# Acknowledgement of country

The Queensland University of Technology (QUT) acknowledges the Turrbal and Yugara, as the First Nations owners of the lands where QUT now stands. We pay respect to their Elders, lores, customs and creation spirits. We recognise that these lands have always been places of teaching, research and learning. QUT acknowledges the important role Aboriginal and Torres Strait Islander people play within the QUT community.

# Contributions to Bayesian Transdimensional Algorithms

Laurence Davies[13]

Supervisors: Professor Christopher Drovandi[13],
Dr Matthew Sutton[13], Dr Gentry White[13]

[1]School of Mathematical Sciences
Queensland University of Technology

[3]Center for Data Science
Queensland University of Technology

February 2023

# Table of Contents

# Part I

# Aims/Objectives

# Aims/Objectives

The aims of this thesis are as follows:

1. To identify, explore, and report on a new application transdimensional inference in AEM geophysics whereby Bayesian detection of IP is the primary motivation.
2. To extend the notoriously challenging method of reversible jump Markov chain Monte Carlo (RJMCMC) using state-of-the-art machine-learning methods.
3. To investigate the effects of geometric paths on transdimensional inference in importance sampling algorithms, with a view to propose alternative methods.

# Background: Bayesian Inference

Practitioners consider multiple models for a given data set. Inference over the space of parameters $\boldsymbol{\theta}_k \in \Theta_k$ *and* the space of models $k \in \mathcal{K}$ using Bayes' Theorem is

$$\pi(\boldsymbol{\theta}_k, k|\boldsymbol{y}) = \frac{\mathcal{L}(\boldsymbol{y}|\boldsymbol{\theta}_k, k)p(\boldsymbol{\theta}_k|k)p(k)}{\sum\limits_{k' \in \mathcal{K}} \mathcal{Z}(\boldsymbol{y}|k')p(k')}, \tag{1}$$

where the (often intractable) model marginal likelihood terms $\mathcal{Z}(\boldsymbol{y}|k)$ in the denominator are

$$\mathcal{Z}(\boldsymbol{y}|k) = \int \mathcal{L}(\boldsymbol{y}|\boldsymbol{\theta}_k, k)p(\boldsymbol{\theta}_k|k)d\boldsymbol{\theta}_k. \tag{2}$$

# Background: Inference Algorithms

When inference is intractable, practitioners use algorithms like

1. Markov chain Monte Carlo (MCMC),
2. Importance sampling (IS, including annealed IS and sequential Monte Carlo),
3. Variational inference (approximate inference)

For inference specifically on the space of models $k \in \mathcal{K}$, *Reversible jump MCMC* is a widely accepted *fully Bayesian* approach.

This thesis will apply the above inference algorithms in Project 1, and in Projects 2 and 3 it will focus on extensions to (1) and (2).

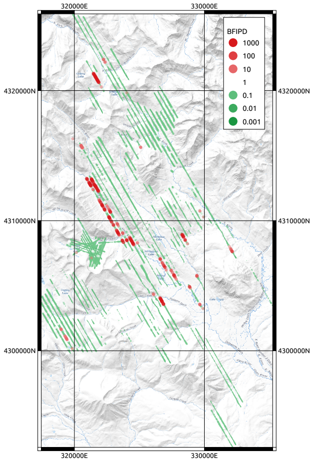
# Part II

# Project 1: Bayesian Detectability of Induced Polarisation in Airborne Electromagnetic Data

# Main contributions and case study

Contributions include

- A novel application of SMC methods and RJMCMC to a challenging problem in geophysics
- Formulated detection of IP as a Bayesian model selection problem (BFIPD).
- Demonstrated effectiveness of BFIPD in a synthetic study.
- Applied to real data in the Colorado East River.




Accepted for publication in Geophysical Journal International (2023) (First submitted 2021).

# Part III

# Project 2: Transport Reversible Jump Proposals

# Table of Contents

# Problem: Sampling a Transdimensional Space

The problem of interest is sampling probability distribution $\pi$ on

$$\mathcal{X} = \bigcup_{k \in \mathcal{K}} (\{k\} \times \Theta_k), \tag{3}$$

with *parameters* $\boldsymbol{\theta}_k \in \Theta_k \subseteq \mathbb{R}^{n_k}$ and *model index* (or indicator) $k \in \mathcal{K}$.
We want to make inference on the joint distribution (or conditional factorization)

$$\pi(k, \boldsymbol{\theta}_k) = \pi(k)\pi(\boldsymbol{\theta}_k | k).$$

# Reversible Jump Markov Chain Monte Carlo

Noting $\boldsymbol{x} = (k, \boldsymbol{\theta}_k)$, we want to propose from point $\boldsymbol{x}$ to point $\boldsymbol{x}'$, noting $\boldsymbol{\theta}_k$, $\boldsymbol{\theta}'_{k'}$ have dimensions $n_k$, $n_{k'}$ respectively.

- Require dimensions match: introduce auxiliary variables $\boldsymbol{u}_k \in \mathcal{U}_{k,k'} \subseteq \mathbb{R}^{w_k}$ and $\boldsymbol{u}_{k'} \in \mathcal{U}_{k',k} \subseteq \mathbb{R}^{w_{k'}}$ such that $n_k + w_k = n_{k'} + w_{k'}$.
- **Choose** a diffeomorphism e. $\boldsymbol{\theta}_{k'}, \boldsymbol{u}_{k'} = h_{k,k'}(\boldsymbol{\theta}_k, \boldsymbol{u}_k)$.

