Bayesian inference for aerosol vertical profiles

Shahine Bouabid[1] Duncan Watson-Parris[2] Dino Sejdinovic[1]

[1] Department of Statistics, University of Oxford
[2] Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford

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

Motivation

Motivation

- Uncertainty in magnitude of forcing due to ACIs comes from:
  1. Uncertainty in estimation of pre-industrial forcing
  2. Uncertainty in estimation of present day forcing

Motivation

- Uncertainty in magnitude of forcing due to ACIs comes from:
  1. Uncertainty in estimation of pre-industrial forcing
  2. **Uncertainty in estimation of present day forcing**

Motivation

- Uncertainty in magnitude of forcing due to ACIs comes from:
  1. Uncertainty in estimation of pre-industrial forcing
  2. **Uncertainty in estimation of present day forcing**
- Uncertainty in present-day forcing due (in part) to **difficulty to get informative measurements of aerosol at global scale**

Motivation

- Uncertainty in magnitude of forcing due to ACIs comes from:
  1. Uncertainty in estimation of pre-industrial forcing
  2. **Uncertainty in estimation of present day forcing**
- Uncertainty in present-day forcing due (in part) to **difficulty to get informative measurements of aerosol at global scale** $\longrightarrow$ best we can do is AOD.

Motivation

- Uncertainty in magnitude of forcing due to ACIs comes from:
  1. Uncertainty in estimation of pre-industrial forcing
  2. **Uncertainty in estimation of present day forcing**
- Uncertainty in present-day forcing due (in part) to **difficulty to get informative measurements of aerosol at global scale** $\longrightarrow$ best we can do is AOD.
- AOD is a 2D quantity which does **not inform about vertical distribution of aerosols**.

Motivation

- Uncertainty in magnitude of forcing due to ACIs comes from:
  1. Uncertainty in estimation of pre-industrial forcing
  2. **Uncertainty in estimation of present day forcing**
- Uncertainty in present-day forcing due (in part) to **difficulty to get informative measurements of aerosol at global scale** $\longrightarrow$ best we can do is AOD.
- AOD is a 2D quantity which does **not inform about vertical distribution of aerosols**.
- Vertical distribution of aerosols **changes magnitude and even sign of the forcing** .

Motivation

- Uncertainty in magnitude of forcing due to ACIs comes from:
  1. Uncertainty in estimation of pre-industrial forcing
  2. **Uncertainty in estimation of present day forcing**
- Uncertainty in present-day forcing due (in part) to **difficulty to get informative measurements of aerosol at global scale** $\longrightarrow$ best we can do is AOD.
- AOD is a 2D quantity which does **not inform about vertical distribution of aerosols**.
- Vertical distribution of aerosols **changes magnitude and even sign of the forcing** .

Objective

Try to reconstruct aerosol vertical profiles using AOD

2D Spatial Disaggregation

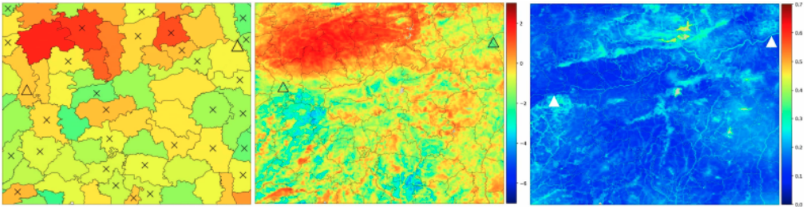


Figure 1: **Left**: regional Malaria incidence rate; **Center**: Spatially disaggregated mean Malaria incidence rate; **Right**: Standard deviation over spatially disaggreagted rate; Law et al. [1].

