Scalable Inference for the Inverse Temperature of a Hidden Potts Model

Matt Moores

Department of Statistics
University of Warwick, UK

ACEMS International Workshop
on Monte Carlo Methods for Spatial Stochastic Systems
July 21, 2015

Acknowledgements

Queensland University of Technology

- Kerrie Mengersen
- Tony Pettitt
- Chris Drovandi
- Clair Alston

University of Warwick & Université Paris Dauphine

- Christian Robert

Funded by an APA (Australian Department of Education & Training), QUT Postgraduate Award, UK EPSRC platform grant (ref: EP/L014165/1) and the EPSRC Network on Computational Statistics & Machine Learning (ref: EP/K009788/2)

Outline

Background

Image analysis often involves:

- Large datasets, with millions of pixels
- Multiple images with similar characteristics

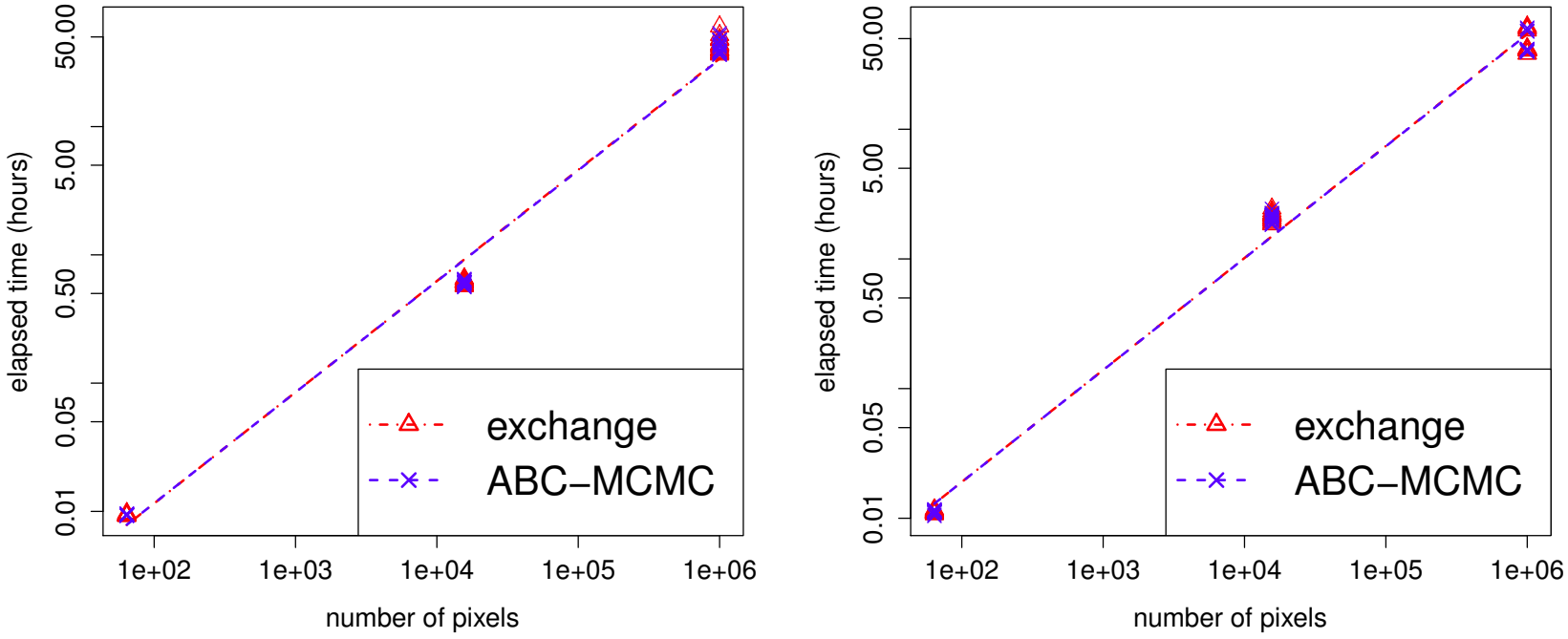
For example: satellite remote sensing (Landsat, MODIS), medical imaging (CT scans, MRI)

Table: Scale of common types of images

Number of pixels	Landsat ($90\text{m}^2/\text{px}$)	CT slices (512×512)
2^6	0.06km^2	...
5^6	14.06km^2	0.1
10^6	900.00km^2	3.8
15^6	10251.56km^2	43.5

Scalability

Computational cost is dominated by simulation of auxiliary variables (exchange algorithm) or pseudo-data (ABC)



(a) 2D images, $k = 3$ (b) 3D images, $k = 3$

Moores, Pettitt & Mengersen (2015) arXiv:1503.08066 [stat.CO]

hidden Markov random field

Joint distribution of observed pixel intensities $y_i \in \mathbf{y}$ and latent labels $z_i \in \mathbf{z}$:

$$\Pr(\mathbf{y}, \mathbf{z} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}^2, \beta) \;=\; \mathcal{L}(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \mathbf{z})\pi(\mathbf{z}|\beta) \tag{1}$$

Additive Gaussian noise:

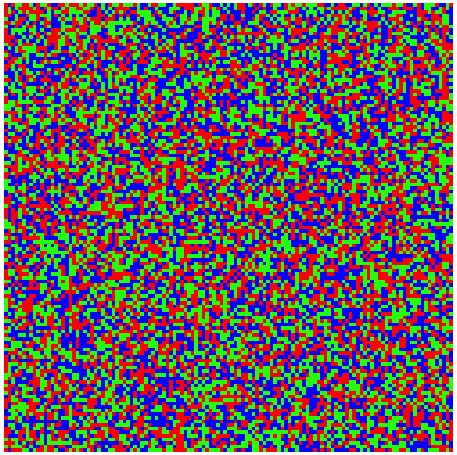
$$y_i \mid z_i = j \quad \overset{iid}{\sim} \quad \mathcal{N}\left(\mu_j, \sigma_j^2\right) \tag{2}$$

Potts model:

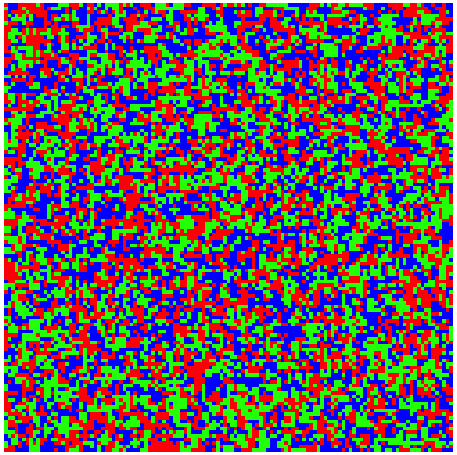
$$\pi(z_i \mid z_{\setminus i}, \beta) \;=\; \frac{\exp\left\{\beta \sum_{i\sim\ell} \delta(z_i, z_\ell)\right\}}{\sum_{j=1}^{k} \exp\left\{\beta \sum_{i\sim\ell} \delta(j, z_\ell)\right\}} \tag{3}$$

Potts (1952) *Proceedings of the Cambridge Philosophical Society* **48**(1)