A (simplified) RJMCMC Algorithm when $n_{k'} > n_k$ is:

1. Propose model index $k' \sim j_k(\ \cdot\ )$
2. Propose auxiliary variables $\boldsymbol{u}_k \sim g_{k,k'}(\ \cdot\ )$
3. Accept with probability

$$\alpha(\boldsymbol{x}, \boldsymbol{x}') = 1 \wedge \frac{\pi(\boldsymbol{x}') j_{k'}(k) g_{k',k}(\boldsymbol{u}'_{k'})}{\pi(\boldsymbol{x}) j_k(k') g_{k,k'}(\boldsymbol{u}_k)} \left| J_{h_{k,k'}}(\boldsymbol{\theta}_k, \boldsymbol{u}_k) \right|. \qquad (4)$$

# Motivation: RJMCMC Proposal Performance

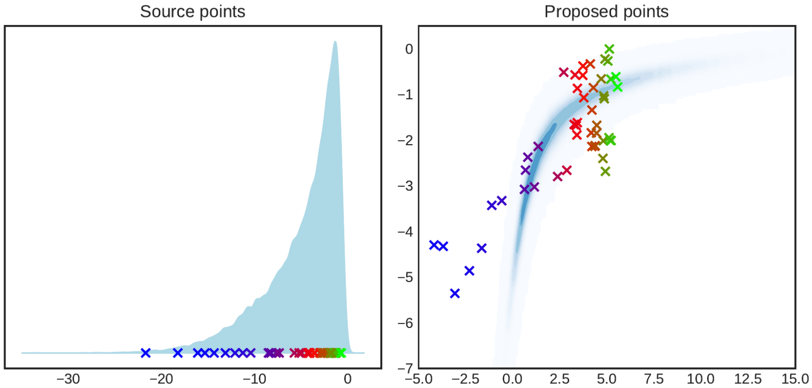



Figure: Points in a 1D model (*left*), proposed via RJMCMC to points in a 2D model (*right*).

# Transport Maps and Normalising Flows

## Transport Map (TM)

A function $T : \mathbb{R}^n \to \mathbb{R}^n$ is called a *transport map* from distribution $\mu_{\boldsymbol{\theta}}$ to distribution $\mu_{\boldsymbol{Z}}$ if $\mu_{\boldsymbol{Z}}$ is the *pushforward* of $\mu_{\boldsymbol{\theta}}$ using the measurable function $T$, i.e., $\mu_{\boldsymbol{Z}} = T\sharp\mu_{\boldsymbol{\theta}}$.

## Normalising Flows (NF) and Flow-Based Models

Let $\{T_{\boldsymbol{\psi}}\}$ be a family of bijective, differentiable transformations[1] whose inverse $T_{\psi}^{-1}$ has domain on the support of some arbitrary *base* distribution $\mu_{\boldsymbol{Z}}$. Then, for fixed parameters $\boldsymbol{\psi}$, the PDF of the random vector $\boldsymbol{Z} = T_{\boldsymbol{\psi}}(\boldsymbol{\theta})$ is

$$\mu_{\boldsymbol{\theta}}(\boldsymbol{\theta};\boldsymbol{\psi}) = \mu_{\boldsymbol{z}}(T_{\boldsymbol{\psi}}(\boldsymbol{\theta}))|J_{T_{\boldsymbol{\psi}}}(\boldsymbol{\theta})|,\ \boldsymbol{\theta} \in \mathbb{R}^n. \tag{5}$$

Distributions $\mu_{\boldsymbol{\theta}}$ are *flow-based models*, where $\{T_{\boldsymbol{\psi}}\}$ are the *normalising flows*.

With finite samples $\boldsymbol{s} \sim \pi$, we obtain an *approximate* TM $\hat{T}$ via density estimation, minimising the KLD from $\{\boldsymbol{s}\}$ to $\mu_{\boldsymbol{\theta}}$.

[1] *diffeomorphisms*

# Table of Contents

# Proposed Method: Transport Reversible Jump Proposals

Let $\nu$ be a univariate distribution.

# Proposed Method: Transport Reversible Jump Proposals

Let $\nu$ be a univariate distribution.
For each model $k \in \mathcal{K}$:

# Proposed Method: Transport Reversible Jump Proposals

Let $\nu$ be a univariate distribution.
For each model $k \in \mathcal{K}$:

- Define reference distributions $\otimes_{n_k}\nu$.

# Proposed Method: Transport Reversible Jump Proposals

Let $\nu$ be a univariate distribution.
For each model $k \in \mathcal{K}$:

- Define reference distributions $\otimes_{n_k} \nu$.
- Train flows on samples $\boldsymbol{s}_k \sim \pi(\boldsymbol{\theta}_k | k)$ to obtain $\hat{T}_k$.

# Proposed Method: Transport Reversible Jump Proposals

Let $\nu$ be a univariate distribution.
For each model $k \in \mathcal{K}$:

- Define reference distributions $\otimes_{n_k} \nu$.
- Train flows on samples $\boldsymbol{s}_k \sim \pi(\boldsymbol{\theta}_k | k)$ to obtain $\hat{T}_k$.

Then, a transdimensional proposal where $n_{k'} > n_k$ is

$$
\begin{aligned}
\boldsymbol{z}_k &\leftarrow T_k(\boldsymbol{\theta}_k), \\
\boldsymbol{z}'_{k'} &\leftarrow \bar{h}_{k,k'}(\boldsymbol{z}_k, \boldsymbol{u}_k), \\
\boldsymbol{\theta}_{k'} &\leftarrow T_{k'}^{-1}(\boldsymbol{z}'_{k'}),
\end{aligned} \tag{6}
$$

# Proposed Method: Transport Reversible Jump Proposals

Let $\nu$ be a univariate distribution.
For each model $k \in \mathcal{K}$:

- Define reference distributions $\otimes_{n_k} \nu$.
- Train flows on samples $\boldsymbol{s}_k \sim \pi(\boldsymbol{\theta}_k | k)$ to obtain $\hat{T}_k$.

Then, a transdimensional proposal where $n_{k'} > n_k$ is

$$
\begin{aligned}
\boldsymbol{z}_k &\leftarrow T_k(\boldsymbol{\theta}_k), \\
\boldsymbol{z}'_{k'} &\leftarrow \bar{h}_{k,k'}(\boldsymbol{z}_k, \boldsymbol{u}_k), \\
\boldsymbol{\theta}_{k'} &\leftarrow T_{k'}^{-1}(\boldsymbol{z}'_{k'}),
\end{aligned} \tag{6}
$$

where $\bar{h}_{k,k'}$ is a *volume-preserving* diffeomorphism on $\otimes_{n_k} \nu$.

