Bayesian inference and mathematical imaging. Part IV: mixture, random fields, and hierarchical models.

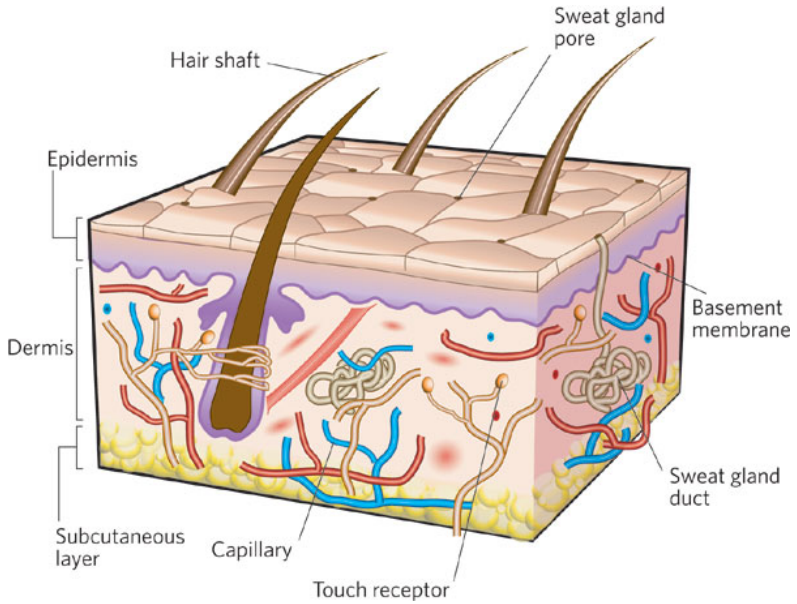
Dr. Marcelo Pereyra
http://www.macs.hw.ac.uk/~mp71/

Maxwell Institute for Mathematical Sciences, Heriot-Watt University

January 2019, CIRM, Marseille.

Skin

- Skin cancer is the most common form of cancer
- Skin melanoma kills 14 000 in Europe every year
- Diagnosis and treatment are main public health issues

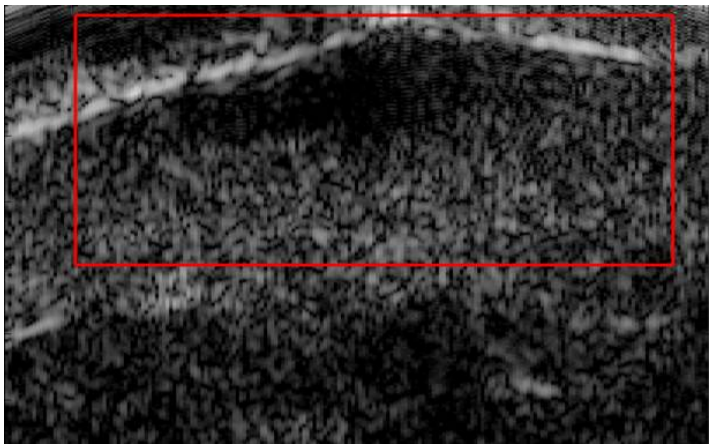


Human Skin Layers (MacNeil, 2007)

Ultrasound imaging

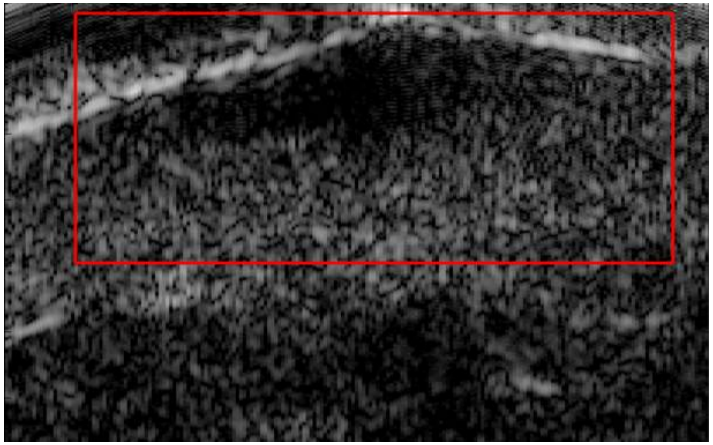
Ultrasound imaging

- US: diagnostics, routine tests, therapy and surgery
- US imaging of skin: new high frequency 3D ultrasound probes
- Study skin diseases & improve diagnosis
- Assess lesion boundaries prior to surgery (measure depth)



Dermis view with skin lesion outlined by the red rectangle

Ultrasound imaging



Limitations:

- Manual annotation of 3D images is time-consuming
- Strong speckle noise ($\mathrm{SNR} < 5.9 dB$), poor contrast & edges
- Segmentation is extremely operator-dependant

Objective:

Automatic and reliable segmentation of skin layers & lesions in 3D.

Outline

Outline

Physical Signal Model

Point scattering framework (Morse and Ingard, 1987)

$$x_n \triangleq x(t_n) = \sum_{i=1}^{M} a_i p(t_n - \tau_i) \tag{1}$$

$$r_n \triangleq r(t_n) = \left| \sum_{i=1}^{M} a_i \left[p(t_n - \tau_i) + \jmath \tilde{p}(t - \tau_i) \right] \right| \tag{2}$$

- M: total number of point scatterers
- a_i: cross-section of ith scatterer
- τ_i: time of arrival of ith backscattered wave
- $p(t) + \jmath \tilde{p}(t)$: analytic extension of the interrogating pulse $p(t)$

Medical ultrasound Imaging: M, $\mathbf{a}$ and $\boldsymbol{\tau}$ are unknown quantities

Statistical Signal Model

Important questions

- What are the possible statistical distributions of x_n and r_n ?
- What information about M, $\mathbf{a}$ and $\boldsymbol{\tau}$ in $f(x_n)$ and $f(r_n)$?

Conventional analytical answer: central limit theorem (M is very large) (Wagner et al., 1983)

$$x_n = \sum_{i=1}^{M} a_i p(t_n - \tau_i) \sim \mathcal{N}(0, \sigma_n^2)$$

$$r_n = \left| \sum_{i=1}^{M} a_i \left[p(t_n - \tau_i) + \jmath \tilde{p}(t_n - \tau_i) \right] \right| \sim \mathcal{R}ayleigh(\sigma_n)$$

$\sigma_n^2 \propto M\langle a_i^2 \rangle$ is the power backscattered by the nth resolution cell

Statistical Signal Model

For many biol. tissues $x_n \sim \mathcal{N}(0, \sigma^2)$ and $r_n \sim \mathcal{R}ayleigh(\sigma)$ are poor models, the empirical tails are not well modeled (Shankar et al., 1993; Shankar, 2000, 2003; Raju and Srinivasan, 2002)

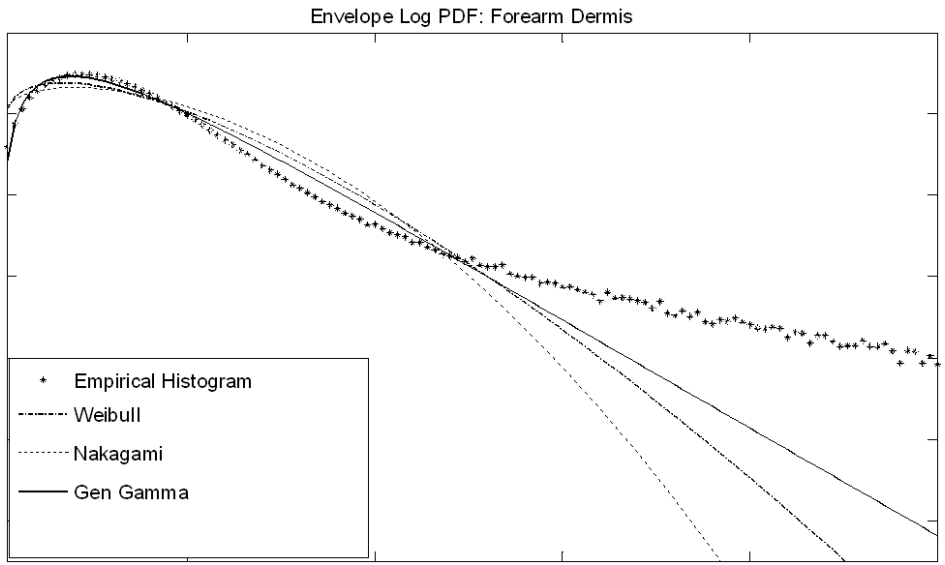


Figure : Comparison of the empirical envelope pdf obtained from forearm dermis, and the corresponding estimations using the generalized gamma, Weibull and Nakagami distributions

Statistical Signal Model

How are these non-Gaussian statistics explained ?

- M is not large enough to enforce the CLT
- M is large and a_i has very high variance
- σ_n^2 fluctuates strongly within homogenous regions
- The signal formation model is inaccurate

What is an appropriate non-Gaussian distribution for x_n and r_n ?
$\rightarrow$ *to answer these questions we study the limit distributions of x_n and r_n.*
Limit distribution: domain of attraction (equilibrium point) in the space of probability density functions

The Gaussian distribution (CLT) - is only a particular case (finite variance). There are infinitely other equilibrium points.