2D Spatial Disaggregation

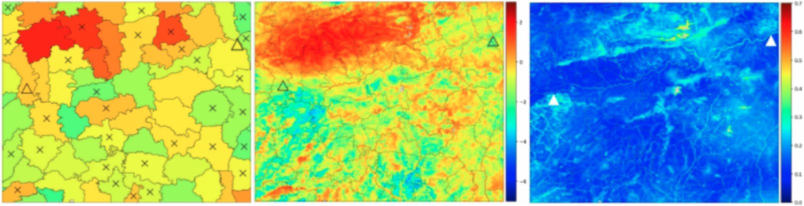


Figure 1: **Left**: regional Malaria incidence rate; **Center**: Spatially disaggregated mean Malaria incidence rate; **Right**: Standard deviation over spatially disaggreagted rate; Law et al. [1].

$$\text{rate}_{\text{region}} = \text{Aggregate}_{\text{region}} \left\{ \text{rate}_{\text{fine-grid}} \left(x_{\text{fine-grid}} \right) \right\}$$

2D Spatial Disaggregation

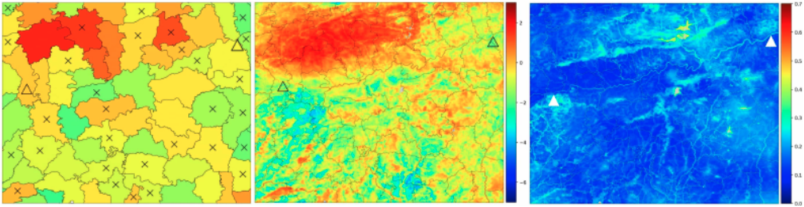


Figure 1: **Left**: regional Malaria incidence rate; **Center**: Spatially disaggregated mean Malaria incidence rate; **Right**: Standard deviation over spatially disaggreagted rate; Law et al. [1].

$$\text{rate}_{\text{region}} = \text{Aggregate}_{\text{region}} \left\{ \text{rate}_{\text{fine-grid}}(x_{\text{fine-grid}}) \right\}$$

- **Observations**: $\text{rate}_{\text{region}}$ and $x_{\text{fine-grid}}$

2D Spatial Disaggregation

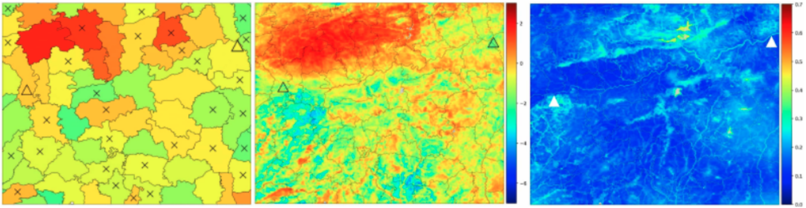
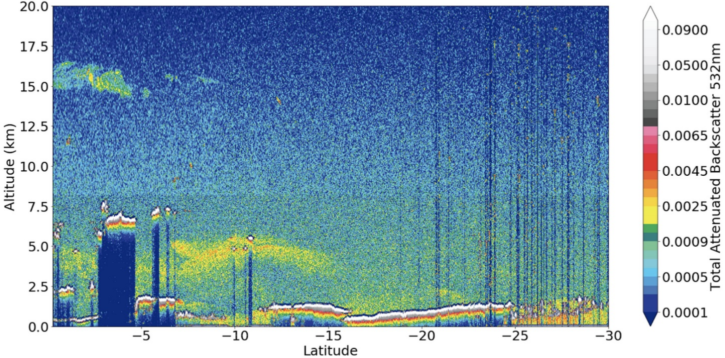


Figure 1: **Left**: regional Malaria incidence rate; **Center**: Spatially disaggregated mean Malaria incidence rate; **Right**: Standard deviation over spatially disaggreagted rate; Law et al. [1].

$$\text{rate}_{\text{region}} = \text{Aggregate}_{\text{region}} \left\{ \text{rate}_{\text{fine-grid}}(x_{\text{fine-grid}}) \right\}$$

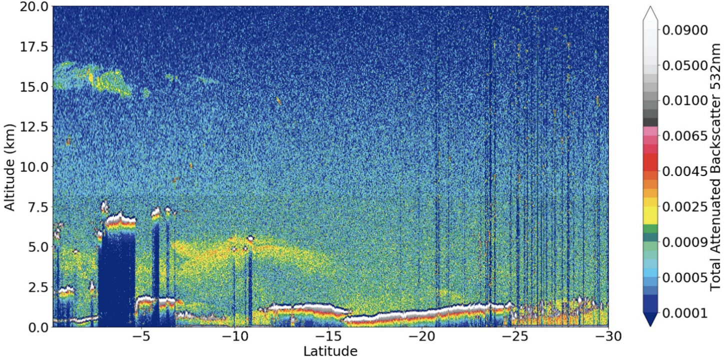
- **Observations**: $\text{rate}_{\text{region}}$ and $x_{\text{fine-grid}}$
- **Goal**: Infer $\text{rate}_{\text{fine-grid}}$ as a function of $x_{\text{fine-grid}}$

