Suitably impressive thesis title

Lennart Golks

Department of Physics
University of Otago

A thesis submitted for the degree of
Doctor of Philosophy

November 2025

Abstract

Your abstract text goes here. Check your departmental regulations, but generally this should be less than 300 words. See the beginning of Chapter **??** for more.

Suitably impressive thesis title

Lennart Golks

Department of Physics

University of Otago

A thesis submitted for the degree of

Doctor of Philosophy

November 2025

Acknowledgements

Personal

I would like to thank Alex Elliott for his wonderful help and support. None of this would be possible otherwise.

Institutional

If you want to separate out your thanks for funding and institutional support, I don't think there's any rule against it. Of course, you could also just remove the subsections and do one big traditional acknowledgement section.

Abstract

Your abstract text goes here. Check your departmental regulations, but generally this should be less than 300 words. See the beginning of Chapter **??** for more.

Contents

List of Figures

List of Abbreviations

i.i.d. . . . . . . independent and identically distributed

MRF . . . . . . Markov Random Field

GMRF . . . . Gaussian Markov Random Field

MTC . . . . . Marginal Then Conditional sampler

GOMOS . . . Global Ozone Monitoring by Occultation of Stars

MCMC . . . . Markov Chain Monte-Carlo

MH . . . . . . Metropolis-Hastings

columnwidth 421.10046pt

1 Introduction

1.1 What is going on?, 3 facts, What is new in this thesis?

- hierachical Bayesian model, sampling to TT approx
- RTE as an example
- nonLinear to Linear Affine funciton (affine RTO)

1.2 What has been published?

2

Building a physics based hierarchical Linear model

- two hyperparameters, marginal and then conditonal, MTC, use as a building block [1, 2]
- sampling, Gibbs-MH, t-walk [3]
- increase hyperparameters, Temperature and pressure, tt-approx, SIRT [4, 5]

2.1 Linear model

- define model [6]
- explain parameters and constants [7]
- how pressure to height and temperature hydro-static equilibrium equation [5]

2.2 two hyperameters, fast sampling paper

- similar to MTC paper, [1]
- can compare to regularization and can integrate easily with less solves to regularization or RTO, [8]

- define priors make table
- how many steps, integrated autocorrealtion time , ref
- relative error
- how I found max curvature, [9]
- how many taylor series

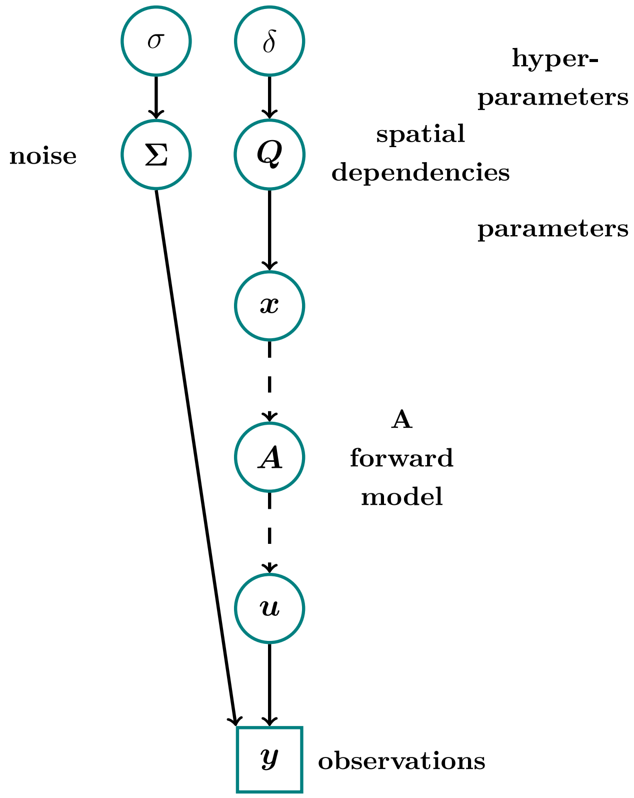

Figure 2.1

2.3 four hyperameters, t-walk, TT approx, and RTO

- t-walk ref
- motivation why more hyper-parameters, explain parabula
- how sample from them, compared to TT approx

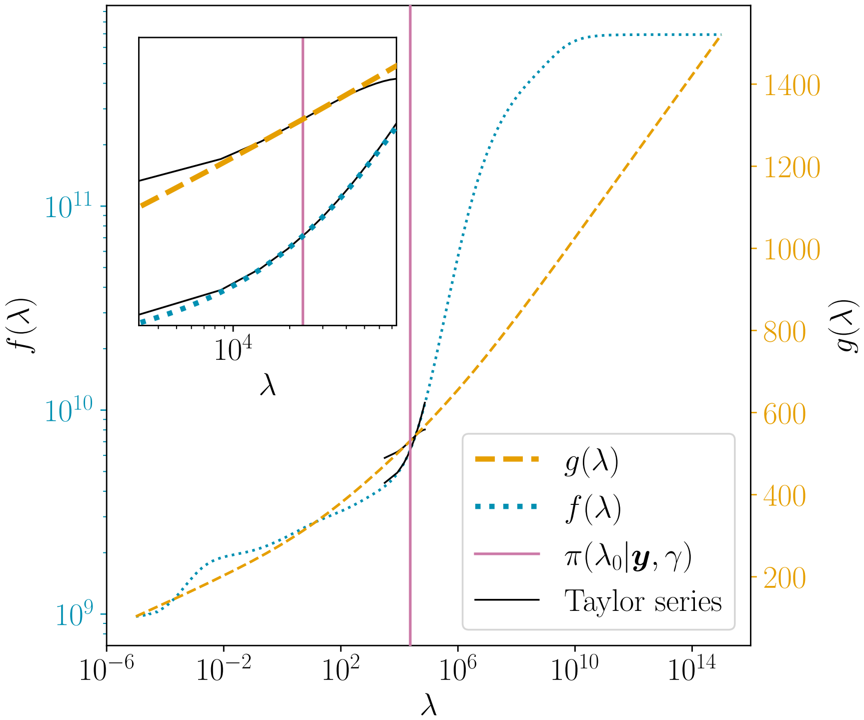

Figure 2.2: Functions $f(\lambda)$, dotted, and $g(\lambda)$, dashed, of the marginal posterior distribution for the specific forward model used in this study. Both functions are well-behaved over a large range of λ. In the support region of the MWG the pink square refers to the mode of the marginal posterior. Additionally, we plot the Taylor series of fourth order for $f(\lambda)$ and $g(\lambda)$ around the mode, see black line.

- explain Rosenblatt and SIRT trasnport

- RTO

- define priors

- how solve inverse and determinant

- how many steps, integrated autocorrealtion time

- relative error

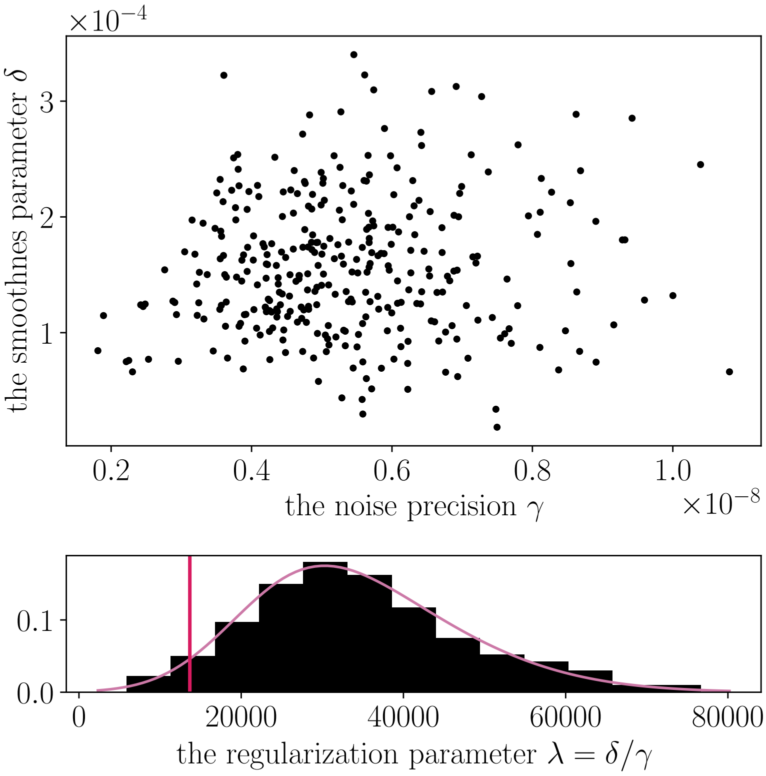

Figure 2.3: The scatter plot shows independent samples of δ and γ as the result of the MWG algorithm. The histogram displays independent samples of $\lambda \sim \pi(\lambda|\boldsymbol{y}, \gamma)$. The vertical line corresponds to the optimal regularization parameter.