Main results

Main results:

If $f(x_n)$ converges as $M \to \infty$ to a non-Gaussian distribution then

1. x_n has a symmetric α-stable limiting distribution
   $x_n \sim S\alpha S(\alpha, \gamma)$ with $\alpha \in (0, 2)$ and $\gamma \in \mathbb{R}^+$

2. The distribution of the scattering cross-section a_i is heavy-tailed with the same characteristic exponent α
   $f_A(a_i) \propto a_i^{-(\alpha+1)}$

3. r_n is the envelope of a complex $S\alpha S$ random variable
   $r_n \sim \alpha\mathcal{R}ayleigh(\alpha, \gamma)$

Result 1: $S\alpha S$ statistical model

If x_n converges in distribution as $M \to \infty$, then it converges to a $S\alpha S(\alpha, \gamma)$ distribution with $\alpha \in (0, 2)$ and $\gamma \in \mathbb{R}^+$

1. x_n is a sequence of random summands $a_i p(t - \tau_i)$
   Its limit distribution must be invariant to addition
2. The characteristic function must be closed under exponentiation, only the α-stable family has this property
3. $a_i p(t - \tau_i)$ is statistically symmetric, $f_X(x_n)$ converges to a symmetric α-stable distribution with parameters $\alpha \in (0, 2]$ and $\gamma \in \mathbb{R}^+$
4. The case $\alpha = 2$ corresponds to the Gaussian distribution and x_n is known to be not-Gaussian, we conclude that $\alpha \in (0, 2)$

Result 2: Power-law scattering cross-section distribution

Given that $a_i \in \mathbb{R}^+$ and $p(t-\tau_i) \in [-P, P]$ is bounded, if $x_n \sim S\alpha S(\alpha, \gamma)$ with $\alpha < 2$, then a_i follows a heavy-tailed distribution with exponent α

$$f_A(a_i) \propto a_i^{-(\alpha+1)}$$

Key idea:

use necessary conditions for convergence to infer the class of $f_A(a_i)$.

x_n *is in the domain of attraction of a* $S\alpha S$ *distribution with* $\alpha < 2$ *only if* $F_Z(z_i = a_i p(t_n - \tau_i))$ *verifies the Doebling & Gnedenko conditions (Samorodnitsky and Taqqu, 2000)*

C1

$$\lim_{z_i \to \infty} \frac{F_Z(-z_i)}{1 - F_Z(z_i)} = \frac{C_+}{C_-} = 1$$

C2

$$\lim_{z_i \to \infty} \frac{1 - F_Z(z_i) + F_Z(-z_i)}{1 - F_Z(lz_i) + F_Z(-lz_i)} = l^\alpha, \quad \forall l > 0$$

C1 & C2 imply that $F_Z(z_i) \propto |z_i|^{-\alpha} + o\left(|z_i|^{-\alpha}\right)$.

Result 2: Power-law scattering cross-section distribution

Also, z_i is a product of random variables $z_i = a_i u_i$ (Rohatgi, 1976)

$$F_Z(z_i) = \begin{cases} \int_{-P}^{0} f_U(u_i)[1 - F_A(z_i/u_i)]\, du_i & \text{if } z_i < 0 \\ F_Z(0^-) + \int_0^P f_U(u_i) F_A(z_i/u_i)\, du_i & \text{if } z_i \geq 0 \end{cases}$$

with $u_i = p(t_n - \tau_i)$. Then, for $z_i \gg 0$

$$\int_0^P f_U(u_i) F_A(z_i/u_i)\, du_i \approx c z_i^{-\alpha}$$

This condition is verified by all power-law distributions

$$f_A(a_i) = L(a_i) a_i^{-(\alpha+1)}$$

where $L(a_i)$ is a *slow varying* function (i.e., $\lim_{s\to\infty} \frac{L(ks)}{L(s)} = 1$)

Result 3: αRayleigh envelope distribution

The envelope r_n is the amplitude of the analytic extension of x_n

$$r_n \cos(\varphi_n) = x_n + \jmath y_n, \quad r_n = |x_n + \jmath y_n|$$

Assuming that the position of scatterers is purely random

$$x_n \sim S\alpha S(\alpha, \gamma) \Rightarrow y_n \sim S\alpha S(\alpha, \gamma), \quad y_n \perp x_n$$

By deriving $f(r_n, \varphi_n)$ from $f(x_n, y_n)$ and marginalizing w.r.t. φ_n

$$r_n \sim \alpha\mathcal{R}ayleigh(r_n|\alpha, \gamma)$$

where

$$\begin{aligned}\alpha\mathcal{R}ayleigh(r_n|\alpha, \gamma) &\triangleq \int_0^\infty \mathcal{R}ayleigh(r_n|\sigma) S_{\frac{\alpha}{2}}(\sigma^2|\gamma \cos\left(\frac{\pi\alpha}{4}\right)^{\frac{2}{\alpha}}, 1, 0) d\sigma \\ &= \int_0^\infty r_n \lambda \exp[-(\gamma\lambda)^\alpha] J_0(r_n\lambda)\, d\lambda\end{aligned}$$

Interpretation of α and γ

For modeling and physical interpretation purposes the scattering cross-sections can be assumed to follow a Pareto distribution

$$f_A(a_i) \cong \alpha a_m^{\alpha} a_i^{-(\alpha+1)}$$

a_m is given by $a_m = \lim_{a_i \to \infty} a_i^{\alpha} F_A(a_i)$

Moreover, γ is a scale or spread parameter

$$\gamma = D^*(\alpha) \sqrt[\alpha]{M} a_m$$

where $D^*(\alpha) = \sqrt[\alpha]{\frac{2\pi \langle p_i^{\alpha} \rangle}{\Gamma(\alpha) \sin\left(\frac{\pi\alpha}{2}\right)}}$, M is the number of scatters and $\langle p_i^{\alpha} \rangle$ is the α-th fractional moment of $p(t - \tau_i)$

Experimental validation

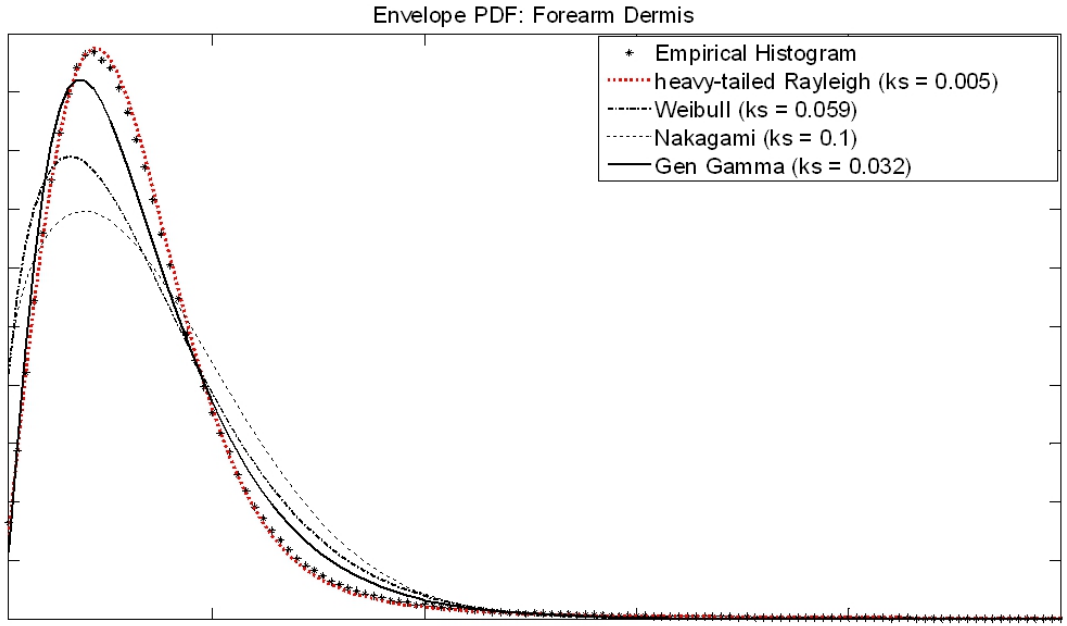
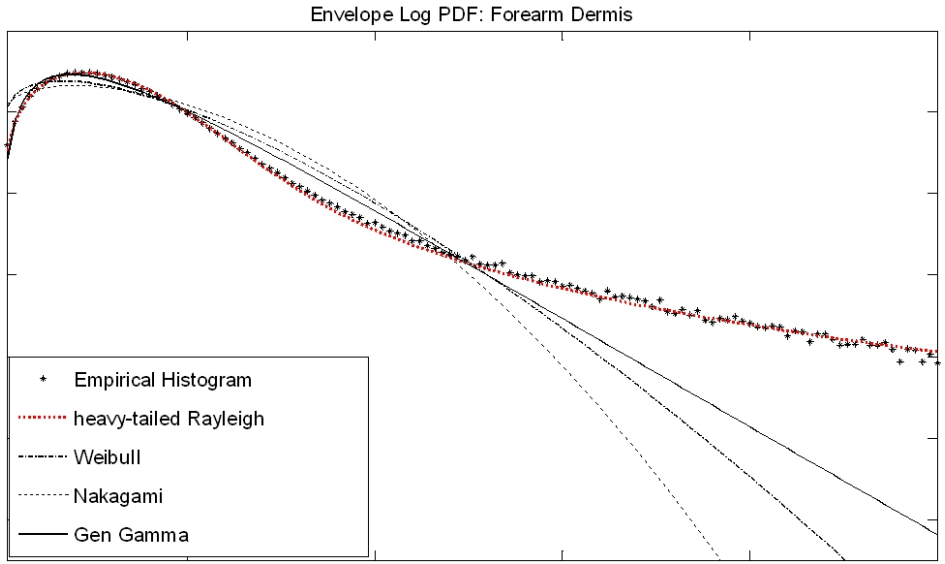


Figure : Comparison of the empirical envelope pdf obtained from forearm dermis, and the corresponding estimations using the heavy-tailed Rayleigh, generalized gamma, Weibull and Nakagami distribution