# Illustration between 1D and 2D distributions

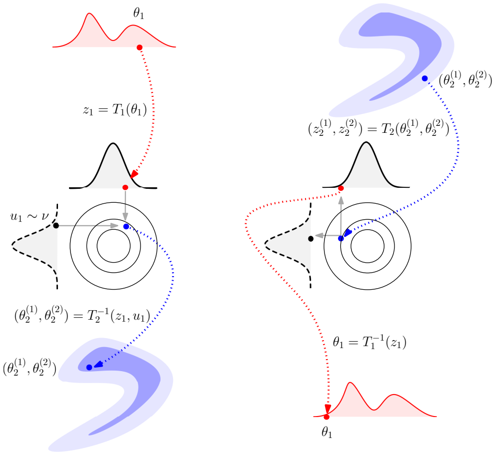



Figure: Illustration of the proposal class. Here, the reference $\nu$ is Gaussian. The diffeomorphisms $(\bar{h}_{k,k'})$ on the reference distributions simply concatenate or extract coordinates as required.

# Proposition: RJMCMC with Exact TMs

**Proposition 1**

Suppose that RJMCMC proposals are of the form described in (6), and for each $k \in \mathcal{K}$, satisfy $T_k \sharp \pi_k = \otimes_{n_k} \nu$. Then, (4) reduces to

$$\alpha(\boldsymbol{x}, \boldsymbol{x}') = 1 \wedge \frac{\pi(k')}{\pi(k)} \frac{j_{k'}(k)}{j_k(k')}. \tag{7}$$

Corollary

Provided the conditions of Proposition 1 are satisfied, choosing $\{j_k\}$ such that

$$\pi(k') j_{k'}(k) = \pi(k) j_k(k'), \quad \forall k, k' \in \mathcal{K}, \tag{8}$$

leads to a rejection-free proposal.

# Table of Contents

# Illustrative Sinh Arcsinh 1D 2D Example

Denote $S_{\boldsymbol{\epsilon},\boldsymbol{\delta}}.$ as the (element-wise) inverse sinh-arcsinh transformation of [Jones and Pewsey, 2009].
The PDF for $\boldsymbol{\theta} = T(\boldsymbol{Z})$ takes the form in (5). The target of interest for this example, where $\boldsymbol{\theta}_1 = (\theta_1^{(1)})$ and $\boldsymbol{\theta}_2 = (\theta_2^{(1)}, \theta_2^{(2)})$, is

$$\pi(k, \boldsymbol{\theta}_k) = \begin{cases} \frac{1}{4} p_{\epsilon_1, \delta_1, 1}\big(\boldsymbol{\theta}_1\big), & k = 1, \\ \frac{3}{4} p_{\boldsymbol{\epsilon}_2, \boldsymbol{\delta}_2, \mathrm{L}}\big(\boldsymbol{\theta}_2\big), & k = 2, \end{cases} \tag{9}$$

The "perfect" transform is

$$T(\boldsymbol{Z}) = S_{\boldsymbol{\epsilon},\boldsymbol{\delta}}(\mathrm{L}\boldsymbol{Z}), \text{ i.e. } T^{-1}(\cdot) = \mathrm{L}^{-1} S^{-1}_{\boldsymbol{\epsilon},\boldsymbol{\delta}}(\cdot), \tag{10}$$

for chosen reference distributions $\phi_n$, $n_k = k$, where $\boldsymbol{Z} \sim \mathcal{N}(\mathbf{0}_n, \mathrm{I}_{n\times n})$ and lower triangular $n \times n$ matrix $\mathrm{L}$.

# Example: Sinh Arcsinh Target with Transport RJ Proposal

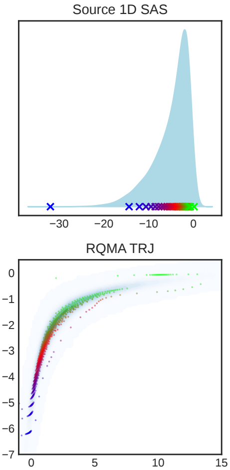



Figure: Systematic draws from conditional target $\pi(x_1|k=1)$ propose jumps to $\pi(\boldsymbol{x}_2|k=2)$.

# Example: Sinh Arcsinh Target with Transport RJ Proposal

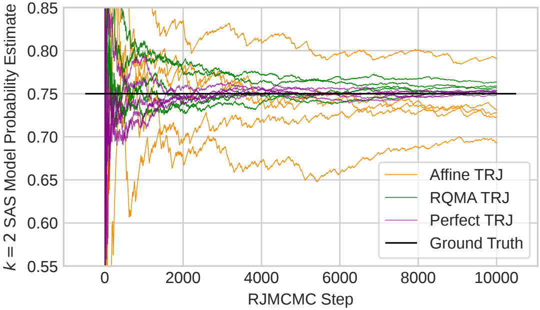



Figure: Running estimates of the model probabilities for the $k = 2$ component of the Sinh-Arcsinh target. Proposal are all TRJ with input TMs (1) Affine, (2) RQMA-NF, (3) Perfect. Ten chains on each proposal type are depicted, where alternating within-model proposals are a simple normal random walk.

# Table of Contents

# Bartolucci Bridge-Sampling Estimator

For an RJMCMC chain, [Bartolucci et al., 2006] showed that the Bayes factor $B_{k,k'}$ (ratio of marginal likelihoods) is estimated via

$$\hat{B}_{k,k'} = \frac{N_{k'}^{-1} \sum_{i=1}^{N_{k'}} \alpha_i'}{N_k^{-1} \sum_{i=1}^{N_k} \alpha_i}, \tag{11}$$

where $N_{k'}$ and $N_k$ are the number of proposed moves from model $k'$ to $k$, and from $k$ to $k'$, respectively in the run of the chain. When prior model probabilities are uniform, we obtain estimates of posterior model probabilities via

$$\hat{\pi}(k) = \hat{B}_{j,k}^{-1} \left(1 + \sum_{i \in \mathcal{K} \setminus \{j\}} \hat{B}_{i,j}\right)^{-1}, \text{ for arbitrary } j \in \mathcal{K}. \tag{12}$$

The **Bartolucci Bridge Estimator** (BBE) simply adopts the above for proposals from *samples* of the conditional targets.