2.3.1 Sampling

2.3.2 Rosenblatt Transform

2.3.3 (Squared) Inverse Rosenblatt Transform

2.3.4 Tensor-Train Approximation

2.4 Temperature and pressure hyperameters, tt-approx

- updating scheme

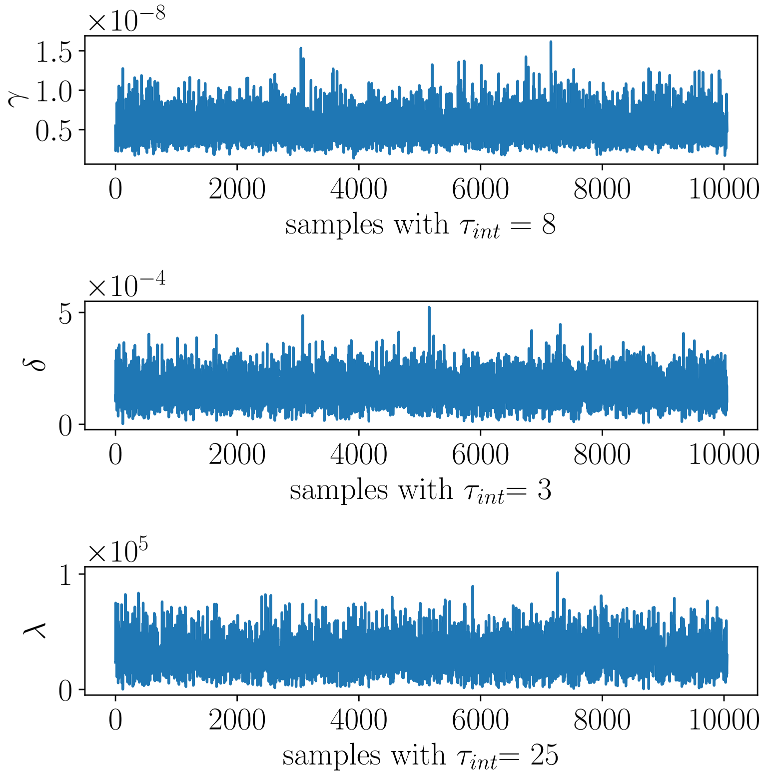

Figure 2.4: text

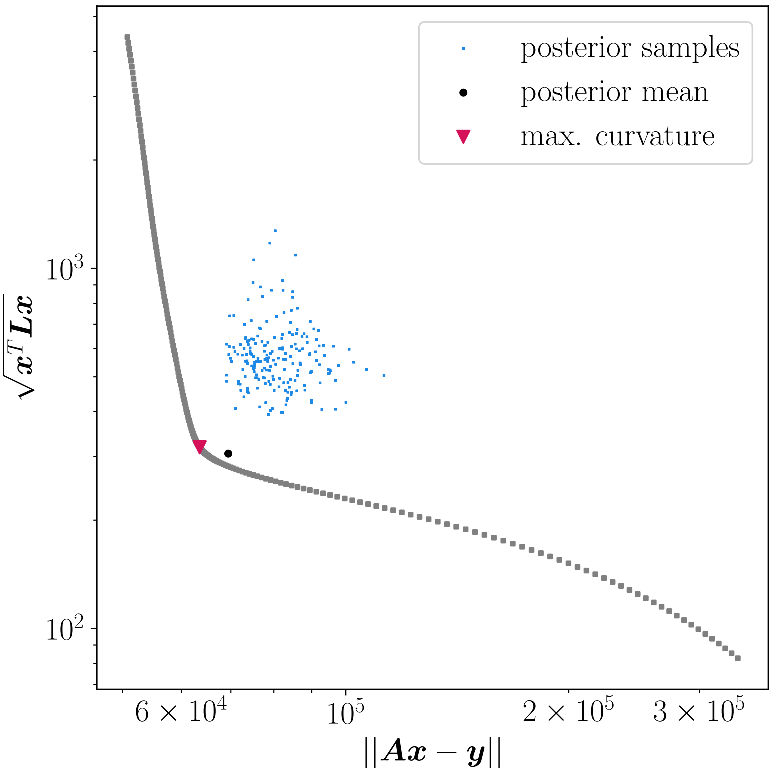

Figure 2.5: For varying λ we plot the seminorm $\sqrt{\boldsymbol{x}_\lambda^T \boldsymbol{L} \boldsymbol{x}_\lambda}$ against data misfit $\|\boldsymbol{A}\boldsymbol{x}_\lambda - \boldsymbol{y}\|$ of the regularised profiles. The triangle marks the point of maximum curvature closest to the origin of the L-curve. We plot the seminorm and the data misfit of the conditional posterior samples as well as of the posterior mean.

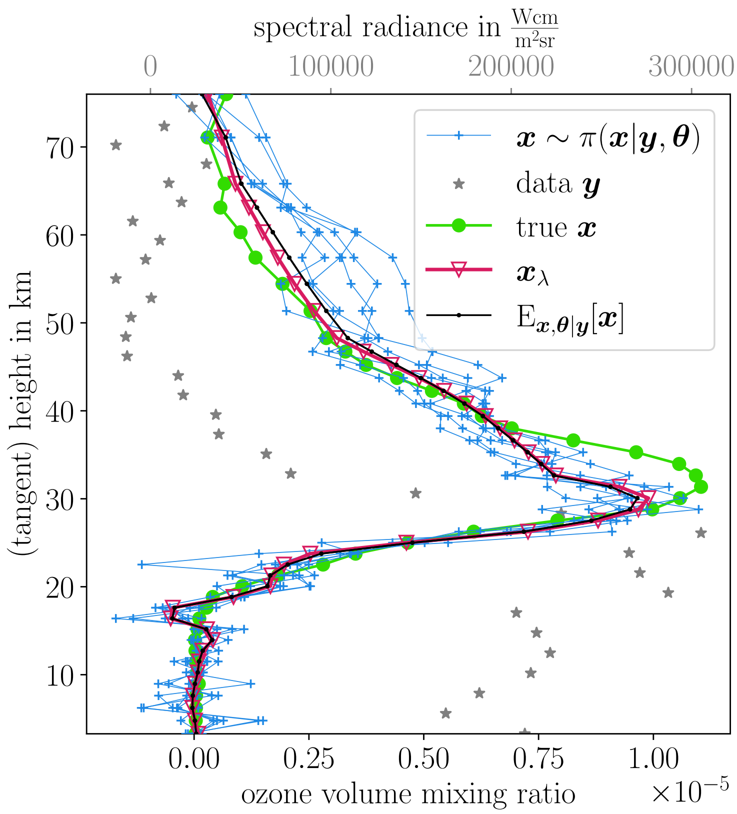

Figure 2.6: Plot of the true ozone profile (●), posterior samples (+), and posterior mean (●). We display the optimal regularised solution (▽) and the simulated data (∗) in spectral radiance.