Experimental validation



Comparison of distributions tails by means of a logarithmic pdfs

Outline

Outline

Segmentation Problem

- $\mathbf{r} = (r_1, \ldots, r_n, \ldots, r_N)^T \in \mathbb{R}^N$ is a 2D or 3D B-mode image
- $\mathbf{r}$ is made up by K regions (biological tissues) $\mathcal{C}_1, \ldots, \mathcal{C}_K$
- $\mathbf{z} = \{z_1, \ldots, z_N\}$ is a label vector that maps observations $r_1, \ldots, r_N$ to tissues or classes $C_1, \ldots, C_K$

$$z_n = k \Leftrightarrow n \in C_k$$

- Each region is characterized by its own α-Rayleigh statistics

$$z_n = k \Rightarrow r_n \sim p_{\alpha\mathcal{R}}(r_n|\alpha_k, \gamma_k)$$

We consider the maximum a posteriori (MAP) segmentation problem

$$\hat{\mathbf{z}} = \underset{\mathbf{z}}{\text{argmax}}\ f(\mathbf{z}|\mathbf{r}) = \underset{\mathbf{z}}{\text{argmax}} \int\int f(\mathbf{z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}|\mathbf{r}) d\boldsymbol{\alpha} d\boldsymbol{\gamma}$$

- K: the number of classes (considered known),
- $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_K\}$ and $\boldsymbol{\gamma} = \{\gamma_1, \ldots, \gamma_K\}$ (unknown)

Bayesian Model

We define a Bayesian model (likelihood and priors) associated to the unknown parameter vector $(\mathbf{z}^T, \boldsymbol{\alpha}^T, \boldsymbol{\gamma}^T)^T$

Likelihood $f(\mathbf{r}|\mathbf{z}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$
We use an α-Rayleigh observation model

$$f(r_n|z_n = k, \alpha_k, \gamma_k) = p_{\alpha\mathcal{R}}(r_n|\alpha_k, \gamma_k)$$

Assuming observations r_n are independent

$$f(\mathbf{r}|\mathbf{z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) = \prod_{k=1}^{K} \prod_{\{n|z_n=k\}} p_{\alpha\mathcal{R}}(r_n|\alpha_k, \gamma_k)$$

which is closely related to an α-Rayleigh mixture model

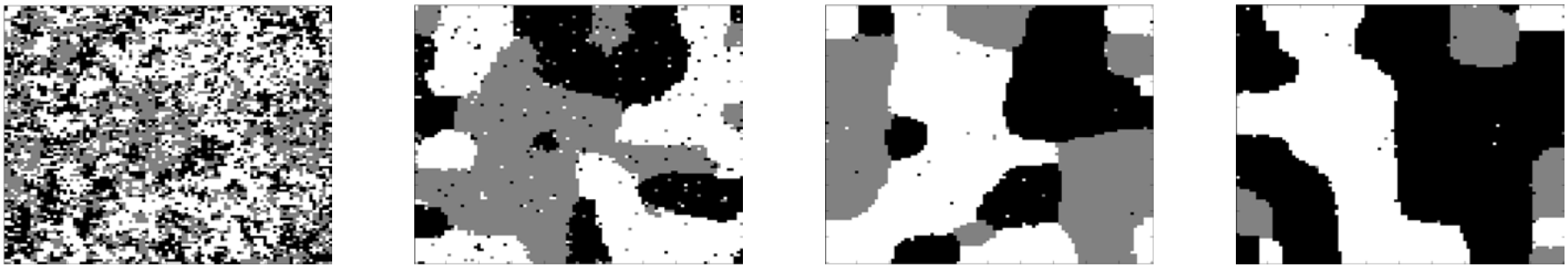
$$f(\mathbf{r}|\boldsymbol{\alpha}, \boldsymbol{\gamma}) = \prod_{\mathbf{n}} \sum_{k=1}^{K} \omega_k f_{\alpha\mathcal{R}}(r_n|\alpha_k, \gamma_k) \approx \sum_{\mathbf{z}} f(\mathbf{r}|\mathbf{z}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$$

Label vector $\mathbf{z}$

We consider 2D and 3D Potts Markov fields as prior distributions for $\mathbf{z}$ (Wu, 1982)

$$\mathrm{P}(\mathbf{z}) = \frac{1}{C(\beta)} \exp\left[\sum_{n=1}^{N} \sum_{n' \in \mathcal{V}(n)} \beta \delta(z_n - z_{n'})\right]$$

where β is the granularity coefficient, $\delta(\cdot)$ is the Kronecker function and $\mathcal{V}(n)$ is the field's neighborhood structure

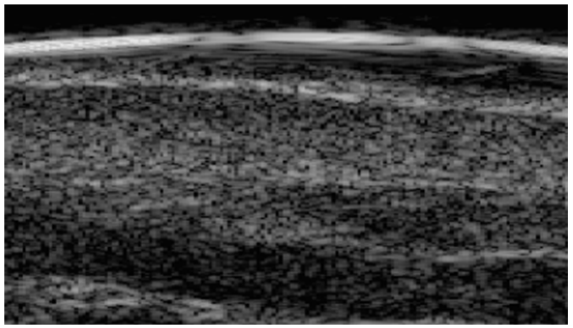


(a) $\beta = 0.6$ (b) $\beta = 1$ (c) $\beta = 1.2$ (d) $\beta = 1.4$

Figure : Synthetic images of a 3D Potts-Markov model generated using different granularity coefficients (1 slice)

Label vector $\mathbf{z}$

The proposed segmentation method uses 2D MRFs for single-slice 2D ultrasound images and 3D MRFs for multiple-slice 3D images



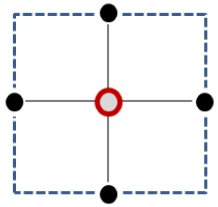
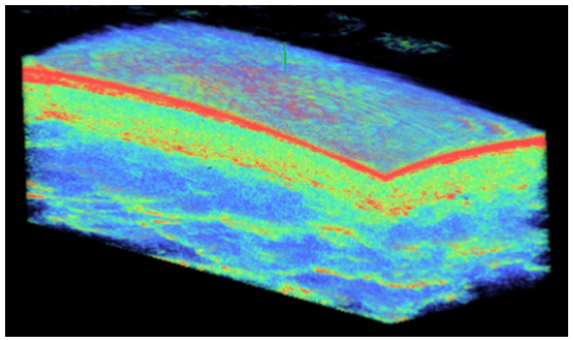


Figure : ultrasound image (left) and neighborhood structure $\mathcal{V}(n)$ (right) in 2D



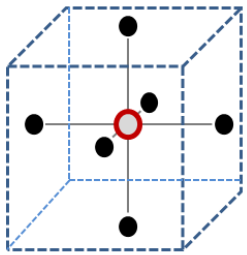


Figure : ultrasound image (left) and neighborhood structure $\mathcal{V}(n)$ (right) in 3D

α-Rayleigh parameters

characteristic index α_k

Non-informative prior on α_k $(k = 1, \ldots, K)$

$$\alpha_k \sim \mathcal{U}(0, 2)$$

this interval $(0, 2)$ covers all possible values of α_k

spread γ_k

Inverse gamma prior on γ_k with hyperparameters a_0 and b_0

$$\gamma_k \sim \mathcal{IG}(a_0, b_0), \quad k = 1, \ldots, K$$

where $a_0 = 1$ and $b_0 = 1$ to obtain a vague prior

Prior on unknown parameter vector $(\boldsymbol{\theta}, \mathbf{z})$

Assuming a priori independence between the parameters α_k, γ_k and the labels $\mathbf{z}$ the joint prior is

$$f(\mathbf{z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) = \mathrm{P}(\mathbf{z})\, p(\boldsymbol{\alpha}) p(\boldsymbol{\gamma}) = \mathrm{P}(\mathbf{z}) \prod_{k=1}^{K} f(\alpha_k) f(\gamma_k)$$

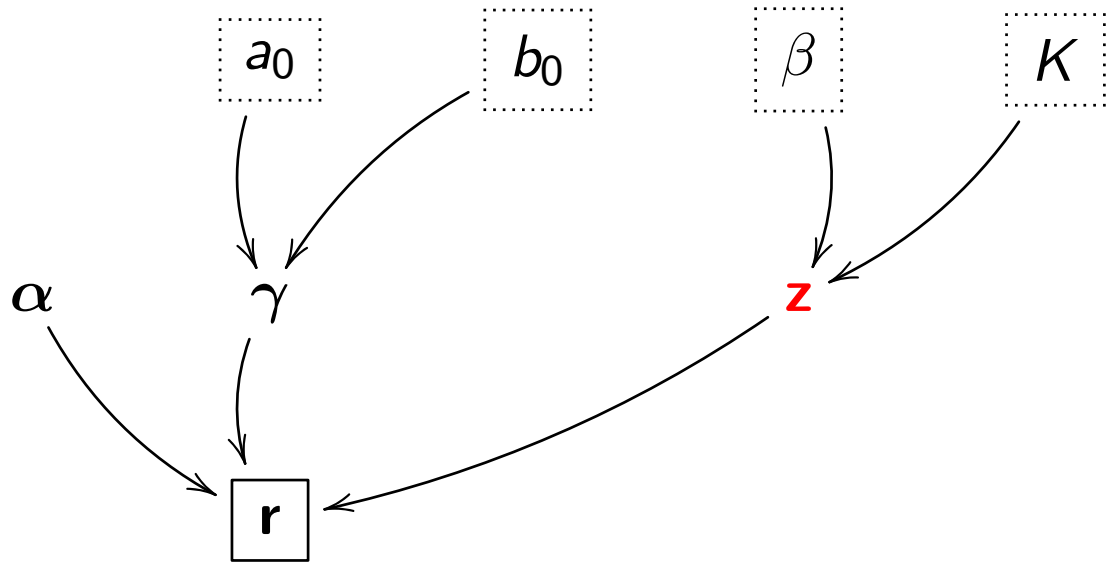


Figure : Directed acyclic graph (DAG) for the α-Rayleigh mixture model (the fixed nonrandom hyperparameters appear in dashed boxes)

Posterior density

Using Bayes theorem, the posterior distribution of $(\mathbf{z}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$ is

$$f(\mathbf{z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}|\mathbf{r}) \propto f(\mathbf{r}|\mathbf{z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) \mathrm{P}(\mathbf{z}) f(\boldsymbol{\alpha}) f(\boldsymbol{\gamma})$$

We are interested in the marginal

$$f(\mathbf{z}|\mathbf{r}) = \int \int f(\mathbf{z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}|\mathbf{r}) \, d\boldsymbol{\alpha} d\boldsymbol{\gamma}$$

MCMC Method

1. We use a Hybrid Gibbs sampler to generate samples asymptotically distributed according to $f(\mathbf{z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}|\mathbf{r})$
2. We marginalize implicitly by discarding samples $\boldsymbol{\alpha}^{(t)}, \boldsymbol{\gamma}^{(t)}$
3. The samples $\mathbf{z}^{(t)}$ are then used to approximate the voxel-wise MAP estimators $\hat{z}_n = \text{argmax}_{z_n} \; f(z_n|\mathbf{r})$ for $n = 1, \ldots, N$.