Disaggregating along a 3$^{\text{rd}}$ dimension?



$$\tau_{\text{column}} = \text{Aggregate}_{\text{column}} \left\{ b_{\text{ext}} \right\} = \int_0^H b_{\text{ext}}(h)\, \mathrm{d}h$$

Disaggregating along a 3rd dimension?



$$\tau_{\text{column}} = \text{Aggregate}_{\text{column}} \{b_{\text{ext}}\} = \int_0^H b_{\text{ext}}(h)\, \mathrm{d}h$$

- **Observations**: τ_{column} and x_{3D}
- **Goal**: Infer b_{ext} as a function of x_{3D}

Which x_{3D}?

- Use **simple, readily available predictors** such as pressure, temperature, humidity $\longrightarrow$ reanalysis data.

For example, for a given altitude h we can take

$$x = (t, \mathrm{lat}, \mathrm{lon}, P, T, \mathrm{RH}) \tag{1}$$

Which x_{3D}?

- Use **simple, readily available predictors** such as pressure, temperature, humidity ⟶ reanalysis data.

For example, for a given altitude h we can take

$$x = (t, \text{lat}, \text{lon}, P, T, \text{RH}) \tag{1}$$

Objective

Using observations of AOD and vertically-resolved meteorological predictors, we want to estimate aerosol extinction

Design of a prior over b_{ext}

An idealized vertical prior

- Idealized profiles assumed in remote sensing products $b_{\text{ext}}(h) \propto e^{-h/L}$.

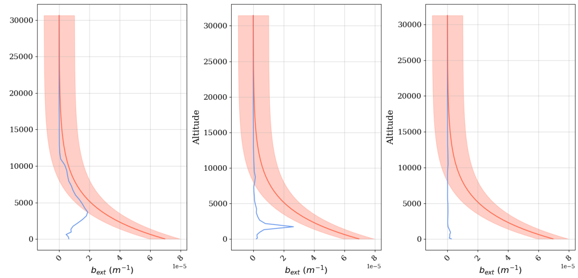


Figure 2: Examples of idealized exponential vertical profiles

An idealized vertical prior

- Idealized profiles assumed in remote sensing products $b_{\text{ext}}(h) \propto e^{-h/L}$.

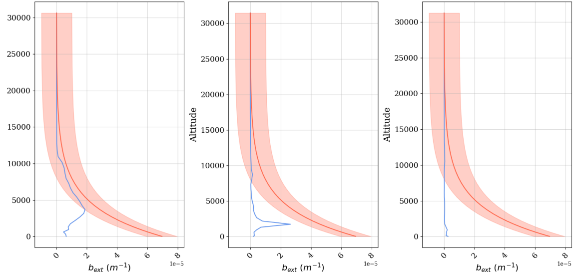


Figure 2: Examples of idealized exponential vertical profiles

- Rough approximation but captures a key structure: **most aerosol lie in *boundary layer*** ($< 2\,km$)

Weighting of the ideal prior

- Propose to weight the idealized exponential profile with a positive weight function $w(x|h) > 0$

$$\varphi(x|h) = w(x|h)e^{-h/L} \tag{2}$$

Weighting of the ideal prior

- Propose to weight the idealized exponential profile with a positive weight function $w(x|h) > 0$

$$\varphi(x|h) = w(x|h)e^{-h/L} \tag{2}$$

- Capture finer details of variability **putting more mass where meteorological predictors suggest higher aerosol loading**

Weighting of the ideal prior

- Propose to weight the idealized exponential profile with a positive weight function $w(x|h) > 0$

$$\varphi(x|h) = w(x|h)e^{-h/L} \tag{2}$$

- Capture finer details of variability **putting more mass where meteorological predictors suggest higher aerosol loading**

Expect relationship between $x|h$ and $b_{\text{ext}}(h)$ to be non-trivial and highly non-linear $\Rightarrow$ learn the weighting $w(x|h)$

