

STATISTICS & MCMC COMPUTATION OF VORONOI
TESSELLATIONS

Jesper Møller
Dept. of Mathematical Sciences
Aalborg University, Denmark

www.math.aau.dk/~jm

Aim of this talk

Fitting Voronoi tessellation models to actual data

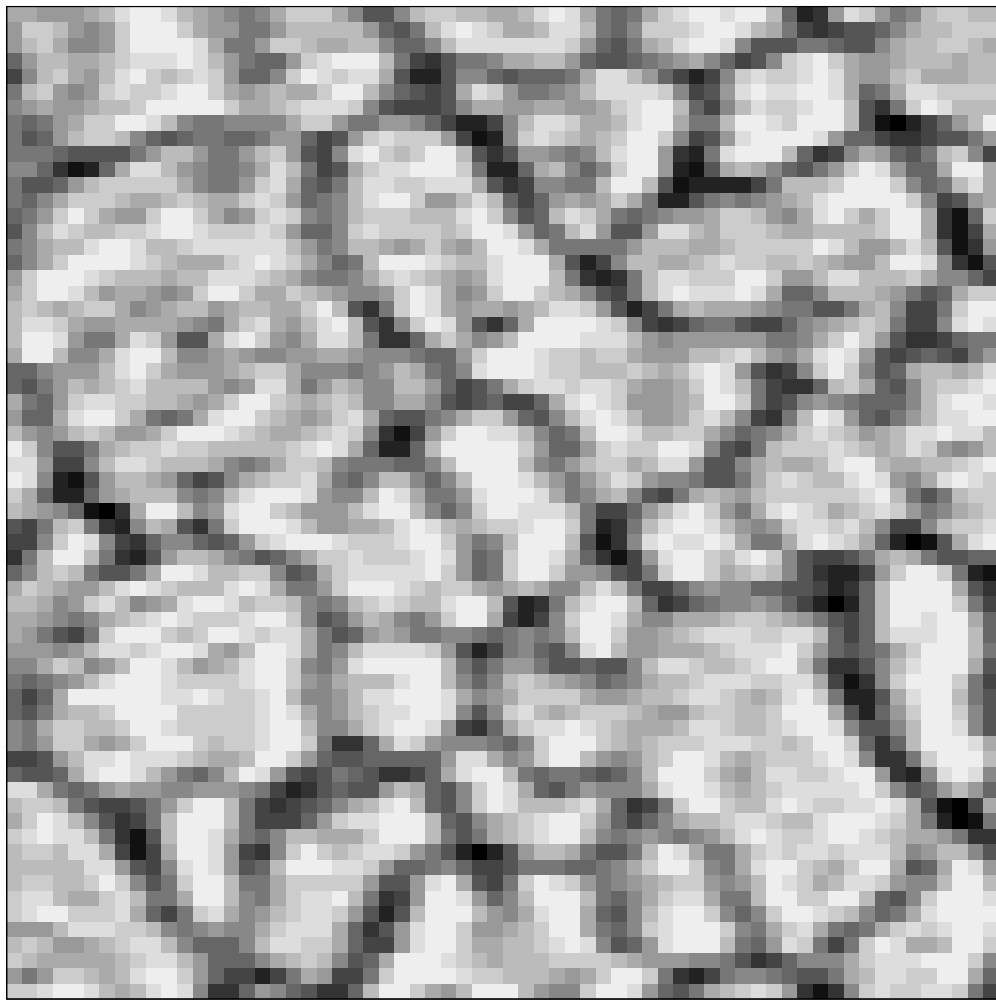
- using parametric statistical models
- and a simulation-based approach for inference.