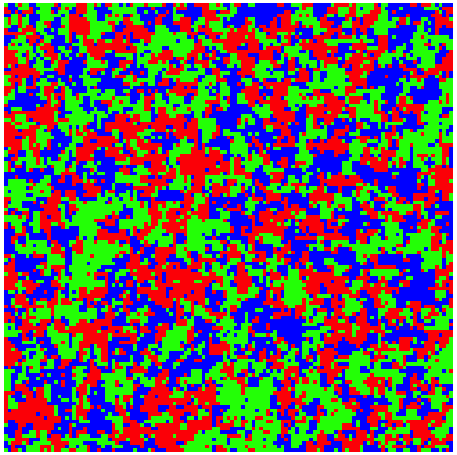
Inverse Temperature



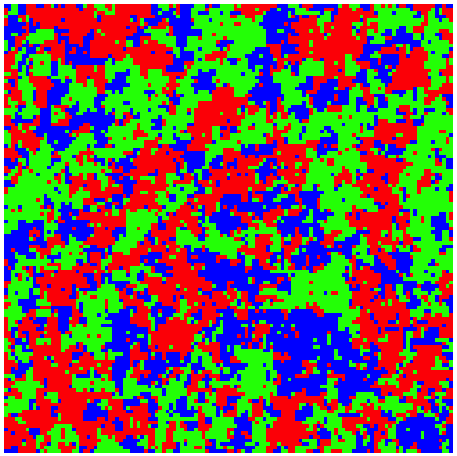
(c) $\beta = 0.1$



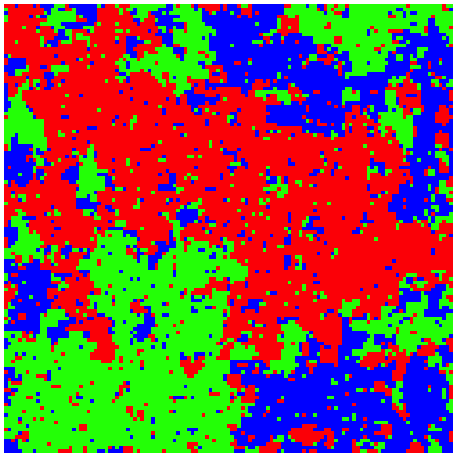
(d) $\beta = 0.5$



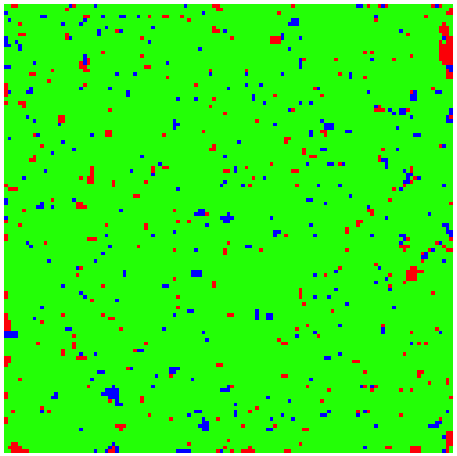
(e) $\beta = 0.85$



(f) $\beta = 0.95$



(g) $\beta = 1.005$



(h) $\beta = 1.15$

Doubly-intractable posterior

$$p(\beta|\mathbf{z}) = \frac{\mathcal{C}^{-1}(\beta)e^{\beta \mathrm{S}(\mathbf{z})}\pi(\beta)}{\int_\beta \mathcal{C}^{-1}(\beta)e^{\beta \mathrm{S}(\mathbf{z})}\pi(d\beta)} \tag{4}$$

The normalising constant has computational complexity $\mathcal{O}(nk^n)$, since it involves a sum over all possible combinations of the labels $\mathbf{z} \in \mathcal{Z}$:

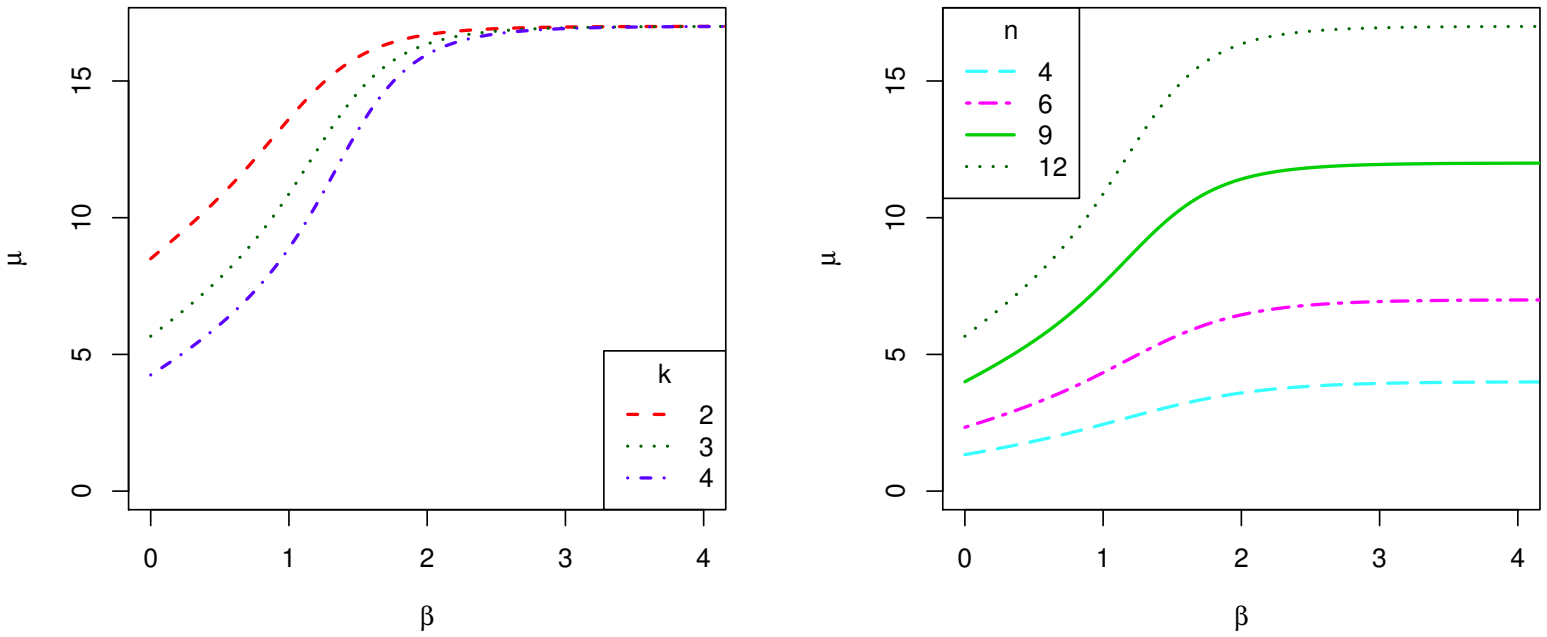
$$\mathcal{C}(\beta) = \sum_{\mathbf{z}\in\mathcal{Z}} e^{\beta \mathrm{S}(\mathbf{z})} \tag{5}$$

$\mathrm{S}(\mathbf{z})$ is the sufficient statistic of the Potts model:

$$\mathrm{S}(\mathbf{z}) = \sum_{i\sim\ell\in\mathcal{E}} \delta(z_i, z_\ell) \tag{6}$$

where $\mathcal{E}$ is the set of all unique neighbour pairs.

