

# **Suitably impressive thesis title**

Lennart Golks

Department of Physics  
University of Otago

*A thesis submitted for the degree of  
Doctor of Philosophy*

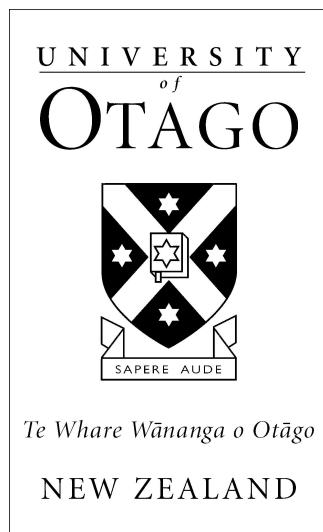
November 2025

## **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

<b>List of Figures</b>	<b>ix</b>
<b>List of Abbreviations</b>	<b>xi</b>
<b>1 Introduction</b>	<b>3</b>
1.1 What is going on?, 3 facts, What is new in this thesis? . . . . .	3
1.2 What has been published? . . . . .	3
<b>2 Building a physics based hierarchical Linear model</b>	<b>5</b>
2.1 Linear model . . . . .	5
2.2 two hyperameters, fast sampling paper . . . . .	6
2.3 four hyperameters, t-walk, TT approx, and RTO . . . . .	7
2.3.1 Sampling . . . . .	8
2.3.2 Rosenblatt Transform . . . . .	8
2.3.3 (Squared) Inverse Rosenblatt Transform . . . . .	8
2.3.4 Tensor-Train Approximation . . . . .	8
2.4 Temperature and pressure hyperameters, tt-approx . . . . .	8
<b>3 Nonlinear Forward model</b>	<b>31</b>
3.1 Sampling . . . . .	31
3.2 local linear Map and strategy . . . . .	31
3.2.1 Machine learning vs Gaussian elimination . . . . .	31
3.3 affine RTO . . . . .	31
<b>4 Introduction</b>	<b>37</b>
4.1 What is going on?, 3 facts, What is new in this thesis? . . . . .	37
4.2 What has been published? . . . . .	37
<b>Appendices</b>	
<b>A Posterior of Bayesian Hierachical model</b>	<b>41</b>
<b>B Convergence of the Metropolis-Hastings</b>	<b>43</b>

<b>C Randomize then Optimize - RTO</b>	<b>45</b>
<b>D Inverting Matrices - QR factorization</b>	<b>47</b>
<b>E Taylor expansion of <math>g(\lambda)</math></b>	<b>49</b>
<b>F Radiation transfer and absorption line shape</b>	<b>51</b>
<b>G whispering gallery resonator</b>	<b>53</b>
<b>References</b>	<b>55</b>

# List of Figures

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 $\lambda$ . 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. . . . .	7
2.3 The scatter plot shows independent samples of $\delta$ and $\gamma$ as the result of the MWG algorithm. The histogram displays independent samples of $\lambda \sim \pi(\lambda \mathbf{y}, \gamma)$ . The vertical line corresponds to the optimal regularization parameter. . . . .	8
2.5 For varying $\lambda$ we plot the seminorm $\sqrt{\mathbf{x}_\lambda^T \mathbf{L} \mathbf{x}_\lambda}$ against data misfit $\ \mathbf{A}\mathbf{x}_\lambda - \mathbf{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. . . . .	10
2.6 Plot of the true ozone profile ( $\bullet$ ), posterior samples ( $+$ ), and posterior mean ( $\bullet$ ). We display the optimal regularised solution ( $\nabla$ ) and the simulated data ( $*$ ) in spectral radiance. . . . .	11
2.7 short text . . . . .	12
2.14 short text . . . . .	19
2.15 show that t-walk works for posterior . . . . .	21
2.16 show that TT does not work for posterior . . . . .	22
2.17 show that TT works for prior . . . . .	23
2.18 . . . . .	24
G.1 whispering gallery resonator . . . . .	54

*x*

## List of Abbreviations

<b>i.i.d.</b>	independent and identically distributed
<b>MRF</b>	Markov Random Field
<b>GMRF</b>	Gaussian Markov Random Field
<b>MTC</b>	Marginal Then Conditional sampler
<b>GOMOS</b>	Global Ozone Monitoring by Occultation of Stars
<b>MCMC</b>	Markov Chain Monte-Carlo
<b>MH</b>	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]

parameter and context	$m$ , number of measurements	pointing accuracy	$n$ , number of atmospheric layers	signal-to-noise voltage ratio
$\mathbf{y}$ , data	33	0.009°	43	10

**Table 2.1:** Data Spec

## 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

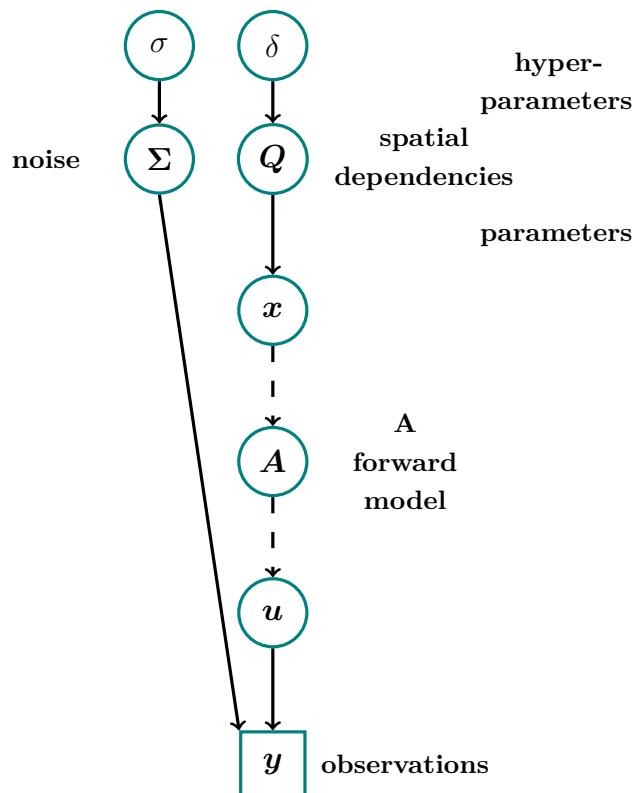
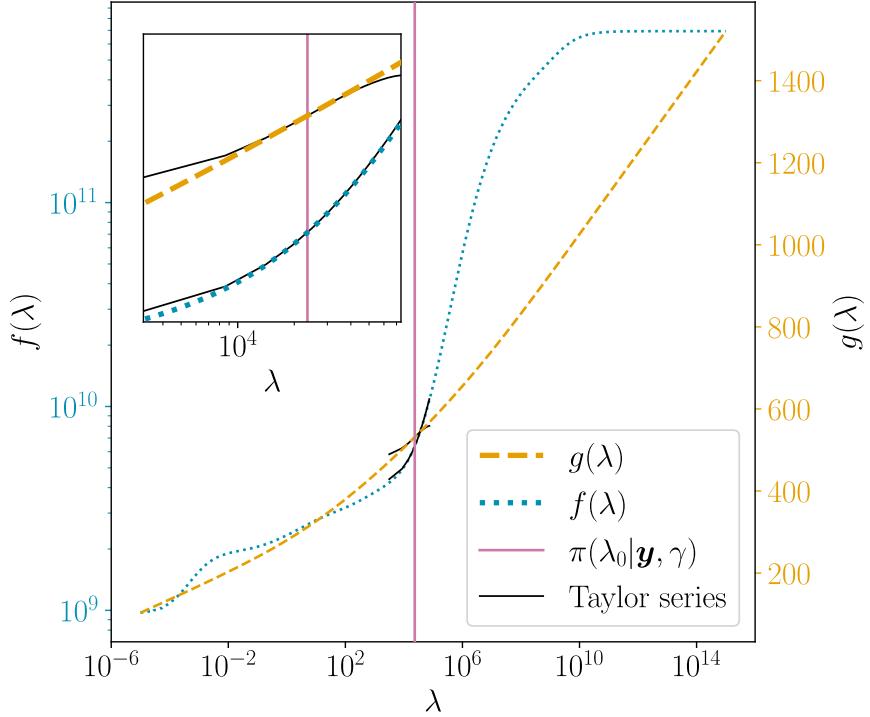


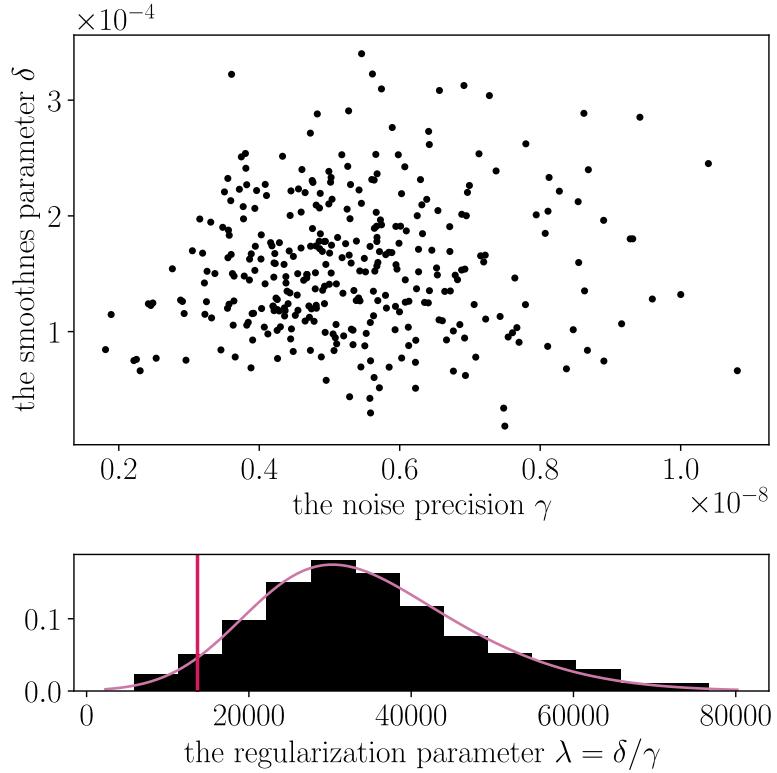
Figure 2.1



**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  $\lambda$ . 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.