In particular, using a Bayesian approach, we shall

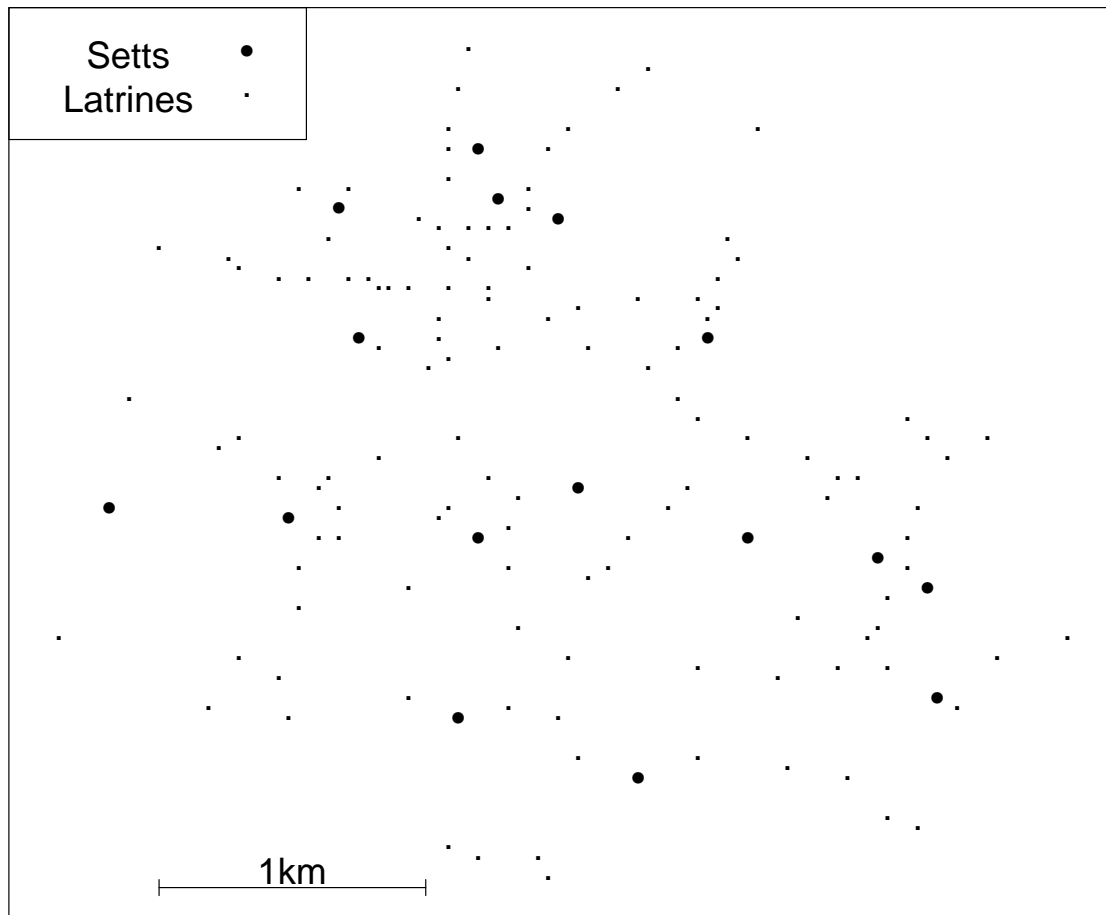
- reconstruct unobserved Voronoi tessellations
- indicate the uncertainty in the reconstruction.

Data examples...

A hidden tessellation in a noisy image



Badgers



Many application areas

Material science, ecology, reservoir modelling, astrophysics,

Happy message of this talk

Markov chain Monte Carlo (MCMC) methods provide useful computational tools for flexible (hierarchical) Bayesian model constructions, where

- one important ingredient is an underlying Voronoi tessellation,
- ‘noise’ can be modelled,
- and uncertainty can be quantified.

General setting...

Bayes theorem

posterior \propto prior \times likelihood

i.e.

$$\text{prob}(x, \theta | y) \propto \text{prob}(x, \theta) \times \text{prob}(y | \theta, x)$$

where

$y = \text{data}$, $x = \text{nuclei}$, $\theta = \text{other parameters (incl. 'errors')}$.