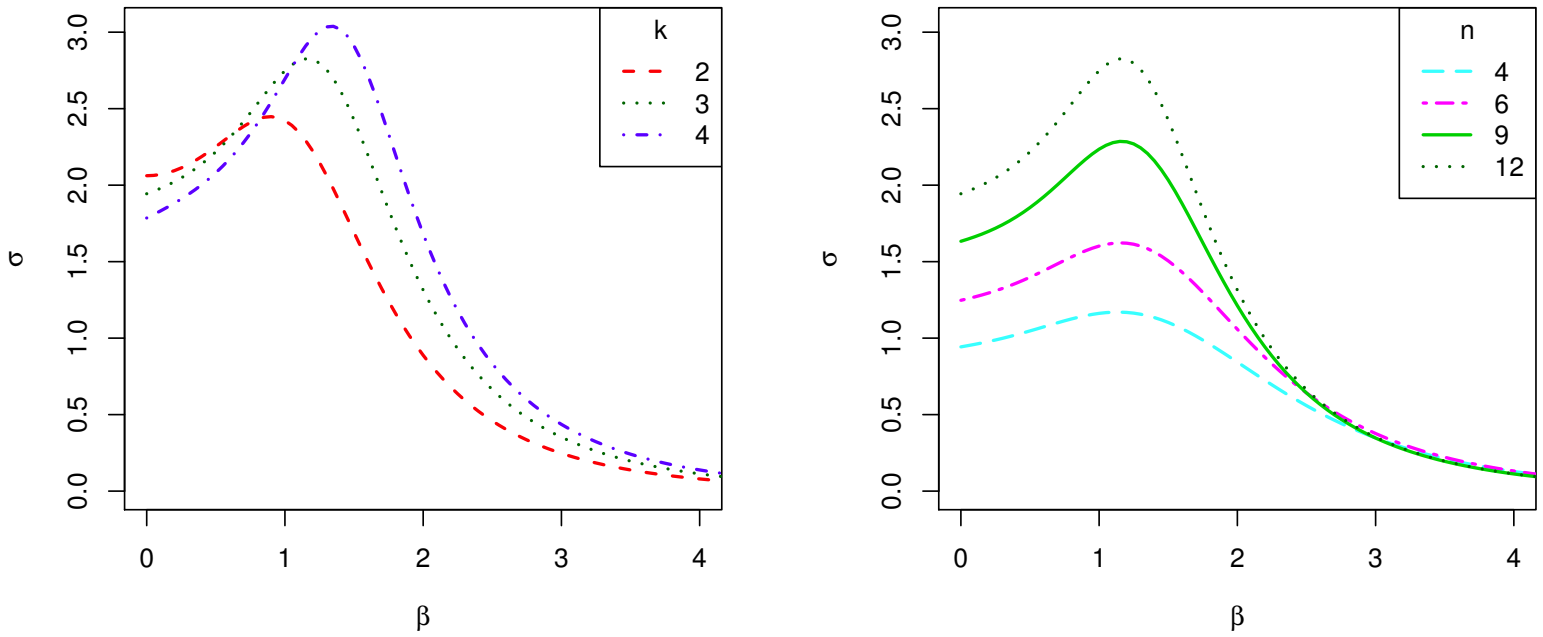
Expectation of $\mathrm{S}(\mathbf{z})$



(a) $n = 12$ & $k \in \{2, 3, 4\}$ (b) $k = 3$ & $n \in \{4, 6, 9, 12\}$

Figure: Distribution of $\mathbb{E}_{\mathbf{z}|\beta}[S(\mathbf{z})]$

Standard deviation of $\mathrm{S}(\mathbf{z})$



(a) $n = 12$ & $k \in \{2, 3, 4\}$ (b) $k = 3$ & $n \in \{4, 6, 9, 12\}$

Figure: Distribution of $\sigma_{\mathbf{z}|\beta}[S(\mathbf{z})]$

Distribution of $\mathrm{S}(\mathbf{z})$

In terms of the normalising constant:

$$\mathbb{E}_{\mathbf{z}|\beta}[\mathrm{S}(\mathbf{z})] = \frac{\mathrm{d}}{\mathrm{d}\beta} \log\{\mathcal{C}(\beta)\} \tag{7}$$

Score function:

$$\frac{\mathrm{d}}{\mathrm{d}\beta} \log\{p(\mathrm{S}(\mathbf{z}) \mid \beta)\} = \mathrm{S}(\mathbf{z}) - \mathbb{E}_{\mathbf{z}|\beta}[\mathrm{S}(\mathbf{z})] \tag{8}$$

Fisher information:

$$\mathcal{I}(\beta) = \mathbb{E}_{\mathbf{z}|\beta}\left[\left(\frac{\mathrm{d}}{\mathrm{d}\beta} \log\{p(\mathrm{S}(\mathbf{z}) \mid \beta)\}\right)^2\right] \tag{9}$$

$$= \mathbb{E}_{\mathbf{z}|\beta}\left[\frac{\mathrm{d}^2}{\mathrm{d}\beta^2} \log\{\mathcal{C}(\beta)\}\right] \tag{10}$$

Special cases

When $\beta = 0$ the labels are independent, hence:

$$\mathbb{E}_0 = \mathbb{E}_{\mathbf{z}|\beta=0}[\mathrm{S}(\mathbf{z})] = \tfrac{1}{k}|\mathcal{E}|$$
$$\mathbb{V}_0 = \mathbb{V}_{\mathbf{z}|\beta=0}[\mathrm{S}(\mathbf{z})] = |\mathcal{E}|\left(\tfrac{1}{k}\right)\left(1 - \tfrac{1}{k}\right)$$

where $|\mathcal{E}|$ is the total number of edges in the image lattice.

As $\beta \to \infty$,

$$\mathbb{E}_\infty = \lim_{\beta\to\infty} \mathbb{E}_{\mathbf{z}|\beta}[\mathrm{S}(\mathbf{z})] = |\mathcal{E}|$$
$$\mathbb{V}_\infty = \lim_{\beta\to\infty} \mathbb{V}_{\mathbf{z}|\beta}[\mathrm{S}(\mathbf{z})] = 0$$

For a 2D lattice with asymptotically large n:

$$\beta_{crit} = \log\{1 + \sqrt{k}\}$$
$$\mathbb{V}_{max} = \mathbb{V}_{\mathbf{z}|\beta=\beta_{crit}}[\mathrm{S}(\mathbf{z})] = \tfrac{2}{\pi}|\mathcal{E}|\log\{|\mathcal{E}|\}$$

Pickard (1987) *JASA* **82**(397)
Moores, Pettitt & Mengersen (2015) arXiv:1503.08066 [stat.CO]

Exchange Algorithm

Algorithm 1 Exchange Algorithm

for all iterations $t = 1, \ldots, T$ **do**
  Draw proposed parameter value $\beta' \sim q(\beta'|\beta_{t-1})$
  Generate $\mathbf{w}|\beta'$ by (perfect) sampling from Eq. (3)
  Calculate the Radon-Nikodym derivative:

$$\rho = \frac{q(\beta_{t-1}|\beta')\pi(\beta')\mathcal{C}(\beta_{t-1})e^{\beta'\mathrm{S}(\mathbf{z})}}{q(\beta'|\beta_{t-1})\pi(\beta_{t-1})\mathcal{C}(\beta')e^{\beta_{t-1}\mathrm{S}(\mathbf{z})}} \frac{\mathcal{C}(\beta')e^{\beta_{t-1}\mathrm{S}(\mathbf{w})}}{\mathcal{C}(\beta_{t-1})e^{\beta'\mathrm{S}(\mathbf{w})}}$$

  Draw $u \sim \mathrm{Uniform}[0, 1]$
  if $u < \min(1, \rho)$ **then**
    $\beta_t \leftarrow \beta'$ **else** $\beta_t \leftarrow \beta_{t-1}$
  end if
end for

Murray, Ghahramani & MacKay (2006) *Proc. 22nd Conf. UAI*

Approximate Bayesian Computation

Algorithm 2 ABC-MCMC

for all iterations $t = 1, \ldots, T$ **do**
  Draw proposed parameter value $\beta' \sim q(\beta'|\beta_{t-1})$
  Generate $\mathbf{w}|\beta'$ by sampling from Eq. (3)
  Draw $u \sim \mathrm{Uniform}[0, 1]$
  if $u < \min\left(1, \frac{\pi(\beta')q(\beta_{t-1}|\beta')}{\pi(\beta_{t-1})q(\beta'|\beta_{t-1})}\right)$ **and** $\|\mathrm{S}(\mathbf{w}) - \mathrm{S}(\mathbf{z})\| < \epsilon$ **then**
    $\beta_t \leftarrow \beta'$ **else** $\beta_t \leftarrow \beta_{t-1}$
  end if
end for

Marjoram, Molitor, Plagnol & Tavaré (2003) *PNAS* **100**(26)
Grelaud, Robert, Marin, Rodolphe & Taly (2009) *Bayesian Analysis* **4**(2)