## 2.3 four hyperparameters, t-walk, TT approx, and RTO

- t-walk ref
- motivation why more hyper-parameters, explain parabula
- how sample from them, compared to TT approx
- explain Rosenblatt and SIRT trasnport
- RTO
- define priors
- how solve inverse and determinant



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

- how many steps, integrated autocorrealtion time
- relative error

### 2.3.1 Sampling

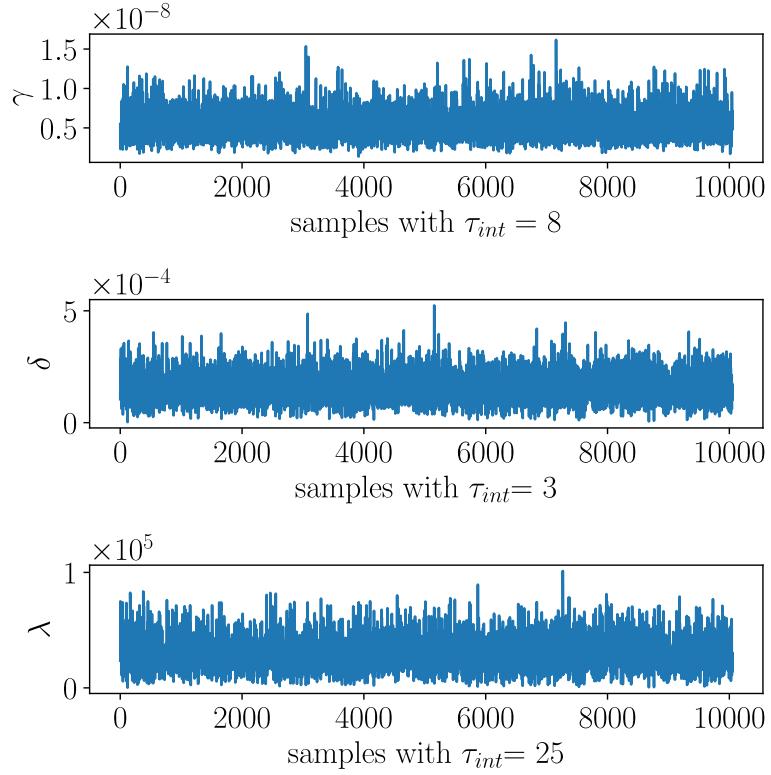
### 2.3.2 Rosenblatt Transform

### 2.3.3 (Squared) Inverse Rosenblatt Transform

### 2.3.4 Tensor-Train Approximation

## 2.4 Temperature and pressure hyperameters, tt-approx

- updating scheme

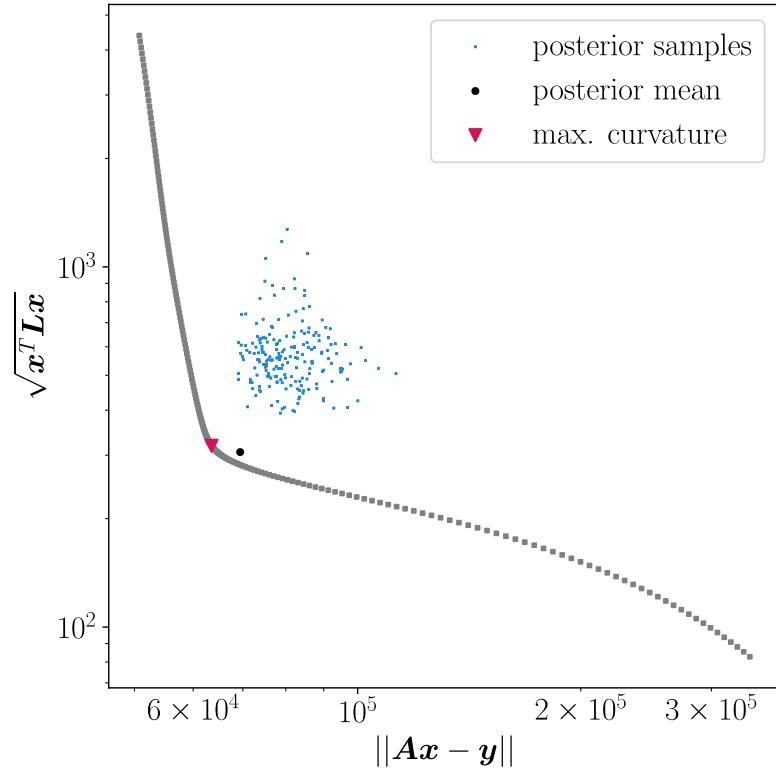
**Figure 2.4:** text

model context	burn in	steps	dimension	time
$\mathbf{x}$ , MTC	4	$10^{-10}$	2	
$\mathbf{x}$ , t-walk	500		4	
$\mathbf{p}$ , t-walk	500		5	
$\mathbf{t}$ , t-walk	2500		13	

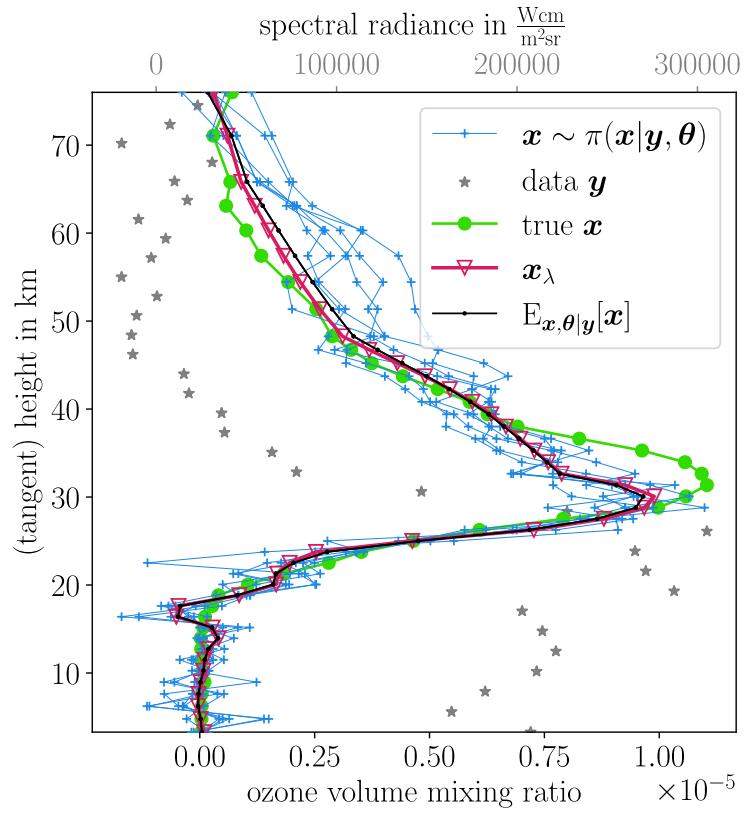
**Table 2.2:** T walk settting

model context	max number of swaps (TT-cross)	maximum relative error (TT-cross)	QMC samples per dimension	SIRT error	time
$\mathbf{x}$	4	$10^{-10}$	4		
$\mathbf{p}$	5		5		
$\mathbf{t}$	5		13		

**Table 2.3:** Sirt and TT settting



**Figure 2.5:** For varying  $\lambda$  we plot the seminorm  $\sqrt{\mathbf{x}_\lambda^T \mathbf{L} \mathbf{x}_\lambda}$  against data misfit  $\|\mathbf{A}\mathbf{x}_\lambda - \mathbf{y}\|$  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 ( $\bullet$ ), posterior samples ( $+$ ), and posterior mean ( $\bullet$ ). We display the optimal regularised solution ( $\nabla$ ) and the simulated data ( $*$ ) in spectral radiance.

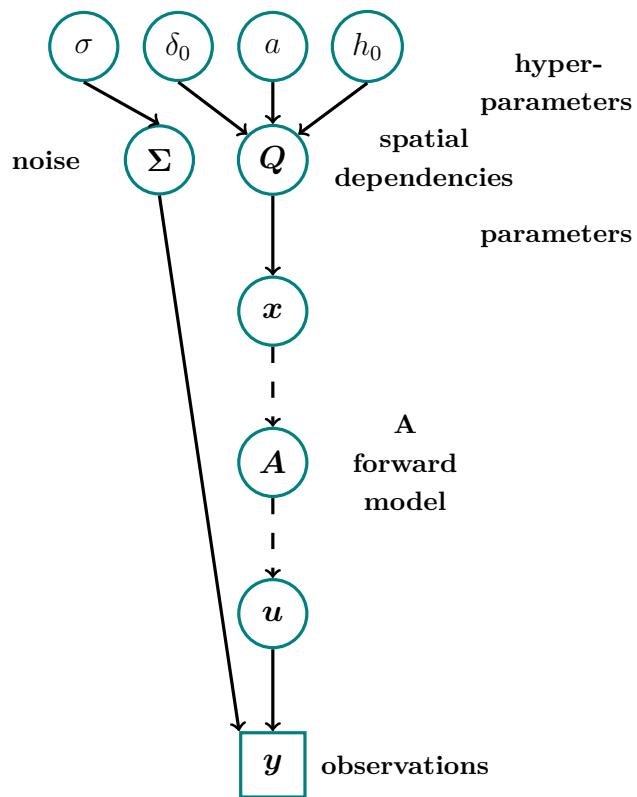
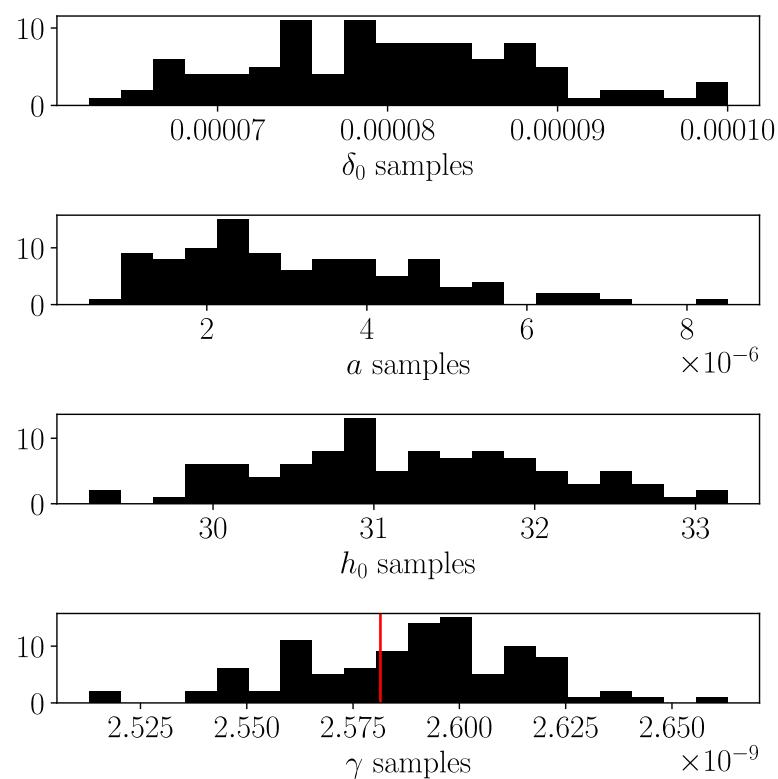
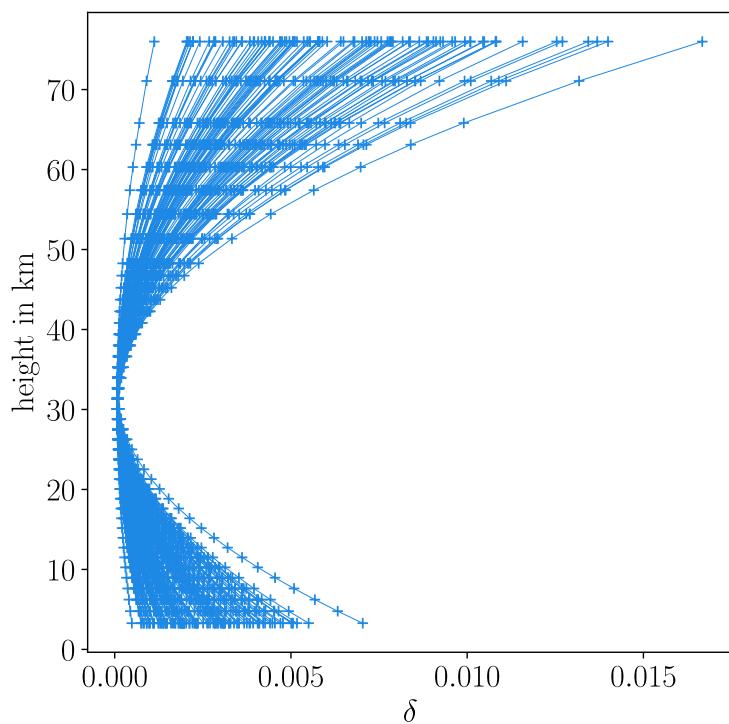