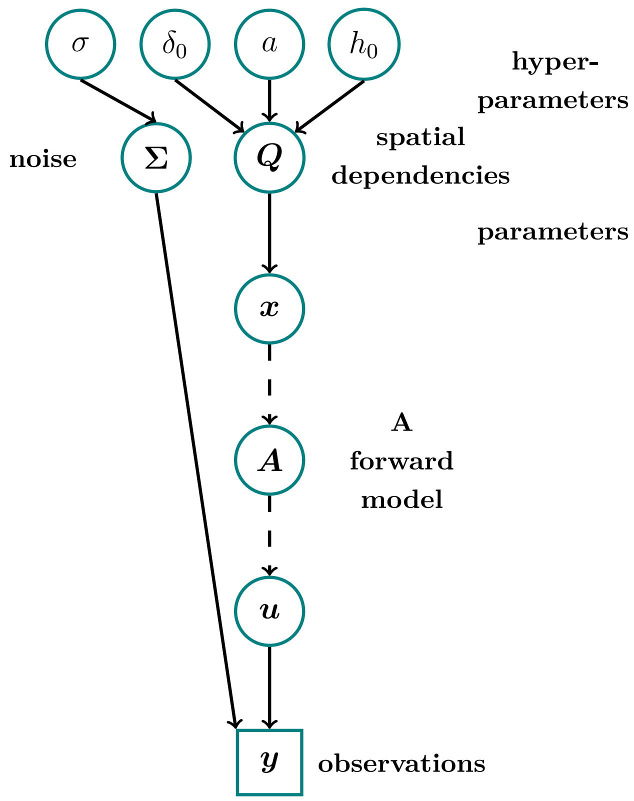

Figure 2.7: text

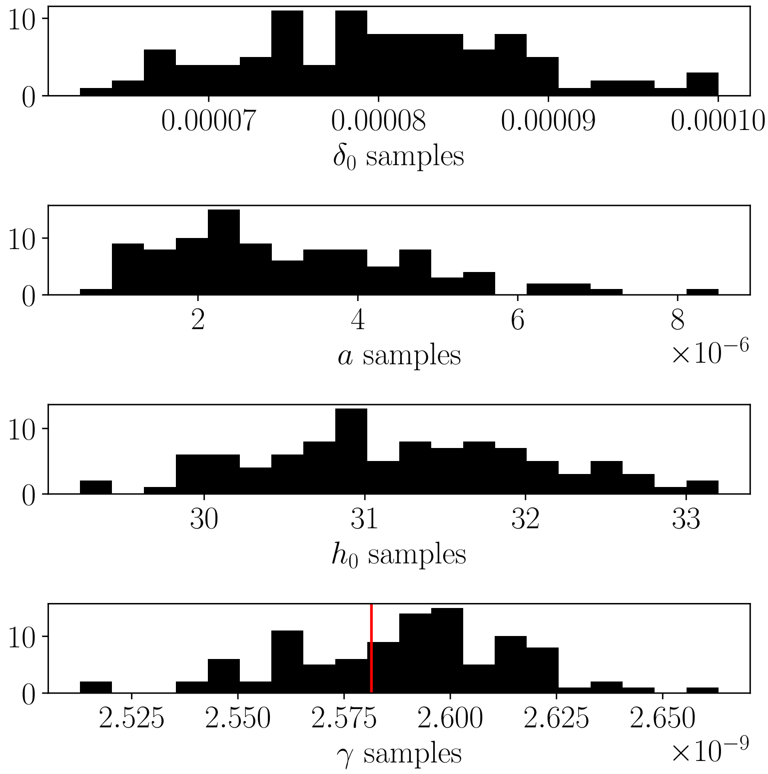

Figure 2.8: text

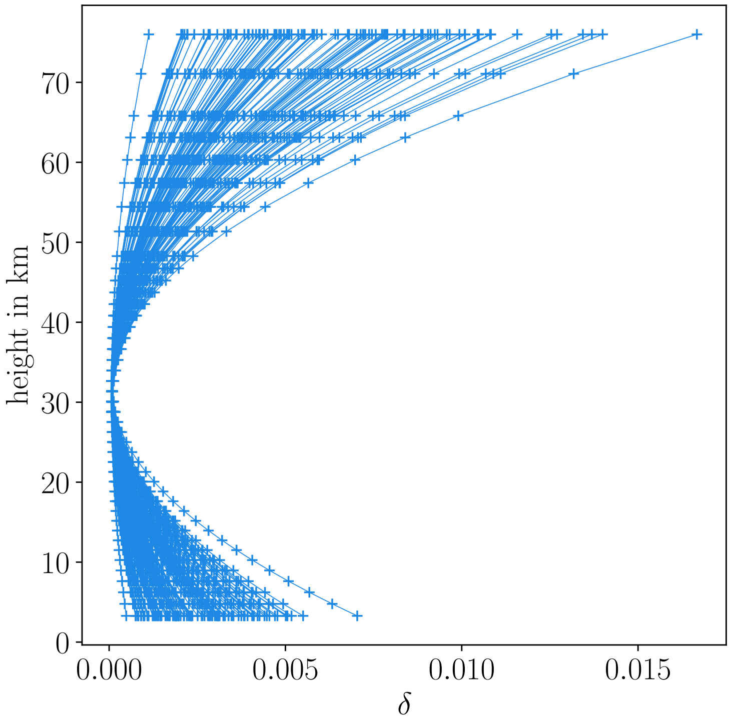

Figure 2.9: text

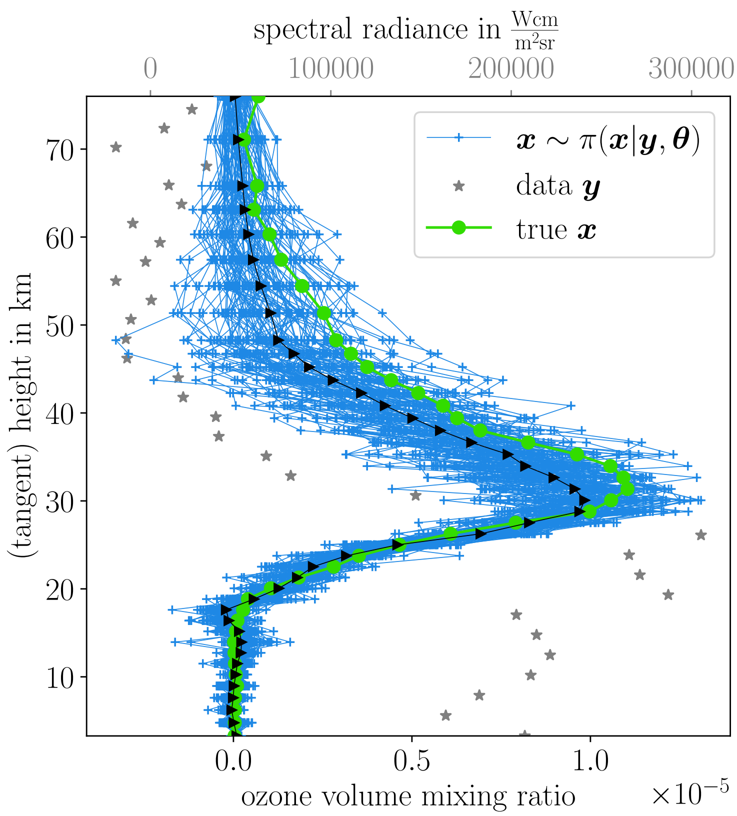

Figure 2.10: text

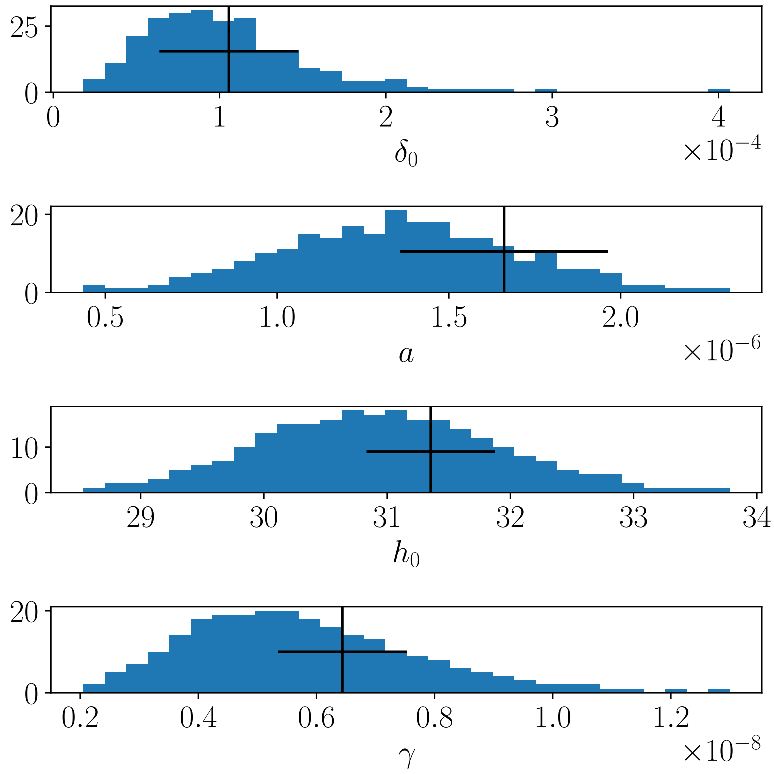

Figure 2.11: TTSIRT output

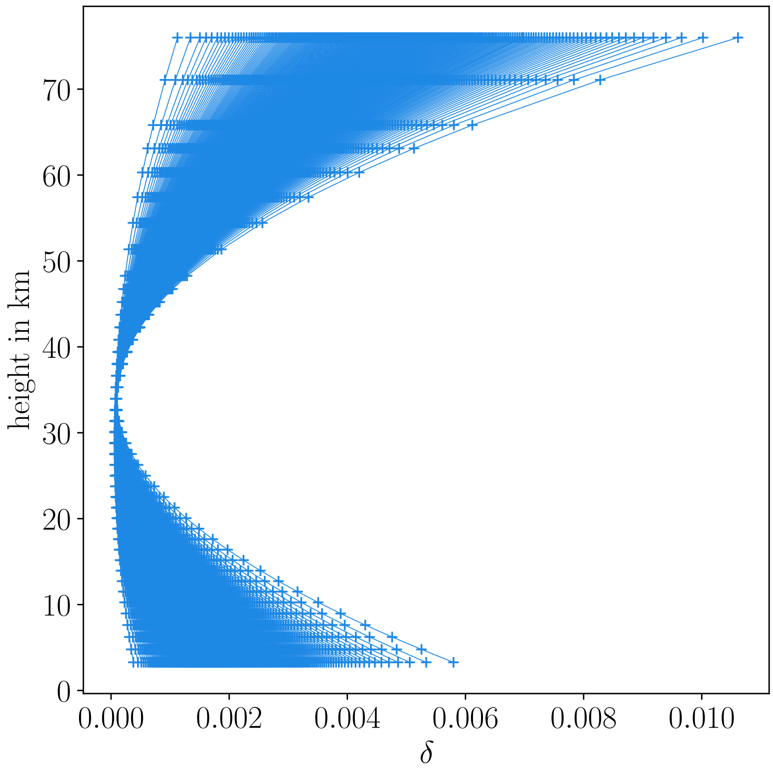

Figure 2.12: text

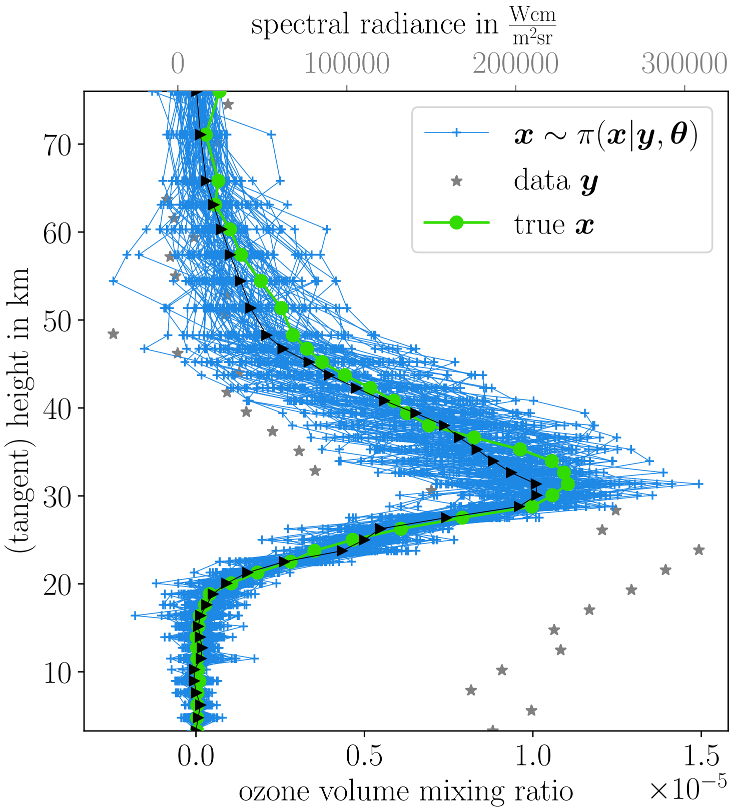

Figure 2.13: text

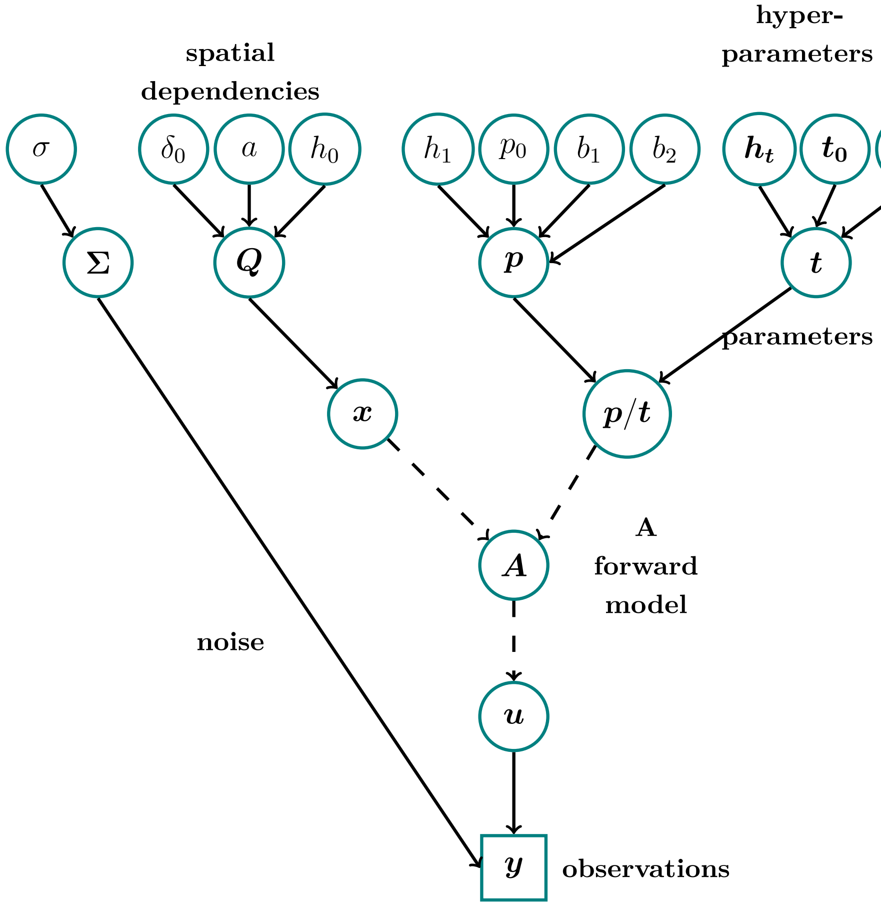

Figure 2.14: text

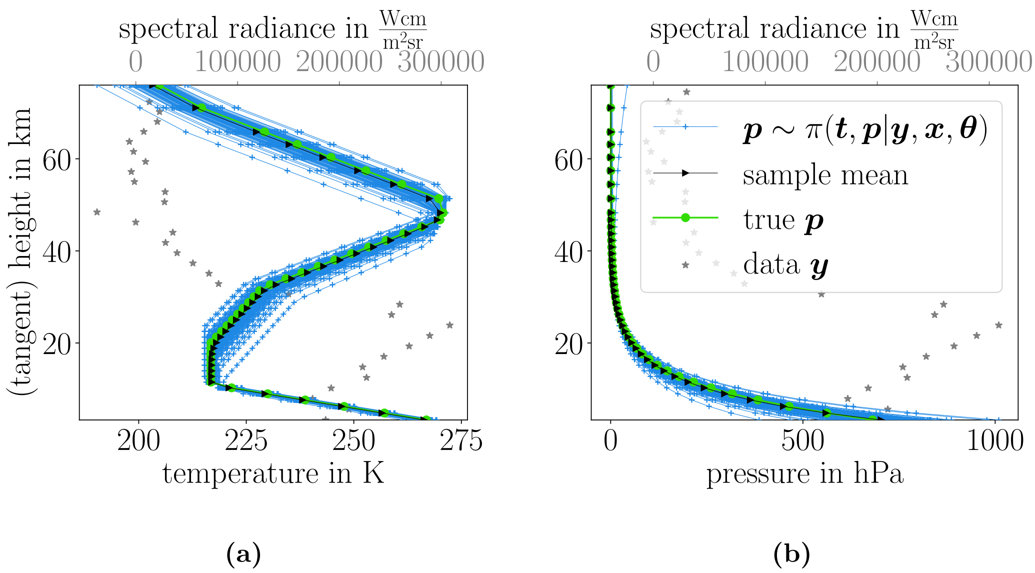

Figure 2.15

		TT bounds		
model parameters	priors	lower	upper	Context
γ	$\mathcal{T}(1, 1e-10)$	-	-	$\boldsymbol{x}$
δ	$\mathcal{T}(1, 1e-10)$	-	-	$\boldsymbol{x}$
a_0	$\mathcal{T}(3, 1e6)$	1e-15	1e-5	$\boldsymbol{x}$
h_0	$\mathcal{N}(31.35, 1)$	27	35	$\boldsymbol{x}$
$h_{p,1}$	$\mathcal{N}(34.3, 0.5)$	32.8	35.64	$\boldsymbol{p/t}$
p_0	$\mathcal{N}(6.5, 0.1)$	6.17	6.73	$\boldsymbol{p/t}$