Precomputation Step

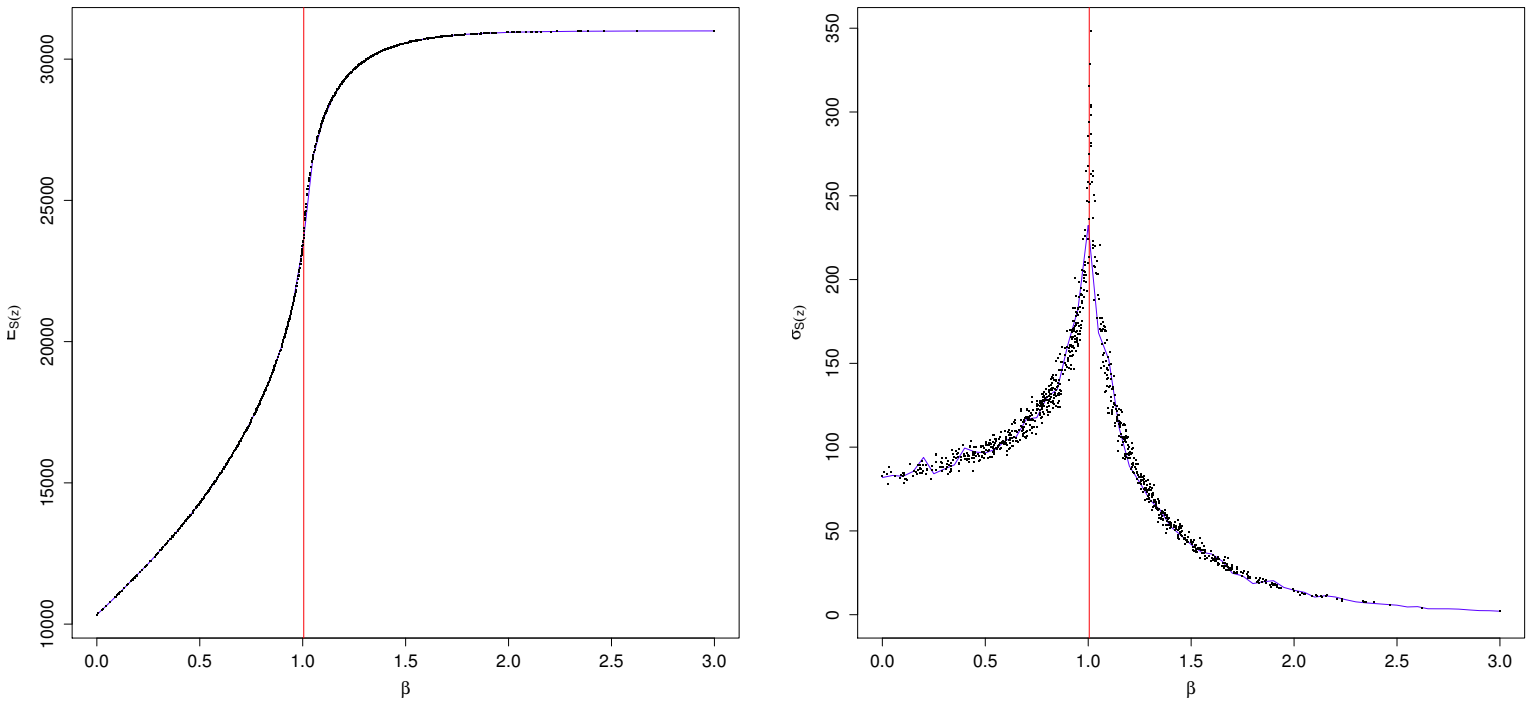
The distribution of the summary statistics $f(\mathrm{S}(\mathbf{w}) \mid \beta)$ is independent of the observed data $\mathbf{y}$ and the labels $\mathbf{z}$

- By simulating pseudo-data for values of β, we can create a binding function $\phi(\beta)$ for an auxiliary model $f_A(\mathrm{S}(\mathbf{w}) \mid \phi(\beta))$
- This binding function can be reused across multiple datasets, amortising its computational cost

By replacing $\mathrm{S}(\mathbf{w})$ with our auxiliary model, we avoid the need to simulate pseudo-data or auxiliary variables during model fitting.

Moores, Drovandi, Mengsersen & Robert (2015) *Statistics & Computing* **25**(1)
Drovandi, Pettitt & Lee (2015) *Statistical Science* **30**(1)

Piecewise linear model

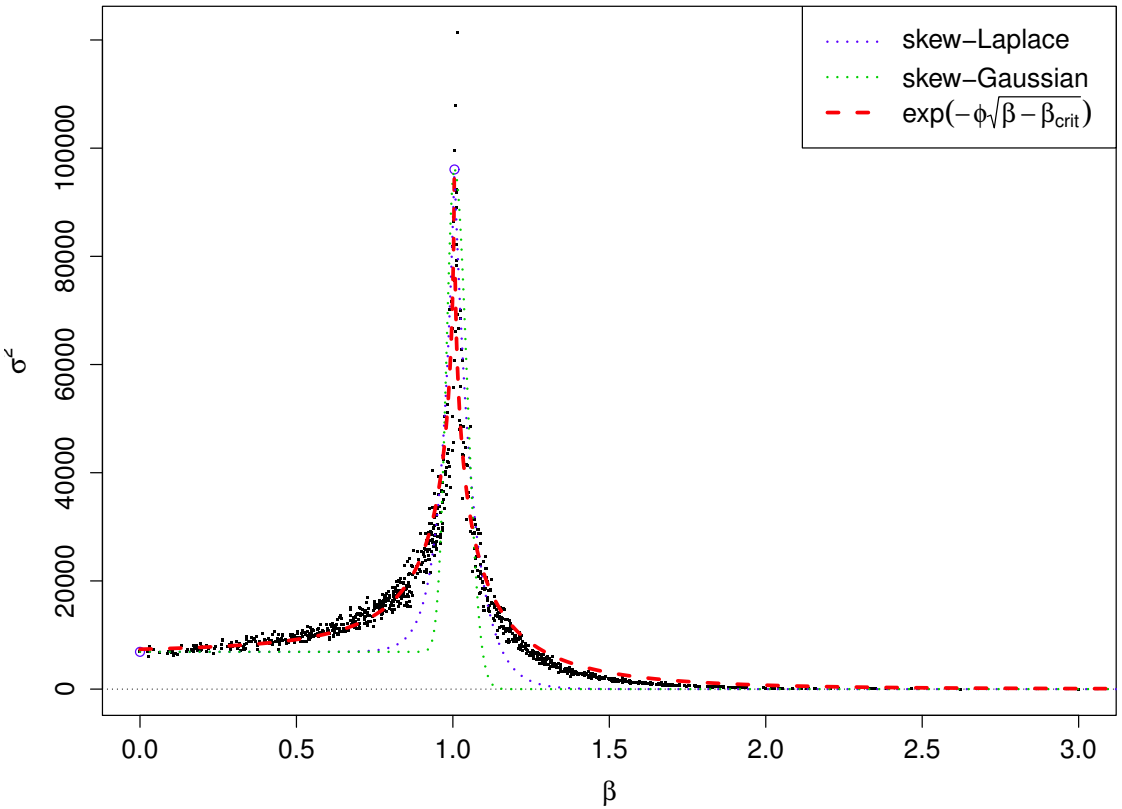


(a) $\hat{\phi}_\mu(\beta)$ (b) $\hat{\phi}_\sigma(\beta)$

Figure: Binding functions for $\mathrm{S}(\mathbf{w}) \mid \beta$ with $n = 5^6, k = 3$

Parametric auxiliary model for $\mathbb{V}_{\mathbf{z}|\beta}[\mathrm{S}(\mathbf{z})]$

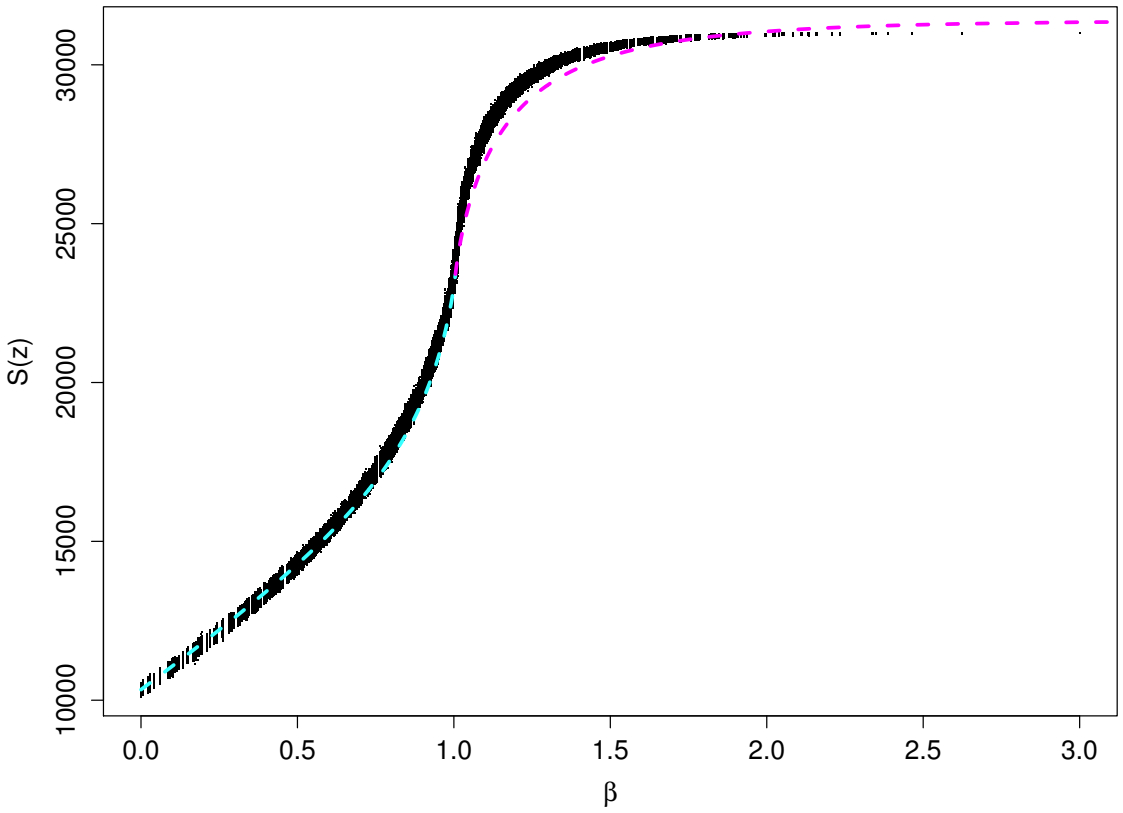
$$\hat{\phi}_{\sigma^2}(\beta) = \begin{cases} \mathbb{V}_0 + (\mathbb{V}_{max} - \mathbb{V}_0)e^{-\phi_1\sqrt{\beta_{crit}-\beta}} & : 0 \leq \beta < \beta_{crit} \\ \mathbb{V}_{max}e^{-\phi_2\sqrt{\beta-\beta_{crit}}} & : \beta \geq \beta_{crit} \end{cases} \tag{11}$$