Markov chain Monte Carlo

- Generate a sample

$$(X_1, \Theta_1), \dots, (X_n, \Theta_n)$$

from a Markov chain with equilibrium distribution
= posterior distribution $\text{prob}(x, \theta|y)$;

- Estimate whatever property of the posterior distribution
(including the Voronoi tessellation) such as ergodic averages

$$\mathbb{E}[h(X, \Theta)|y] \approx \frac{1}{n} \sum_{i=1}^n h(X_i, \Theta_i)$$

and posterior modes (e.g. most likely tessellation).

General hierarchical model

Let $\theta = (\lambda, \psi)$ where

- $\lambda =$ intensity of nuclei, with *priors*

$$\text{prob}(\lambda) \quad \text{and} \quad \text{prob}(x|\lambda),$$

- naturally, the *likelihood* depends not on λ , so

$$\text{prob}(y|\theta, x) = \text{prob}(y|\psi, x).$$

Assuming *a priori* that (λ, x) and ψ are independent,
the *posterior* is

$$\text{prob}(x, \lambda, \psi|y) \propto \text{prob}(x|\lambda) \times \text{prob}(\lambda) \times \text{prob}(\psi) \times \text{prob}(y|\psi, x).$$

MCMC algorithm

Alternate between updating from the “*full conditionals*”:

- nuclei x given parameter (λ, ψ) and data y :

$$\text{prob}(x|\lambda, \psi, y) \propto \text{prob}(x|\lambda) \times \text{prob}(y|\psi, x)$$

- intensity λ given x, ψ, y :

$$\text{prob}(\lambda|x, \psi, y) \propto \text{prob}(x|\lambda) \times \text{prob}(\lambda)$$

- parameter ψ given x, λ, y :

$$\text{prob}(\psi|x, \lambda, y) \propto \text{prob}(\psi) \times \text{prob}(y|\psi, x).$$

- Metropolis-Hastings algorithm for each full conditional.
- Illustrated by various examples of applications from 3 papers.
- See these papers for algorithmic details.
- For instance, for the point process of nuclei, we use a reversible jump MCMC algorithm, along similar lines to Geyer & Møller (1994).

What is a Poisson process?

It is a

- stochastic process $X =$ point pattern in 1D, 2D, 3D,... space;
- defined by its intensity λ , which may depend on location s :
 $\lambda = \lambda(s)$;
- such that
 - for bounded regions A ,

$$N(A) \equiv \#(X \cap A) \sim \text{Poisson} \left(\int_A \lambda(s) \, ds \right)$$