b_1	$\mathcal{N}(0.15, 0.0051)$	0.138	0.167	$\boldsymbol{p/t}$
b_2	$\mathcal{N}(0.13, 0.067)$	0	0.32	$\boldsymbol{p/t}$
$h_{t,0}$	$\mathcal{N}(11, 0.5)$	9.6	12.4	$\boldsymbol{p/t}$
$h_{t,1}$	$\mathcal{N}(20, 3)$	11.6	28.4	$\boldsymbol{p/t}$
$h_{t,2}$	$\mathcal{N}(32, 1)$	29.2	34.8	$\boldsymbol{p/t}$
$h_{t,3}$	$\mathcal{N}(47, 2)$	41.4	52.6	$\boldsymbol{p/t}$
$h_{t,4}$	$\mathcal{N}(51, 2)$	45.4	56.6	$\boldsymbol{p/t}$
$h_{t,5}$	$\mathcal{N}(71, 2)$	65.4	76.6	$\boldsymbol{p/t}$
$a_{t,1}$	$\mathcal{N}(-6.5, 0.01)$	-6.528	-6.472	$\boldsymbol{p/t}$
$a_{t,2}$	$\mathcal{N}(1, 0.01)$	0.972	1.028	$\boldsymbol{p/t}$
$a_{t,3}$	$\mathcal{N}(2.8, 0.1)$	2.52	3.078	$\boldsymbol{p/t}$
$a_{t,4}$	$\mathcal{N}(-2.8, 0.01)$	-2.828	-2.772	$\boldsymbol{p/t}$
$a_{t,5}$	$\mathcal{N}(-2, 0.01)$	-2.028	-1.972	$\boldsymbol{p/t}$
t_0	$\mathcal{N}(288, 2)$	282.54	293.75	$\boldsymbol{p/t}$