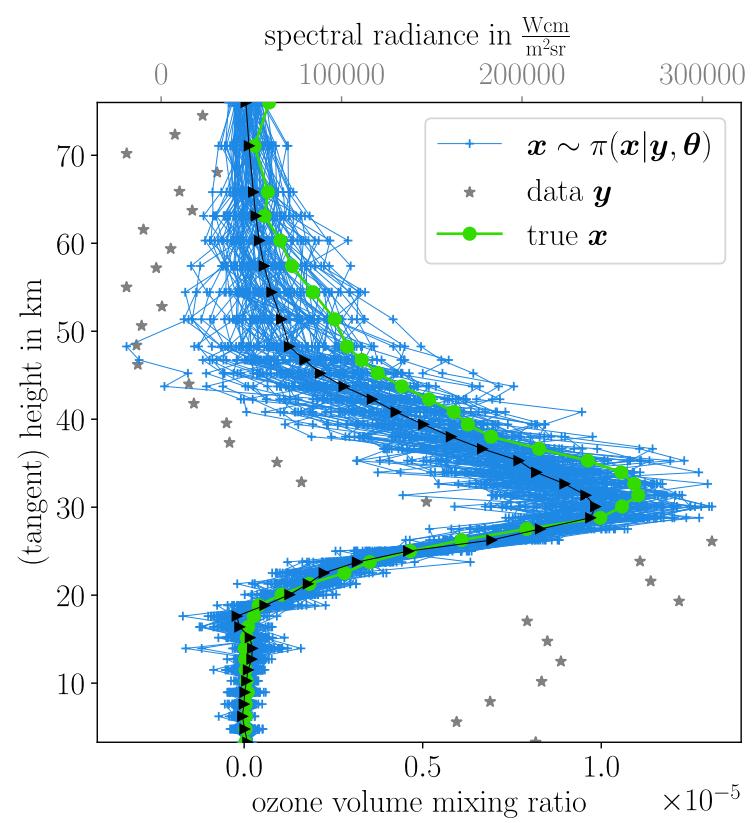
Figure 2.7: text



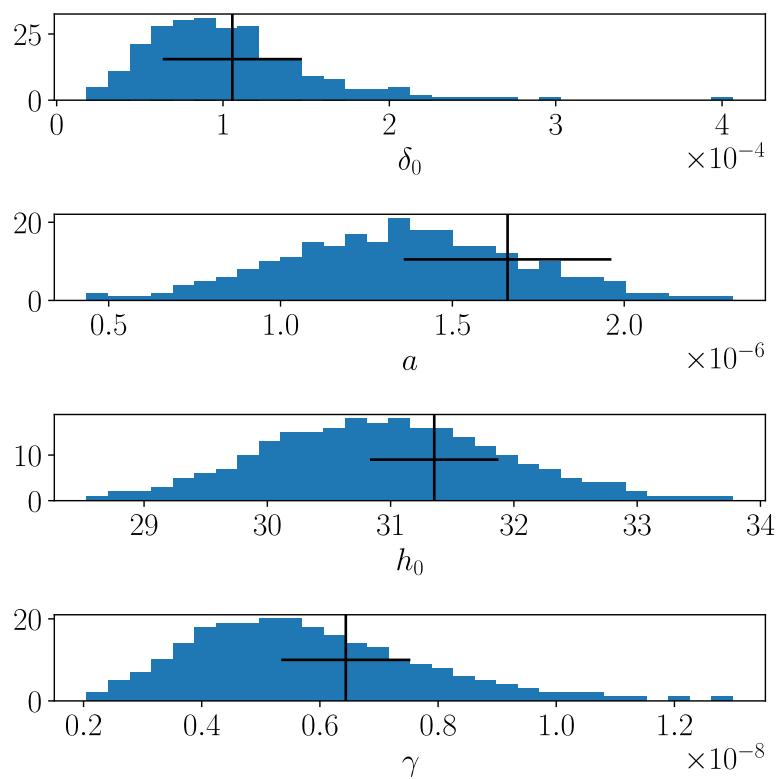
**Figure 2.8:** text



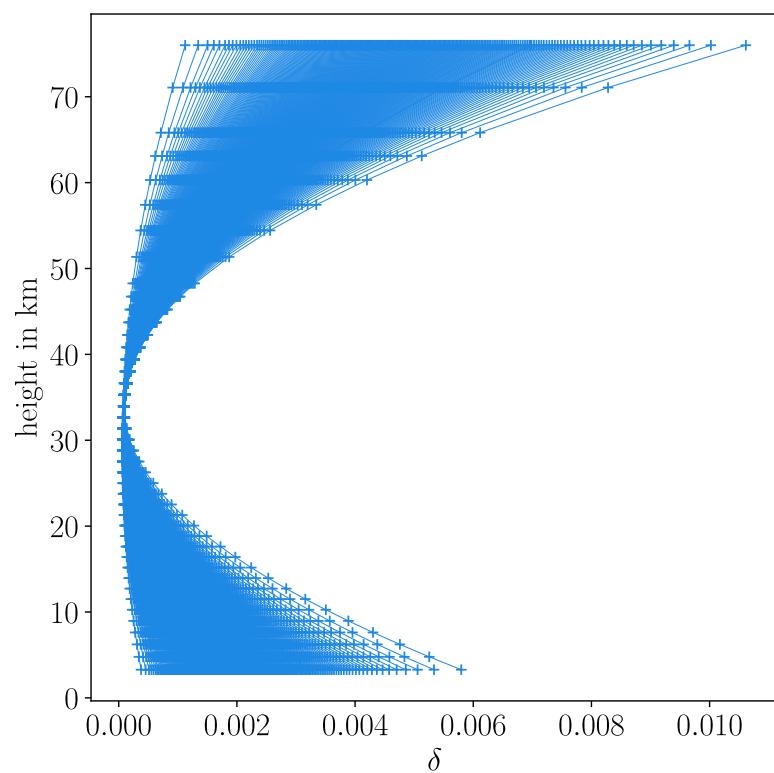
**Figure 2.9:** text



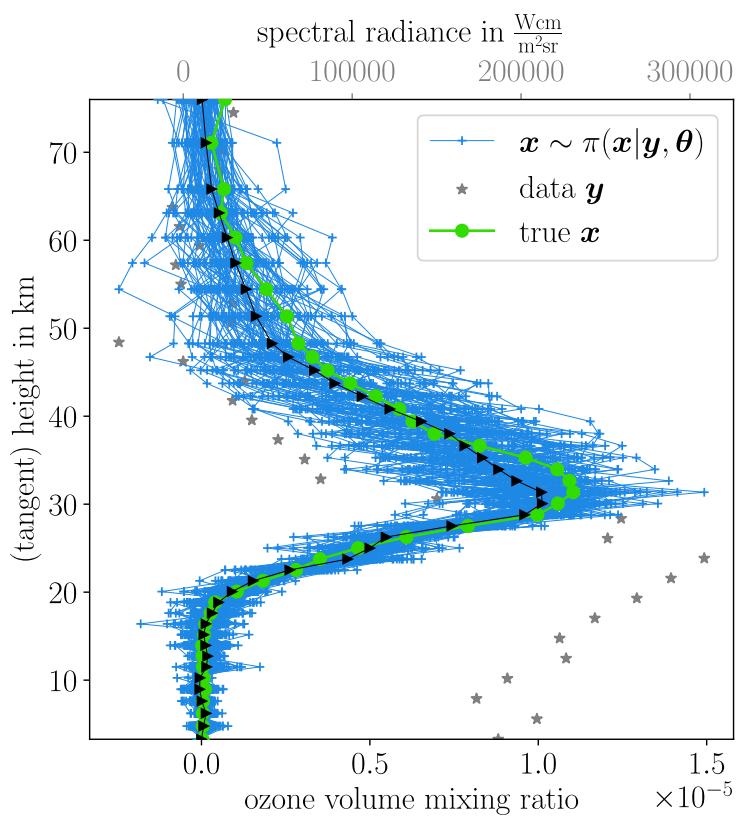
**Figure 2.10:** text



**Figure 2.11:** TTSIRT output



**Figure 2.12:** text



**Figure 2.13:** text

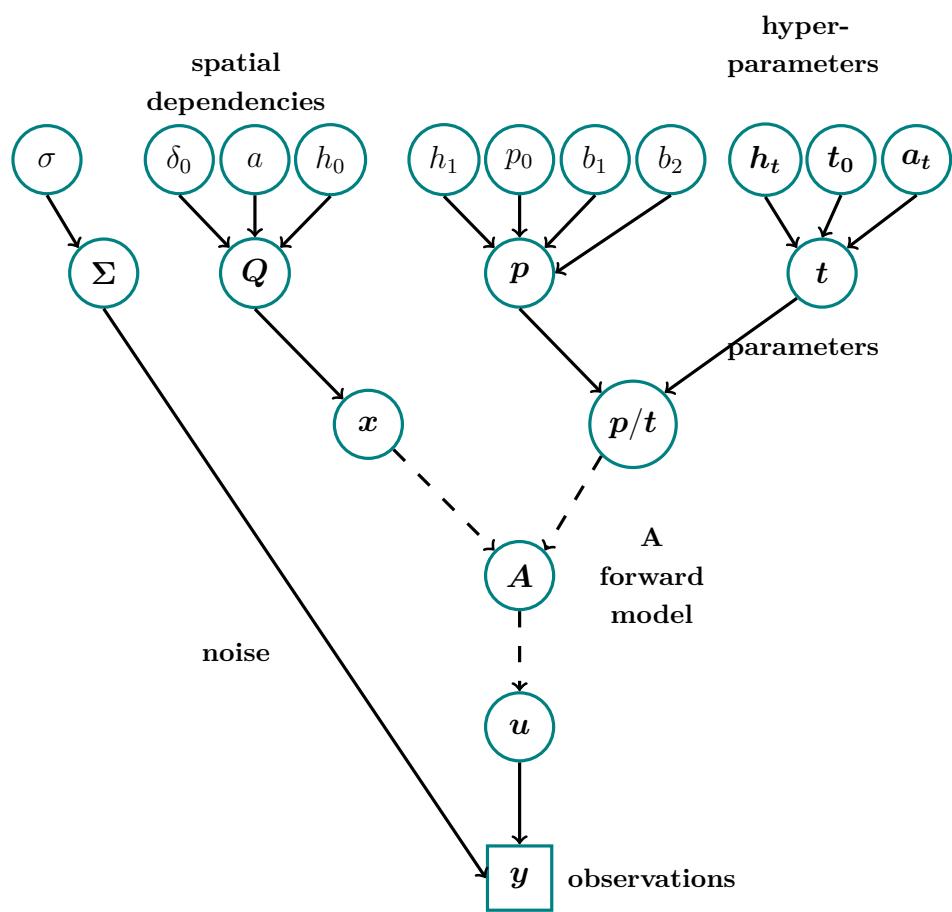
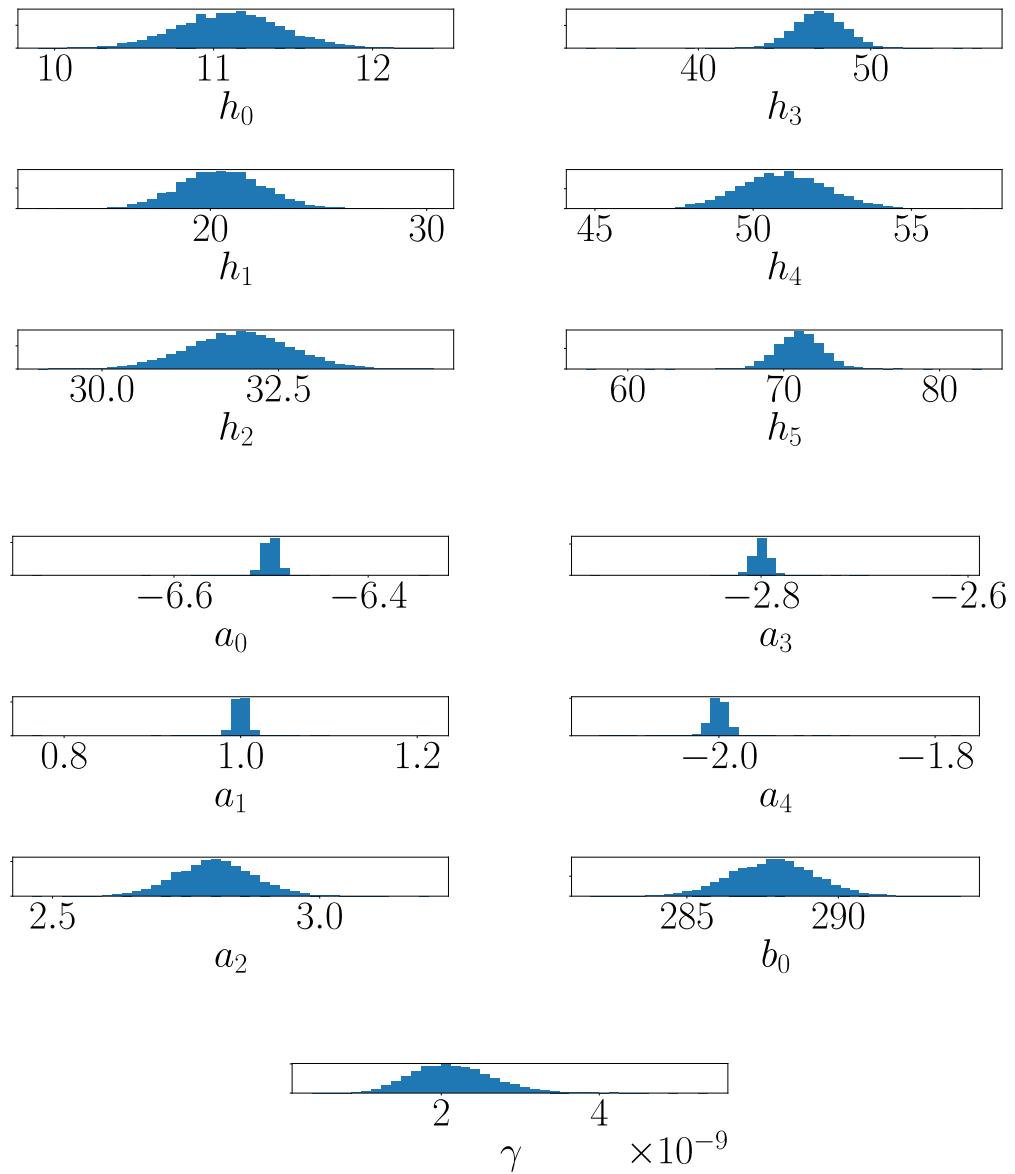


