mathbf{u}$, which is given in closed-form by

$$q(\mathrm{f}) = \mathcal{N}(\mathrm{f}|\bar{\boldsymbol{\mu}}_{x|h}, \bar{\boldsymbol{\Sigma}}_{x|h}) \tag{5}$$

$$\bar{\boldsymbol{\mu}}_{x|h} = k(x|h, \mathbf{w})\mathbf{K}_{\mathbf{ww}}^{-1}\boldsymbol{\mu}_\mathbf{w} \tag{6}$$

$$\bar{\boldsymbol{\Sigma}}_{x|h} = k(x|h, x|h) - k(x|h, \mathbf{w})\left(\mathbf{K}_{\mathbf{ww}}^{-1} - \mathbf{K}_{\mathbf{ww}}^{-1}\boldsymbol{\Sigma}_\mathbf{w}\mathbf{K}_{\mathbf{ww}}^{-1}\right)k(\mathbf{w}, x|h). \tag{7}$$

Naturally, (5) can be extended to describe a variational posterior over multiple GP entries. Namely, let $\mathbf{x}^* \in \mathbb{R}^D$ denote a vector of input entries. If $\mathbf{f}^* \in \mathbb{R}^D$ denotes a realization of $f(\mathbf{x}^*)$, then the associated variational posterior is given by

$$q(\mathbf{f}^*) = \mathcal{N}(\mathbf{f}^*|\bar{\boldsymbol{\mu}}_{\mathbf{x}^*}, \bar{\boldsymbol{\Sigma}}_{\mathbf{x}^*}) \tag{8}$$

$$\bar{\boldsymbol{\mu}}_{\mathbf{x}^*} = k(\mathbf{x}^*, \mathbf{w})\mathbf{K}_{\mathbf{ww}}^{-1}\boldsymbol{\mu}_\mathbf{w} \tag{9}$$

$$\bar{\boldsymbol{\Sigma}}_{\mathbf{x}^*} = k(\mathbf{x}^*, \mathbf{x}^*) - k(\mathbf{x}^*, \mathbf{w})\left(\mathbf{K}_{\mathbf{ww}}^{-1} - \mathbf{K}_{\mathbf{ww}}^{-1}\boldsymbol{\Sigma}_\mathbf{w}\mathbf{K}_{\mathbf{ww}}^{-1}\right)k(\mathbf{w}, \mathbf{x}^*). \tag{10}$$

The sparse nature of this approach becomes apparent in (9) and (10). Indeed, regardless of the number of samples we wish to evaluate the variational posterior over, we only need to invert a $p \times p$ matrix, incurring a $\mathcal{O}(p^3)$ computational cost.

A.3 Learning the variational parameters $\boldsymbol{\mu}_{\mathbf{w}}, \boldsymbol{\Sigma}_{\mathbf{w}}$

As mentioned above, the variational parameters $\boldsymbol{\mu}_{\mathbf{w}}, \boldsymbol{\Sigma}_{\mathbf{w}}$ need to be tuned such that $q(\mathbf{u})$ best approximates the posterior $p(\mathbf{u}|\boldsymbol{\tau})$, which is intractable. This problem is casted as the maximisation of an objective called the evidence lower-bound (ELBO), given by

$$\mathrm{ELBO}(q) = \mathbb{E}_{q(\mathbf{f})}[\log p(\boldsymbol{\tau}|\mathbf{f})] - \mathrm{KL}(q(\mathbf{u})||p(\mathbf{u})). \tag{11}$$

The ELBO is a lower-bound to the marginal log-likelihood $\log p(\boldsymbol{\tau}) \geq \mathrm{ELBO}(q)$. It can thus be used as a proxy of $\log p(\boldsymbol{\tau})$ to also tune the model hyperparameters.