Table 2.1: Gaussian $\mathcal{N}(\mu, \sigma)$ and gamma distribution $\mathcal{T}(\alpha = \text{scale}, \beta = \text{rate})$ Bounds for t and p 2.8 times the variance around the mean round pressure approx and test if would work with previous gamma prior or fix gamma prior with set values

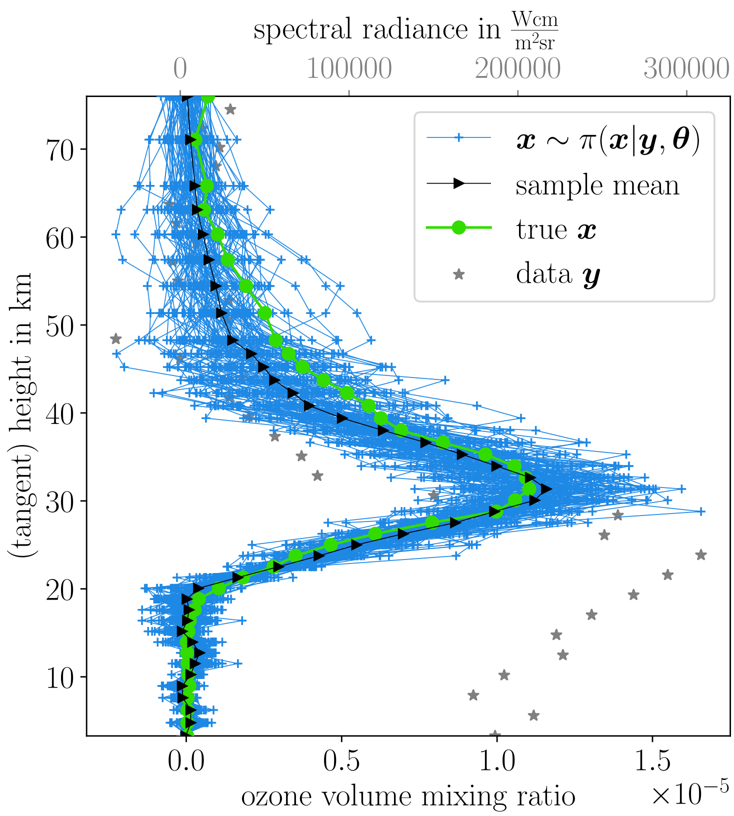

Figure 2.16: text

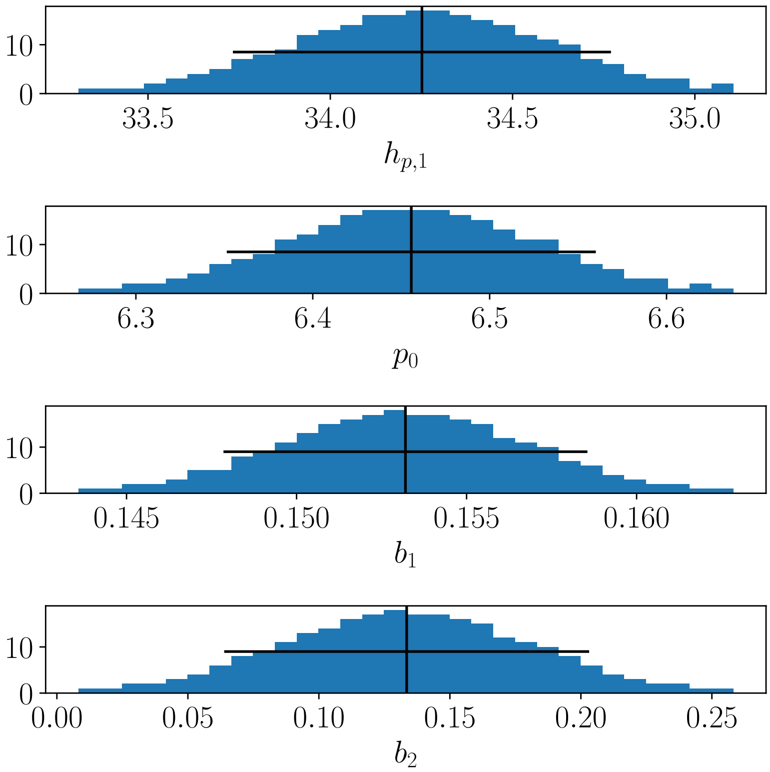

Figure 2.17: pritor

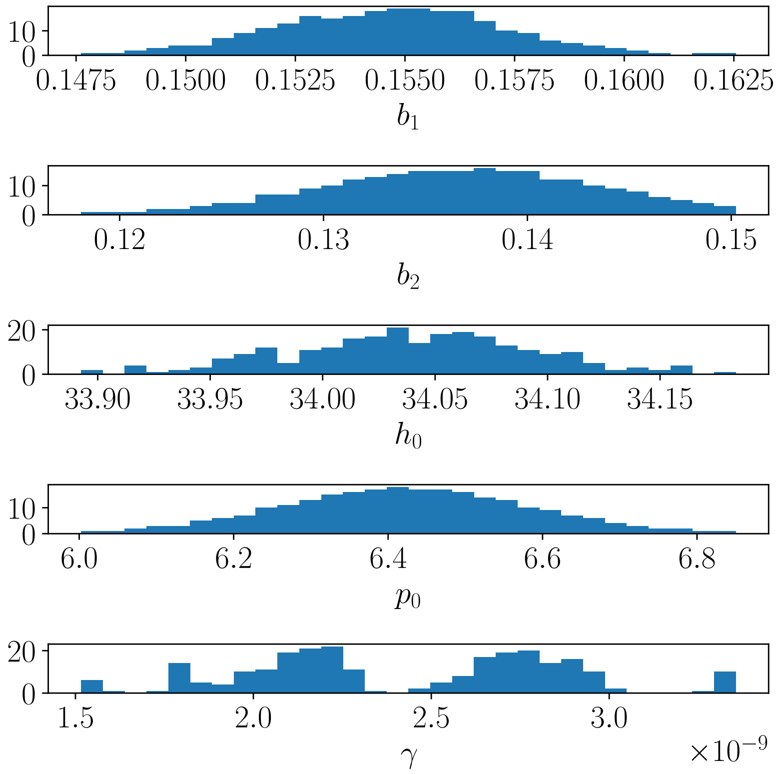

Figure 2.18: posteriore doesnt work with TT format

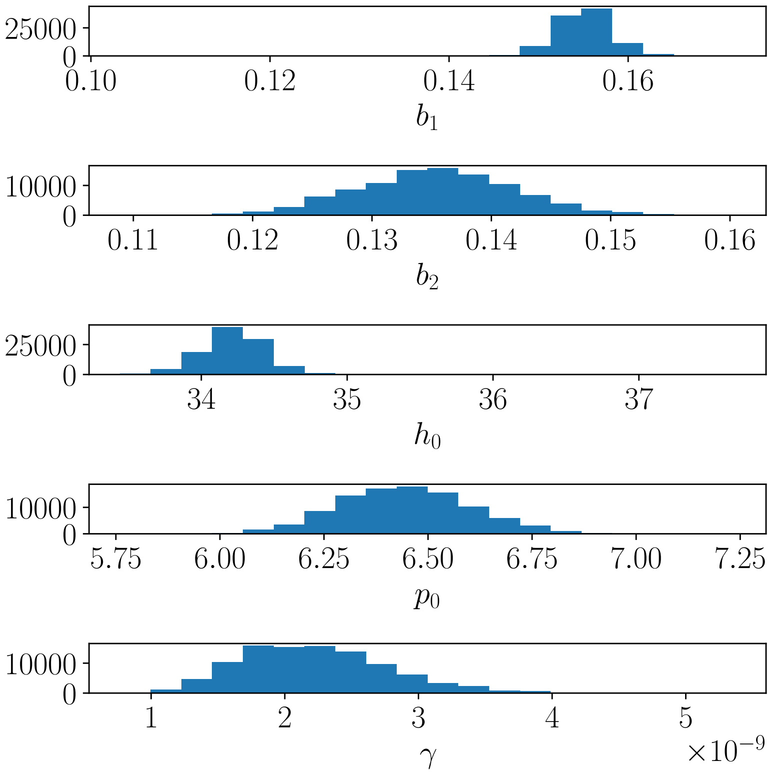