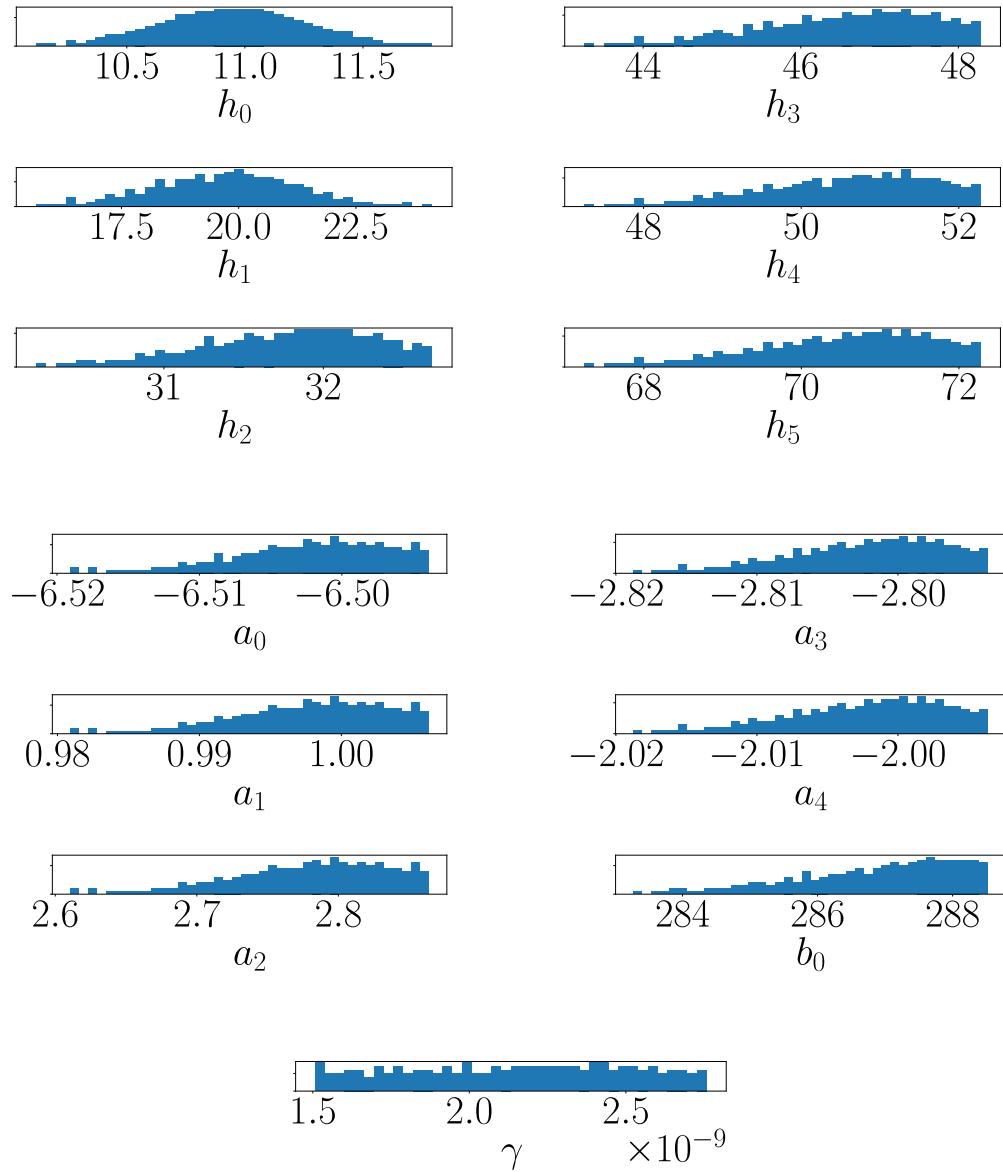
Figure 2.14: text

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

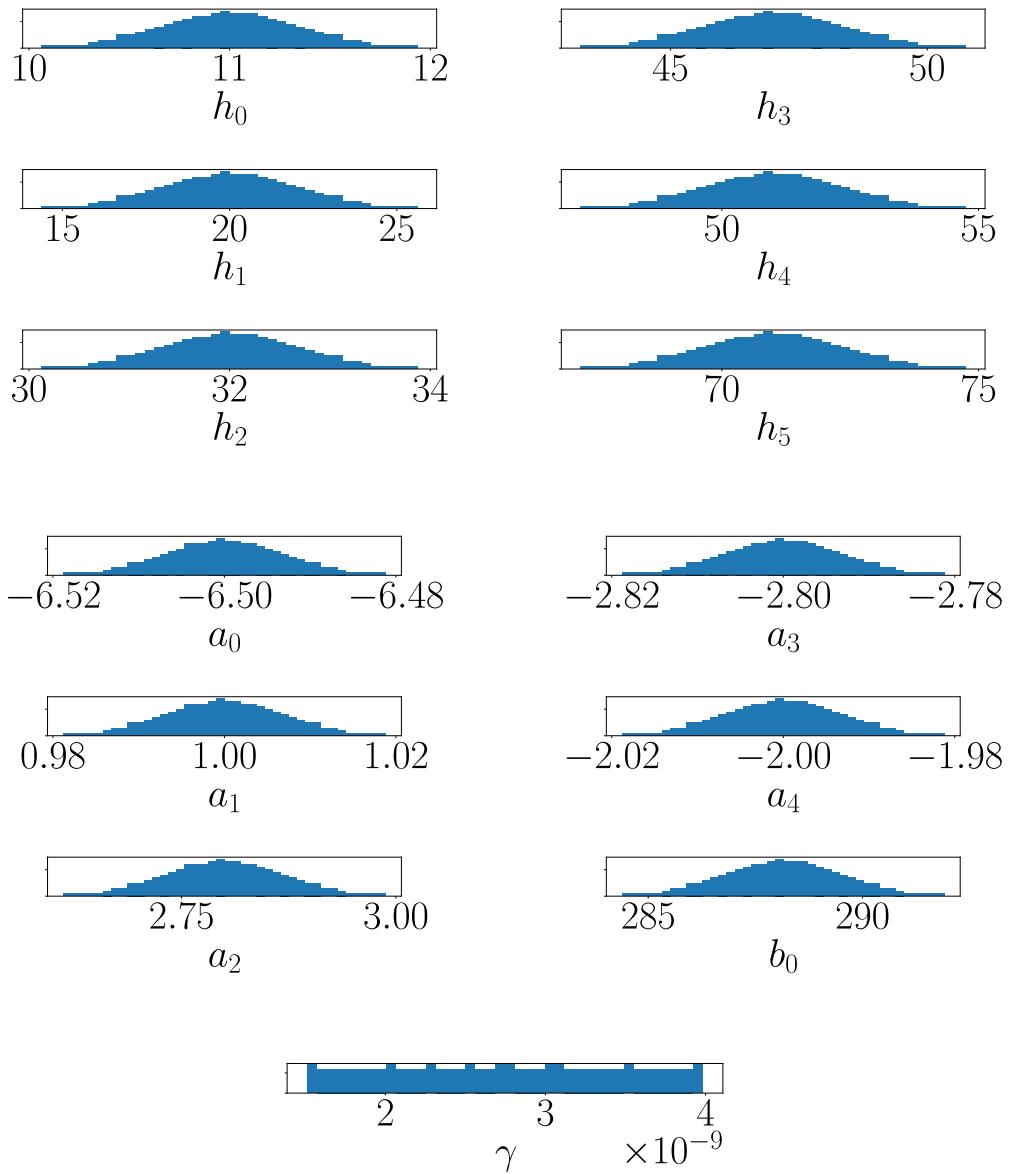
**Table 2.4:** Gaussian  $\mathcal{N}(\mu, \sigma)$  and gamma distribution  $\mathcal{T}(\alpha = \text{scale}, \beta = \text{rate})$  Bounds for t and p 2.8 times the variance around the mean round pressure approx and test if would work with