The second term in (11) is the Kullback-Leibler divergence between two multivariate normal distributions. It admits a closed-form expression and can be computed. The first term, on the other hand, is an expected log-likelihood under the variational posterior which cannot be analytically computed. It can be decomposed into column-wise terms

$$\mathbb{E}_{q(\mathbf{f})}[\log p(\boldsymbol{\tau}|\mathbf{f})] = \sum_{i=1}^{n} \mathbb{E}_{q(\mathbf{f}_i)} \left[\log \mathcal{LN}\left(\tau_i | \log \eta_i - \frac{\sigma^2}{2}, \sigma\right)\right], \tag{12}$$

where $\mathbf{f}_i = \begin{bmatrix} \mathbf{f}_i^{(1)} & \dots & \mathbf{f}_i^{(m_i)} \end{bmatrix}^\top \in \mathbb{R}^{m_i}$ corresponds to the GP realization over the i^{th} column only.

To estimate (12), we must first evaluate the mean parameter $\eta_i = \int_0^H \psi(f(x_i|h))e^{-h/L}\,\mathrm{d}h$ with the finite number of GP evaluations $\mathbf{f}_i$ we have access to. We propose to use the trapezoidal integration scheme given by

$$\hat{\eta}_i = \sum_{j=1}^{m_i-1} \frac{\exp(\mathbf{f}_i^{(j+1)} - h_i^{(j+1)}/L) - \exp(\mathbf{f}_i^{(j)} - h_i^{(j)}/L)}{2} \left(h_i^{(j+1)} - h_i^{(j)}\right). \tag{13}$$

While we choose the trapezoidal rule for simplicity, we note that alternative finite-sample integration schemes can be chosen here in accordance with the needs.

Second, because of the log-normal observation model, the expected log-likelihood remains intractable. To estimate it, while allowing backpropagation through the variational parameters, we use a reparametrization trick [5]. Namely, we sample $\boldsymbol{\epsilon}_i \sim \mathcal{N}(0, \mathbf{I}_{m_i})$ and compute $\mathbf{f}_i = \bar{\boldsymbol{\mu}}_i + \bar{\boldsymbol{\Sigma}}_i^{1/2}\boldsymbol{\epsilon}_i$, where $\bar{\boldsymbol{\mu}}_i, \bar{\boldsymbol{\Sigma}}_i$ are the variational posterior parameters for the i^{th} column and are obtained by application of (9 and (10) over the predictors of the i^{th} column. The resulting GP sample $\mathbf{f}_i$ is then used to estimate the mean parameter $\hat{\eta}_i$ following (13) and we can approximate the expected log-likelihood with its one-sample estimate

$$\mathbb{E}_{q(\mathbf{f}_i)} \left[\log \mathcal{LN}\left(\tau_i | \log \eta_i - \frac{\sigma^2}{2}, \sigma\right)\right] \approx \log \mathcal{LN}\left(\tau_i | \hat{\eta}_i - \frac{\sigma^2}{2}, \sigma\right). \tag{14}$$

This method allows to estimate the ELBO objective, which in turn can be maximised with respect to the variational parameters $\boldsymbol{\mu}_{\mathbf{w}}, \boldsymbol{\Sigma}_{\mathbf{w}}$ using a stochastic gradient approach for example. The model hyperparameters can also be tuned jointly with this objective. These include the log-normal scale σ or the kernel k hyperparameters (e.g. variances and lengthscales), with an option to parametrize these kernels using feature maps given by deep neural networks [6]. As it is standard in sparse variational GPs [19], we will also learn the inducing locations $\mathbf{w}$.

B Dataset

	Name	Notation	Dimensions
Predictors	Temperature	T	(t, lat, lon, lev)
	Pressure	P	(t, lat, lon, lev)
	Relative humidity	RH	(t, lat, lon, lev)
	Vertical velocity	ω	(t, lat, lon, lev)
Response	AOD 550nm	τ	(t, lat, lon)
Groundtruth	Extinction coefficient 533nm	b_{ext}	(t, lat, lon, lev)

Table 2: Gridded variables from ECHAM-HAM simulation data. The grid includes 8 time steps (t), 96 latitude levels (lat), 192 longitude levels (lon) and 31 vertical pressure levels (lev) — which is a proxy for h. Our objective is to vertically disaggregate the response τ using the vertically resolved predictors (T, P, RH, ω) and spatiotemporal columns locations (t, lat, lon).