Figure 2.19: t-walk does work with posterioer

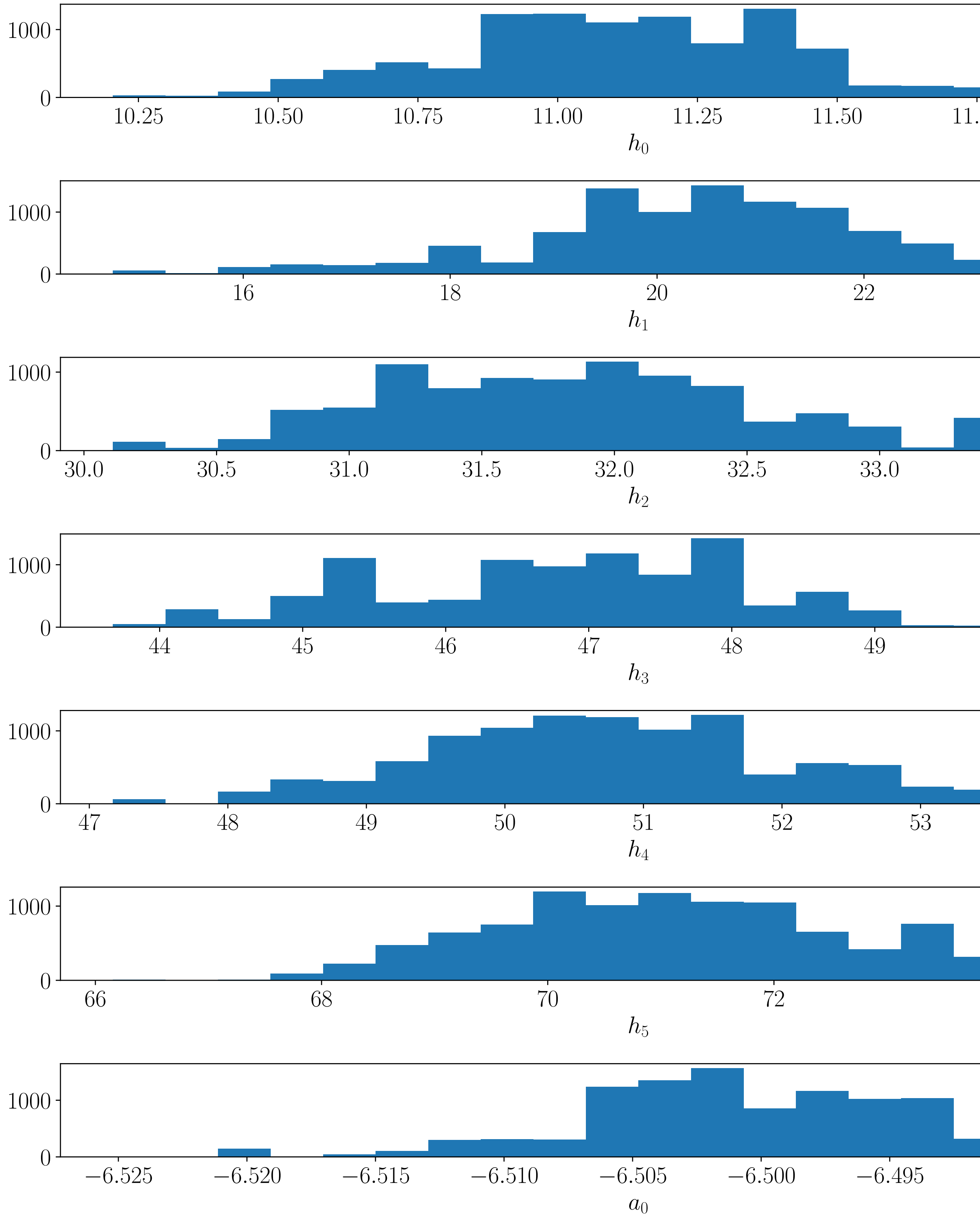

Figure 2.20: t-walk does work with posterioer

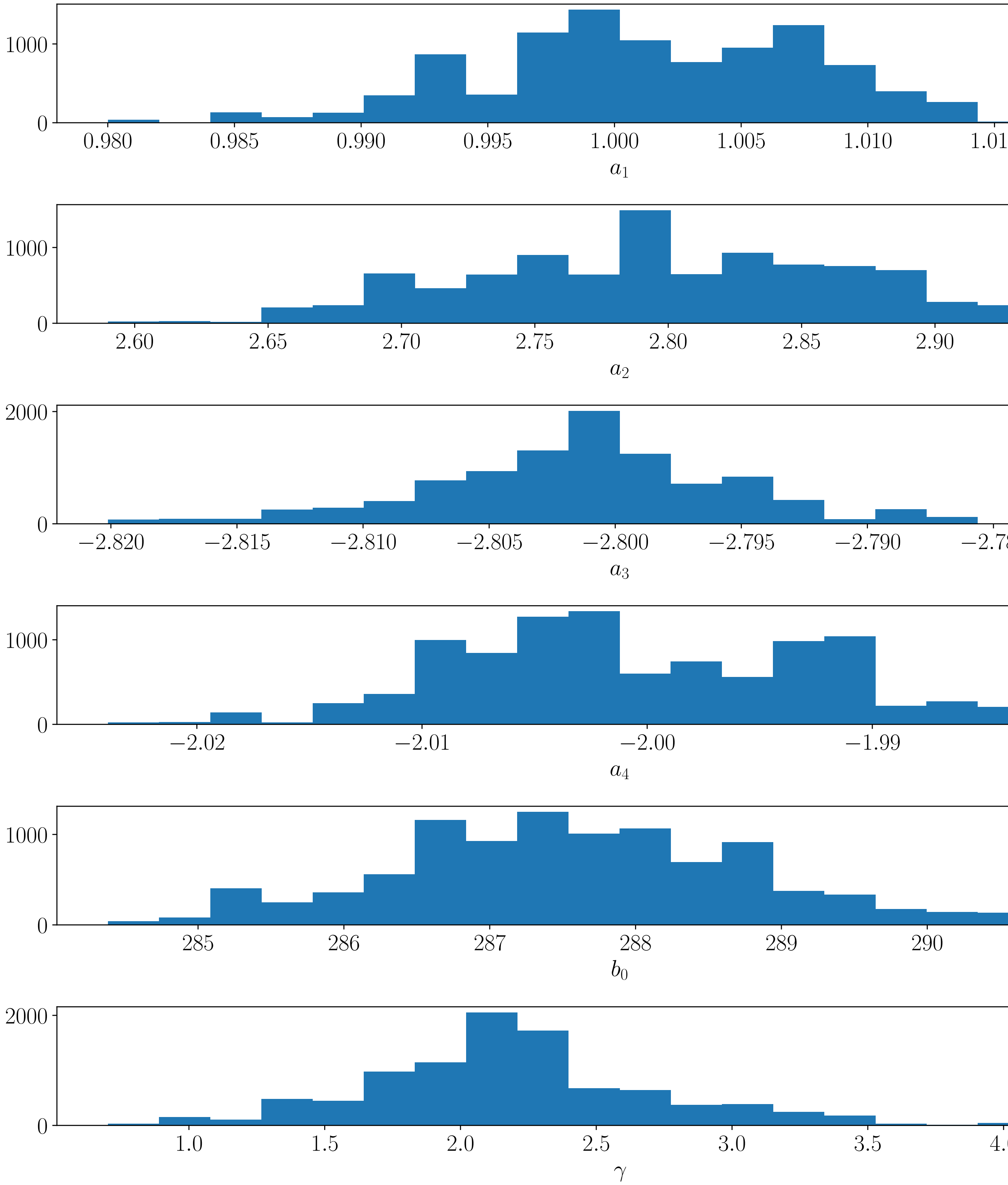

Figure 2.21: t-walk does work with posterioer

Figure 2.22: tt does not work woth posterior

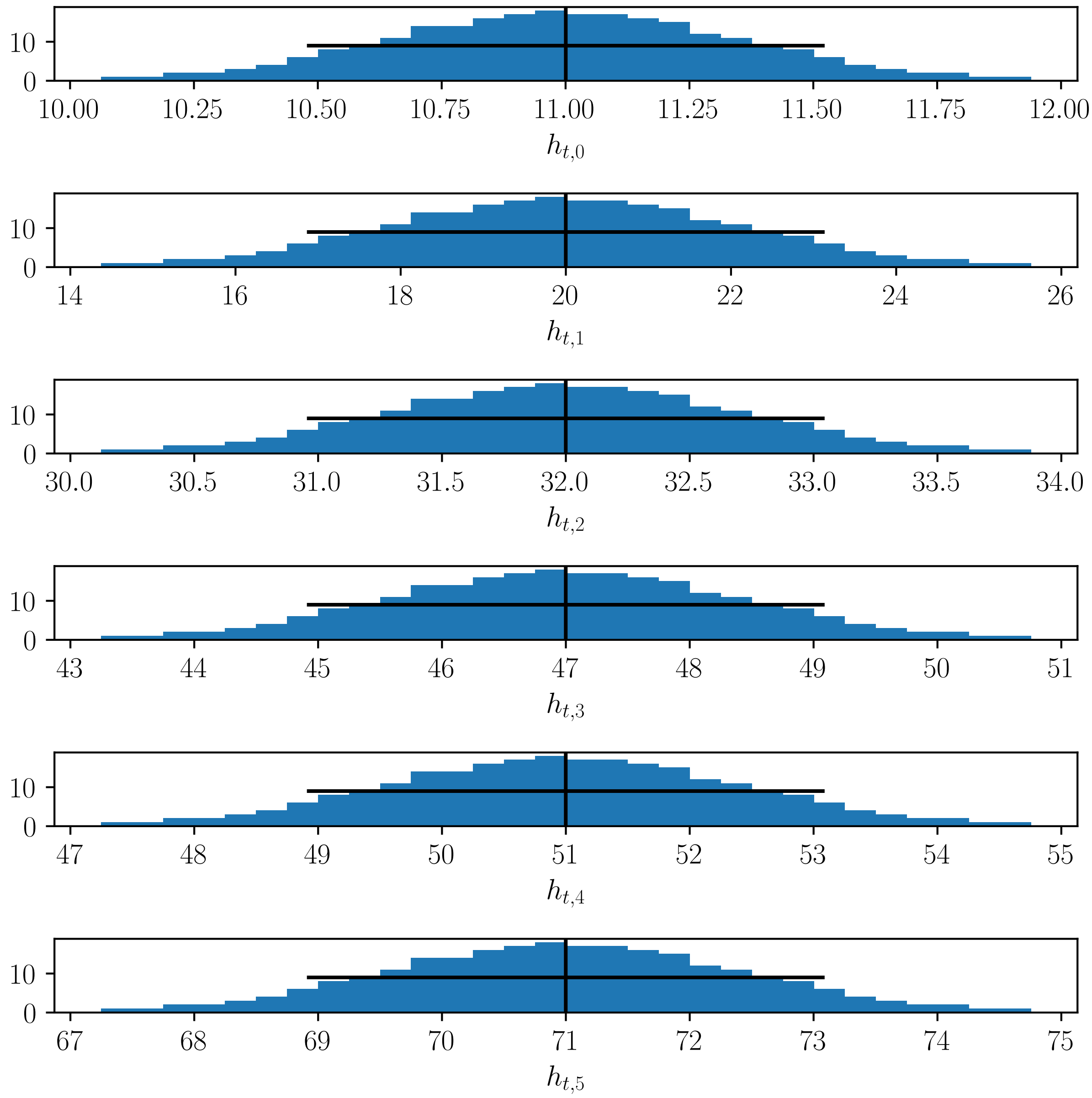

Figure 2.23: priors work

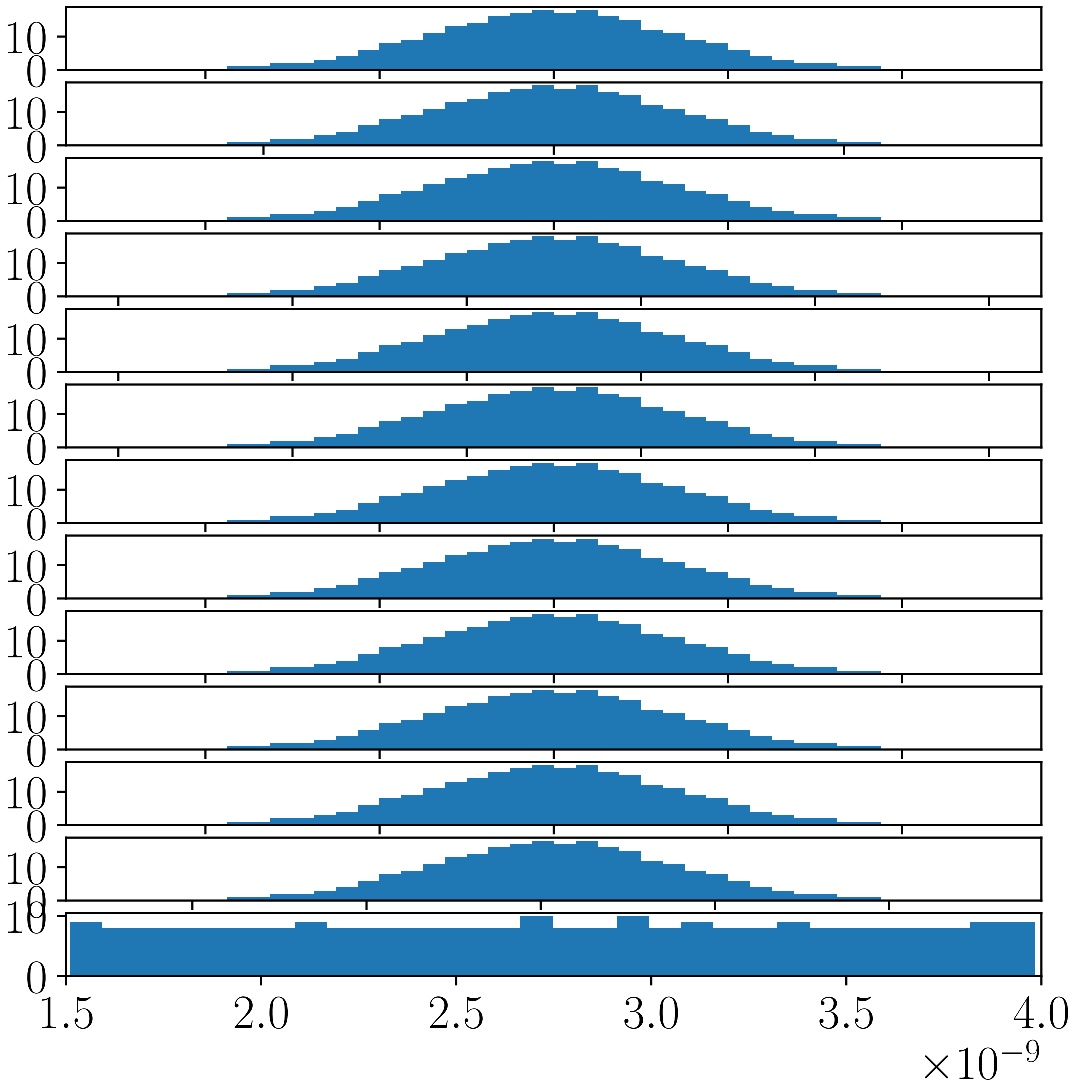

Figure 2.24: priorsTT work

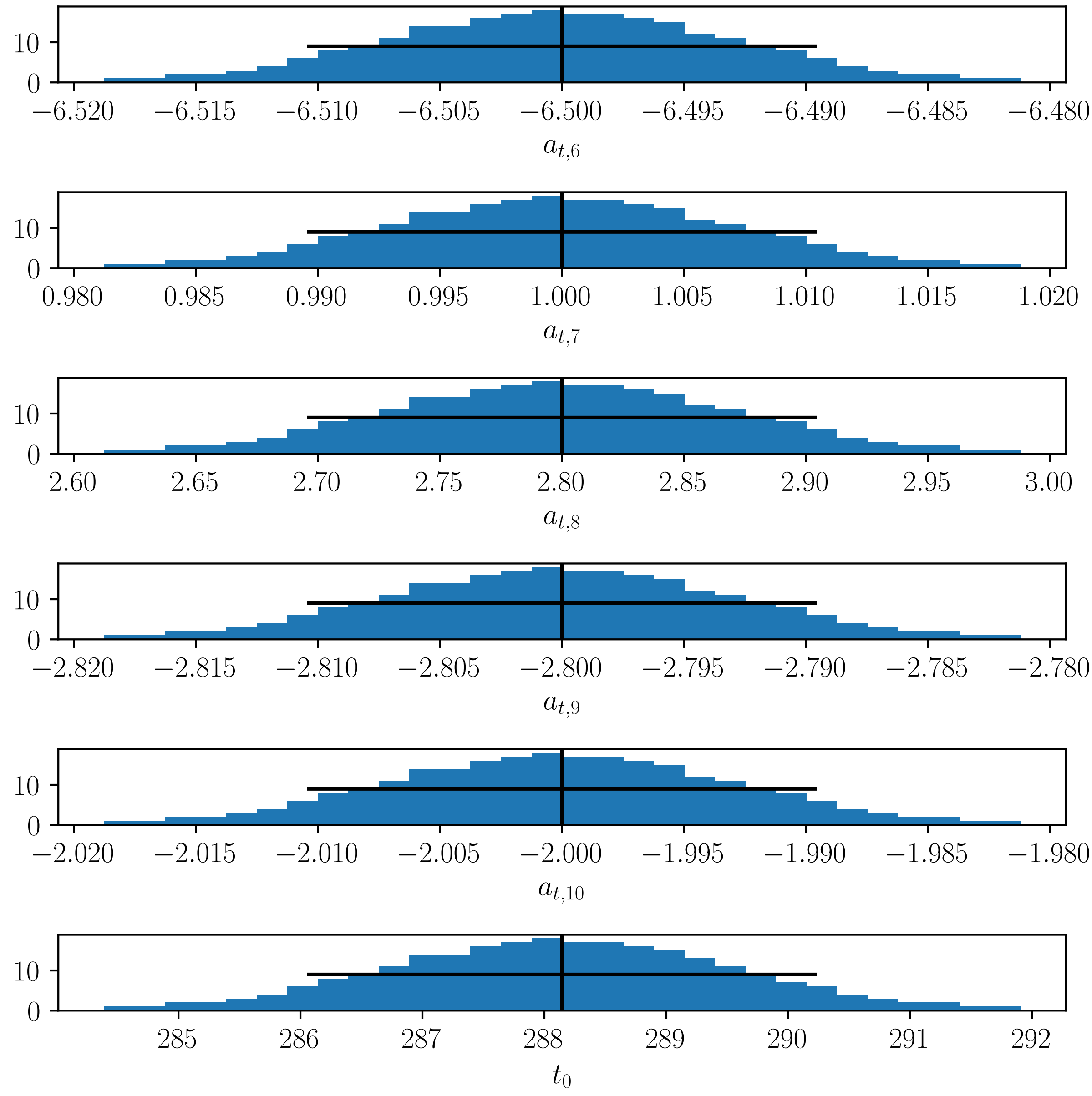

Figure 2.25: priors work

3 Nonlinear Forward model

- updating scheme, slow
- local linear map, strategy, schematic
- affine function, RTO