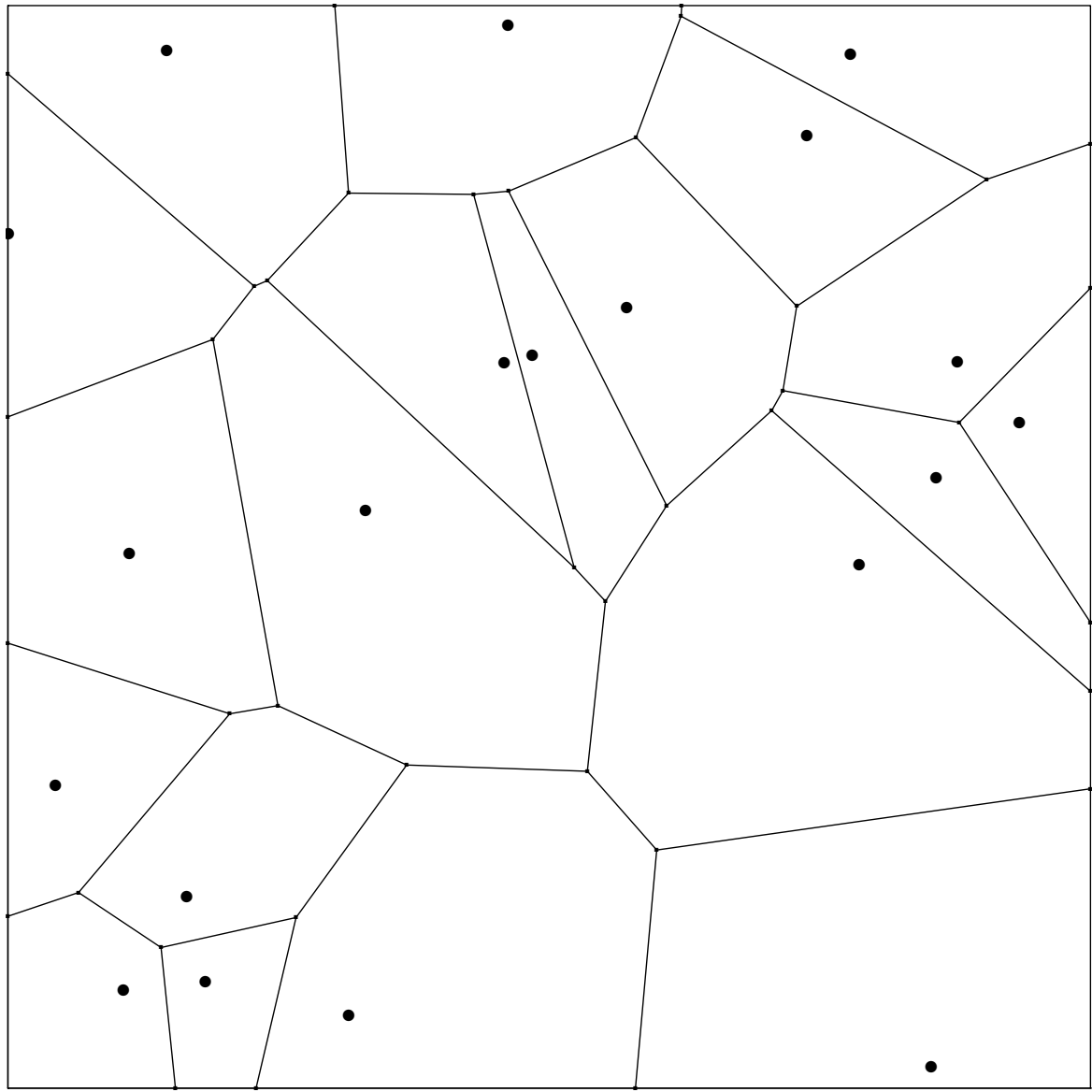
- for disjoint regions A and B ,

$N(A)$ and $N(B)$ are independent.