# Table of Contents

# Bayesian Factor Analysis

We model monthly exchange rates of six currencies relative to the British pound, spanning January 1975 to December 1986 [West and Harrison, 1997, ], denoted as $\boldsymbol{y}_i \in \mathbb{R}^6$ for $i = 1, ..., 143$, of the random vector $\boldsymbol{Y}$.
We assume $\boldsymbol{Y} \sim \mathcal{N}(\mathbf{0}_6, \boldsymbol{\Sigma})$, where

- $\boldsymbol{\Sigma} = \boldsymbol{\beta}_k \boldsymbol{\beta}_k^\top + \boldsymbol{\Lambda}$,
- $\Lambda$ is a $6 \times 6$ positive diagonal matrix,
- $\boldsymbol{\beta}_k$ is a $6 \times k$ lower-triangular matrix with a positive diagonal,
- $k$ is the number of factors, $\boldsymbol{\theta}_k$ dimension $6(k+1) - k(k-1)/2$.

# Bayesian Factor Analysis: Model Configuration

Following [Lopes and West, 2004], for each $\boldsymbol{\beta}_k = [\beta_{ij}]$ with $i = 1, \ldots, 6$, $j = 1, \ldots, k$, the priors are

$$\begin{aligned} \beta_{ij} &\sim \mathcal{N}(0,1), \quad i < j \\ \beta_{ii} &\sim \mathcal{N}_+(0,1), \\ \Lambda_{ii} &\sim \mathcal{IG}(1.1, 0.05), \end{aligned} \tag{13}$$

We are interested in the posterior probability of $\boldsymbol{\theta}_k = (\boldsymbol{\beta}_k, \boldsymbol{\Lambda})$ for $k = 2$ or $3$ factors, with $\boldsymbol{\theta}_k$ dimensions 17 and 21 respectively. Via Bayes' Theorem the posterior is

$$\pi(k, \boldsymbol{\theta}_k | \boldsymbol{y}) \propto p(k) p(\boldsymbol{\beta}_k | k) p(\boldsymbol{\Lambda}) \prod_{i=1}^{143} \phi_{\boldsymbol{\beta}\boldsymbol{\beta}^\top + \boldsymbol{\Lambda}}(\boldsymbol{y}_i), \tag{14}$$

where $\boldsymbol{y} = (\boldsymbol{y}_1, ..., \boldsymbol{y}_{143})$.

# Bayesian Factor Analysis: Proposal Design

## Original [Lopes and West, 2004] Independence Proposal

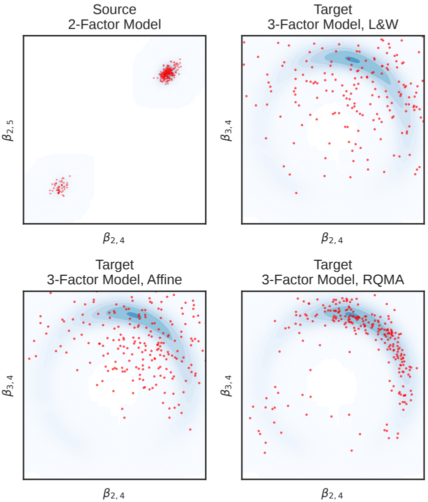
Write $\boldsymbol{\mu}_{\boldsymbol{\beta}_k}$, $\boldsymbol{B}_k$ as the posterior mean and covariance of $\boldsymbol{\beta}_k$. Denoting $\boldsymbol{\theta}_k = (\boldsymbol{\beta}_k, \boldsymbol{\Lambda})$, the independence proposal is

$$q_k(\boldsymbol{\theta}_k) = q_k(\boldsymbol{\beta}_k) \prod_{i=1}^{6} q_k(\Lambda_{ii}), \tag{15}$$

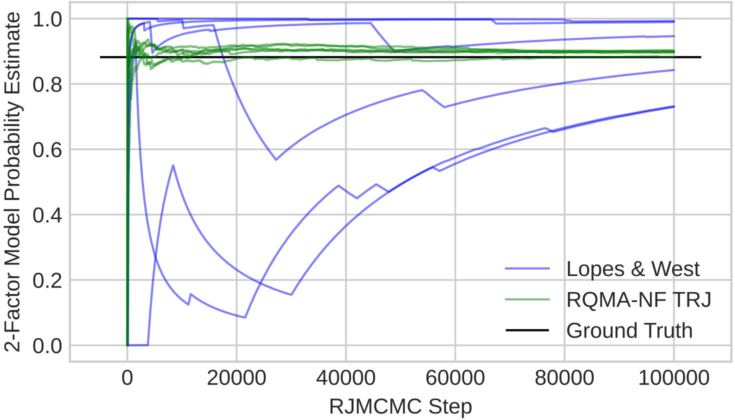
where for $k \in \mathcal{K}$, $q_k(\boldsymbol{\beta}_k) = \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}_k}, 2\boldsymbol{B}_k)$, and $q_k(\Lambda_{ii}) = \mathcal{IG}(18, 18v_{k,i}^2)$ where $v_{k,i}^2$ is the approximate conditional posterior mode of $\Lambda_{ii}$ given $k$.

We compare the [Lopes and West, 2004] proposal to Affine and RQMA-NF TRJ trained on finite draws $\boldsymbol{s} \sim \pi(\boldsymbol{\theta}_k|k)$ obtained via HMC-NUTS (for $k = 3$) and SMC (for $k = 2$).

# Bayesian Factor Analysis: Proposal Comparison



# Bayesian Factor Analysis: Running Estimates from RJMCMC Chain



# Bayesian Factor Analysis: BBE Study

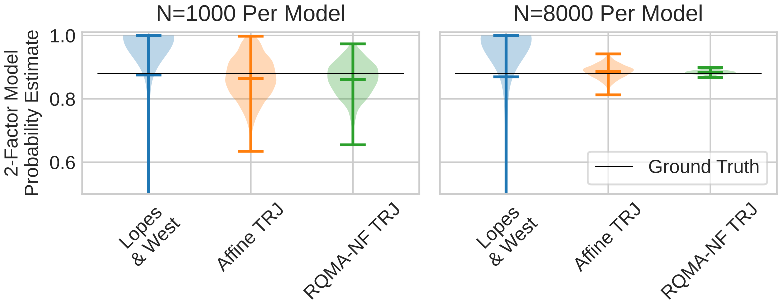



Figure: Violin plot showing the variability of the 2-factor model probability estimates in the case where only the 2-factor and 3-factor models are compared. Model probability estimates are obtained via the BBE. Ground truth is estimated via extended individual SMC runs ($N = 5 \cdot 10^4$).