Outline

Hybrid Metropolis-within-Gibbs sampler

Generate samples asymptotically distributed according to $f(\mathbf{z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}|\mathbf{r})$ by iteratively sampling $\mathrm{P}(\mathbf{z}|\boldsymbol{\alpha}, \boldsymbol{\gamma}, \mathbf{r})$, $f(\boldsymbol{\alpha}|\mathbf{z}, \boldsymbol{\gamma}, \mathbf{r})$ and $f(\boldsymbol{\gamma}|\mathbf{z}, \boldsymbol{\alpha}, \mathbf{r})$

for $t = 1, 2, ..$ to T **do**

— *Update* the label vector $\mathbf{z}$ —

for $n = 1, 2, ..$ to N **do**

1. Draw $z_n^{(t)}$ from $\mathrm{P}(z_n = k|\mathbf{z}_{1:n-1}^{(t)}, \mathbf{z}_{n+1:N}^{(t-1)}, r_n, \alpha_k^{(t)}, \gamma_k^{(t)})$

end for

— *Update* the α-Rayleigh parameters —

for $k = 1, 2, ..$ to K **do**

2. Sample $\alpha_k^{(t)}$ from $f(\alpha_k|\boldsymbol{\gamma}^{(t-1)}, \mathbf{z}, \mathbf{r})$
3. Sample $\gamma_k^{(t)}$ from $f(\gamma_k|\boldsymbol{\alpha}^{(t)}, \mathbf{z}, \mathbf{r})$

end for

end for

Output samples $\mathbf{z}^{(t)}$

Conditional probability mass function $\mathrm{P}(\mathbf{z}|\boldsymbol{\alpha}, \boldsymbol{\gamma}, \mathbf{r})$

Generation of samples according to $\mathrm{P}[\mathbf{z}|\boldsymbol{\alpha}, \boldsymbol{\gamma}, \mathbf{r}]$
Sample $\mathbf{z}$ coordinate-by-coordinate using Gibbs moves

$$\mathrm{P}(z_n = k|z_{-n}, r_n, \alpha_k, \gamma_k) \propto f(r_n|z_n = k, \boldsymbol{\alpha}, \boldsymbol{\gamma})\mathrm{P}(z_n|\mathbf{z}_{-n})$$

where $k = 1, \ldots, K$ and $\mathbf{z}_{-n}$ is the vector $\mathbf{z}$ whose nth element has been removed.

The resulting Markov random field has the following conditionals

$$\mathrm{P}(z_n = k|z_{-n}, r_n, \alpha_k, \gamma_k) \propto \exp\left[\sum_{n' \in \mathcal{V}(n)} \beta\delta(k - z_{n'})\right] \times r_n \int_0^\infty \lambda \exp\left[-(\gamma_k\lambda)^{\alpha_k}\right] J_0(r_n\lambda)\, d\lambda.$$

This step has been parallelized to reduce computing time.

Hybrid Metropolis-within-Gibbs sampler

Update $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$

The conventional Gibbs sampler requires sampling from $f(\alpha_k|\boldsymbol{\gamma}, \mathbf{z}, \mathbf{r})$ and $f(\gamma_k|\boldsymbol{\alpha}, \mathbf{z}, \mathbf{r})$. **However, sampling from these distributions is not easy.**

Generation using Metropolis-Hastings

Instead, we generate samples asymptotically distributed according to $f(\alpha_k|\boldsymbol{\gamma}, \mathbf{z}, \mathbf{r})$ and $f(\gamma_k|\boldsymbol{\alpha}, \mathbf{z}, \mathbf{r})$ using Metropolis-Hasting moves. This results in a Metropolis-within-Gibbs sampler which also converges to the desired posterior density.

Metroplis-Hastings move

1. Generate a sample according to an appropriate proposal distribution
2. Accept or reject that sample with a given probability

Conditional probability density function $p(\boldsymbol{\alpha}|\boldsymbol{\gamma}, \mathbf{z}, \mathbf{r})$

Generation of samples according to $f(\boldsymbol{\alpha}|\boldsymbol{\gamma}, \mathbf{z}, \mathbf{r})$

We sample $\boldsymbol{\alpha}$ coordinate-by-coordinate using Metropolis-Hastings moves

Posterior density

$$f(\alpha_k|\boldsymbol{\gamma}, \mathbf{z}, \mathbf{r}) \propto \prod_{\{n|z_n=k\}}^{N} p_{\alpha\mathcal{R}}(r_n|\alpha_k^*, \gamma_k) p(\alpha_k)$$

Random walk proposal

$$\alpha_k^* \sim \mathcal{N}_{(0,2)}(\alpha_k^{(t-1)}, \sigma_{\alpha,k}^2)$$

Resulting acceptance probability

$$1 \vee \left\{ \frac{\mathcal{N}_{(0,2)}(\alpha_k^{(t-1)}|\alpha_k^*, \sigma_{\alpha,k}^2)}{\mathcal{N}_{(0,2)}(\alpha_k^*|\alpha_k^{(t-1)}, \sigma_{\alpha,k}^2)} \times \prod_{\{n|z_n=k\}}^{N} \frac{p_{\alpha\mathcal{R}}(r_n|\alpha_k^*, \gamma_k)}{p_{\alpha\mathcal{R}}(r_n|\alpha_k^{(t-1)}, \gamma_k)} \right\}$$

Conditional probability density function $p(\boldsymbol{\gamma}|\boldsymbol{\alpha}, \mathbf{z}, \mathbf{r})$

Generation of samples according to $f(\boldsymbol{\gamma}|\boldsymbol{\alpha}, \mathbf{z}, \mathbf{r})$

We sample $\boldsymbol{\gamma}$ coordinate-by-coordinate using Metropolis-Hastings moves

Posterior density

$$f(\gamma_k|\boldsymbol{\alpha}, \mathbf{z}, \mathbf{r}) \propto \prod_{\{n|z_n=k\}}^{N} p_{\alpha\mathcal{R}}(r_n|\alpha_k^*, \gamma_k) p(\gamma_k)$$

Random walk proposal

$$\gamma_k^* \sim \mathcal{N}_{\mathbb{R}^+}\left(\gamma_k^{(t-1)}, \sigma_{\gamma,k}^2\right)$$

Resulting acceptance probability

$$1 \vee \left\{ \frac{\mathcal{N}_{\mathbb{R}^+}\left(\gamma_k^{(t-1)}|\gamma_k^*, \sigma_{\gamma,k}^2\right)}{\mathcal{N}_{\mathbb{R}^+}\left(\gamma_k^*|\gamma_k^{(t-1)}, \sigma_{\gamma,k}^2\right)} \times \prod_{\{n|z_n=k\}}^{N} \frac{p_{\alpha\mathcal{R}}(r_n|\alpha_k, \gamma_k^*) f(\gamma_k^*|a_0, b_0)}{p_{\alpha\mathcal{R}}(r_n|\alpha_k, \gamma_k^{(t-1)}) f(\gamma_k^{(t-1)}|a_0, b_0)} \right\}$$

Approximation of the Likelihood

Evaluating the likelihood is very time-consuming and is required at every step of the sampler and for every observation

$$p_{\alpha\mathcal{R}}(r_n|\alpha_k,\gamma_k) = r_n \int_0^\infty \lambda \exp\left[-(\gamma_k\lambda)^{\alpha_k}\right] J_0(r_n\lambda)\, d\lambda$$

An efficient alternative is to approximate the likelihood using the following asymptotic expansions (Sun and Han, 2008)

$$p_{\alpha\mathcal{R}}(r_n|\alpha_k,\gamma_k) = \sum_{p=0}^{P} a_p(\alpha_k,\gamma_k) r_n^{2p+1} + o\left(r_n^{2(P+1)+1}\right) \quad r_n \to 0$$

and

$$p_{\alpha\mathcal{R}}(r_n|\alpha_k,\gamma_k) = \sum_{p=1}^{P} b_p(\alpha_k,\gamma_k) r_n^{-\alpha_k p-1} + o\left(r_n^{-\alpha_k(P+1)-1}\right) \quad r_n \to \infty$$

This function has been parallelized to reduce computing time.