Paper 1

Bayesian analysis of deformed tessellation models

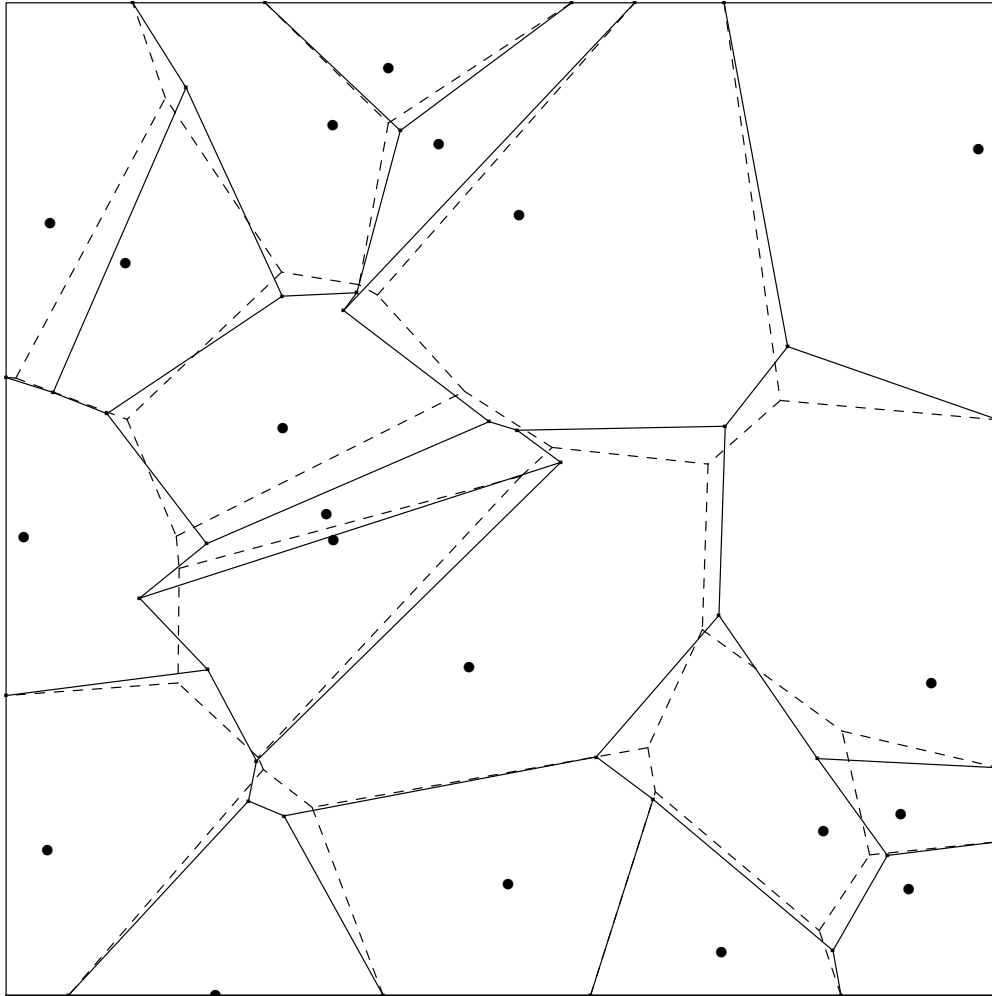
- Joint with Paul Blackwell.
- *Advances in Applied Probability*, 2003.



To *define a prior over Voronoi tessellations* on a finite planar region,

- define a Poisson process of nuclei over the region, with intensity $\lambda > 0$;
- take a Gamma prior for λ .

It is often useful to broaden the class of tessellation models by allowing (usually small) perturbations of individual vertices.



Deformed tessellation model

- Perturbations to vertices are independent, circular bivariate normal displacements, with unknown variance σ^2 .
- To ensure that this gives a new tessellation with edges in one-to-one correspondence with the old one, we condition on there being no extra crossings of edges.
- Further prior assumptions... For example, given λ , the variance σ^2 has an Inverse-Gamma distribution, with mean inversely proportional to λ , so that ‘shape’ is scale-invariant.