C Evaluation metrics

The integrated calibration index is defined as

$$\text{ICI} = \int_0^1 |N(\alpha) - (1-\alpha)| \, \mathrm{d}\alpha, \tag{15}$$

where $N(\alpha)$ is the percentage of ECHAM-HAM extinction coefficient observations that fall within the $(1-\alpha)$ credible interval of the distribution of $b_{\text{ext}}|\boldsymbol{\tau}$.

Table 3: Evaluation metrics; Deterministic metrics compare the predicted posterior mean $\mathbb{E}[b_{\text{ext}}|\boldsymbol{\tau}]$ to the ECHAM-HAM extinction coefficient; Probabilistic metrics evaluate the complete predicted posterior probability distribution of $b_{\text{ext}}|\boldsymbol{\tau}$ against ECHAM-HAM extinction coefficient.

	Metric	Description	Best when
Deterministic	RMSE	Root mean square error	close to 0
	Bias	Mean bias	close to 0
	Bias98	Bias in the 98th percentile	close to 0
Probabilistic	ELBO	Evidence lower-bound of groundtruth b_{ext}	higher
	ICI	Integrated calibration index — see (15)	close to 0

D Vertical slices plots

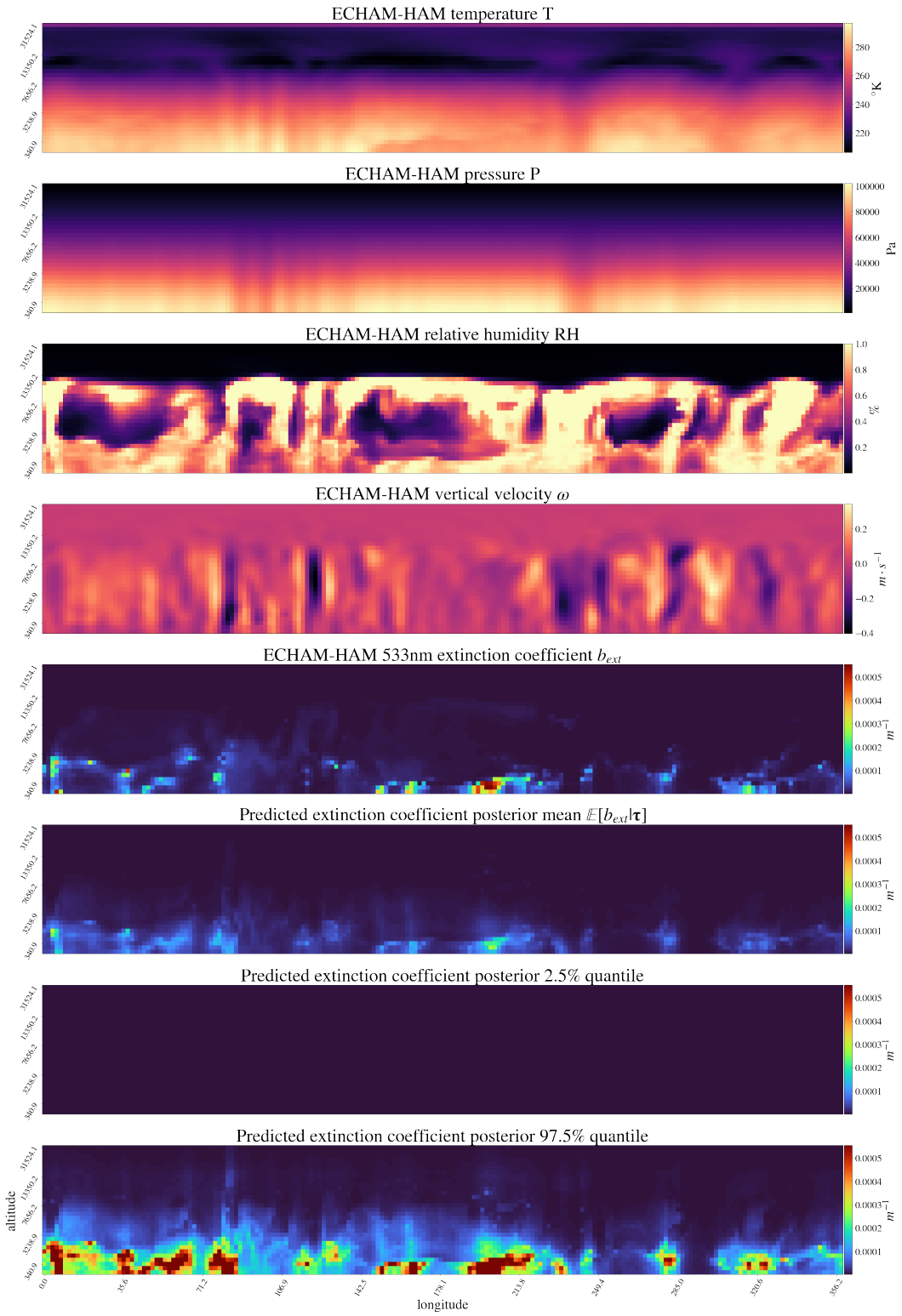

Figure 2: Vertical slices at latitude 51.29° of meteorological predictors (T, P, RH, ω), groundtruth extinction coefficient, predicted extinction coefficient posterior mean, 2.5% and 97.5% quantiles of the predicted extinction coefficient posterior distribution