Outline

Experimental Results - Synthetic Data

Synthetic Data

- Generate a spatially coherent 3-component α-Rayleigh mixture with $\boldsymbol{\alpha} = [1.99; 1.99; 1.8]$ and $\boldsymbol{\gamma} = [1; 5; 10]$
- Estimate the posterior density $f(\mathbf{z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}|\mathbf{r})$ using the proposed Gibbs sampler in 2D (4-pixel neighborhood, $25,000$ iterations)
- Compute MAP estimate of $\mathbf{z}|\mathbf{r}$ and MMSE estimate of $\boldsymbol{\alpha}, \boldsymbol{\gamma}|\mathbf{r}$

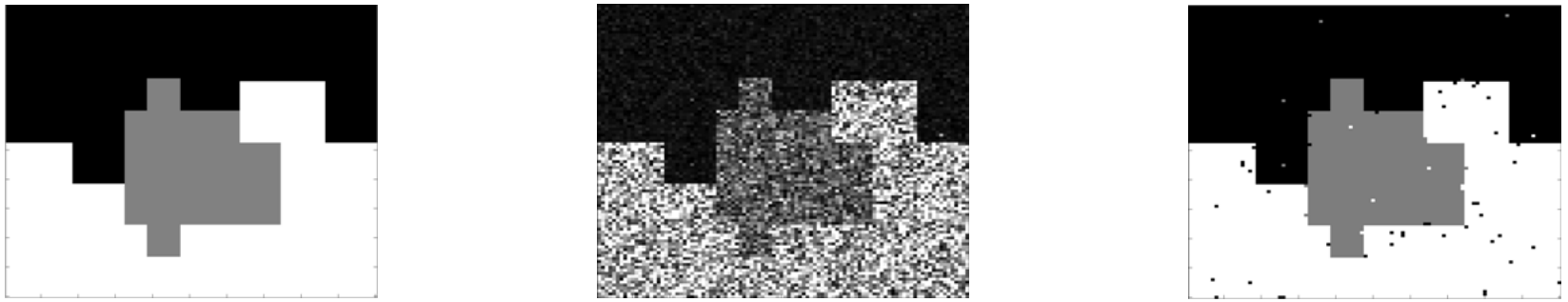


Figure : True labels, observations $\mathbf{r}$ (only input to the algorithm), and MAP estimates for a 3-class mixture

Experimental Results - Synthetic Data

Table : Parameter estimation

	true value	MMSE estimates	standard deviation
α_1	1.99	1.99	0.002
α_2	1.99	1.99	0.003
α_3	1.80	1.79	0.006
γ_1	1.00	1.00	0.003
γ_2	5.00	5.01	0.025
γ_3	10.00	9.96	0.036

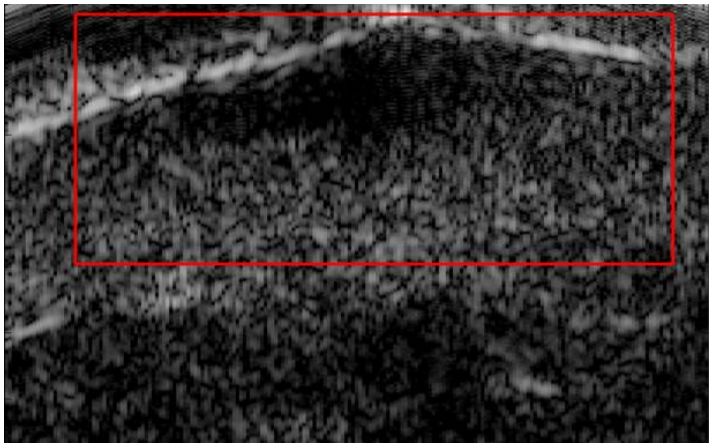
Experimental Results - in vivo data

Application to real data

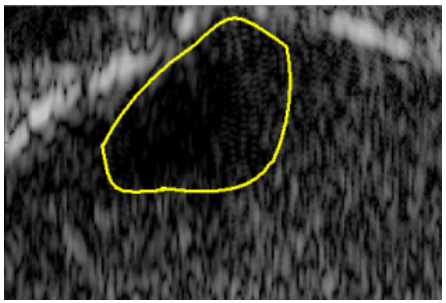
Segmentation of an in-vivo skin lesion in a 3D high frequency ultrasound

- Image acquired at 100MHz with a focalized 25MHz 3D probe
- The number of classes K was identified visually by the clinician
- The granularity coefficient was fixed heuristically to $\beta = 1$
- Algorithm convergence was measured using the "between-within variance criterion" (Gelman and Rubin, 1992)
- Results were computed using 1.000 iterations of the proposed method (900 burn-in period)
- Algorithm implemented in MATLAB with C-MEX and OpenMP

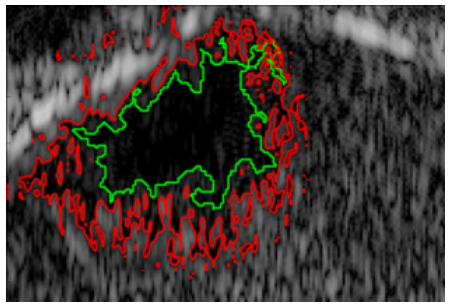
Experimental Results - in vivo data



(a) Dermis view with skin lesion (ROI = $450 \times 200 \times 16$)



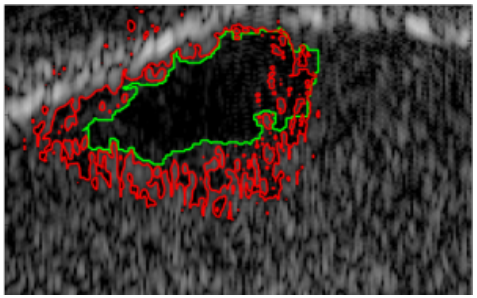
(b) ROI (slice 7)



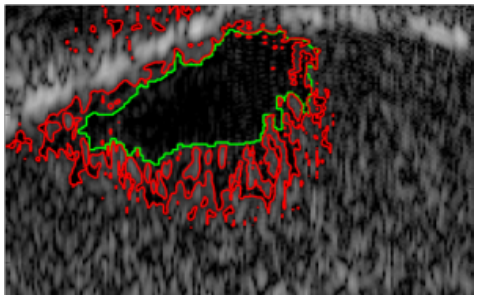
(c) 2D Segmentation contour

Figure : Log-compressed US images of skin tumor and the estimated segmentation contours. Yellow: expert annotation, green: *proposed*, red: (*Sarti,2005*)