# Table of Contents

# Example: Block Variable Selection in Robust Regression

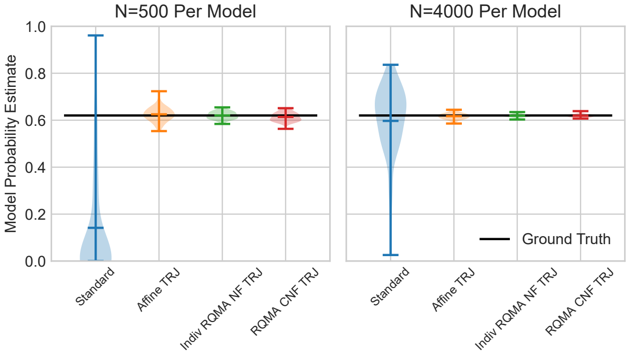



Figure: Violin plot showing the variability of the $k = (1, 1, 1, 1)$ model probability estimate for each proposal type using the BBE vs ground truth individual SMC ($N = 5 \cdot 10^4$). Individual SMC with $N = 500, 4000$ particles sampled conditional targets split into training/test samples for a total of 80 passes.

# Table of Contents

# Conclusions and Future Work

- We have introduced a theoretically rejection-free approach for RJMCMC proposal design utilizing transport maps.

# Conclusions and Future Work

- We have introduced a theoretically rejection-free approach for RJMCMC proposal design utilizing transport maps.
- Proposals using approximate transport maps yield good results when compared to baseline approaches.

# Conclusions and Future Work

- We have introduced a theoretically rejection-free approach for RJMCMC proposal design utilizing transport maps.
- Proposals using approximate transport maps yield good results when compared to baseline approaches.
- There is the caveat of requiring samples to train the approximate TMs (but pilot runs are also required in many other approaches).
- We have introduced the idea of using a conditional normalising flow to reduce training time. This would be useful for large model spaces!

# Conclusions and Future Work

- We have introduced a theoretically rejection-free approach for RJMCMC proposal design utilizing transport maps.
- Proposals using approximate transport maps yield good results when compared to baseline approaches.
- There is the caveat of requiring samples to train the approximate TMs (but pilot runs are also required in many other approaches).
- We have introduced the idea of using a conditional normalising flow to reduce training time. This would be useful for large model spaces!
- Efforts are justified in expensive-likelihood scenarios.

# Conclusions and Future Work

- We have introduced a theoretically rejection-free approach for RJMCMC proposal design utilizing transport maps.
- Proposals using approximate transport maps yield good results when compared to baseline approaches.
- There is the caveat of requiring samples to train the approximate TMs (but pilot runs are also required in many other approaches).
- We have introduced the idea of using a conditional normalising flow to reduce training time. This would be useful for large model spaces!
- Efforts are justified in expensive-likelihood scenarios.
- Finally, whilst the BBE benchmark was used to assess cross-model proposal quality, the results seem promising and justify further investigation in lieu of standard RJMCMC.

# Part IV

# Project 3: Weight-Stabilised Sequential Monte Carlo

# The problem with geometric paths

We consider a sequence of $T$ geometrically-annealed distributions

$$\pi_t(\boldsymbol{x}) = \mathcal{Z}_t^{-1}\pi(\boldsymbol{x})^{\gamma_t} g(\boldsymbol{x})^{1-\gamma_t}, \qquad \gamma_0 = 0 < \gamma_t < 1 = \gamma_T. \tag{16}$$

Using $\boldsymbol{x} = (k, \boldsymbol{\theta}_k)$, we set the target

$$\pi(k, \boldsymbol{\theta}_k) = \begin{cases} \frac{1}{4}\mathcal{N}(\theta_1; 0, 1), & k = 1, \\ \frac{3}{4}\mathcal{N}(\boldsymbol{\theta}_2; \mathbf{0}, \beta^2 \Sigma_{\pi,k=2}), & k = 2, \end{cases} \tag{17}$$

where $\Sigma_{\pi,k=2} = [[1, 0.999], [0.999, 1]]$, and importance sampling distribution

$$g(k, \boldsymbol{\theta}_k) = \begin{cases} \frac{1}{2}\mathcal{N}(\theta_1; 0, 4^2), & k = 1, \\ \frac{1}{2}\mathcal{N}(\boldsymbol{\theta}_2; \mathbf{0}, 4^2 \mathrm{I}), & k = 2. \end{cases} \tag{18}$$

What does it look like for different values of $\beta$ versus $\gamma_t$?

# Gaussian example: the problem with geometric paths

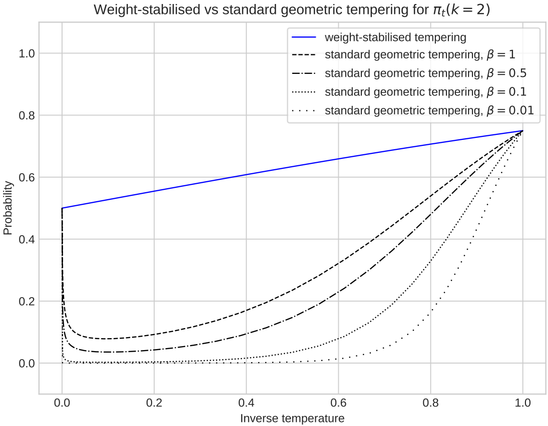



Figure: Comparison between Weight Stabilised and Standard Geometric Tempering, visualising the $\pi_t(k = 2)$ intermediate model probability.