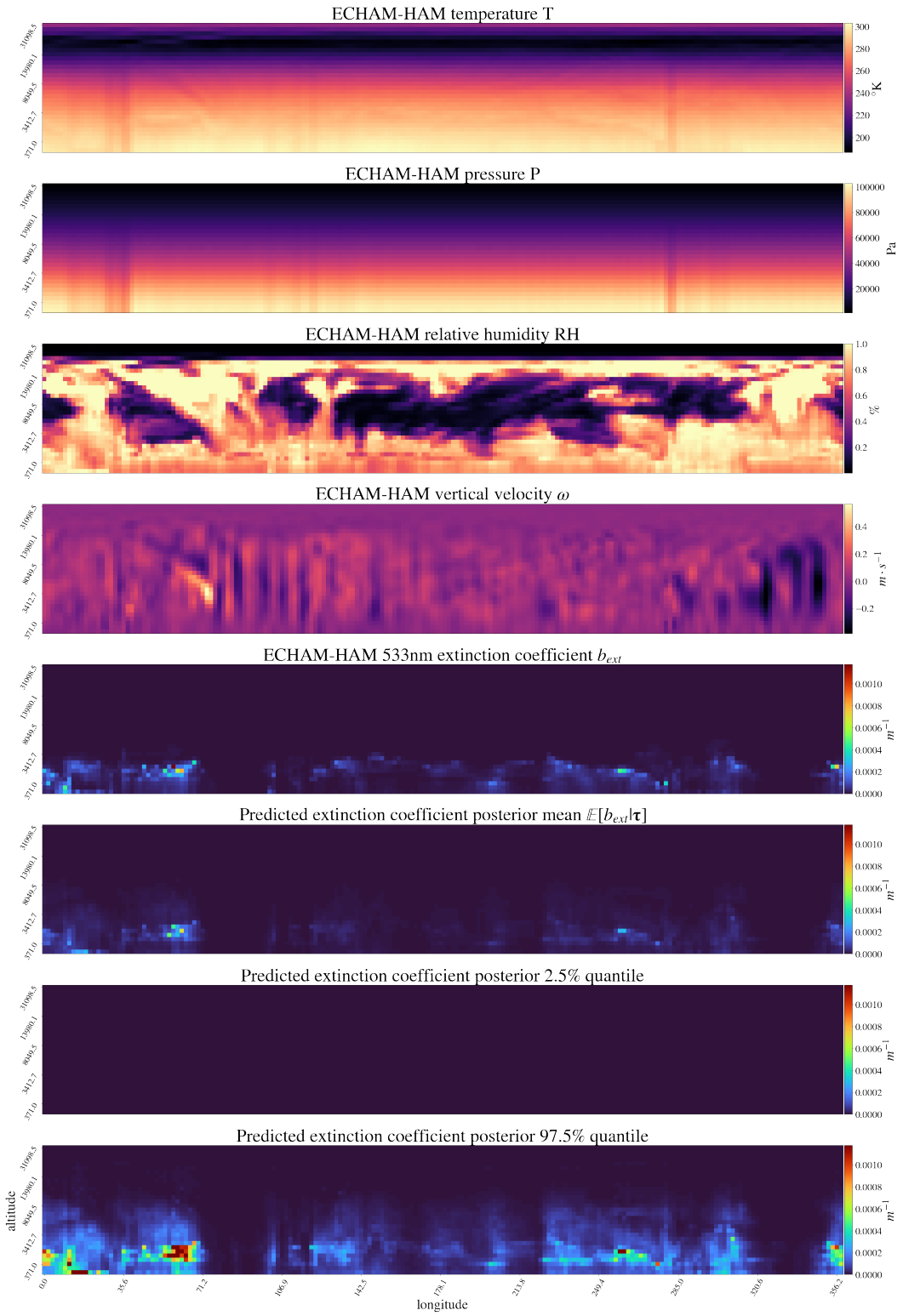

Figure 3: Vertical slices at latitude -0.93° of meteorological predictors (T, P, RH, ω), groundtruth extinction coefficient, predicted extinction coefficient posterior mean, 2.5% and 97.5% quantiles of the predicted extinction coefficient posterior distribution