previous gamma prior or fix gamma prior with set values



**Figure 2.15:** show that t- walk works for posterior



**Figure 2.16:** show that TT does not work for posterior

**Figure 2.17:** show that TT works for prior

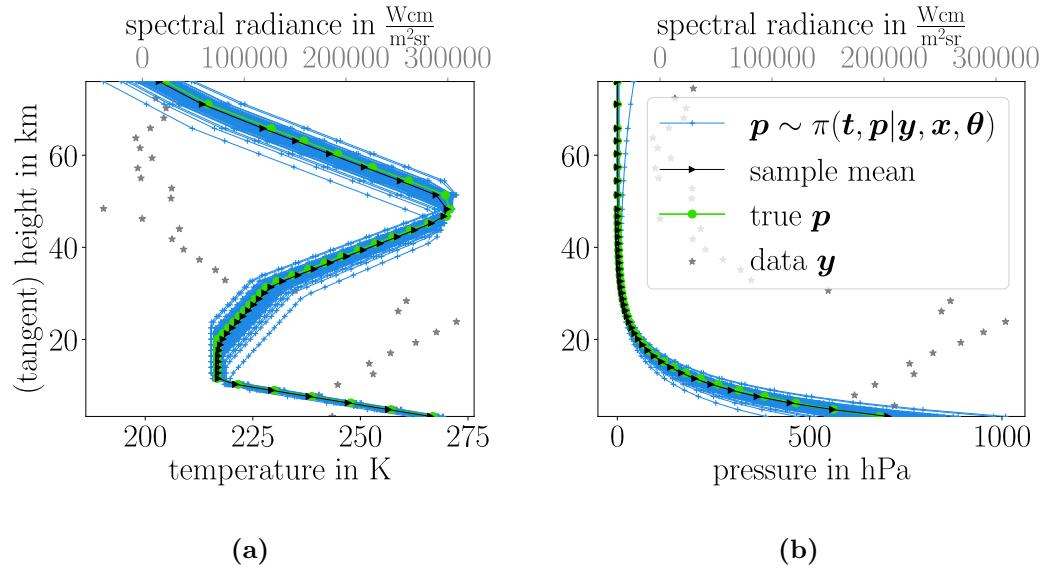


Figure 2.18

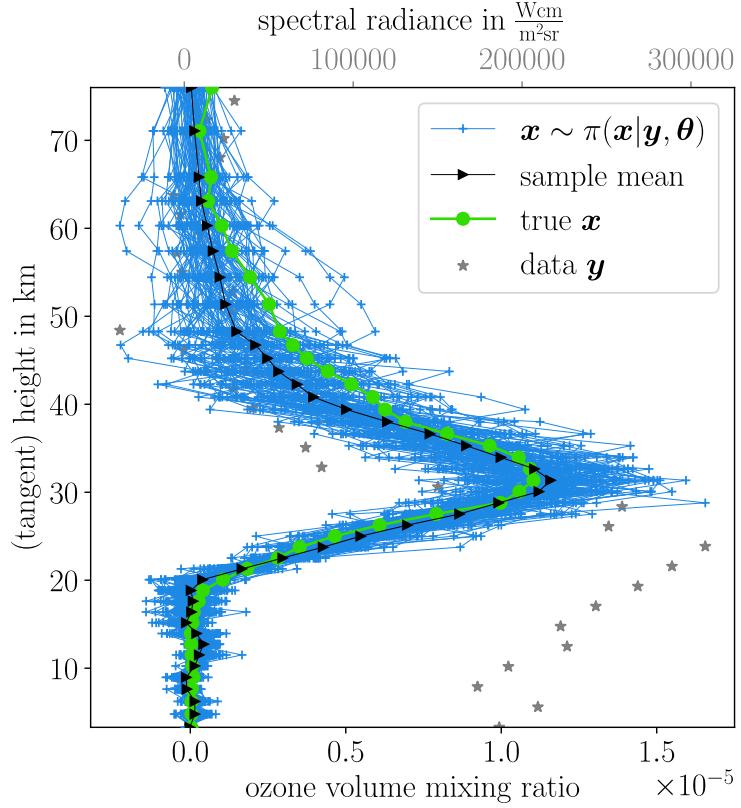
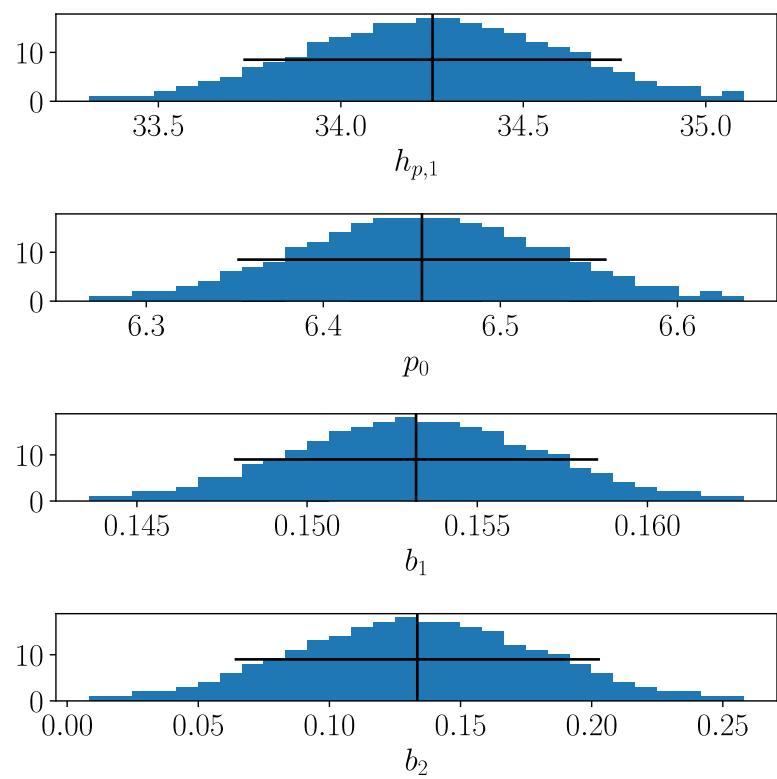
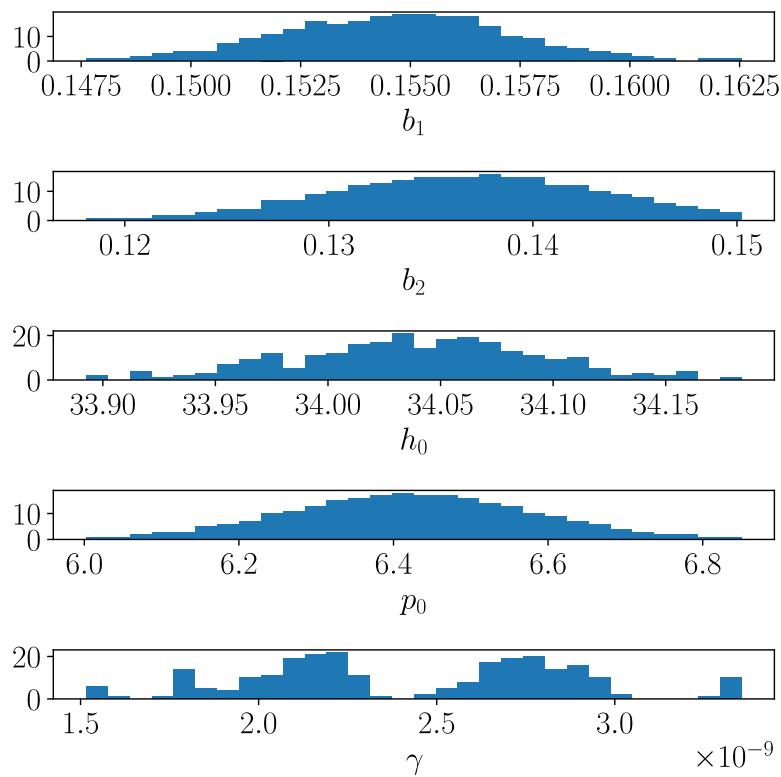