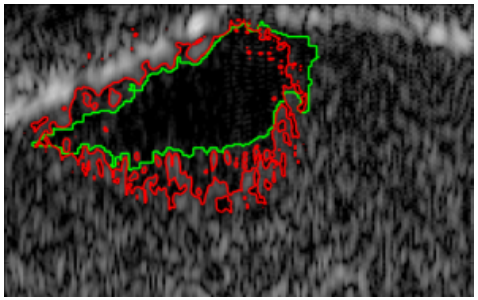
Segmentation results in 3D



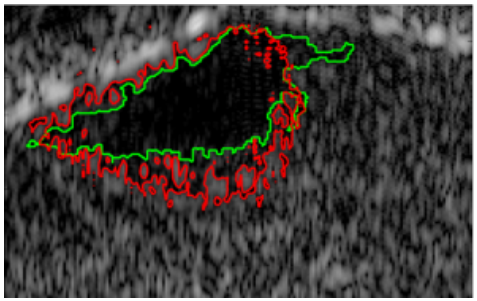
(a) Slice 1



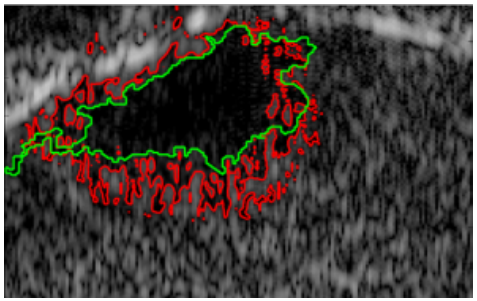
(b) Slice 3



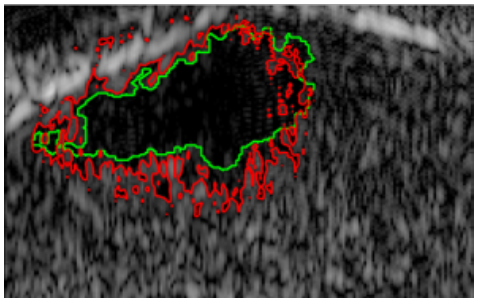
(c) Slice 5



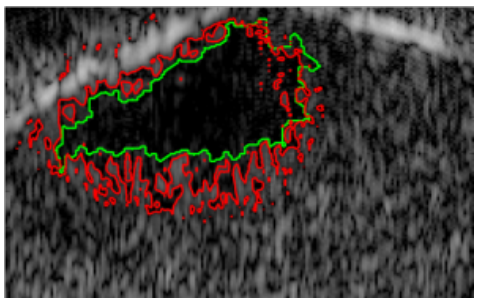
(d) Slice 7



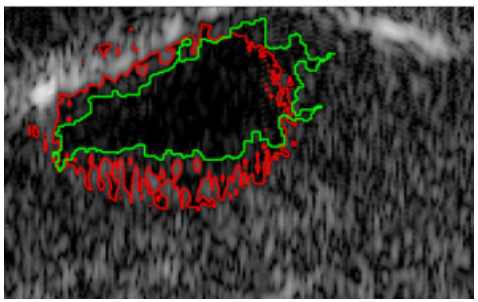
(e) Slice 9



(f) Slice 11



(g) Slice 13



(h) Slice 15

Lesion reconstructed in 3D

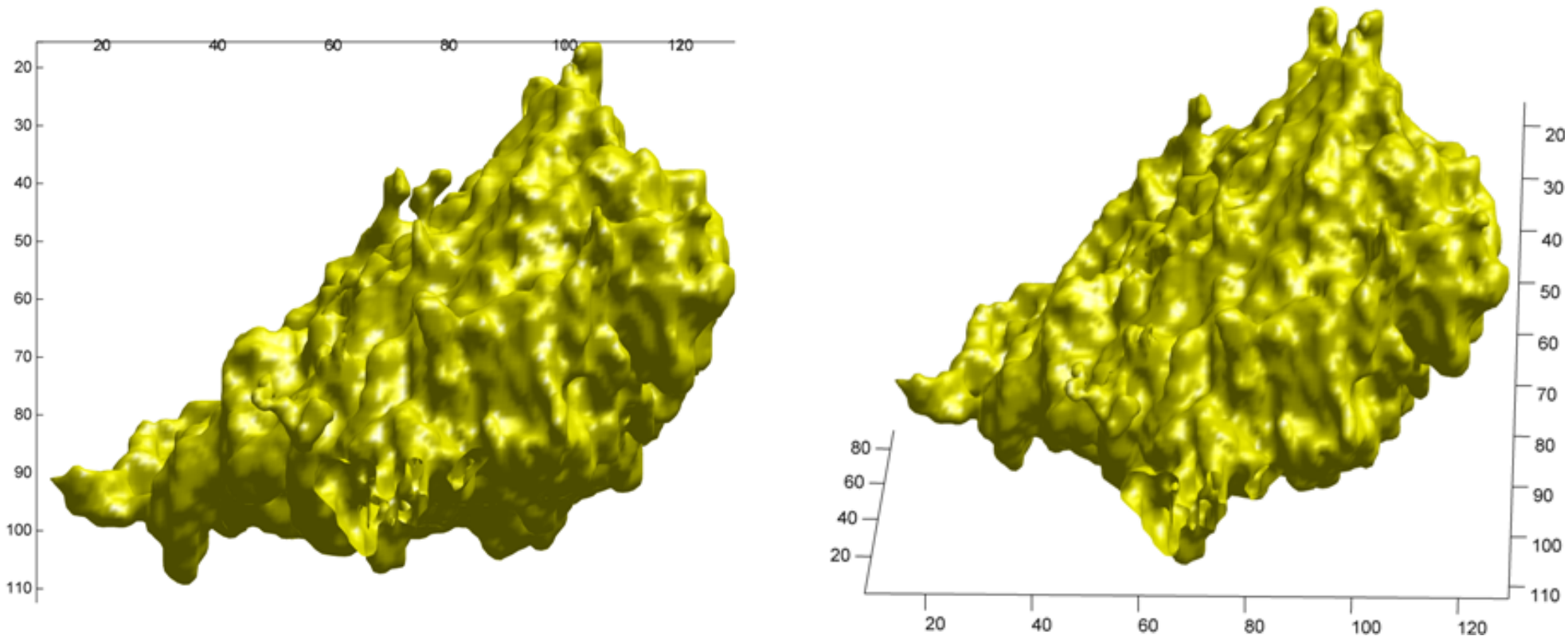
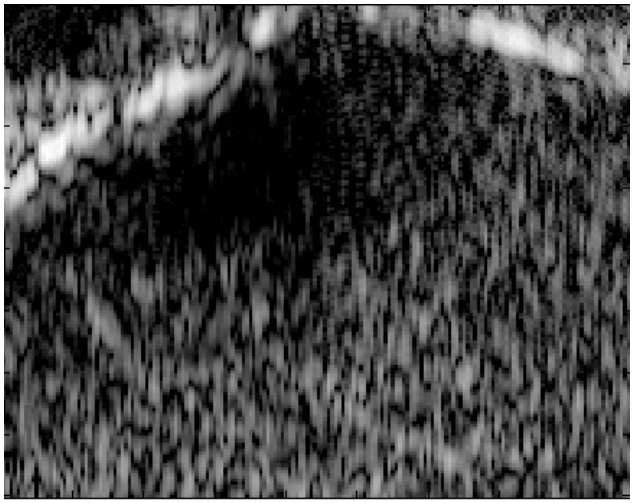


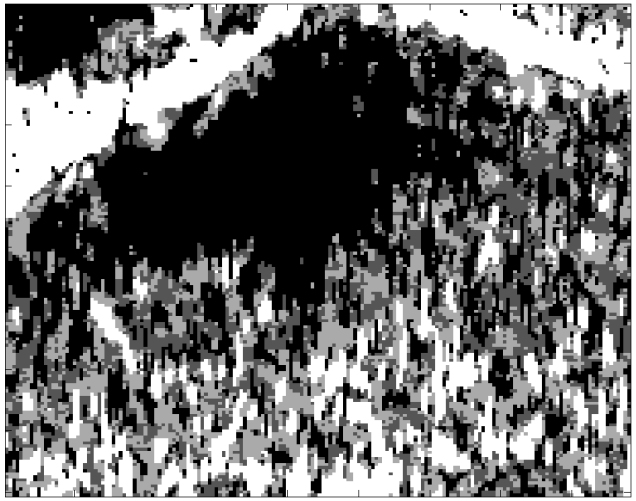
Figure : 3D reconstruction of the melanoma tumor

Outline

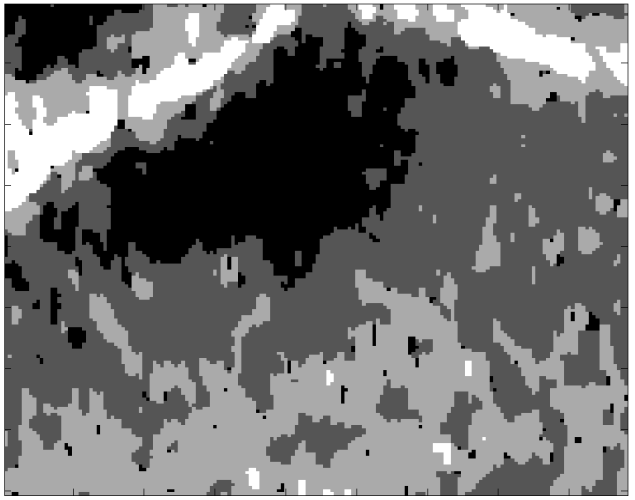
Which β value should be used?



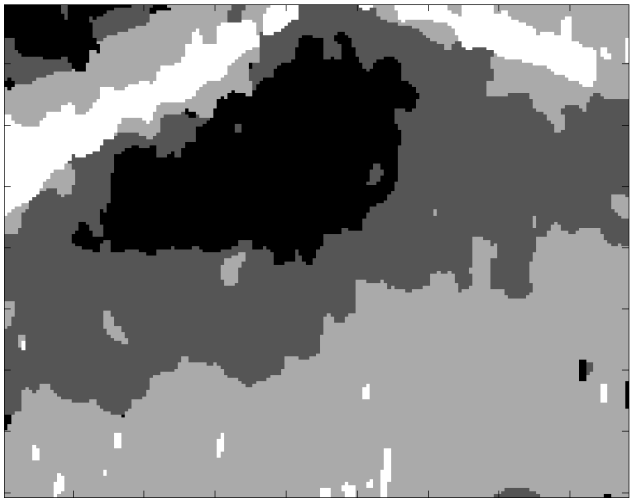
B-mode US ROI)



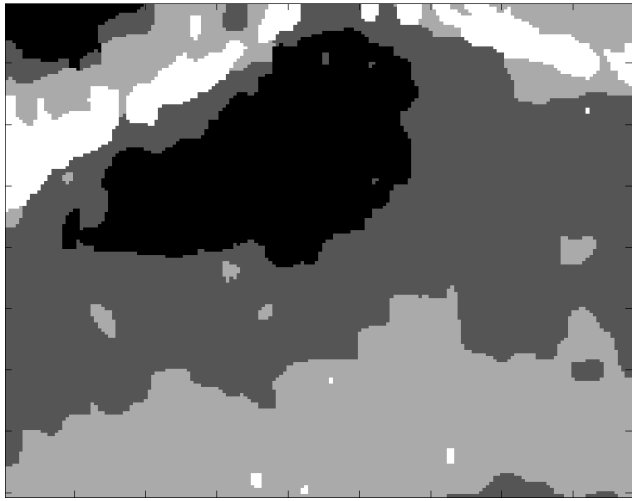
MAP $\mathbf{z}$ ($\beta = 0.5$)



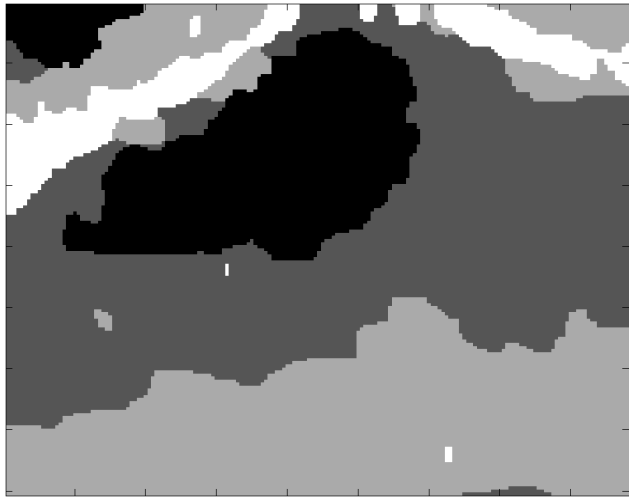
MAP $\mathbf{z}$ ($\beta = 0.75$)



MAP $\mathbf{z}$ ($\beta = 1.0$)