Parametric auxiliary model for $\mathbb{E}_{\mathbf{z}|\beta}[\mathrm{S}(\mathbf{z})]$

The binding function for the expectation is available in closed form:

$$\hat{\phi}_\mu(\beta) = \begin{cases} \mathbb{E}_0 + \beta\mathbb{V}_0 + \int_0^\beta (\mathbb{V}_{max} - \mathbb{V}_0) e^{-\phi_1\sqrt{\beta_{crit}-\beta}} d\beta & : 0 \leq \beta < \beta_{crit} \\ \mathbb{E}_{\beta_{crit}} + \int_{\beta_{crit}}^\beta \mathbb{V}_{max} e^{-\phi_2\sqrt{\beta-\beta_{crit}}} d\beta & : \beta \geq \beta_{crit} \end{cases}$$



Algorithm using the auxiliary model $f_A(\mathrm{S}(\mathbf{w}) \mid \phi(\beta))$

Algorithm 3 Bayesian Indirect Inference

Generate $\mathbf{w}_s|\beta_s$ for sample points β_s, where $s = 1, \ldots, S$
Fit the binding functions $\hat{\phi}_{\sigma^2}(\beta)$ & $\hat{\phi}_{\mu}(\beta)$
for all iterations $t = 1, \ldots, T$ **do**
    Draw proposed parameter value $\beta' \sim q(\beta'|\beta_{t-1})$
    Approximate the Radon-Nikodym derivative:

$$\rho = \frac{q(\beta_{t-1}|\beta')\pi(\beta')\, f_A\left(\mathrm{S}(\mathbf{z}) \mid \hat{\phi}_{\mu}(\beta'), \hat{\phi}_{\sigma^2}(\beta')\right)}{q(\beta'|\beta_{t-1})\pi(\beta_{t-1})\, f_A\left(\mathrm{S}(\mathbf{z}) \mid \hat{\phi}_{\mu}(\beta_{t-1}), \hat{\phi}_{\sigma^2}(\beta_{t-1})\right)}$$

    Draw $u \sim \mathrm{Uniform}[0, 1]$
    if $u < \min(1, \rho)$ **then**
        $\beta_t \leftarrow \beta'$ **else** $\beta_t \leftarrow \beta_{t-1}$
    end if
end for

Simulation Study

20 images, $n = 125 \times 125$, $k = 3$:

- $\beta \sim \mathcal{U}(0, 1.307)$
- $\mathbf{z} \sim f(\cdot|\beta)$ using 2000 iterations of Swendsen-Wang
- $\mu_j \in \left\{\mathcal{N}\left(-0.15, 0.05^2\right), \mathcal{N}\left(0.05, 0.05^2\right), \mathcal{N}\left(0.25, 0.05^2\right)\right\}$
- $\frac{1}{\sigma_j^2} \sim \Gamma\left(\frac{3}{2}, \frac{0.015}{2}\right)$

Comparison of 3 algorithms:

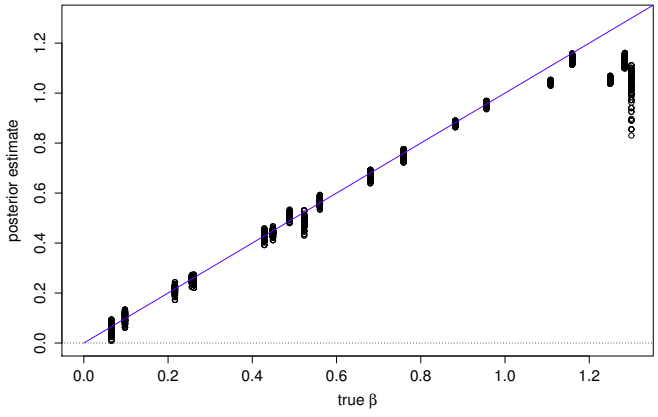
exchange approximate exchange algorithm (using 500 iters of Gibbs sampling)

ABC-MCMC with pseudo-data

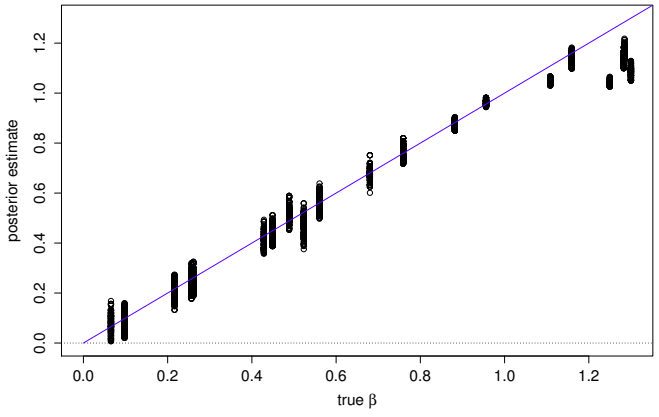
auxiliary indirect inference using $f_A(\mathrm{S}(\mathbf{w}) \mid \phi(\beta))$

Swendsen & Wang (1987) *Physical Review Letters* **58**
Feng & Tierney (2014) PottsUtils, *R package version 0.3-2*