Probabilistic modelling of the weight function

- **Lack of knowledge** about influence of meteorological variables
  $\rightarrow$ epistemic uncertainty

Probabilistic modelling of the weight function

- **Lack of knowledge** about influence of meteorological variables $\rightarrow$ epistemic uncertainty
- Reflect this with **Bayesian design** of $w(x|h)$

$$
\begin{aligned}
w(x|h) &= \psi(f(x|h)) \qquad (3) \\
f &\sim \mathrm{GP}(m, k) \qquad (4) \\
\psi &> 0 \qquad (5)
\end{aligned}
$$

Probabilistic modelling of the weight function

- **Lack of knowledge** about influence of meteorological variables $\rightarrow$ epistemic uncertainty
- Reflect this with **Bayesian design** of $w(x|h)$

$$w(x|h) = \psi(f(x|h)) \tag{3}$$
$$f \sim \mathrm{GP}(m, k) \tag{4}$$
$$\psi > 0 \tag{5}$$

- Simple choice $\psi = \exp$
- $\psi \circ f$ describes expressive range of probability distribution over complex positive functions
- Remains interpretable (kernel user-specified determines covariance and functional smoothness)

Connecting $\varphi(x|h)$ to observations

Need for an observation model

- We observe AOD $\tau \to$ ideally we want exactly $\tau = \int_0^H \varphi(x|h)\,\mathrm{d}h$
- Unrealistic because observations are likely noisy

Need for an observation model

- We observe AOD $\tau \to$ ideally we want exactly $\tau = \int_0^H \varphi(x|h)\,\mathrm{d}h$
- Unrealistic because observations are likely noisy

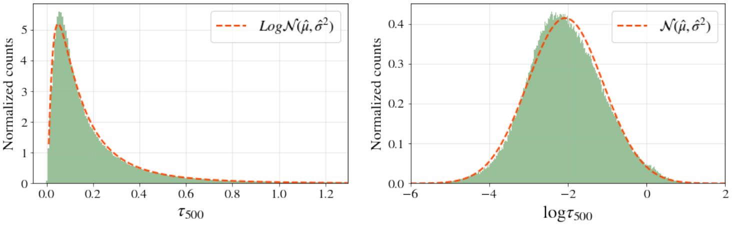


Figure 3: **Left**: empirical distribution of AOD retrievals from AERONET stations; **Right**: logspace version of left plot

Need for an observation model

- We observe AOD $\tau \rightarrow$ ideally we want exactly $\tau = \int_0^H \varphi(x|h)\,\mathrm{d}h$
- Unrealistic because observations are likely noisy

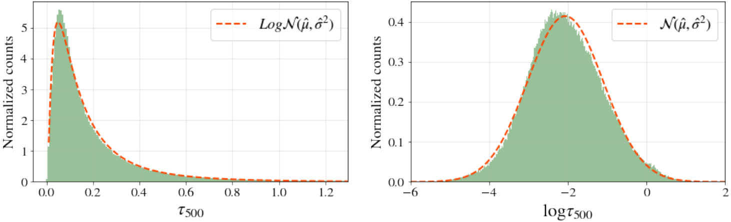


Figure 3: **Left**: empirical distribution of AOD retrievals from AERONET stations; **Right**: logspace version of left plot

Assume **log-normal** model $\tau|\mu, \sigma \sim \mathcal{LN}(\mu, \sigma)$.

Connect with $\varphi(x|h)$

- Use mean reparametrization of $\mathcal{LN}(\mu, \sigma)$.

Connect with $\varphi(x|h)$

- Use mean reparametrization of $\mathcal{LN}(\mu, \sigma)$.