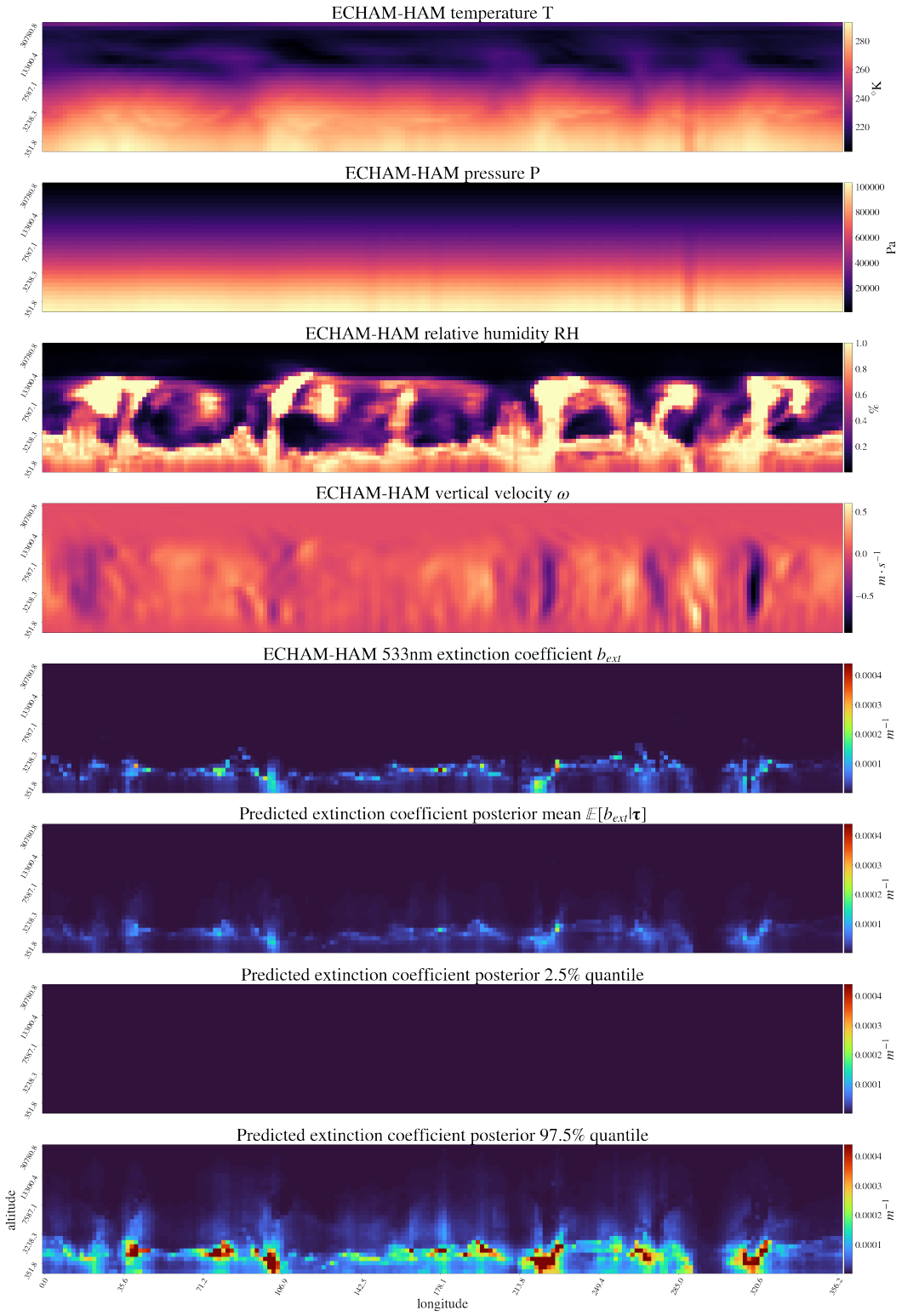

Figure 4: Vertical slices at latitude -38.2° of meteorological predictors (T, P, RH, ω), groundtruth extinction coefficient, predicted extinction coefficient posterior mean, 2.5% and 97.5% quantiles of the predicted extinction coefficient posterior distribution