# Weight-Stabilising the sequence: Approach 1 (Gaussian)

From [Tawn et al., 2020], an the normalising constant of an annealed Gaussian changes dependent on its covariance

$$\mathcal{N}(\boldsymbol{\theta}_k; \mu, \Sigma)^\gamma \propto |\Sigma|^{\frac{1-\gamma}{2}} \mathcal{N}(\boldsymbol{\theta}_k; \mu, \gamma^{-1}\Sigma). \tag{19}$$

We propose the $\mathcal{W}$-sequence of distributions with density functions

$$\pi_t^{\mathcal{W}}(k, \boldsymbol{\theta}_k) = (\mathcal{C}_t^{\mathcal{W}})^{-1} \mathcal{W}_{t,k}^{-1} \pi(k, \boldsymbol{\theta}_k)^{\gamma_t} g(k, \boldsymbol{\theta}_k)^{1-\gamma_t}, \tag{20}$$

where

$$\mathcal{C}_t^{\mathcal{W}} = \sum_{k \in \mathcal{K}} \pi(k)^{\gamma_t} g(k)^{1-\gamma_t}. \tag{21}$$

**Note:** Here, we use $\mathcal{C}_t$ for the normalising constant instead of $\mathcal{Z}_t$ because $\pi$ and $g$ are *both normalised*.

# Weight-Stabilising the sequence: Approach 1 (Gaussian)

With the $\mathcal{W}$-sequence of Gaussian targets

$$\pi_t^{\mathcal{W}}(k, \boldsymbol{\theta}_k) = (\mathcal{C}_t^{\mathcal{W}})^{-1} \mathcal{W}_{t,k}^{-1} \pi(k, \boldsymbol{\theta}_k)^{\gamma_t} g(k, \boldsymbol{\theta}_k)^{1-\gamma_t}, \tag{22}$$

the weights $\mathcal{W}^{-1}$ are known.

**Proposition 1: Gaussian-based model-stabilising weight**

Given the annealed sequence (22), for each $k \in \mathcal{K}$ and $0 < t < T$ the inverse model-stabilising weight is

$$\mathcal{W}_{t,k} = \left[ \frac{|\Sigma_{\pi,k}|^{1-\gamma_t} |\Sigma_{g,k}|^{\gamma_t}}{|\Sigma_{\gamma_t,k}|} \right]^{1/2} \exp\left( -\frac{\gamma_t(1-\gamma_t)}{2} \Delta\mu_k^\top \Sigma_{\gamma_t,k}^{-1} \Delta\mu_k \right), \tag{23}$$

where $\Sigma_{\gamma_t,k} = (1-\gamma_t)\Sigma_{\pi,k} + \gamma_t \Sigma_{g,k}$ and $\Delta\mu_k = \mu_{\pi,k} - \mu_{g,k}$. For cases $t = 0$ and $t = T$, we have $\mathcal{W}_{0,k} = \mathcal{W}_{T,k} = 1$.

# Weight-Stabilising Approach 2 (Unnormalised)

When we have unknown normalising constants, using $\eta$ for our unnormalised target, define the evidence-based model weight-stabilised sequence

$$\pi_t^{\mathcal{W}}(k, \boldsymbol{\theta}_k) = (\mathcal{Z}_t^{\mathcal{W}})^{-1} \mathcal{W}_{t,k}^{-1} \eta(k, \boldsymbol{\theta}_k)^{\gamma_t} g(k, \boldsymbol{\theta}_k)^{1-\gamma_t}, \tag{24}$$

where

$$\mathcal{Z}_t^{\mathcal{W}} = \sum_{k \in \mathcal{K}} \eta(k)^{\gamma_t} g(k)^{1-\gamma_t}. \tag{25}$$

## The exact model stabilising weight

Via rearrangement (see appendix) we have

$$\mathcal{W}_{t,k} = \mathcal{Z}_{t,k} (\mathcal{Z}_{T,k})^{-\gamma_t}. \tag{26}$$

# Weight-Stabilising Approach 2 (Unnormalised)

**Definition: An evidence-based approximate model-stabilising weight**

Given an SMC sampler on the sequence of distributions in (45), we define

$$\bar{\mathcal{W}}_{t,k} = \mathcal{Z}_{t,k}(\mathcal{Z}_{t-1,k})^{-\gamma_t}\mathcal{S}(\gamma_{t-1}, \gamma_p, \gamma_t) \tag{27}$$

where

$$\mathcal{S}(\gamma_{t-1}, \gamma_p, \gamma_t) = \left(\frac{\mathcal{Z}_{p,k}}{\mathcal{Z}_{t-1,k}}\right)^{\frac{-\gamma_t(1-\gamma_{t-1})}{\gamma_p-\gamma_{t-1}}}, \quad \gamma_{t-1} < \gamma_p \leq 1, \tag{28}$$

where $\mathcal{Z}_{p,k}$ is the normalising constant defined in (39) at inverse temperature $\gamma_p$ for an arbitrary step $t-1 < p < T$.

# Weight-Stabilising Approach 2 (Unnormalised)

**Proposition 2: Lower bound approximation to the MSW**

Given the above definition,

$$\mathcal{S}(\gamma_{t-1}, \gamma_p, \gamma_t) \geq \left( \frac{\mathcal{Z}_{T,k}}{\mathcal{Z}_{t-1,k}} \right)^{-\gamma_t}, \tag{29}$$

and thus

$$\bar{\mathcal{W}}_{t,k}^{-1} \leq \mathcal{W}_{t,k}^{-1}. \tag{30}$$

# Example 1: Robust Regression (8 models)

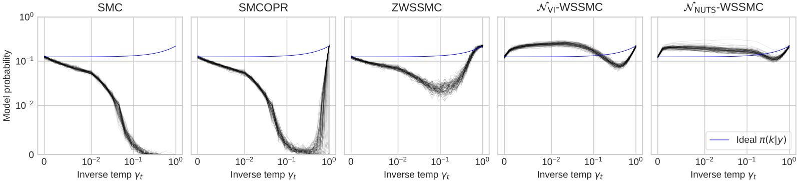



Figure: Comparison of the performance of weight-stabilising SMC samplers versus the standard SMC sampler and SMC with online particle recycling (SMCOPR) on the running model probability $\pi_t(k|\boldsymbol{y})$ vs inverse temperature $\gamma_t$ for model $k = (1, 0, 1, 1)$ in the robust variable selection example. The weight-stabilising SMC samplers in order are evidence-weight-stabilised SMC (ZWSSMC), WSSMC with final conditional target moments determined using variational inference ($\mathcal{N}_{\mathrm{VI}}$-WSSMC), and WSSMC with final conditional target moments determined using a no-U-turn HMC sampler ($\mathcal{N}_{\mathrm{NUTS}}$-WSSMC).