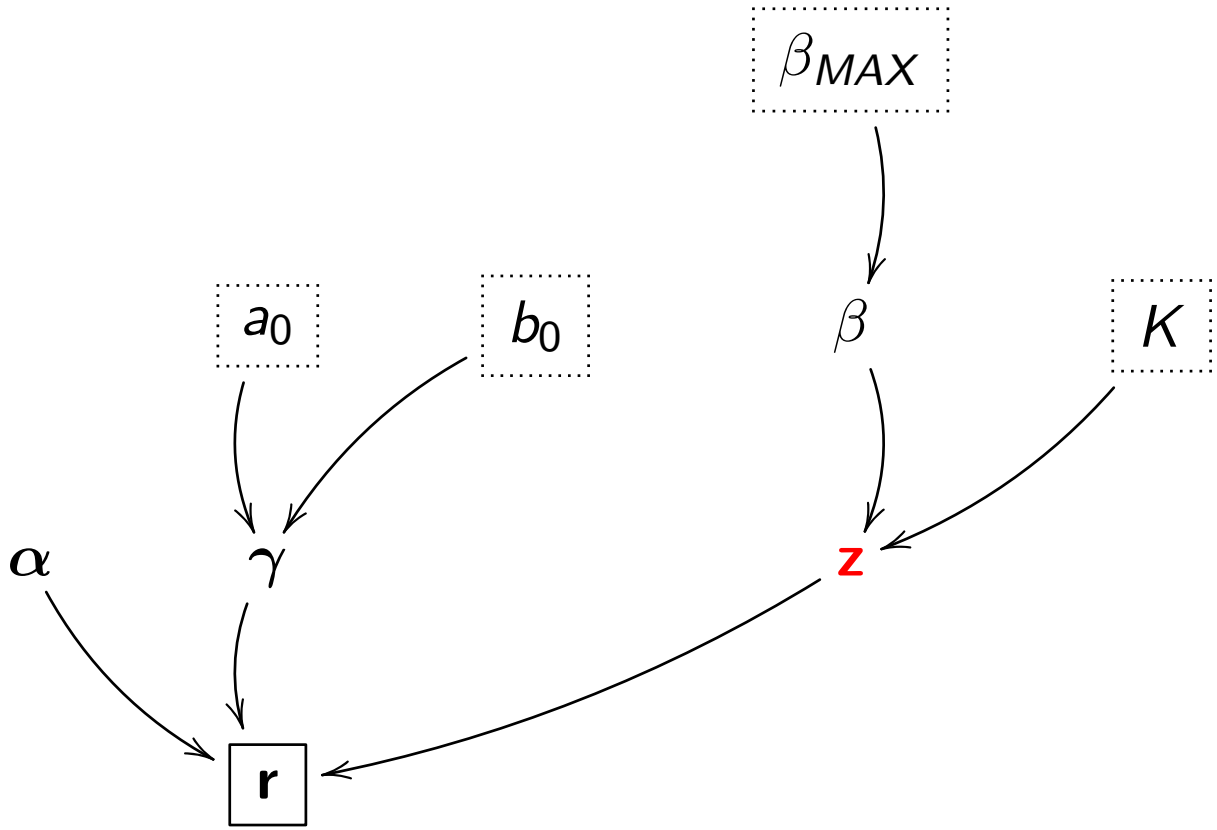
MAP $\mathbf{z}$ ($\beta = 1.25$)



MAP $\mathbf{z}$ ($\beta = 1.5$)

Granularity coefficient β

- Previous experiments used $\beta = 1$, but $\beta > 1$ could improve results
- Estimate β jointly with $\mathbf{z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}$ from the data
- Inference on hierarchical Bayesian models $f(\mathbf{z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \beta | \mathbf{r})$
- Marginalize w.r.t. β: $f(\mathbf{z}|\mathbf{r}) = \int \int \int f(\mathbf{z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \beta | \mathbf{r}) \, d\boldsymbol{\alpha} d\boldsymbol{\gamma} d\beta$



Granularity coefficient β

Conditional density $f(\beta|\boldsymbol{\theta}, \mathbf{z}, \mathbf{r})$

$$\begin{aligned} f(\beta|\boldsymbol{\theta}, \mathbf{z}, \mathbf{r}) &\propto f(\mathbf{r}|\boldsymbol{\theta}, \mathbf{z}, \beta) f(\theta) f(\mathbf{z}|\beta) f(\beta) \\ &\propto f(\mathbf{z}|\beta) f(\beta) \end{aligned}$$

- $f(\mathbf{z}|\beta)$: Potts Markov field
- $f(\beta)$: prior on β

$$\beta \sim \mathcal{U}(0, \beta_{MAX})$$

Sampling β using MH moves requires computing the ratio

$$\text{ratio} = \min\{1, \xi\} \tag{3}$$

with

$$\xi = \frac{f(\mathbf{z}|\beta^*)}{f(\mathbf{z}|\beta^{(t-1)})} \frac{f(\beta^*)}{f(\beta^{(t-1)})} \frac{q(\beta^*|\beta^{(t-1)})}{q(\beta^{(t-1)}|\beta^*)}$$

$\beta^* \sim q(\beta^*|\beta^{(t-1)})$ is an appropriate proposal distribution

Granularity coefficient β

Replacing $f(\mathbf{z}|\beta) = \frac{1}{C(\beta)} \exp\left[\Phi_\beta(\mathbf{z})\right]$ in ξ

$$\xi = \frac{C(\beta^{(t-1)})}{C(\beta^*)} \frac{\exp\left[\Phi_{\beta^*}(\mathbf{z})\right]}{\exp\left[\Phi_{\beta^{(t-1)}}(\mathbf{z})\right]} \frac{f(\beta^*)}{f(\beta^{(t-1)})} \frac{q(\beta^*|\beta^{(t-1)})}{q(\beta^{(t-1)}|\beta^*)}$$

However the ratio $\frac{C(\beta^{(t-1)})}{C(\beta^*)}$ is intractable

Possible solutions:

- Pseudo-likelihood estimators (Besag, 1975)
- Approximation of $C(\beta)$ (Gelman and Meng, 1998; Descombes et al., 1999; Risser et al., 2009)
- Auxiliary variables and perfect sampling (Moller et al., 2006; Murray et al., 2006; Del Moral et al., 2006; Andrieu et al., 2010)
- **Likelihood-free (ABC) methods** (Marjoram et al., 2003; Marin et al., 2011)

Likelihood-free (ABC) Sampling

Idea: Replace $f(\mathbf{z}|\beta)$ (intractable) by a tractable sufficient statistic $\eta(\mathbf{z})$

$$f(\beta|\mathbf{z}) = f\left(\beta|\eta(\mathbf{z})\right)$$

1. Generate an auxiliary variable $\mathbf{w} \sim \mathrm{P}_{\mathbf{Z}}(\mathbf{w}|\beta)$
2. Accept $\mathbf{w}$ if $\eta(\mathbf{w}) = \eta(\mathbf{z})$

Indeed, $\eta(\mathbf{w}) = \eta(\mathbf{z})$ occurs with probability $\mathrm{P}_{\mathbf{Z}}(\mathbf{z}|\beta)$

The Gibbs potential of a Markov random fields is a sufficient statistic, i.e.,

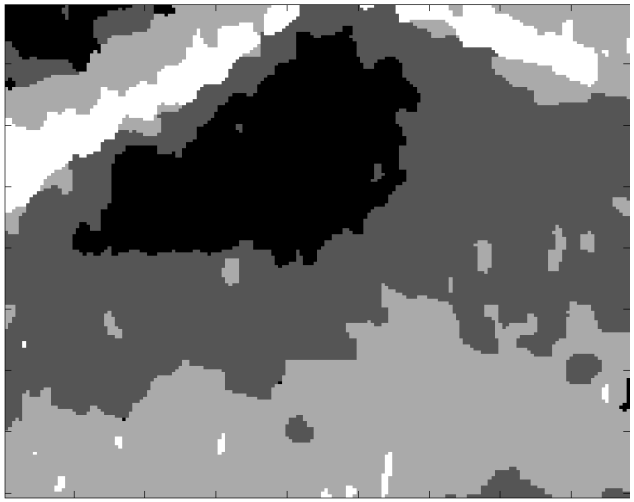
$$\eta(\mathbf{z}) = \sum_{n=1}^{N} \sum_{n' \in \mathcal{V}(n)} \delta(z_n - z_{n'})$$

which is a scalar

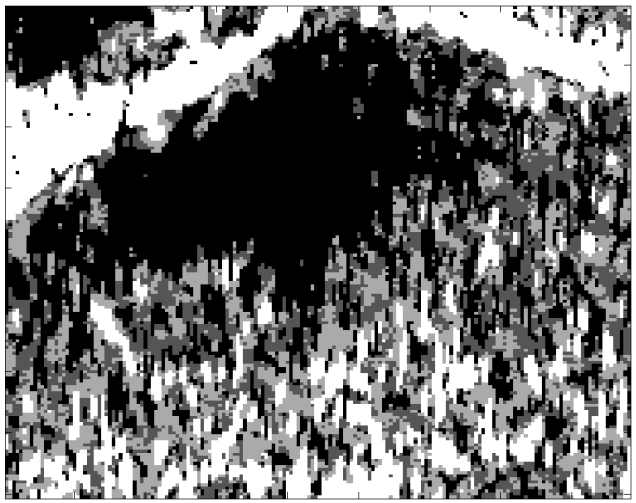
Proposed Likelihood-free Metropolis Hastings Move

Input: $\{\beta^{(t-1)}, \mathbf{z}^{(t)}, \nu, s_\beta^2\}$, number of moves M.
Generate $\beta^* \sim \mathcal{N}_{(0,B)}\left(\beta^{(t-1)}, s_\beta^2\right)$
Generate $\mathbf{w} \sim \mathrm{P}_{\mathbf{Z}}(\mathbf{w}|\beta^*)$ through M Gibbs moves with initial state $\mathbf{z}^{(t)}$
if $|\eta(\mathbf{z}^{(t)}) - \eta(\mathbf{w})| < \epsilon$ **then**
    Set ratio $= \frac{f(\beta^*)}{f(\beta^{(t-1)})} \frac{q(\beta^{(t-1)}|\beta^*)}{q(\beta^*|\beta^{(t-1)})}$
    Draw $u \sim \mathcal{U}_{(0,1)}$
    if $(u < \text{ratio})$ **then**
        Set $\beta^{(t)} = \beta^*$
    else
        Set $\beta^{(t)} = \beta^{(t-1)}$
    end if
else
    Set $\beta^{(t)} = \beta^{(t-1)}$
end if