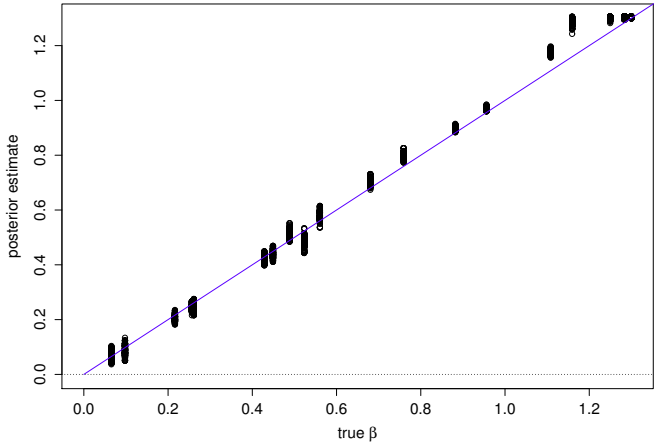
Posterior Samples



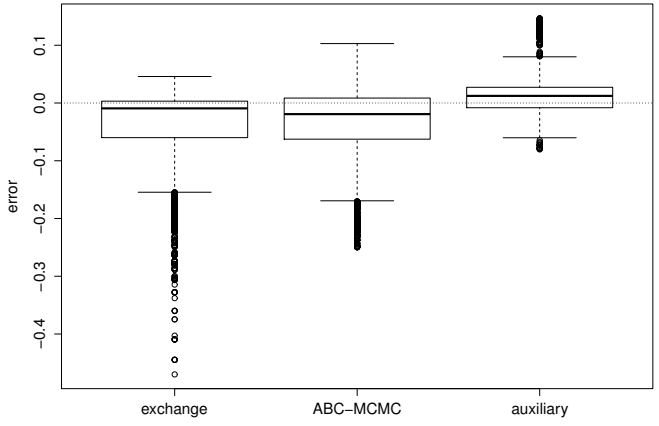
(a) exchange



(b) ABC-MCMC



(c) auxiliary

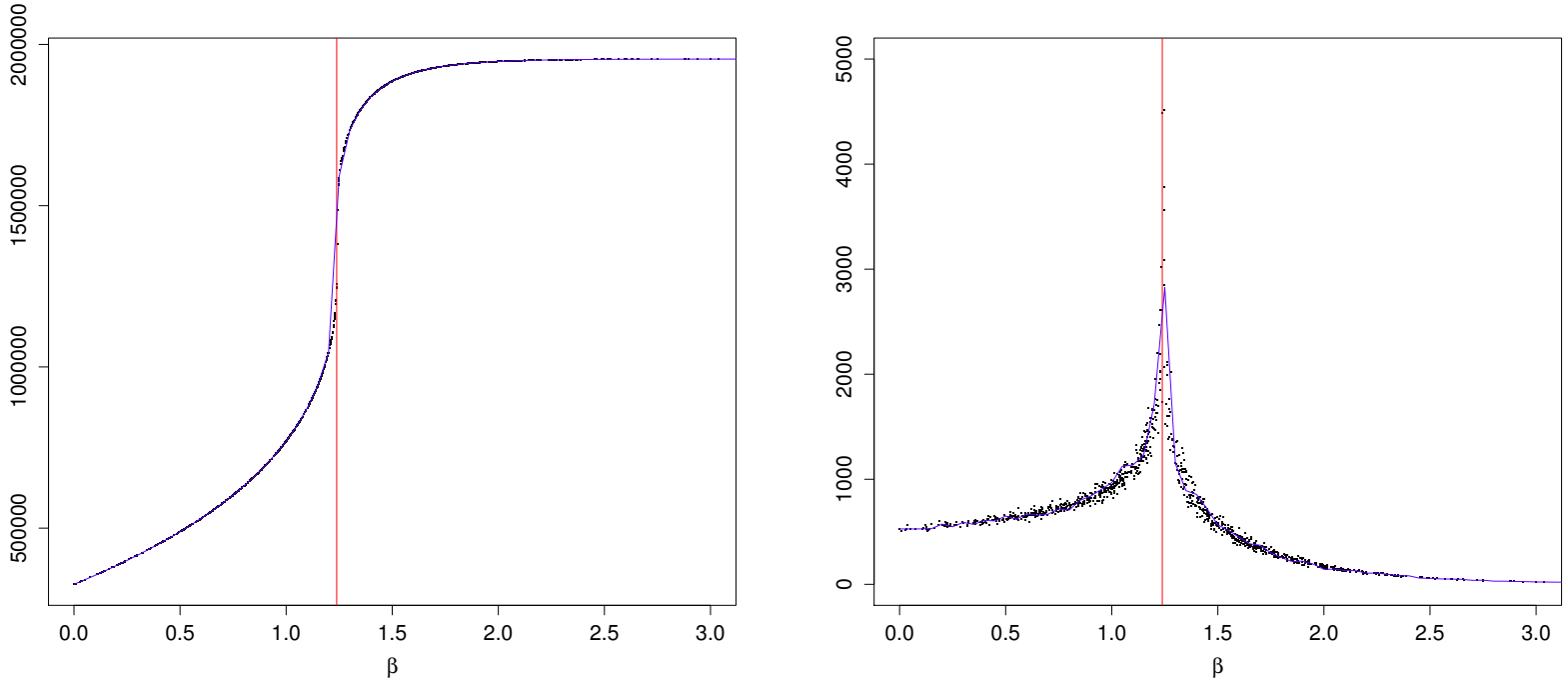


(d) error

Satellite image of southwest Brisbane

Figure: Normalised difference vegetation index (NDVI)