# Example 1: Robust Regression (8 models)

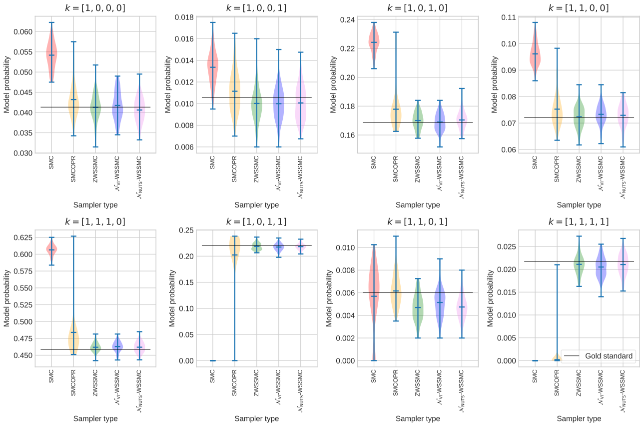



Figure: Model probabilities for each model in the robust variable selection example, over $100$ simulations for each sampler type. Gold standard probabilities were obtained using a standard SMC sampler with $N = 50,000$ particles.

# Example 2: Gaussian Mixture (Challenging, 19 models)

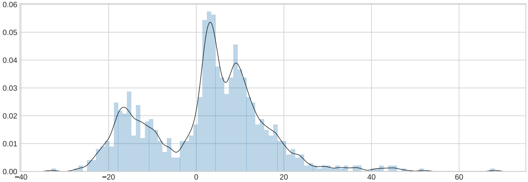



Figure: A histogram of the data used in the Gaussian mixture model example. This data was generated using a six component mixture.

$$\mathcal{L}(\boldsymbol{y}|\boldsymbol{p}_k, \boldsymbol{\mu}_k, \boldsymbol{\sigma}_k) = \sum_{c=1}^{k} p_c \mathcal{N}(\boldsymbol{y}; \mu_c, \sigma_c^2), \tag{31}$$

$$\boldsymbol{p}_k = \text{SymmetricDirichlet}(\alpha = 1), \qquad \text{weights, } k \tag{32}$$

$$\boldsymbol{\nu}_k = \text{SymmetricDirichlet}(\alpha = 1), \qquad \mu_{c+1} - \mu_c,\ k+1 \tag{33}$$

$$\boldsymbol{\sigma}_k = \text{HalfNormal}(\sigma = 10). \qquad \text{standard deviations, } k \tag{34}$$

# Example 2: Gaussian Mixture (Challenging, 19 models)

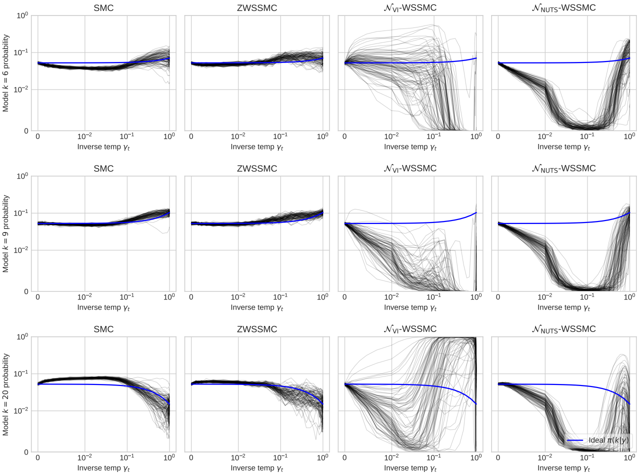



Figure: Model probability vs inverse temperature for models $k = 6, 9, 20$ in GMM example. Standard SMC is compared to evidence-weight-stabilised SMC (ZWSSMC), WSSMC using variational inference ($\mathcal{N}_{\mathrm{VI}}$-WSSMC), and WSSMC using NUTS ($\mathcal{N}_{\mathrm{NUTS}}$-WSSMC).

# Example 2: Gaussian Mixture (Challenging, 19 models)

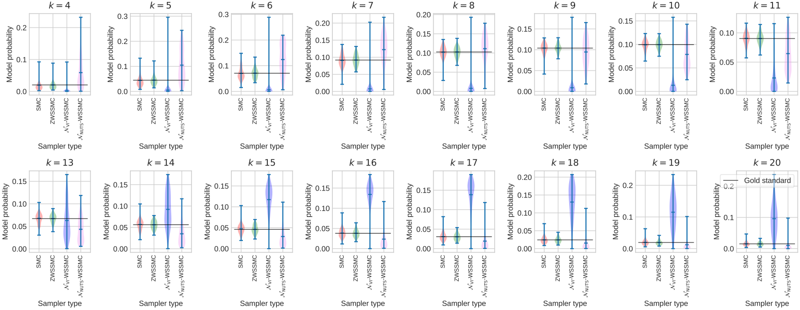



Figure: Model probabilities for each model in the GMM example, over $100$ simulations for each sampler type. Gold standard probabilities were obtained by averaging model probabilities obtained using ten standard SMC samplers with $N = 50,000$ particles.

# Weight-Stabilising SMC Conclusions

- Gaussian approach is very limited, but

# Weight-Stabilising SMC Conclusions

- Gaussian approach is very limited, but
- The evidence-derived model stabilising weights seems to be broadly applicable.

# Weight-Stabilising SMC Conclusions

- Gaussian approach is very limited, but
- The evidence-derived model stabilising weights seems to be broadly applicable.
- Future work to improve on the secant approximation to $\mathcal{Z}_{T,k}/\mathcal{Z}_{t-1,k}$.

# Weight-Stabilising SMC Conclusions

- Gaussian approach is very limited, but
- The evidence-derived model stabilising weights seems to be broadly applicable.
- Future work to improve on the secant approximation to $\mathcal{Z}_{T,k}/\mathcal{Z}_{t-1,k}$.
- Note that model are a natural partitioning - could we extend this approach to improve within-model inference? e.g. by formulating a deterministic $\mathcal{W}_t(\boldsymbol{\theta})$.