  $\eta = \mathbb{E}[\mathcal{LN}(\mu, \sigma)] = e^{\mu + \sigma^2/2} \Rightarrow \eta = e^{\mu + \sigma^2/2}$

Connect with $\varphi(x|h)$

- Use mean reparametrization of $\mathcal{LN}(\mu, \sigma)$.

$$\eta = \mathbb{E}[\mathcal{LN}(\mu, \sigma)] = e^{\mu + \sigma^2/2} \Rightarrow \eta = e^{\mu + \sigma^2/2}$$
$$\Rightarrow \mu = \log \eta - \frac{\sigma^2}{2}$$

Connect with $\varphi(x|h)$

- Use mean reparametrization of $\mathcal{LN}(\mu, \sigma)$.

$$\eta = \mathbb{E}[\mathcal{LN}(\mu, \sigma)] = e^{\mu+\sigma^2/2} \Rightarrow \eta = e^{\mu+\sigma^2/2}$$
$$\Rightarrow \mu = \log \eta - \frac{\sigma^2}{2}$$
$$\Rightarrow \mathcal{LN}(\mu, \sigma) = \mathcal{LN}\left(\log \eta - \frac{\sigma^2}{2}, \sigma\right)$$

Connect with $\varphi(x|h)$

- Use mean reparametrization of $\mathcal{LN}(\mu, \sigma)$.

$$\eta = \mathbb{E}[\mathcal{LN}(\mu, \sigma)] = e^{\mu + \sigma^2/2} \Rightarrow \eta = e^{\mu + \sigma^2/2}$$

$$\Rightarrow \mu = \log \eta - \frac{\sigma^2}{2}$$

$$\Rightarrow \mathcal{LN}(\mu, \sigma) = \mathcal{LN}\left(\log \eta - \frac{\sigma^2}{2}, \sigma\right)$$

Connect with $\varphi(x|h)$

- Use mean reparametrization of $\mathcal{LN}(\mu, \sigma)$.

$$
\begin{aligned}
\eta = \mathbb{E}[\mathcal{LN}(\mu, \sigma)] = e^{\mu+\sigma^2/2} &\Rightarrow \eta = e^{\mu+\sigma^2/2} \\
&\Rightarrow \mu = \log \eta - \frac{\sigma^2}{2} \\
&\Rightarrow \mathcal{LN}(\mu, \sigma) = \mathcal{LN}\left(\log \eta - \frac{\sigma^2}{2}, \sigma\right)
\end{aligned}
$$

Observation model

$$\tau|\eta \sim \mathcal{LN}\left(\log \eta - \frac{\sigma^2}{2}, \sigma\right) \tag{6}$$

$$\eta = \int_0^H \varphi(x|h)\,\mathrm{d}h \tag{7}$$

With multiple observations $\tau_1, ..., \tau_n$, scale parameter $\sigma > 0$ assumed shared among columns but η (or μ) is column-specific.

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:

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

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

τ_i	Observed AOD
$\mathcal{LN}$	Log-normal distribution
η_i, σ	Mean and scale parameters
φ	Prior for b_{ext}
$x_i\|h$	Input covariates at altitude h
H	Atmospheric column height
ψ	Positive link function
L	Idealized heightscale parameter
f	GP prior

Model wrap-up

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:

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

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

τ_i	Observed AOD
$\mathcal{LN}$	Log-normal distribution
η_i, σ	Mean and scale parameters
φ	Prior for b_{ext}
$x_i\|h$	Input covariates at altitude h
H	Atmospheric column height
ψ	Positive link function
L	Idealized heightscale parameter
f	GP prior

- **Objective:** Infer distribution of $\varphi(x|h)|\underbrace{\tau_1, ..., \tau_n}_{\boldsymbol{\tau}}$

Inference

- Actually... $f(x|h)|\boldsymbol{\tau}$
- Access to posterior distribution $p(\mathrm{f}|\boldsymbol{\tau})$ allows to compute predictive mean and variance of φ at input $x|h$ following

$$\mathbb{E}[\varphi(x|h)|\boldsymbol{\tau}] = \int \psi(\mathrm{f}) e^{-h/L} p(\mathrm{f}|\boldsymbol{\tau}) \, \mathrm{df}$$
$$\mathrm{Var}(\varphi(x|h)|\boldsymbol{\tau}) = \mathbb{E}[\varphi(x|h)^2|\boldsymbol{\tau}] - \mathbb{E}[\varphi(x|h)|\boldsymbol{\tau}]^2$$

- Can be estimated with Monte-Carlo (and admits closed form for $\psi = \exp$)

Infer $p(\mathrm{f}|\boldsymbol{\tau})$?