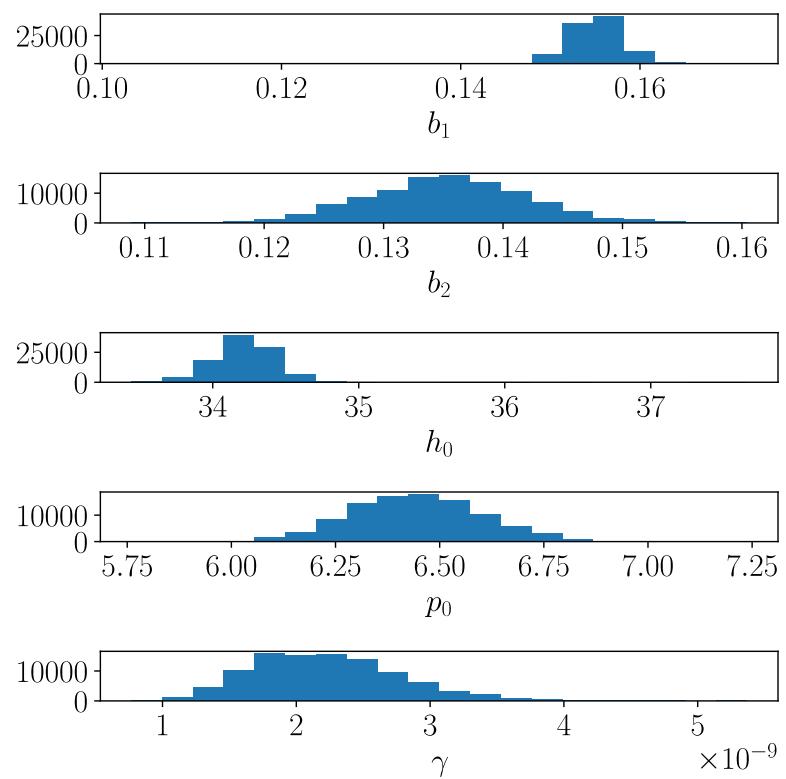
Figure 2.19: text



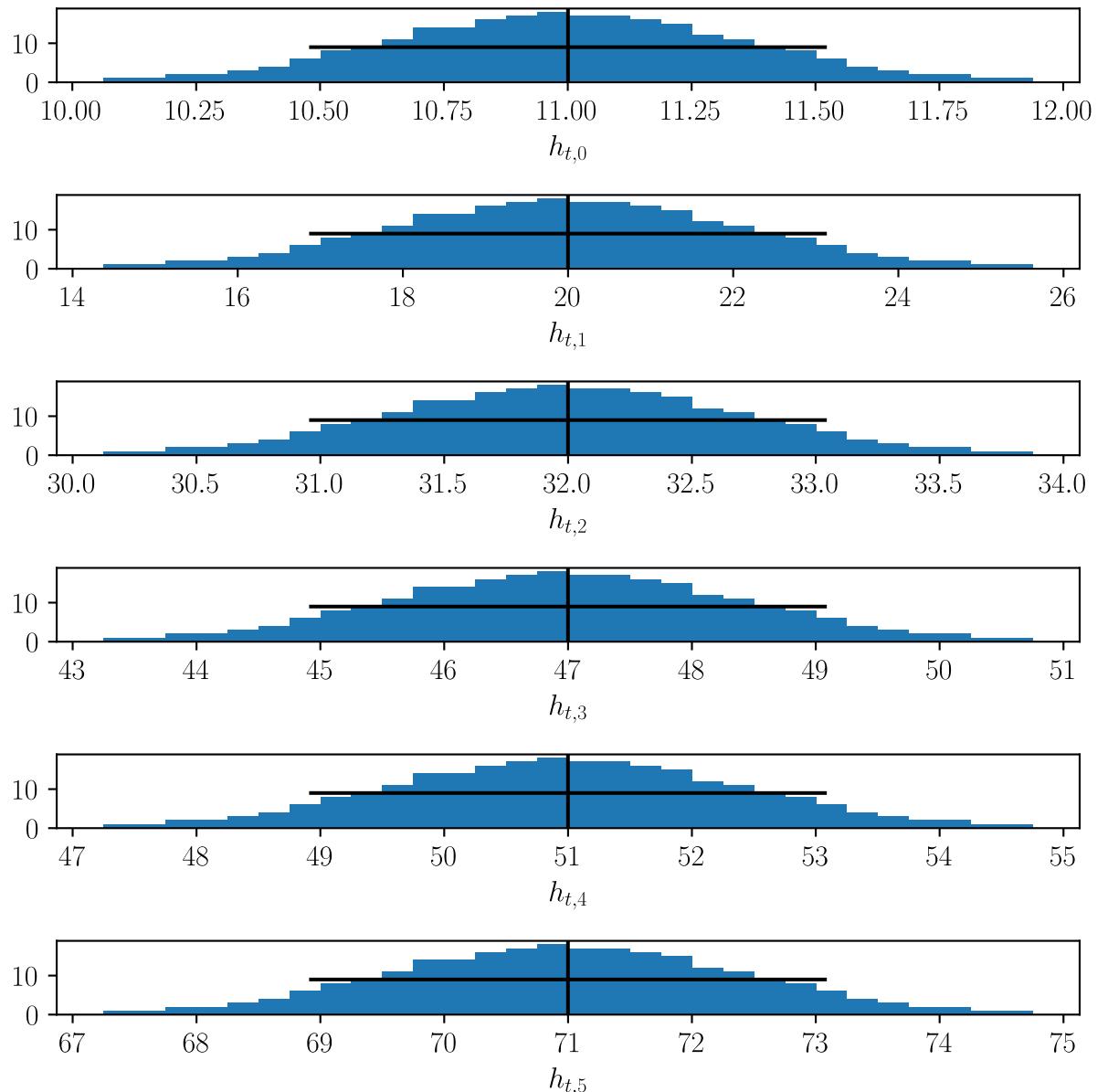
**Figure 2.20:** pritor

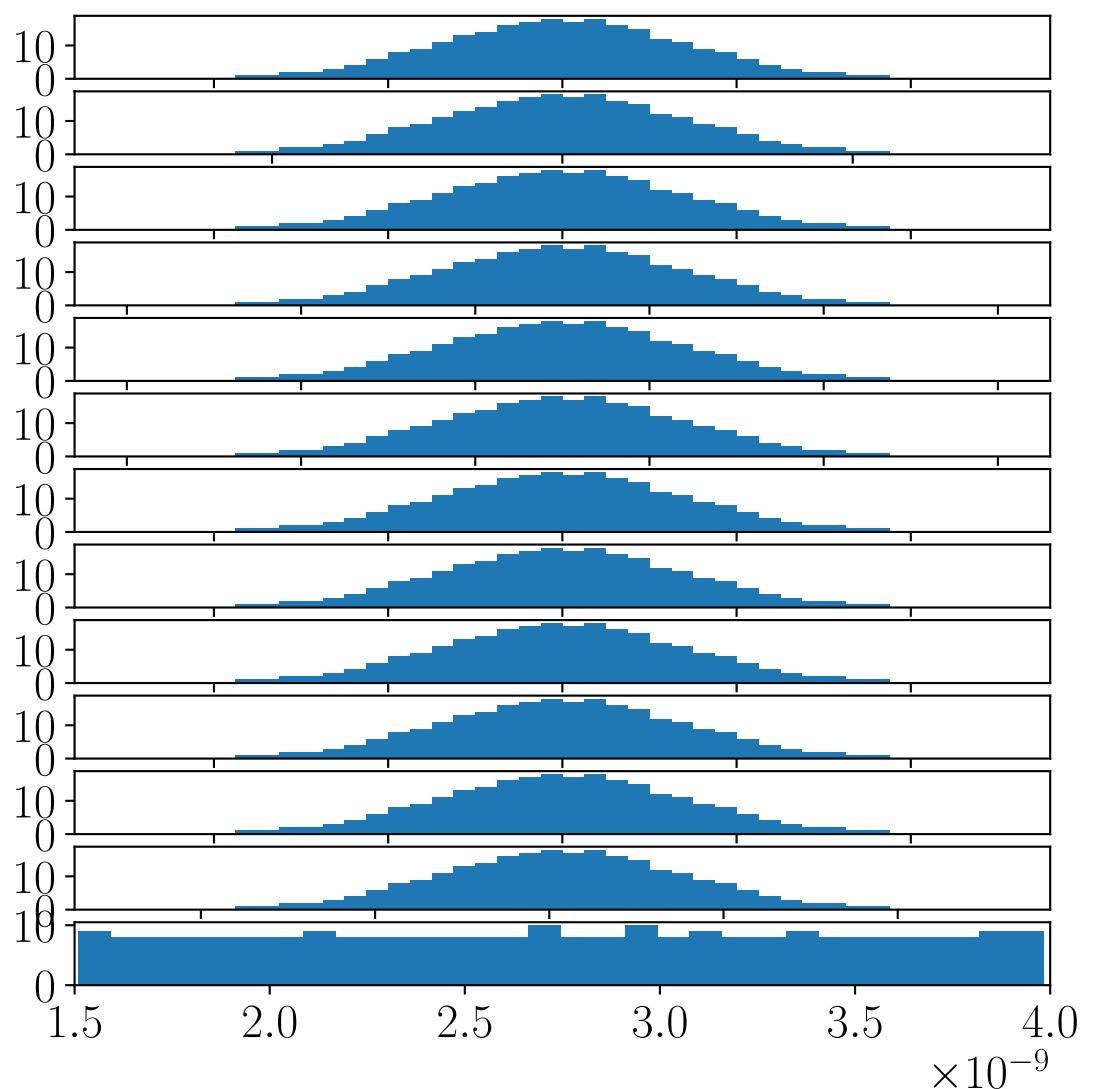


**Figure 2.21:** posteriore doesnt work with TT format

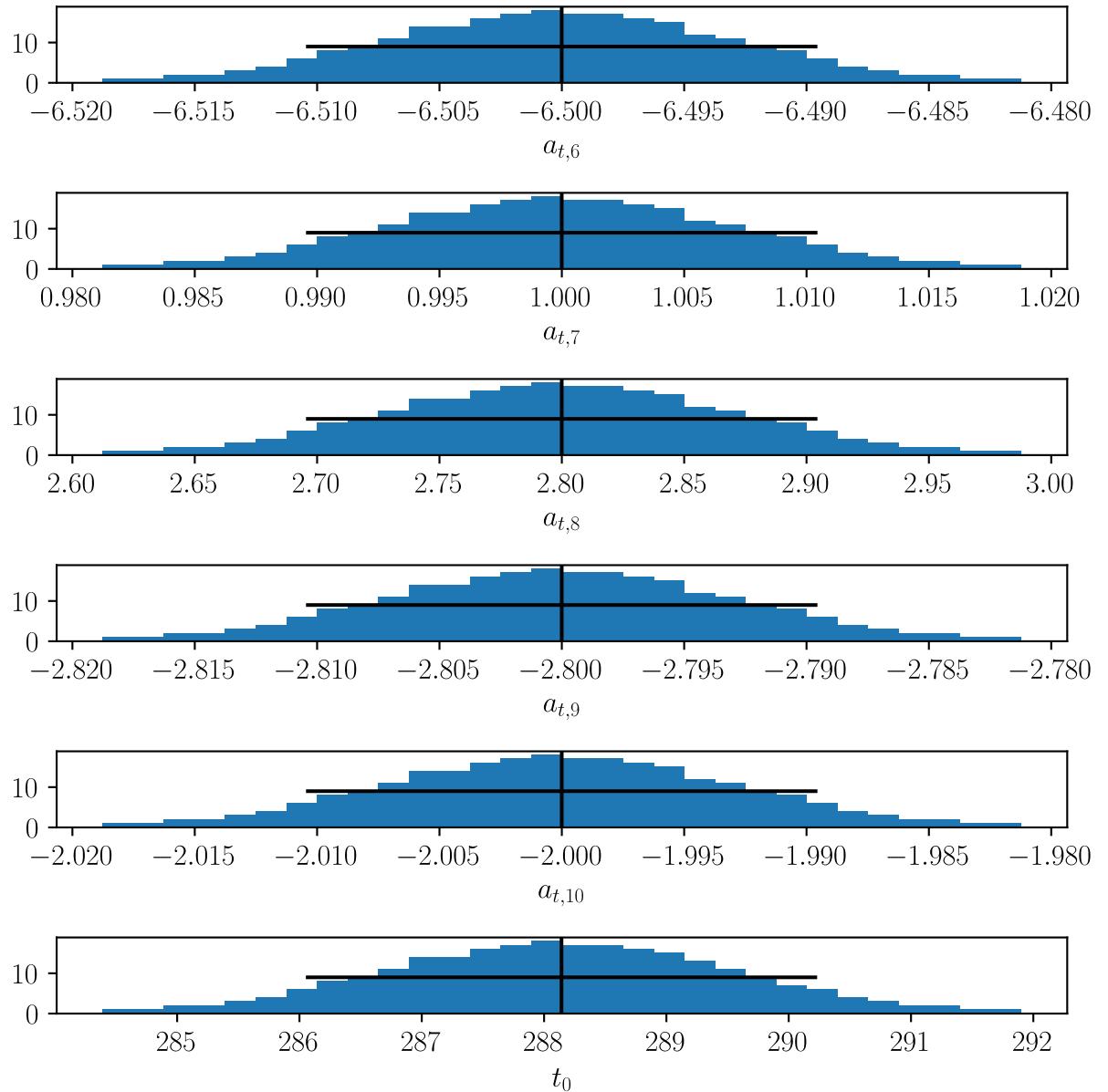


**Figure 2.22:** t-walk does work with posterioeo

**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

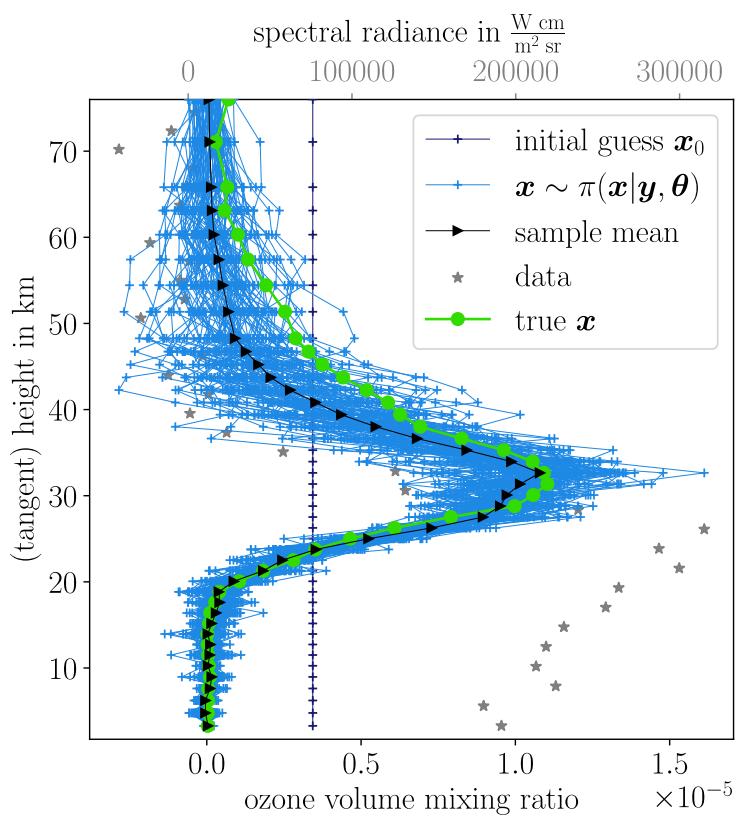