# Part V

# Thesis summary

# Summary

This thesis has contributed two separate and complementary approaches to improve the efficiency of transdimensional inference via sampling methods, preceded by a challenging and motivating application.

## Benefits to proposed contributions

1. The TRJ framework can be adapted to any continuous conditional targets at the required precision.
2. The evidence-derived model-stabilising weights are (so far) widely applicable in improving transdimensional SMC.

## Limitations

1. TRJ proposals require samples of the conditional targets, and flow-based approximate TMs can be expensive to compute.
2. The Gaussian-derived model-stabilising weights are of little benefit for difficult targets.

# Summary

## Future research avenues

1. Conditional TRJ proposals may have potential in amortized model inference.
2. Choosing a different partitioning for stabilising weights could generalise WSSMC.

# Questions

Thank you for participating!

# Part VI

# Appendix

# Evidence-derived model-stabilising weight

When we have unknown normalising constants, define

$$\eta(k, \boldsymbol{\theta}_k) = \mathcal{Z}\pi(k, \boldsymbol{\theta}_k) = \mathcal{Z}\pi(k)\pi(\boldsymbol{\theta}_k|k), \tag{35}$$

where

$$\mathcal{Z} = \sum_{k \in \mathcal{K}} \int_{\Theta_k} \eta(k, \boldsymbol{\theta}_k) d\boldsymbol{\theta}_k, = \sum_{k \in \mathcal{K}} \eta(k)\mathcal{Z}_k, \qquad \mathcal{Z}_k = \int_{\Theta_k} \eta(\boldsymbol{\theta}_k|k) d\boldsymbol{\theta}_k. \tag{36}$$

We note specifically that

$$\eta(\boldsymbol{\theta}_k|k) = \mathcal{Z}_k \pi(\boldsymbol{\theta}_k|k). \tag{37}$$

# Evidence-derived model-stabilising weight (continued)

The SMC geometrically-annealed sequence will be again defined using the initial importance sampling distribution $g$ to be

$$\begin{aligned}\eta_t(k,\boldsymbol{\theta}_k) &= \mathcal{Z}_t\pi_t(k,\boldsymbol{\theta}_k) = \mathcal{Z}_t\pi_t(k)\pi_t(\boldsymbol{\theta}_k|k) \\ &= \eta(k)^{\gamma_t}g(k)^{1-\gamma_t}\eta(\boldsymbol{\theta}_k|k)^{\gamma_t}g(\boldsymbol{\theta}_k|k)^{1-\gamma_t},\end{aligned} \tag{38}$$

for $t = 0,\ldots,T$, where

$$\mathcal{Z}_t = \sum_{k\in\mathcal{K}}\eta(k)^{\gamma_t}g(k)^{1-\gamma_t}\mathcal{Z}_{t,k}, \quad \mathcal{Z}_{t,k} = \int_{\Theta_k}\eta(\boldsymbol{\theta}_k|k)^{\gamma_t}g(\boldsymbol{\theta}_k|k)^{1-\gamma_t}d\boldsymbol{\theta}_k, \tag{39}$$

and noting $\mathcal{Z}_T = \mathcal{Z}$, $\mathcal{Z}_{T,k} = \mathcal{Z}_k$.

# Evidence-derived model-stabilising weight (continued 2)

Substituting (37) using the latter notation into (39), we have

$$\begin{aligned}
\mathcal{Z}_{t,k} &= \int_{\Theta_k} \eta(\boldsymbol{\theta}_k|k)^{\gamma_t} g(\boldsymbol{\theta}_k|k)^{1-\gamma_t} d\boldsymbol{\theta}_k, && (40)\\
&= \int_{\Theta} \big(\mathcal{Z}_{T,k}\pi(\boldsymbol{\theta}_k|k)\big)^{\gamma_t} g(\boldsymbol{\theta}_k|k)^{1-\gamma_t} d\boldsymbol{\theta}_k, && (41)\\
&= (\mathcal{Z}_{T,k})^{\gamma_t} \int_{\Theta} \pi(\boldsymbol{\theta}_k|k)^{\gamma_t} g(\boldsymbol{\theta}_k|k)^{1-\gamma_t} d\boldsymbol{\theta}_k, && (42)\\
&= (\mathcal{Z}_{T,k})^{\gamma_t} \mathcal{W}_{t,k}, && (43)
\end{aligned}$$

That is, we have

$$\mathcal{W}_{t,k} = \mathcal{Z}_{t,k}(\mathcal{Z}_{T,k})^{-\gamma_t}. \quad (44)$$

# Evidence-derived model-stabilising weight (continued 3)

Our model weight-stabilised sequence of targets is familiar, using $\mathcal{Z}$ instead of $\mathcal{C}$, i.e.

$$\pi_t^{\mathcal{W}}(k, \boldsymbol{\theta}_k) = (\mathcal{Z}_t^{\mathcal{W}})^{-1} \mathcal{W}_{t,k}^{-1} \eta(k, \boldsymbol{\theta}_k)^{\gamma_t} g(k, \boldsymbol{\theta}_k)^{1-\gamma_t}, \qquad (45)$$

where

$$\mathcal{Z}_t^{\mathcal{W}} = \sum_{k \in \mathcal{K}} \eta(k)^{\gamma_t} g(k)^{1-\gamma_t}. \qquad (46)$$

# References


Bartolucci, F., Scaccia, L., and Mira, A. (2006).
Efficient Bayes factor estimation from the reversible jump output.
*Biometrika*, 93(1):41–52.

Jones, M. C. and Pewsey, A. (2009).
Sinh-arcsinh distributions.
*Biometrika*, 96(4):761–780.

Lopes, H. F. and West, M. (2004).
Bayesian model assessment in factor analysis.
*Statistica Sinica*, 14(1):41–67.

Tawn, N. G., Roberts, G. O., and Rosenthal, J. S. (2020).
Weight-preserving simulated tempering.
*Statistics and Computing*, 30(1):27–41.

West, M. and Harrison, J. (1997).
*Bayesian forecasting and dynamic models*.
Springer series in statistics. Springer, New York, 2nd ed edition.