Problem

$$p(\mathrm{f}|\boldsymbol{\tau}) = \frac{p(\boldsymbol{\tau}|\mathrm{f})p(\mathrm{f})}{\underbrace{\int p(\boldsymbol{\tau}|\mathrm{f})p(\mathrm{f})\,\mathrm{df}}_{\text{intractable}}}$$

Solution

- Approximate $p(\mathrm{f}|\boldsymbol{\tau})$ (variational approximation)
- Approximation scheme allows for **sparse representation** which **scales to very large number of data points**

Experiments

ECHAM-HAM setup

	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 1: 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). Our objective is to vertically disaggregate the response τ using the vertically resolved predictors $(T, P, \text{RH}, \omega)$ and spatiotemporal columns locations (t, lat, lon).

- Total of $8 \times 96 \times 192 = 147\,456$ columns.

Predictors slices

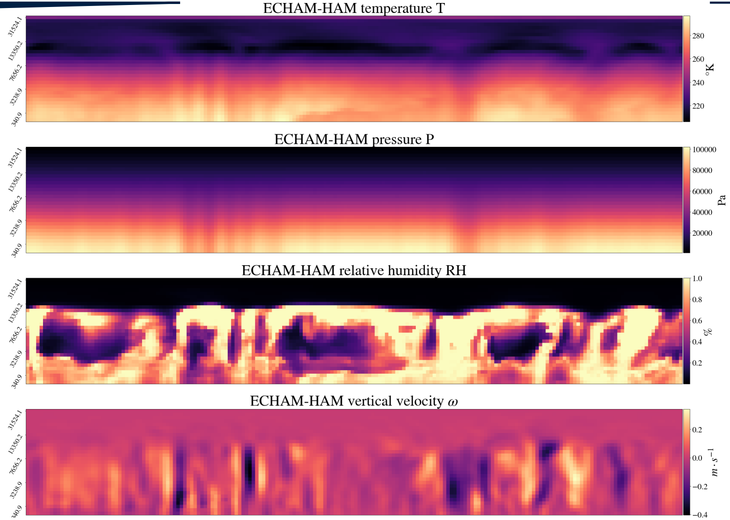


Figure 4: Vertical slices at latitude 51.29° of meteorological predictors

Ideal slices

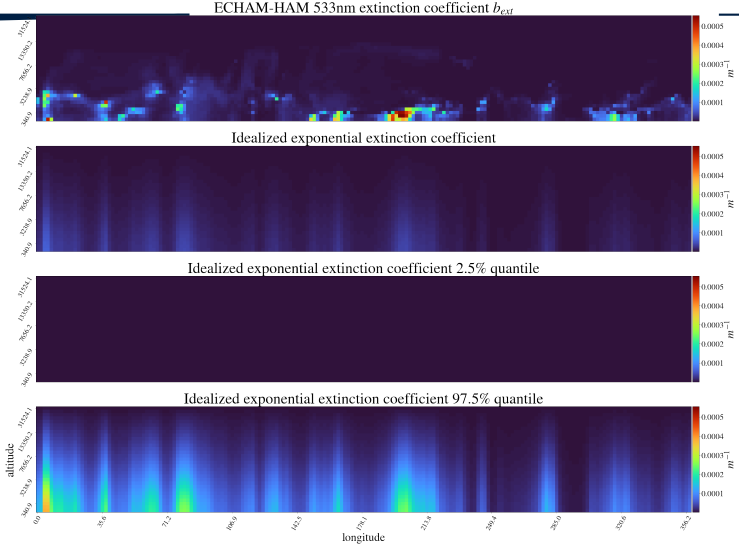


Figure 5: Vertical slices at latitude 51.29° of idealized profiles

Predicted slices

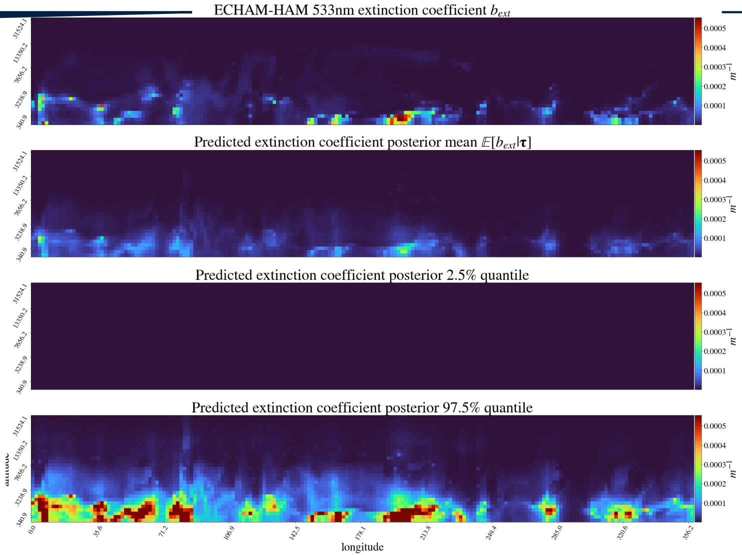


Figure 6: Vertical slices at latitude 51.29° of predicted profiles

Predicted slices

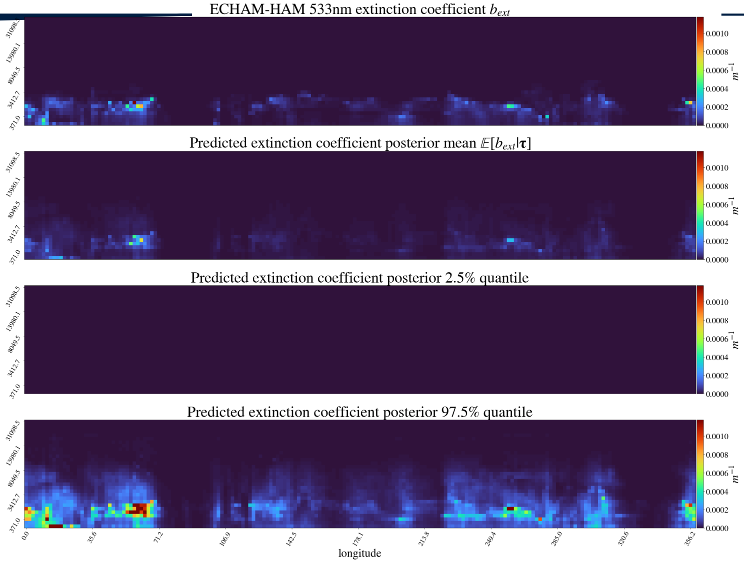


Figure 7: Vertical slices at latitude -0.93° of predicted profiles

Predicted slices

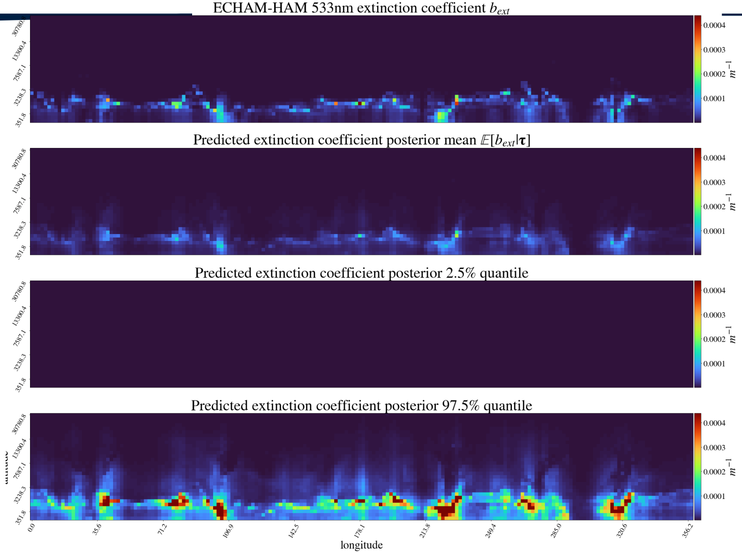


Figure 8: Vertical slices at latitude -38.2° of predicted profiles

Table 2: Scores of our method (Our) compared to an idealized exponential baseline (Ideal)