- one data vector
- strategy to find convergence and local linear map

#### 3.2.1 Machine learning vs Gaussian elimination

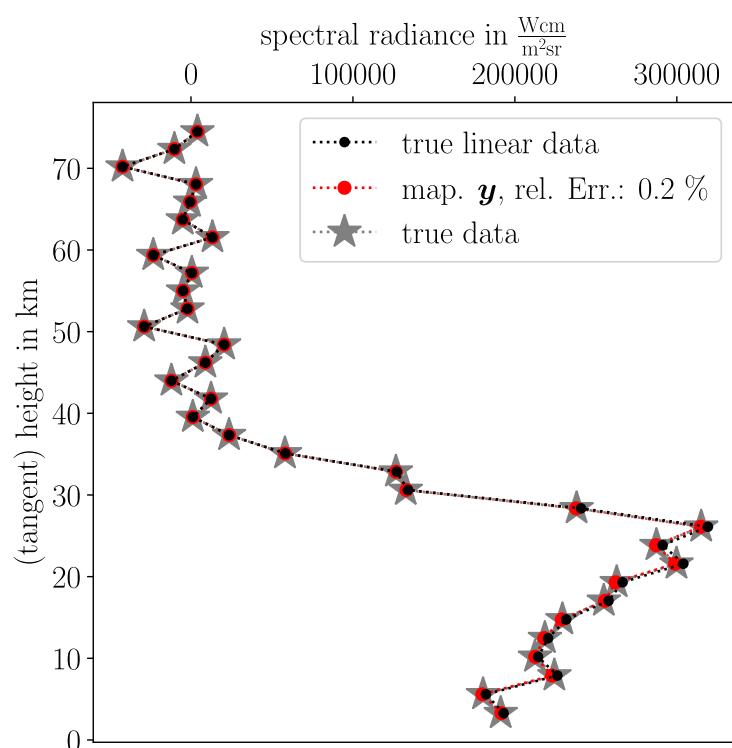
- linear solve
- machine learning class optimizer which package

### 3.3 affine RTO

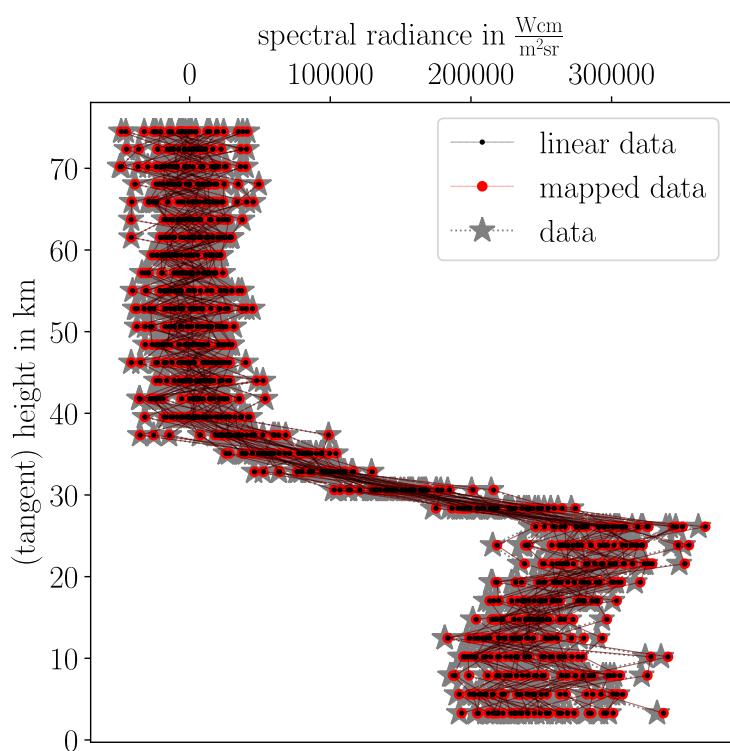
- does it sample from the correct distribution



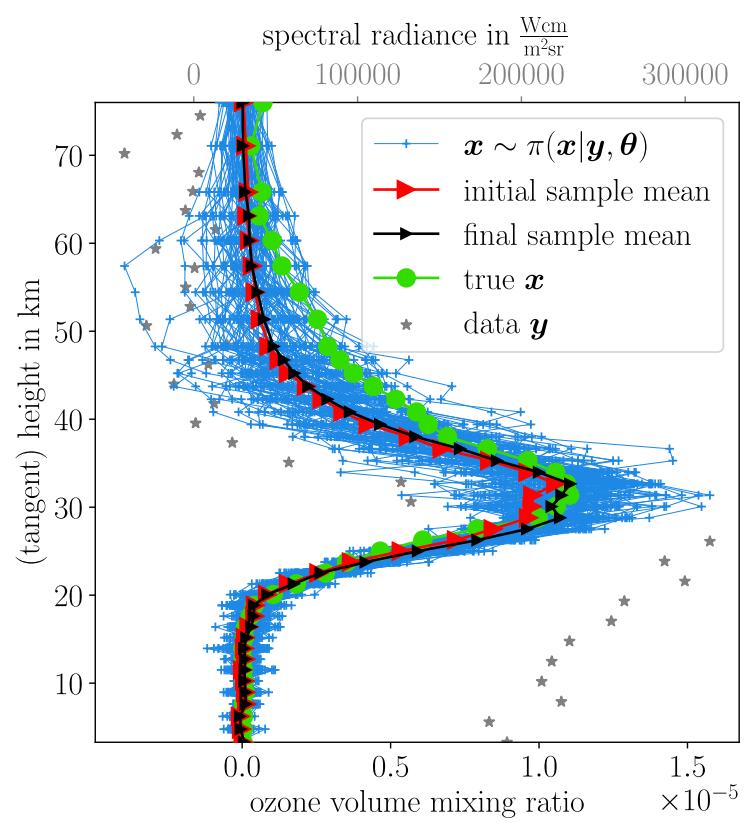
**Figure 3.1:** Non Linear carpet jumping scheme