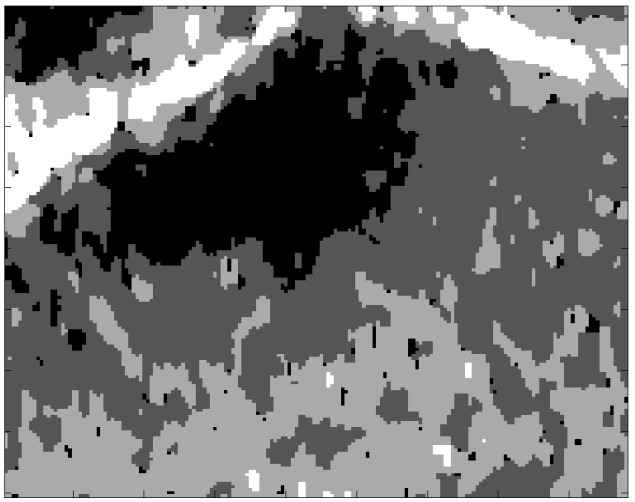
Experimental Results - in vivo data



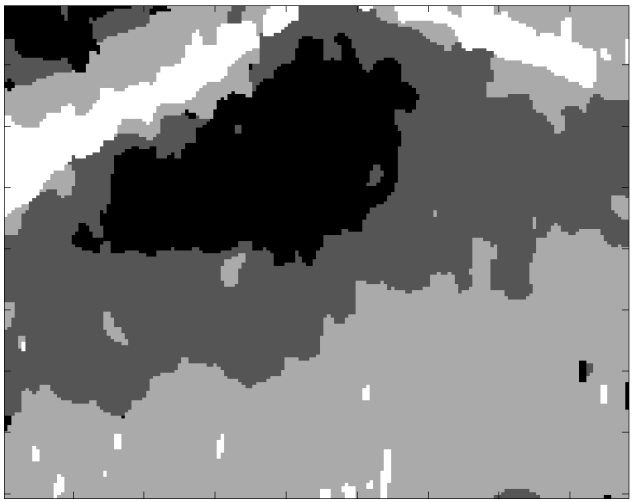
Margin. β ($\hat{\beta}_{MMSE} = 1.02$)



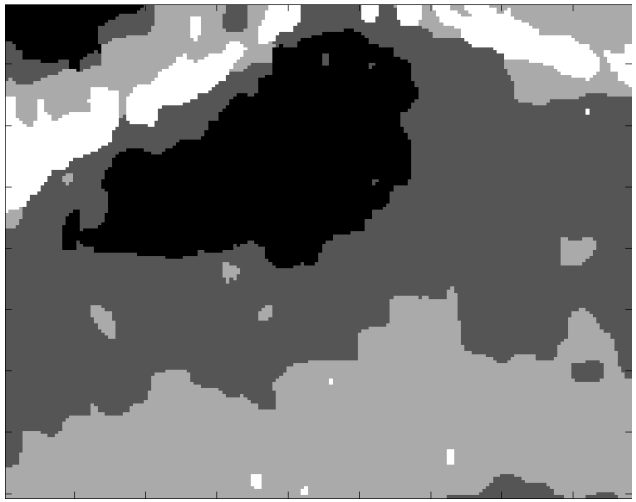
$\beta = 0.5$



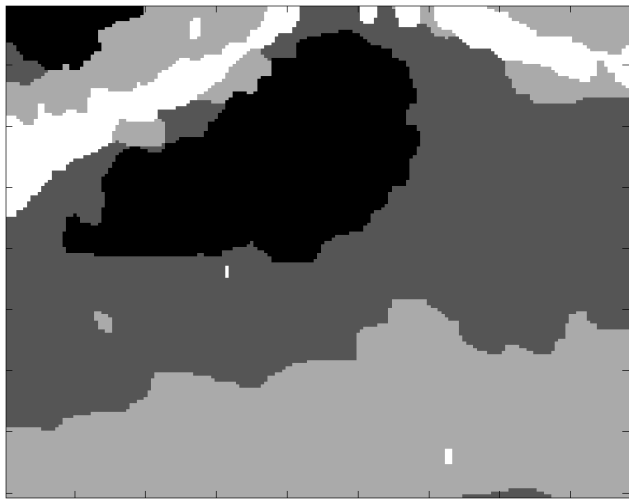
$\beta = 0.75$



$\beta = 1.0$



$\beta = 1.25$



$\beta = 1.5$

Outline

Conclusion

- The challenges facing modern imaging sciences require a methodological paradigm shift to go beyond point estimation.
- In Part I we discussed how the Bayesian framework can support this paradigm shift, provided we significantly accelerate computations.
- In Part II we considered efficiency improvements by integrating modern stochastic and variational computation approaches.
- In Part III we explored methods based on convex optimisation and probability, and developed theory for MAP estimation.

In this talk we studied, though an example, Bayesian models and computation algorithms for models that are more sophisticated than the ones previously considered, and where deterministic approaches fail.

Thank you!

Bibliography I

Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *J. Roy. Stat. Soc. Ser. B*, 72(3).

Besag, J. (1975). Statistical analysis of non-lattice data. *J. Roy. Stat. Soc. Ser. D*, 24(3):179–195.

Del Moral, P., Doucet, A., and Jasra, A. (2006). Sequential Monte Carlo samplers. *J. Roy. Stat. Soc. Ser. B*, 68(3).

Descombes, X., Morris, R., Zerubia, J., and Berthod, M. (1999). Estimation of Markov random field prior parameters using Markov chain Monte Carlo maximum likelihood. *IEEE Trans. Image Process.*, 8(7):945–963.

Gelman, A. and Meng, X. (1998). Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statistical Science*, 13(2):163–185.

Gelman, A. and Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Stat. Sciences*, 7(4):457–511.

MacNeil, S. (2007). Progress and opportunities for tissue-engineered skin. *Nature*, 445(1):874–880.

Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. (2011). Approximate Bayesian Computational methods. *Stat. Comput.*, 21(2):289–291.

Bibliography II

Marjoram, P., Molitor, J., Plagnol, V., and Tavar, S. (2003). Markov chain Monte Carlo without likelihoods. *Proc. Nat. Academy Sci.*, 100(26):15324–15328.

Moller, J., Pettitt, A. N., Reeves, R., and Berthelsen, K. K. (2006). An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. *Biometrika*, 93(2):451–458.

Morse, P. and Ingard, K. (1987). *Theoretical Acoustics*. Princeton University Press, Princeton (NJ).

Murray, I., Ghahramani, Z., and MacKay, D. (2006). MCMC for doubly-intractable distributions. In *Proc. (UAI 06) 22nd Annual Conference on Uncertainty in Artificial Intelligence*, pages 359–366, Cambridge, MA, USA.

Pereyra, M., Dobigeon, N., Batatia, H., and Tourneret, J.-Y. (2012a). Estimating the granularity parameter of a Potts-Markov random field within an MCMC algorithm. submitted to IEEE Trans. Image Process.

Pereyra, M., Dobigeon, N., Batatia, H., and Tourneret, J.-Y. (2012b). Segmentation of skin lesions in 2D and 3D ultrasound images using a spatially coherent generalized Rayleigh mixture model. *IEEE Trans. Med. Imag.* to appear.

Bibliography III

Pereyra, M. A. and Batatia, H. (2012). Modeling ultrasound echoes in skin tissues using symmetric α-stable processes. *IEEE Trans. Ultrason. Ferroelect. Freq. Contr.*, 59(1):60–72.

Raju, B. I. and Srinivasan, M. A. (2002). Statistics of envelope of high-frequency ultrasonic backscatter from human skin in vivo. *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, 49(7):871 –882.

Risser, L., Idier, J., Ciuciu, P., and Vincent, T. (2009). Fast bilinear extrapolation of 3D Ising field partition function. Application to fMRI image analysis. In *Proc. IEEE Int. Conf. Image Proc. (ICIP)*, pages 833–836, Cairo, Egypte.

Rohatgi, V. M. (1976). *An Introduction to Probability Theory and Mathematical Statistics*. Wiley-Interscience, New York (NJ).

Samorodnitsky, G. and Taqqu, M. (2000). *Stable Non-Gaussian Random Processes*. Chapman & Hall/CRC, New York (NJ).

Shankar, P. (2000). A general statistical model for ultrasonic backscattering from tissues. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 47(3):727–736.

Bibliography IV

Shankar, P. M. (2003). A compound scattering pdf for the ultrasonic echo envelope and its relationship to k and nakagami distributions. *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, 50(3):339 –343.

Shankar, P. M., Reid, J. M., Ortega, H., Piccoli, C. W., and Goldberg, B. B. (1993). Use of non-rayleigh statistics for the identification of tumors in ultrasonic b-scans of the breast. *Medical Imaging, IEEE Transactions on*, 12(4):687 –692.

Sun, Z. and Han, C. (2008). Heavy-tailed Rayleigh distribution: A new tool for the modeling of SAR amplitude images. In *Proc. (IGARSS 08). IEEE Int. Geosc. and Remote Sensing Symp.*, volume 4, pages 1253–1256.

Wagner, R. F., Smith, S. W., Sandrik, J. M., and Lopez, H. (1983). Statistics of speckle in ultrasound b-scans. *Sonics and Ultrasonics, IEEE Transactions on*, 30(3):156 – 163.

Wu, F. Y. (1982). The Potts model. *Rev. Mod. Phys.*, 54(1):235–268.