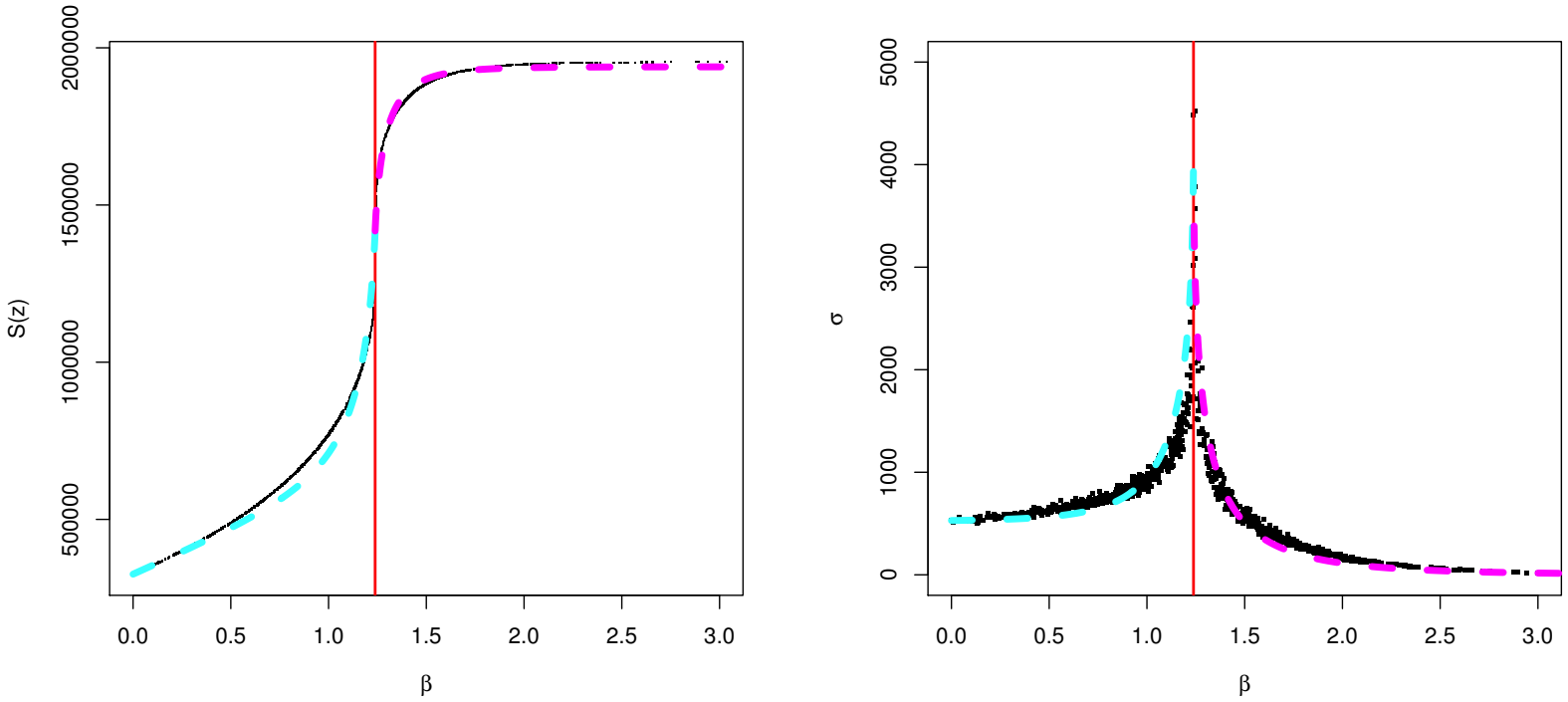
Piecewise linear model



(a) $\hat{\phi}_\mu(\beta)$ (b) $\hat{\phi}_\sigma(\beta)$

Figure: Linear interpolation for $n = 978380, k = 6$

Parametric auxiliary model



(a) $\hat{\phi}_\mu(\beta)$ (b) $\hat{\phi}_\sigma(\beta)$

Figure: Binding functions for $n = 978380, k = 6$

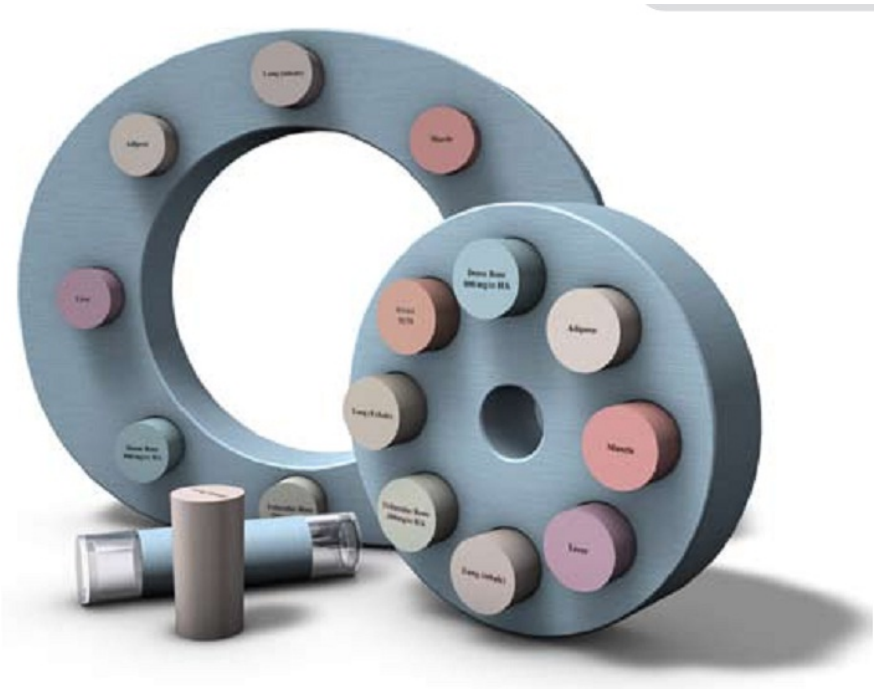
Results

Method	95% CI for β	Iterations	Elapsed	CPU time
Exchange alg.	$[1.275; 1.278]$	5K+5K	$50.8h$	$399.5h$
ABC-MCMC	$[1.316; 1.323]$	5K+5K	$50.6h$	$398.2h$
Piecewise linear	$[1.249; 1.255]$	5K+55K	$1.3h$	$10.0h$
Parametric aux.	$[1.259; 1.265]$	5K+55K	$1.8h$	$7.6h$

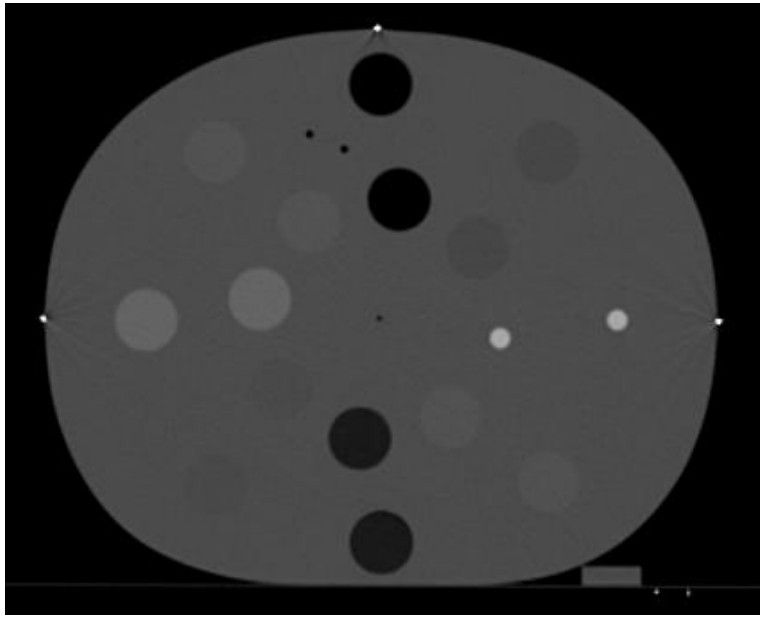
Table: Results for the satellite image of Brisbane, Australia.

Precomputation of $f_A(\mathrm{S}(\mathbf{w}) \mid \phi(\beta))$ took 13h 23m for 987 values of β.