Region	Method	RMSE (10^{-5})	Corr (%)	Bias (10^{-6})	Bias98 (10^{-5})
Entire column	Our	**3.29**±0.02	**70.9**±0.4	**-0.167**±0.105	**-0.646**±0.151
	Ideal	4.10	51.2	-2.40	-4.08
Boundary layer	Our	**6.06**±0.03	**69.8**±0.5	**-1.25**±0.45	**-4.64**±0.32
	Ideal	7.55	53.6	-12.9	-11.7

Region	Method	ELBO	Calib95 (%)	ICI (10^{-2})
Entire column	Our	**13.1**±0.1	**94.9**±0.1	5.29±0.59
	Ideal	**13.1**	96.0	**5.05**
Boundary layer	Our	**10.6**±0.1	98.8±0.1	**8.27**±0.29
	Ideal	10.2	**93.5**	19.1

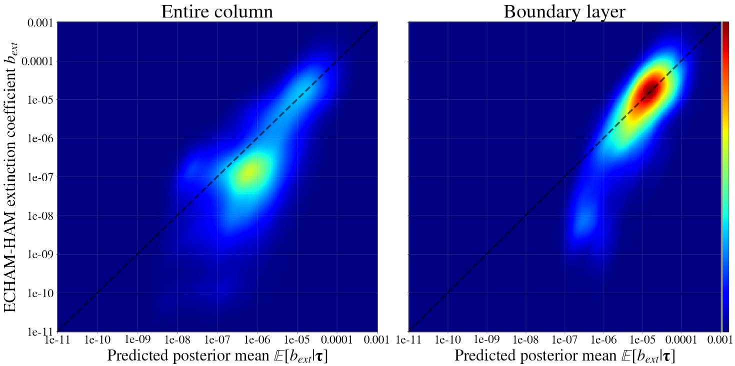


Figure 9: Density plots of groundtruth extinction coefficient values against predicted posterior mean extinction coefficient; **Left:** entire column; **Right:** boundary layer only

Conclusion

Insights

- Possible to reconstruct realistic vertical extinction of aerosol with sound uncertainty quantification using AOD and readily available meteorological predictors

Insights

- Possible to reconstruct realistic vertical extinction of aerosol with sound uncertainty quantification using AOD and readily available meteorological predictors
- Method is simple, computationally efficient, makes assumptions explicit and as such grants control and intepretability

Conclusion

Insights

- Possible to reconstruct realistic vertical extinction of aerosol with sound uncertainty quantification using AOD and readily available meteorological predictors
- Method is simple, computationally efficient, makes assumptions explicit and as such grants control and intepretability
- Can benefit aerosol satellite products, leading to more accurate priors over aerosol vertical profiles

Conclusion

Insights

- Possible to reconstruct realistic vertical extinction of aerosol with sound uncertainty quantification using AOD and readily available meteorological predictors
- Method is simple, computationally efficient, makes assumptions explicit and as such grants control and intepretability
- Can benefit aerosol satellite products, leading to more accurate priors over aerosol vertical profiles

Limitations and Directions

- Can only capture extinction due to aerosol swelling (missing mass concentration, particle size and radiative properties extinction which would require additional predictors harder to obtain)

Conclusion

Insights

- Possible to reconstruct realistic vertical extinction of aerosol with sound uncertainty quantification using AOD and readily available meteorological predictors
- Method is simple, computationally efficient, makes assumptions explicit and as such grants control and intepretability
- Can benefit aerosol satellite products, leading to more accurate priors over aerosol vertical profiles

Limitations and Directions

- Can only capture extinction due to aerosol swelling (missing mass concentration, particle size and radiative properties extinction which would require additional predictors harder to obtain)
- Methodological extensions (use multiple wavelengths, allow unmatched data setting)

Conclusion

Insights

- Possible to reconstruct realistic vertical extinction of aerosol with sound uncertainty quantification using AOD and readily available meteorological predictors
- Method is simple, computationally efficient, makes assumptions explicit and as such grants control and intepretability
- Can benefit aerosol satellite products, leading to more accurate priors over aerosol vertical profiles

Limitations and Directions

- Can only capture extinction due to aerosol swelling (missing mass concentration, particle size and radiative properties extinction which would require additional predictors harder to obtain)
- Methodological extensions (use multiple wavelengths, allow unmatched data setting)
- Different use case: investigation on aerosol mode/species contribution to extinction using model data only

References

[1] Leon Ho Chung 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.