3.1 Sampling

3.2 local linear Map and strategy

- one data vector
- strategy to find convergence and local linear map

3.2.1 Machine learning vs Gaussian elimination

- linear solve
- machine learning class optimizer which package

3.3 affine RTO

- does it sample from the correct distribution

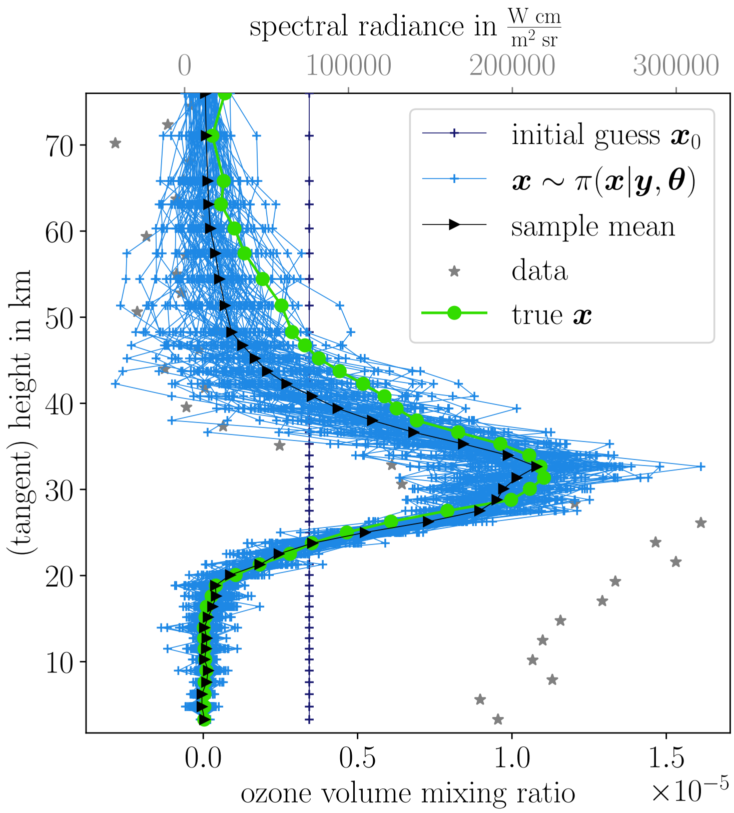

Figure 3.1: Non Linear carpet jumping scheme

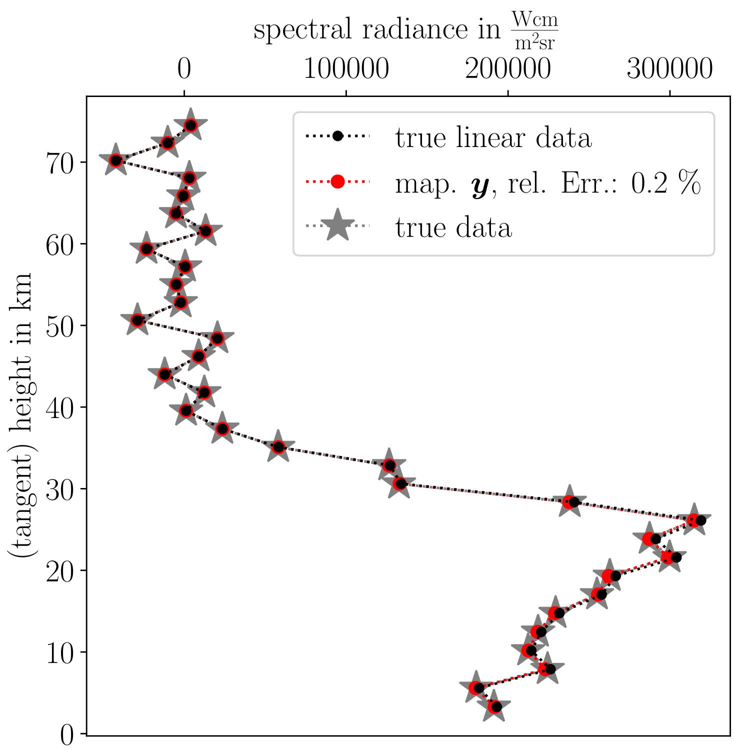

Figure 3.2: Map nonLinear true data to linear data

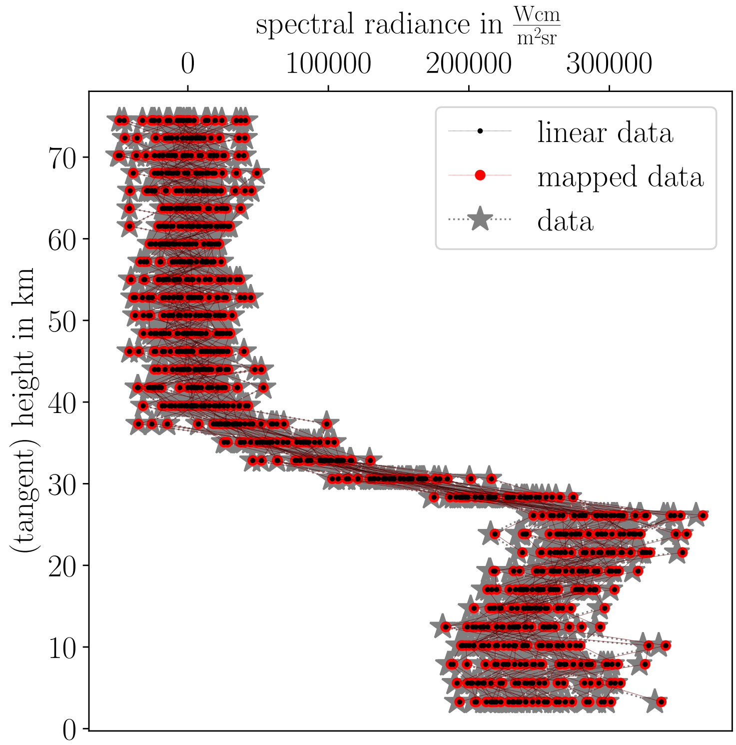

Figure 3.3: Show difference between Lin and NonLin Data including

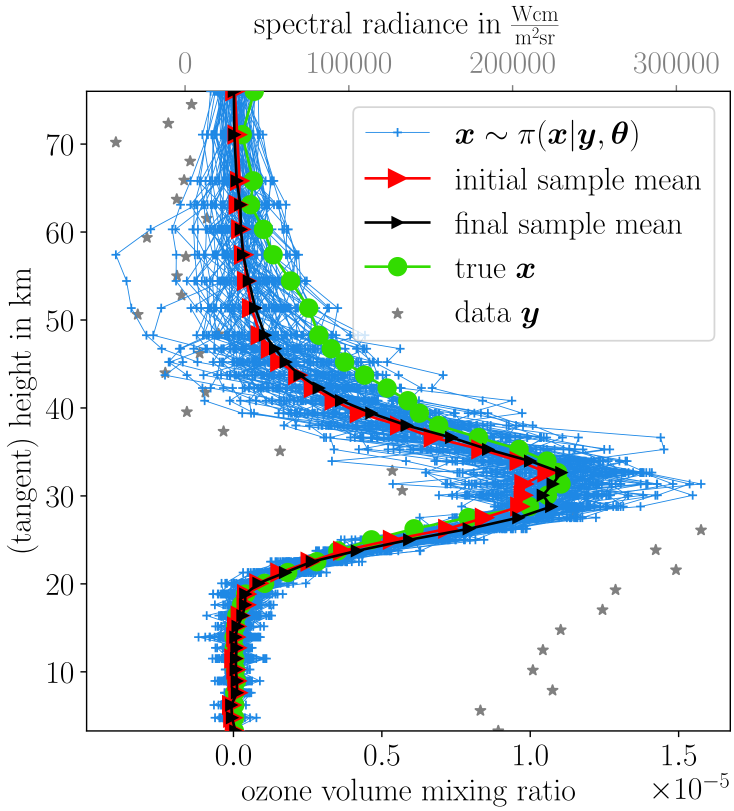

Figure 3.4: Link Func with previous and current

4
Introduction

4.1 What is going on?, 3 facts, What is new in this thesis?

- hierachical Bayesian model, sampling to TT approx
- RTE as an example
- nonLinear to Linear Affine funciton (affine RTO)

4.2 What has been published?

Appendices

A
Posterior of Bayesian Hierachical model

Here we show how to obtain the posterior covariance and mean of our hierarchical Bayesian model in **??** - **??**. We do not consider the hyper-parameters and start with the joint probaility distribution of $(\boldsymbol{x}^T, \boldsymbol{y}^T)^T$, where $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{y} \in \mathcal{Y}$ do not intersect. For more details we refer to Chapter 2 in [10] and to the book of Rue and Held [2].

The exponent of the normal Gaussian can be rewritten into:

$$-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{Q}(\boldsymbol{x} - \boldsymbol{\mu}) = -\frac{1}{2}\boldsymbol{x}^T \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{x}^T \boldsymbol{Q}\boldsymbol{\mu} + \text{const.} \tag{A.1}$$

We like to bring the joint distribution into a similar form so that we can compare the linear and second order terms and find the precision matrix and mean of the joint distribution.

In general the joint distribution to find the experssino for the postiror dostrbution

We can express this posterior through the likelihood and prior probability by Bayesian theorem, with a constant and positive normalization constant:

$$\pi(\boldsymbol{x}|\boldsymbol{y}) \propto \pi(\boldsymbol{y}|\boldsymbol{x})\pi(\boldsymbol{x}) \tag{A.2}$$

Taking the logarithmic function of this formulation we can find an expression for the the posterior covariance, with the $\mathrm{Var}(\boldsymbol{x}) = \boldsymbol{Q}_{\boldsymbol{x}}^{-1}$ and $\mathrm{Var}(\boldsymbol{y}) = \boldsymbol{Q}_{\boldsymbol{y}}^{-1}$.

$$\ln \pi(\boldsymbol{x}|\boldsymbol{y}) \propto \ln \pi(\boldsymbol{y}|\boldsymbol{x}) + \ln \pi(\boldsymbol{x}) \tag{A.3}$$

$$= -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{Q}_{\boldsymbol{x}}(\boldsymbol{x}-\boldsymbol{\mu}) - \frac{1}{2}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})^T \boldsymbol{Q}_{\boldsymbol{y}}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}) \tag{A.4}$$

$$= -\frac{1}{2}\Big[\boldsymbol{x}^T\Big[\boldsymbol{Q}_{\boldsymbol{x}} + \boldsymbol{A}^T\boldsymbol{Q}_{\boldsymbol{y}}\boldsymbol{A}\Big]\boldsymbol{x} + \boldsymbol{x}^T\Big[-\boldsymbol{A}^T\boldsymbol{Q}_{\boldsymbol{y}}\Big]\boldsymbol{y} \tag{A.5}$$

$$+ \boldsymbol{y}^T\Big[-\boldsymbol{Q}_{\boldsymbol{y}}\boldsymbol{A}\Big]\boldsymbol{x} + \boldsymbol{y}^T\Big[\boldsymbol{Q}_{\boldsymbol{y}}\Big]\boldsymbol{y} - 2\boldsymbol{x}^T\boldsymbol{Q}_{\boldsymbol{x}}\boldsymbol{\mu}\Big] + \text{const.} \tag{A.6}$$

Hence we deal with a Gaussian distribution, we consider second order terms only and rearrange to the precision matrix.

$$-\frac{1}{2}\Big[\boldsymbol{x}^T\Big[\boldsymbol{Q}_{\boldsymbol{x}} + \boldsymbol{F}^T\boldsymbol{Q}_{\boldsymbol{y}}\boldsymbol{F}\Big] + \boldsymbol{y}^T\Big[-\boldsymbol{Q}_{\boldsymbol{y}}\boldsymbol{F}\Big] \quad \boldsymbol{y}^T\Big[\boldsymbol{Q}_{\boldsymbol{y}}\Big] + \boldsymbol{x}^T\Big[-\boldsymbol{F}^T\boldsymbol{Q}_{\boldsymbol{y}}\Big]\Big]\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{bmatrix} \tag{A.7}$$

$$= \begin{bmatrix}\boldsymbol{x}^T & \boldsymbol{y}^T\end{bmatrix} \underbrace{\begin{bmatrix}\boldsymbol{Q}_x + \boldsymbol{F}^T\boldsymbol{Q}_y\boldsymbol{F} & -\boldsymbol{F}^T\boldsymbol{Q}_y \\ -\boldsymbol{Q}_y\boldsymbol{F} & \boldsymbol{Q}_y\end{bmatrix}}_{\text{precision matrix}} \begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{bmatrix} \tag{A.8}$$

We denote the precision matrix of the joint field as:

$$\boldsymbol{Q}_{xy} = \begin{bmatrix}\boldsymbol{Q}_{aa} & \boldsymbol{Q}_{ab} \\ \boldsymbol{Q}_{ba} & \boldsymbol{Q}_{bb}\end{bmatrix} = \begin{bmatrix}\boldsymbol{Q}_x + \boldsymbol{F}^T\boldsymbol{Q}_y\boldsymbol{F} & -\boldsymbol{F}^T\boldsymbol{Q}_y \\ -\boldsymbol{Q}_y\boldsymbol{F} & \boldsymbol{Q}_y\end{bmatrix} \tag{A.9}$$

The mean is defined through the linear term.

$$\frac{-2\boldsymbol{x}^T\boldsymbol{Q}_{\boldsymbol{x}}\boldsymbol{\mu}}{-2} = \begin{bmatrix}\boldsymbol{x}^T & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{Q}_{\boldsymbol{x}}\boldsymbol{\mu} \\ 0\end{bmatrix} \tag{A.10}$$

Comparing to the linear term of Equation A.1 we can formulate an expression for the joint mean:

$$\Rightarrow \boldsymbol{\mu}_{\boldsymbol{xy}} = \boldsymbol{Q}_{\boldsymbol{xy}}^{-1}\begin{bmatrix}\boldsymbol{Q}_{\boldsymbol{x}}\boldsymbol{\mu} \\ 0\end{bmatrix} \tag{A.11}$$

The mean of the conditional distribution $\boldsymbol{x}|\boldsymbol{y}$ is given by:

$$\boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{y}} = \boldsymbol{\mu}_{\boldsymbol{x}} + \boldsymbol{Q}_{ba}^{-1}\boldsymbol{Q}_{ab}(\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{y}}) \tag{A.12}$$

$$\boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{y}} = \boldsymbol{\mu} + (\boldsymbol{Q}_x + \boldsymbol{F}^T\boldsymbol{Q}_y\boldsymbol{F})^{-1}\boldsymbol{F}^T\boldsymbol{Q}_y(\boldsymbol{x} - \boldsymbol{F}\boldsymbol{\mu})\,, \tag{A.13}$$

and the covariance of $\boldsymbol{x}|\boldsymbol{y}$ is given by:

$$\boldsymbol{Q}_{\boldsymbol{x}|\boldsymbol{y}} = \boldsymbol{Q}_{aa} = \boldsymbol{Q}_x + \boldsymbol{F}^T\boldsymbol{Q}_y\boldsymbol{F}\,, \tag{A.14}$$

as illustrated through Theorem 2.5 in [2].

B Convergence of the Metropolis-Hastings

If we show that the detailed balance condition holds and that the state space is irreducible and aperiodic under the transition matrix $\boldsymbol{P}$, we generate a Markov chain with a unique stationary distribution proportional to $\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$. Since the posterior is strictly positive $\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y}) \geq 0$ on the finite state space $\Omega(\mathcal{X}, \theta)$ the generated chain is irreducable. Further, it is possible to reject any proposed state and stay in the current state, which leads to aperiodicity. The detailed balance holds for the case that $\boldsymbol{j} = \boldsymbol{i}$, but if $\boldsymbol{j} \neq \boldsymbol{i}$ it is not trivial. In case we accept $\{\boldsymbol{x}, \boldsymbol{\theta}\}^{(n+1)} = \boldsymbol{j}$ as the new state we have $\pi(\boldsymbol{j}|\boldsymbol{y})g(\boldsymbol{i}|\boldsymbol{j}) > \pi(\boldsymbol{i}|\boldsymbol{y})g(\boldsymbol{j}|\boldsymbol{i})$. This gives us $\alpha(\boldsymbol{j}|\boldsymbol{i}) = 1$ and $\alpha(\boldsymbol{i}|\boldsymbol{j}) = \frac{\pi_{\boldsymbol{i}} g(\boldsymbol{j}|\boldsymbol{i})}{\pi_{\boldsymbol{j}} g(\boldsymbol{i}|\boldsymbol{j})}$ and satisfies the detailed balance:

$$\cancel{\pi_{\boldsymbol{j}}} \frac{\pi_{\boldsymbol{i}}}{\cancel{\pi_{\boldsymbol{j}}}} g(\boldsymbol{j}|\boldsymbol{i}) = \pi_{\boldsymbol{i}} g(\boldsymbol{j}|\boldsymbol{i}) \quad .$$

If $\pi(\boldsymbol{j}|\boldsymbol{y})g(\boldsymbol{i}|\boldsymbol{j}) < \pi(\boldsymbol{i}|\boldsymbol{y})g(\boldsymbol{j}|\boldsymbol{i})$ then $\alpha(\boldsymbol{i}|\boldsymbol{j}) = 1$ and $\alpha(\boldsymbol{j}|\boldsymbol{i}) = \frac{\pi_{\boldsymbol{j}} g(\boldsymbol{i}|\boldsymbol{j})}{\pi_{\boldsymbol{i}} g(\boldsymbol{j}|\boldsymbol{i})}$, this satisfies the detailed balance as well.