Electron Density phantom

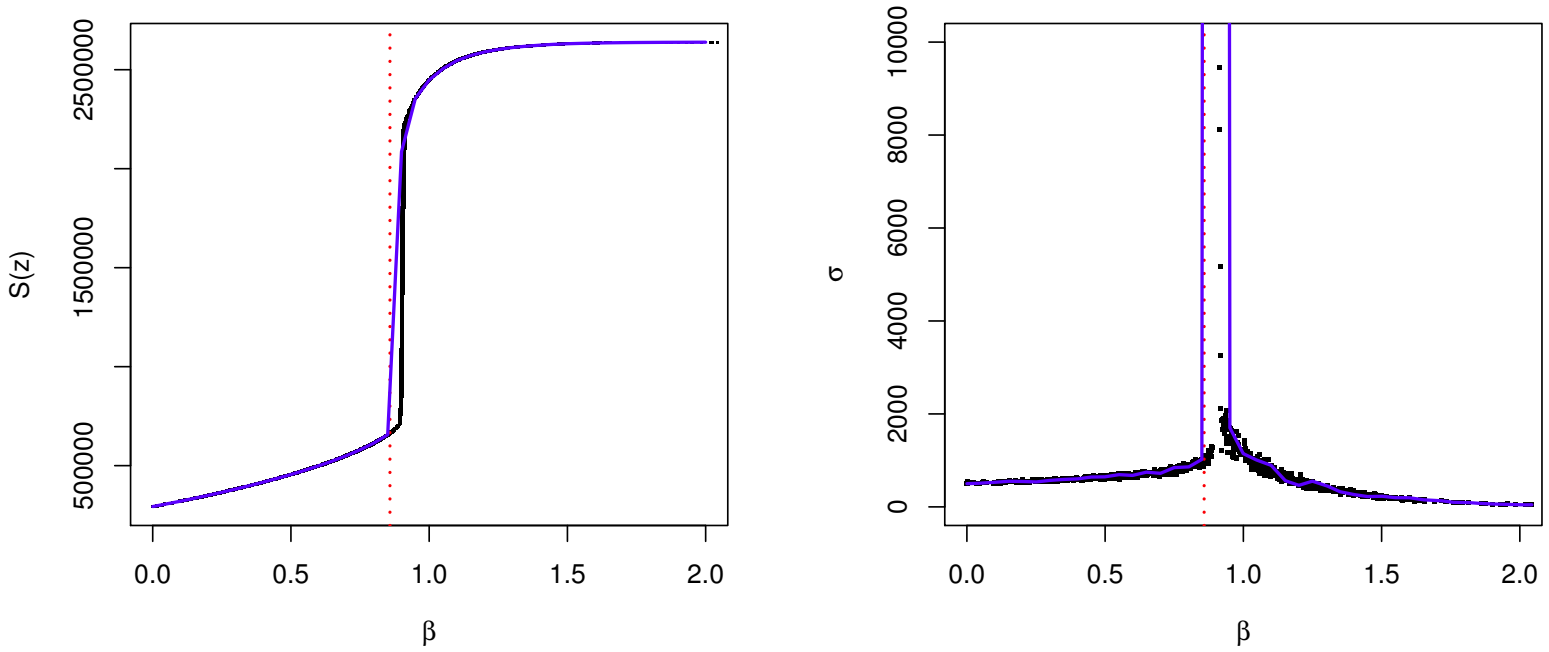


(a) CIRS Model 062 ED phantom



(b) CT scan

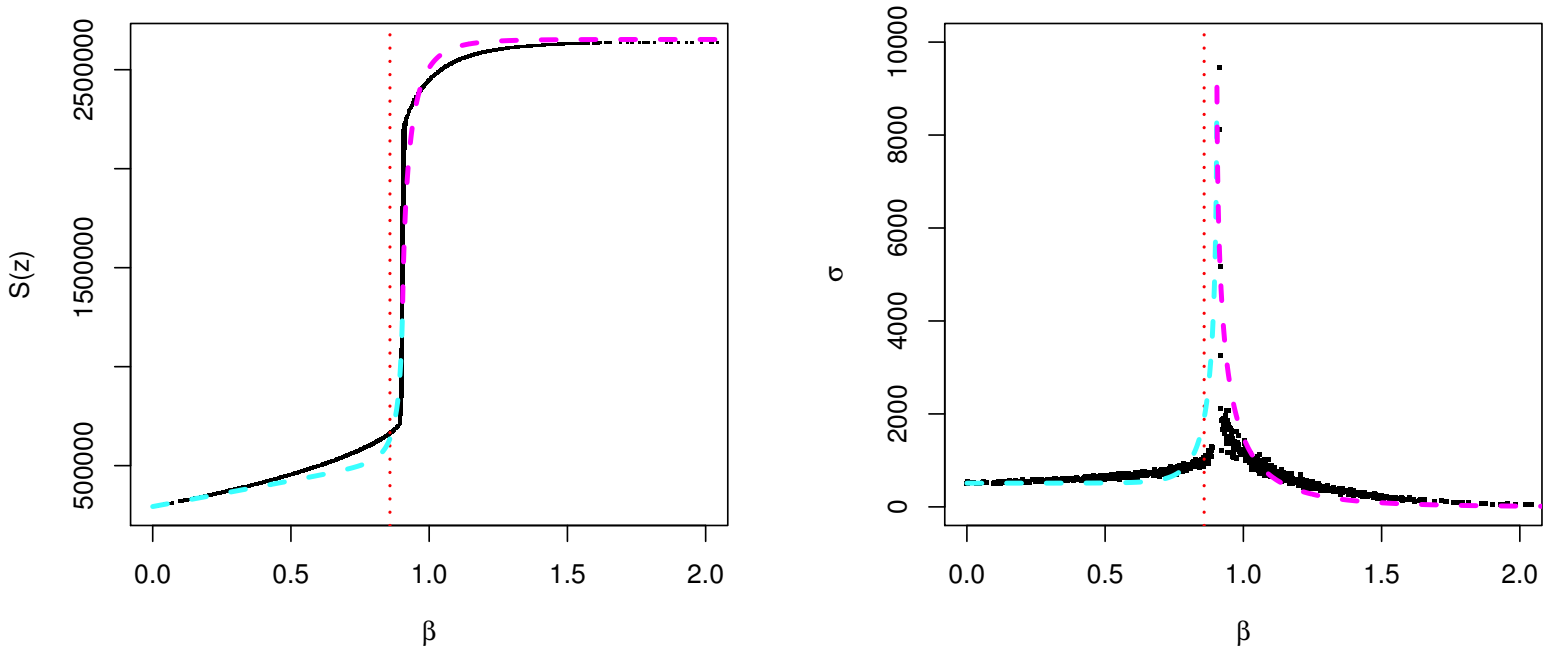
Piecewise linear model



(a) $\hat{\phi}_\mu(\beta)$ (b) $\hat{\phi}_\sigma(\beta)$

Figure: Linear interpolation for $n = 898366, k = 9$

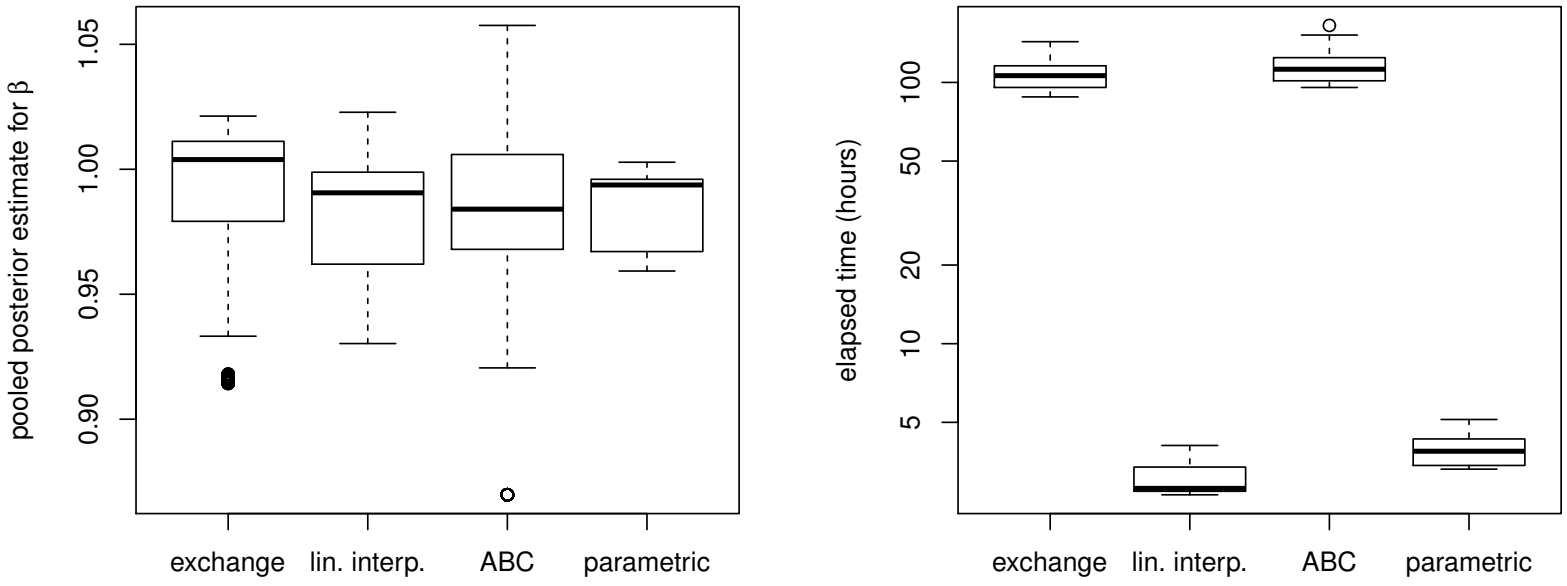
Parametric auxiliary model



(a) $\hat{\phi}_\mu(\beta)$ (b) $\hat{\phi}_\sigma(\beta)$

Figure: Binding functions for $n = 898366, k = 9$

Results



(a) Pooled posterior for β

(b) Elapsed runtime (hours)

Figure: Results for 28 CT scans of the ED phantom

Summary

Scalability of Bayesian computation for intractable likelihoods can be improved by pre-computing an auxiliary model $f_A(\mathrm{S}(\mathbf{w}) \mid \phi(\beta))$

- Pre-computation took 1.4 hours on a 16 core Xeon server for 987 values of β with 15,625 pixels (13.4 hours for 978,380 pixels)
- Average runtime for model fitting improved from 107 hours (exchange algorithm) or 115 hours (ABC-MCMC) to only 4 hours using the parametric auxiliary model

The tractable, parametric approximation for $\hat{\phi}_{\sigma^2}(\beta)$ could be used to design more efficient MCMC proposals, as well as to select the design points for β.

This method could be extended to multivariate applications, such as estimating both β and k for the hidden Potts model, or estimating θ for an exponential random graph model (ERGM).

For Further Reading I

M. Moores, A. N. Pettitt & K. Mengersen
Scalable Bayesian inference for the inverse temperature of a hidden Potts model.
arXiv:1503.08066 [stat.CO], 2015.

M. Moores, C. C. Drovandi, K. Mengersen & C. P. Robert
Pre-processing for approximate Bayesian computation in image analysis.
Statistics & Computing **25**(1): 23–33, 2015.

M. Moores & K. Mengersen
Bayesian approaches to spatial inference: modelling and computational challenges and solutions.
In *Proc.* 33rd *Int. Wkshp MaxEnt*, AIP Conf. Proc. **1636**: 112–117, 2014.

M. Moores & K. Mengersen
`bayesImageS`: Bayesian methods for image segmentation using a hidden Potts model.
R package version 0.3-1
https://researchdatafinder.qut.edu.au/display/n11957

For Further Reading II

C. C. Drovandi, A. N. Pettitt & A. Lee
Bayesian indirect inference using a parametric auxiliary model.
Statist. Sci. **30**(1): 72–95, 2015.

C. C. Drovandi, A. N. Pettitt & M. J. Faddy
Approximate Bayesian computation using indirect inference.
J. R. Stat. Soc. Ser. C **60**(3): 317–37, 2011.

R. G. Everitt
Bayesian Parameter Estimation for Latent Markov Random Fields and Social Networks.
J. Comput. Graph. Stat., **21**(4): 940–60, 2012.

A. Grelaud, C. P. Robert, J.-M. Marin, F. Rodolphe & J.-F. Taly
ABC likelihood-free methods for model choice in Gibbs random fields.
Bayesian Analysis, **4**(2): 317–36, 2009.

D. K. Pickard
Inference for Discrete Markov Fields: The Simplest Nontrivial Case.
J. Am. Stat. Assoc., **82**(397): 90–96, 1987.

For Further Reading III

D. Feng & L. Tierney
`PottsUtils`: Utility Functions of the Potts Models.
R package version 0.3-2
`http://CRAN.R-project.org/package=PottsUtils`

I. Murray, Z. Ghahramani & D. J. C. MacKay
MCMC for Doubly-intractable Distributions.
In *Proc.* 22nd *Conf. UAI*, AUAI Press, 359–366, 2006.

P. Marjoram, J. Molitor, V. Plagnol & S. Tavaré
Markov chain Monte Carlo without likelihoods.
Proc. Natl Acad. Sci., **100**(26): 15324–15328, 2003.

R. B. Potts
Some generalized order-disorder transformations.
Proc. Cambridge Philosophical Society, **48**(1): 106–9, 1952.

R. H. Swendsen & J.-S. Wang
Nonuniversal critical dynamics in Monte Carlo simulations.
Physical Review Letters, **58**: 86–8, 1987.