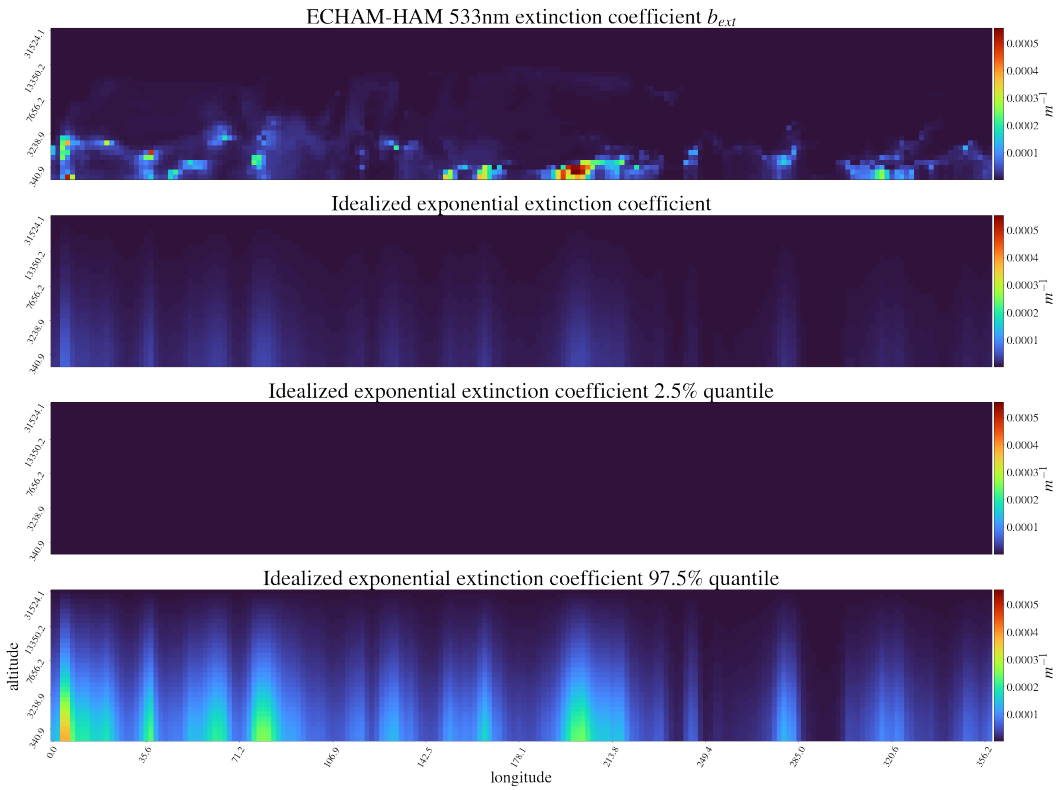

Figure 5: Vertical slices at latitude 51.29° of groundtruth extinction coefficient, idealized exponential extinction coefficient, 2.5% and 97.5% quantiles of the idealized exponential extinction coefficient distribution