**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 Hierarchical 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 probability distribution of  $(\mathbf{x}^T, \mathbf{y}^T)^T$ , where  $\mathbf{x} \in \mathcal{X}$  and  $\mathbf{y} \in \mathcal{Y}$  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:

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

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 expression for the posterior distribution

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

$$\pi(\mathbf{x}|\mathbf{y}) \propto \pi(\mathbf{y}|\mathbf{x})\pi(\mathbf{x}) \quad (\text{A.2})$$

Taking the logarithmic function of this formulation we can find an expression for

the the posterior covariance, with the  $\text{Var}(\mathbf{x}) = \mathbf{Q}_x^{-1}$  and  $\text{Var}(\mathbf{y}) = \mathbf{Q}_y^{-1}$ .

$$\ln \pi(\mathbf{x}|\mathbf{y}) \propto \ln \pi(\mathbf{y}|\mathbf{x}) + \ln \pi(\mathbf{x}) \quad (\text{A.3})$$

$$= -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{Q}_x(\mathbf{x} - \boldsymbol{\mu}) - \frac{1}{2}(\mathbf{y} - \mathbf{Ax})^T \mathbf{Q}_y(\mathbf{y} - \mathbf{Ax}) \quad (\text{A.4})$$

$$= -\frac{1}{2} \left[ \mathbf{x}^T [\mathbf{Q}_x + \mathbf{A}^T \mathbf{Q}_y \mathbf{A}] \mathbf{x} + \mathbf{x}^T [-\mathbf{A}^T \mathbf{Q}_y] \mathbf{y} \right. \quad (\text{A.5})$$

$$\left. + \mathbf{y}^T [-\mathbf{Q}_y \mathbf{A}] \mathbf{x} + \mathbf{y}^T [\mathbf{Q}_y] \mathbf{y} - 2\mathbf{x}^T \mathbf{Q}_x \boldsymbol{\mu} \right] + \text{const.} \quad (\text{A.6})$$

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

$$-\frac{1}{2} \left[ \mathbf{x}^T [\mathbf{Q}_x + \mathbf{F}^T \mathbf{Q}_y \mathbf{F}] + \mathbf{y}^T [-\mathbf{Q}_y \mathbf{F}] \quad \mathbf{y}^T [\mathbf{Q}_y] + \mathbf{x}^T [-\mathbf{F}^T \mathbf{Q}_y] \right] \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \quad (\text{A.7})$$

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

We denote the precision matrix of the joint field as:

$$\mathbf{Q}_{xy} = \begin{bmatrix} \mathbf{Q}_{aa} & \mathbf{Q}_{ab} \\ \mathbf{Q}_{ba} & \mathbf{Q}_{bb} \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_x + \mathbf{F}^T \mathbf{Q}_y \mathbf{F} & -\mathbf{F}^T \mathbf{Q}_y \\ -\mathbf{Q}_y \mathbf{F} & \mathbf{Q}_y \end{bmatrix} \quad (\text{A.9})$$

The mean is defined through the linear term.

$$\frac{-2\mathbf{x}^T \mathbf{Q}_x \boldsymbol{\mu}}{-2} = \begin{bmatrix} \mathbf{x}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{Q}_x \boldsymbol{\mu} \\ 0 \end{bmatrix} \quad (\text{A.10})$$

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

$$\Rightarrow \boldsymbol{\mu}_{xy} = \mathbf{Q}_{xy}^{-1} \begin{bmatrix} \mathbf{Q}_x \boldsymbol{\mu} \\ 0 \end{bmatrix} \quad (\text{A.11})$$

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

$$\boldsymbol{\mu}_{x|y} = \boldsymbol{\mu}_x + \mathbf{Q}_{ba}^{-1} \mathbf{Q}_{ab} (\mathbf{x} - \boldsymbol{\mu}_y) \quad (\text{A.12})$$

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

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

$$\mathbf{Q}_{x|y} = \mathbf{Q}_{aa} = \mathbf{Q}_x + \mathbf{F}^T \mathbf{Q}_y \mathbf{F}, \quad (\text{A.14})$$

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  $\mathbf{P}$ , we generate a Markov chain with a unique stationary distribution proportional to  $\pi(\mathbf{x}, \boldsymbol{\theta} | \mathbf{y})$ . Since the posterior is strictly positive  $\pi(\mathbf{x}, \boldsymbol{\theta} | \mathbf{y}) \geq 0$  on the finite state space  $\Omega(\mathcal{X}, \theta)$  the generated chain is irreducible. 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  $j = i$ , but if  $j \neq i$  it is not trivial. In case we accept  $\{\mathbf{x}, \boldsymbol{\theta}\}^{(n+1)} = j$  as the new state we have  $\pi(j|\mathbf{y})g(i|j) > \pi(i|\mathbf{y})g(j|i)$ . This gives us  $\alpha(j|i) = 1$  and  $\alpha(i|j) = \frac{\pi_i g(j|i)}{\pi_j g(i|j)}$  and satisfies the detailed balance:

$$\frac{\pi_i}{\pi_j} \frac{\pi_i g(j|i)}{\pi_j g(i|j)} g(j|i) = \pi_i g(j|i) \quad .$$

If  $\pi(j|\mathbf{y})g(i|j) < \pi(i|\mathbf{y})g(j|i)$  then  $\alpha(i|j) = 1$  and  $\alpha(j|i) = \frac{\pi_j g(i|j)}{\pi_i g(j|i)}$ , 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

$$\pi(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}) \propto \pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta}) \quad (\text{C.1})$$

$$\propto \exp \left[ (\mathbf{F}\mathbf{x} - \mathbf{y})^T \boldsymbol{\Sigma}^{-1} (\mathbf{F}\mathbf{x} - \mathbf{y}) + (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{Q}(\mathbf{x} - \boldsymbol{\mu}) \right] \quad (\text{C.2})$$

$$= \exp \|\hat{\mathbf{F}}\mathbf{x} - \hat{\mathbf{y}}\|^2 \quad (\text{C.3})$$

where

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

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

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

We can solve this and rewrite to

$$\frac{\partial}{\partial \mathbf{x}} \left[ (\hat{\mathbf{F}}\mathbf{x} - (\hat{\mathbf{y}} + \boldsymbol{\eta}))^T (\hat{\mathbf{F}}\mathbf{x} - (\hat{\mathbf{y}} + \boldsymbol{\eta})) \right] = 0 \quad (\text{C.6})$$

$$\Leftrightarrow \mathbf{x}^T \hat{\mathbf{F}}^T \hat{\mathbf{F}} + \hat{\mathbf{F}}^T \hat{\mathbf{F}} \mathbf{x} - \hat{\mathbf{F}}^T (\hat{\mathbf{y}} + \boldsymbol{\eta}) - (\hat{\mathbf{y}} + \boldsymbol{\eta})^T \hat{\mathbf{F}} \mathbf{x} = 0 \quad (\text{C.7})$$

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

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

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

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

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

where  $\mathbf{v}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{F}^T \boldsymbol{\Sigma}^{-1} \mathbf{F})$  and  $\mathbf{v}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$  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  $\lambda = \lambda' - \Delta\lambda$

$$g(\lambda) = \ln \det \underbrace{(\mathbf{F}^T \mathbf{F} + \lambda \mathbf{L})}_B \quad (\text{E.1})$$

$$g(\lambda') - g(\lambda) = \ln \det(\mathbf{F}^T \mathbf{F} + \lambda' \mathbf{L}) - \ln \det(\mathbf{F}^T \mathbf{F} + \lambda \mathbf{L}) \quad (\text{E.2})$$

$$= \ln \det \left[ \frac{(\mathbf{F}^T \mathbf{F} + (\lambda + \Delta\lambda) \mathbf{L})}{(\mathbf{F}^T \mathbf{F} + \lambda \mathbf{L})} \right] \quad (\text{E.3})$$

$$= \ln \det \left[ 1 + \frac{\Delta\lambda \mathbf{L}}{\mathbf{B}} \right] \quad (\text{E.4})$$

$$= \sum_{r=1}^{\infty} \frac{(-1)^{r+1}}{r!} \text{tr}((\mathbf{B}^{-1} \mathbf{L})^r) (\Delta\lambda)^r \quad (\text{E.5})$$

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

$$g^{(r)}(\lambda) = (-1)^{r+1} \text{tr}((\mathbf{B}^{-1} \mathbf{L})^r) \quad (\text{E.6})$$

$$\approx (-1)^{r+1} \sum_{k=1}^p \mathbf{z}_k^T (\mathbf{B}^{-1} \mathbf{L})^r \mathbf{z}_k \quad (\text{E.7})$$

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



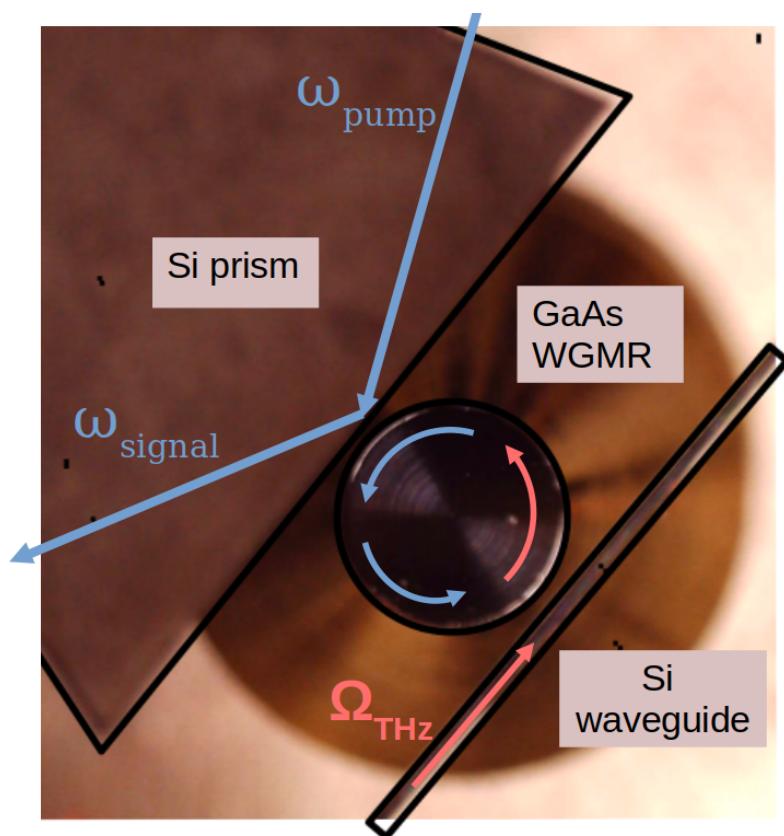
# F

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](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 Satopää 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.