In conclusion the Metropolis-Hastings algorithm samples from a unique distribution proportional to the posterior distribution.

C
Randomize then Optimize - RTO

$$
\begin{aligned}
\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}) &\propto \pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta}) && \text{(C.1)}\\
&\propto \exp\Big[(\boldsymbol{F}\boldsymbol{x}-\boldsymbol{y})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{F}\boldsymbol{x}-\boldsymbol{y}) + (\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{Q}(\boldsymbol{x}-\boldsymbol{\mu})\Big] && \text{(C.2)}\\
&= \exp\|\hat{\boldsymbol{F}}\boldsymbol{x}-\hat{\boldsymbol{y}}\|^2 && \text{(C.3)}
\end{aligned}
$$

where

$$
\hat{\boldsymbol{F}} = \begin{bmatrix}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{F}\\ \boldsymbol{Q}^{1/2}\end{bmatrix}, \quad \hat{\boldsymbol{y}} = \begin{bmatrix}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}\\ \boldsymbol{Q}^{1/2}\boldsymbol{\mu}\end{bmatrix} \tag{C.4}
$$

One sample from the posterior can be computed by minimizing the following with respect to $\boldsymbol{x}$

$$
\boldsymbol{x} = \arg\min_{\hat{\boldsymbol{x}}}\|\hat{\boldsymbol{F}}\hat{\boldsymbol{x}} - (\hat{\boldsymbol{y}}+\boldsymbol{\eta})\|^2, \quad \boldsymbol{\eta}\sim\mathcal{N}(\mathbf{0},\mathbf{I}) \tag{C.5}
$$

We can solve this and rewrite to

$$
\begin{aligned}
\frac{\partial}{\partial\boldsymbol{x}}\Big[(\hat{\boldsymbol{F}}\boldsymbol{x}-(\hat{\boldsymbol{y}}+\boldsymbol{\eta})^T(\hat{\boldsymbol{F}}\boldsymbol{x}-(\hat{\boldsymbol{y}}+\boldsymbol{\eta})\Big] &= 0 && \text{(C.6)}\\
\Leftrightarrow \boldsymbol{x}^T\hat{\boldsymbol{F}}^T\hat{\boldsymbol{F}} + \hat{\boldsymbol{F}}^T\hat{\boldsymbol{F}}\boldsymbol{x} - \hat{\boldsymbol{F}}^T(\hat{\boldsymbol{y}}+\boldsymbol{\eta}) - (\hat{\boldsymbol{y}}+\boldsymbol{\eta})^T\hat{\boldsymbol{F}}\boldsymbol{x} &= 0 && \text{(C.7)}
\end{aligned}
$$

We can argue through the symmetry of the inner product that and the symmetry of the precision matrix

$$
\hat{\boldsymbol{F}}^T\hat{\boldsymbol{F}}\boldsymbol{x} = \hat{\boldsymbol{F}}^T(\hat{\boldsymbol{y}}-\boldsymbol{\eta}) \tag{C.8}
$$

$$
\Leftrightarrow (\boldsymbol{F}^T\boldsymbol{Q}_y\boldsymbol{F}+\boldsymbol{Q})\boldsymbol{x} = \boldsymbol{F}^T\boldsymbol{Q}_y\boldsymbol{y} + \boldsymbol{Q}\boldsymbol{\mu} - \hat{\boldsymbol{F}}^T\boldsymbol{\eta} \tag{C.9}
$$

If we substitute $-\hat{\boldsymbol{F}}^T\boldsymbol{\eta} = \boldsymbol{v}_1 + \boldsymbol{v}_2$ we end up with

$$(\boldsymbol{F}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{F} + \boldsymbol{Q})\boldsymbol{x} = \boldsymbol{F}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y} + \boldsymbol{Q}\boldsymbol{\mu} + \boldsymbol{v}_1 + \boldsymbol{v}_2 \tag{C.10}$$

where $\boldsymbol{v}_1 \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{F}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{F})$ and $\boldsymbol{v}_2 \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q})$ are independent random variables. mayeb introduce... x^2 time nomral variubale

D
Inverting Matrices - QR factorization

E
Taylor expansion of $g(\lambda)$

We Taylor expand the function $g(\lambda)$ around $\lambda = \lambda' - \Delta\lambda$

$$g(\lambda) = \ln\det\underbrace{(\boldsymbol{F}^T\boldsymbol{F} + \lambda\boldsymbol{L})}_{\boldsymbol{B}} \tag{E.1}$$

$$\begin{aligned} g(\lambda') - g(\lambda) &= \ln\det(\boldsymbol{F}^T\boldsymbol{F} + \lambda'\boldsymbol{L}) - \ln\det(\boldsymbol{F}^T\boldsymbol{F} + \lambda\boldsymbol{L}) && \text{(E.2)} \\ &= \ln\det\left[\frac{(\boldsymbol{F}^T\boldsymbol{F} + (\lambda + \Delta\lambda)\boldsymbol{L})}{(\boldsymbol{F}^T\boldsymbol{F} + \lambda\boldsymbol{L})}\right] && \text{(E.3)} \\ &= \ln\det\left[1 + \frac{\Delta\lambda\boldsymbol{L}}{\boldsymbol{B}}\right] && \text{(E.4)} \\ &= \sum_{r=1}^{\infty}\frac{(-1)^{r+1}}{r!}\mathrm{tr}((\boldsymbol{B}^{-1}\boldsymbol{L})^r)(\Delta\lambda)^r && \text{(E.5)} \end{aligned}$$

, where we use the identity from [11] at page 29. So the derivatives of $g(\lambda)$ are:

$$\begin{aligned} g^{(r)}(\lambda) =& (-1)^{r+1}\,\mathrm{tr}\Big((\boldsymbol{B}^{-1}\boldsymbol{L})^r\Big) && \text{(E.6)} \\ \approx& (-1)^{r+1}\sum_{k=1}^{p}\boldsymbol{z}_{\boldsymbol{k}}^T(\boldsymbol{B}^{-1}\boldsymbol{L})^r\boldsymbol{z}_{\boldsymbol{k}} && \text{(E.7)} \end{aligned}$$

Here we use a Monte Carlo estimate and draw p vectors $\boldsymbol{z}_{\boldsymbol{k}} \in \mathbb{R}^n$, where each vector element $z_i \overset{\text{i.i.d.}}{\sim} \mathcal{U}(\{-1, 1\})$ and $i = 1, \ldots, n$.

F
Radiation transfer and absorption line shape

G
whispering gallery resonator

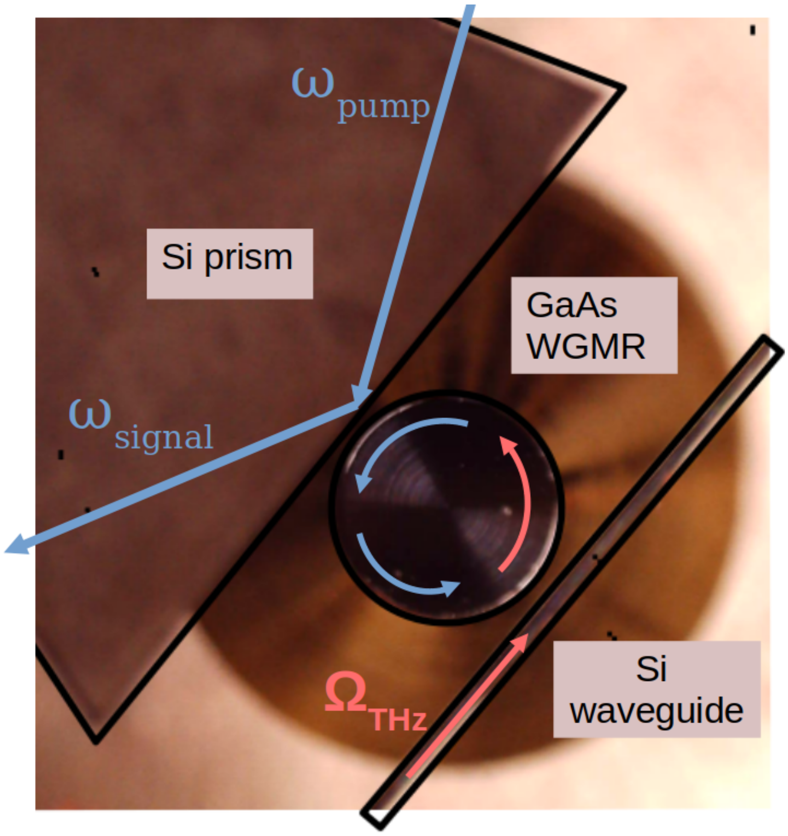

Figure G.1: whispering gallery resonator

References

[1] Colin Fox and Richard A Norton. "Fast sampling in a linear-Gaussian inverse problem". In: *SIAM/ASA Journal on Uncertainty Quantification* 4.1 (2016), pp. 1191–1218.

[2] Havard Rue and Leonhard Held. *Gaussian Markov random fields: theory and applications*. London: CRC press, 2005.

[3] J. Andrés Christen and Colin Fox. "A general purpose sampling algorithm for continuous distributions (the t-walk)". In: *Bayesian Analysis* 5.2 (2010), pp. 263 –281. URL: https://doi.org/10.1214/10-BA603.

[4] Tiangang Cui and Sergey Dolgov. "Deep composition of tensor-trains using squared inverse rosenblatt transports". In: *Foundations of Computational Mathematics* 22.6 (2022), pp. 1863–1922.

[5] *U.S. Standard Atmosphere, 1976*. Washington, D.C.: United States. National Oceanic and Atmospheric Administration, United States Committee on Extension to the Standard Atmosphere, 1976. URL: https://www.ngdc.noaa.gov/stp/space-weather/online-publications/miscellaneous/us-standard-atmosphere-1976/us-standard-atmosphere_st76-1562_noaa.pdf.

[6] C. Readings and R. A. Harris. *Envisat MIPAS an Instrument for Atmospheric Chemistry and Climate Research.* https://earth.esa.int/eogateway/documents/20142/37627/envisat-mipas-instrument-description.pdf. [Online; accessed 16/07/22]. 2000.

[7] Iouli E Gordon et al. "The HITRAN2020 molecular spectroscopic database". In: *Journal of Quantitative Spectroscopy and Radiative Transfer* 277 (2022), p. 107949.

[8] Per Christian Hansen and Dianne Prost O'Leary. "The use of the L-curve in the regularization of discrete ill-posed problems". In: *SIAM Journal on Scientific Computing* 14.6 (1993), pp. 1487–1503.

[9] Ville Satopäa et al. "Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior". In: *2011 31st International Conference on Distributed Computing Systems Workshops*. IEEE. 2011, pp. 166–171.

[10] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. New York: Springer, 2006.

[11] Israel Gohberg, Seymour Goldberg, and Nahum Krupnik. *Traces and Determinants of Linear Operators*. Operator Theory: Advances and Applications. Basel: Birkhäuser Basel, 2012.