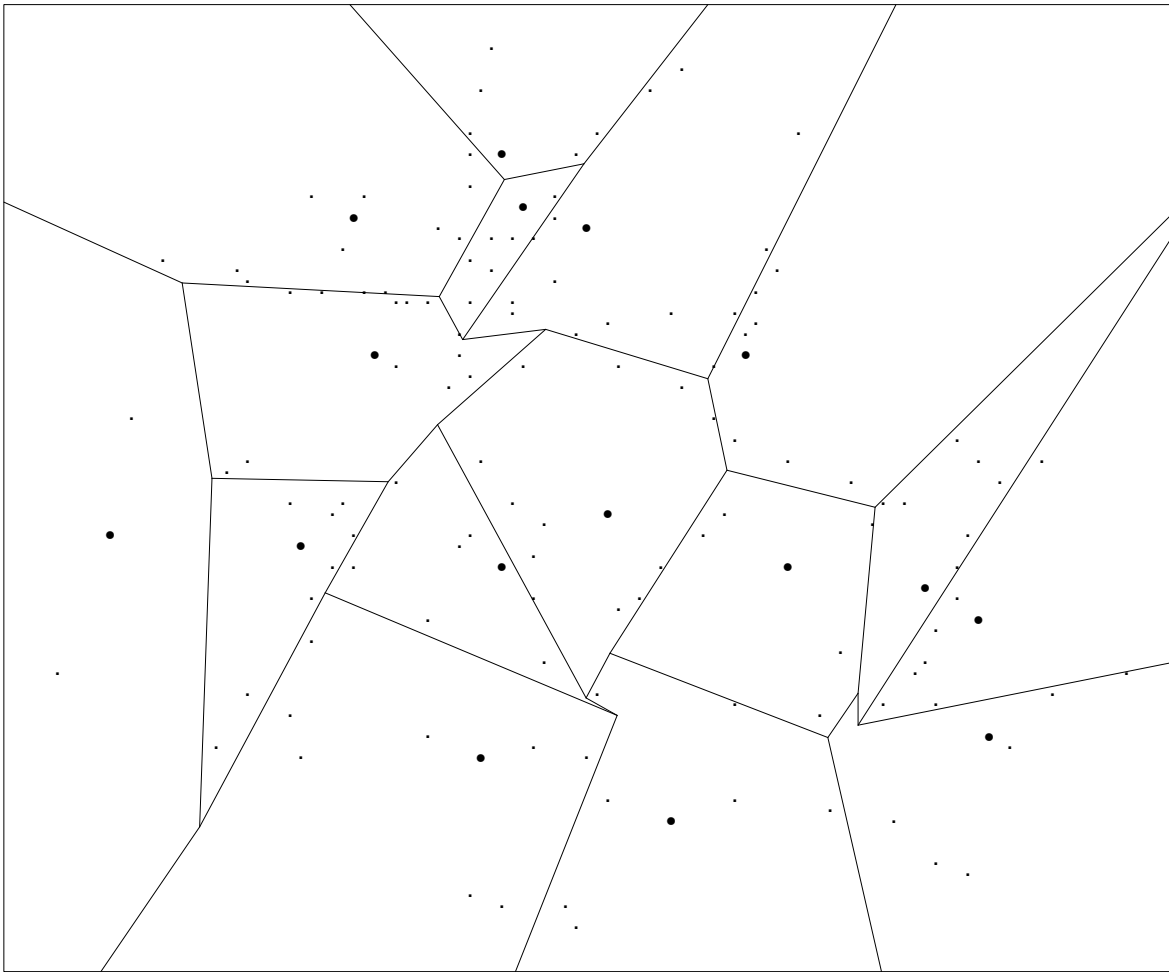
Likelihood

The deformed tessellation is observed ‘with error’. This can happen in various ways, through any random process dependent on the tessellation.

Specific examples:

- observe a realization of a point process with intensity dependent on position within a tile;
- observe a grey-scale image given by a binary image of the tessellation plus noise.

Badgers are territorial, and use latrines partly to mark boundaries.



Observed latrines

We can model the point pattern of observed latrines as forming a Poisson process with increased intensity close to a boundary: Let $\rho(s|E, \alpha, \beta, \gamma)$ be the intensity of latrines, where $E =$ union of the edges of the deformed tessellation, and α, β, γ are certain parameters...

Writing $y = \{y_1, \dots, y_k\}$ for the point pattern of badger latrines, the likelihood is

$$\text{prob}(y|E, \alpha, \beta, \gamma) = \exp\left(-\int \rho(s|E, \alpha, \beta, \gamma) ds\right) \times \prod_{y_i \in y} \rho(y_i|E, \alpha, \beta, \gamma).$$

Intensity of latrines

$$\rho(s|E, \alpha, \beta, \gamma) = \alpha + \beta \times g\left(\frac{d(s, E)}{2\gamma}\right)$$

where

$$g(d) = \begin{cases} 1 / (1 + (d/(1 - d))^2), & 0 \leq d \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

is a decreasing sigmoid function;

$\alpha > 0$ represents the background intensity of points;

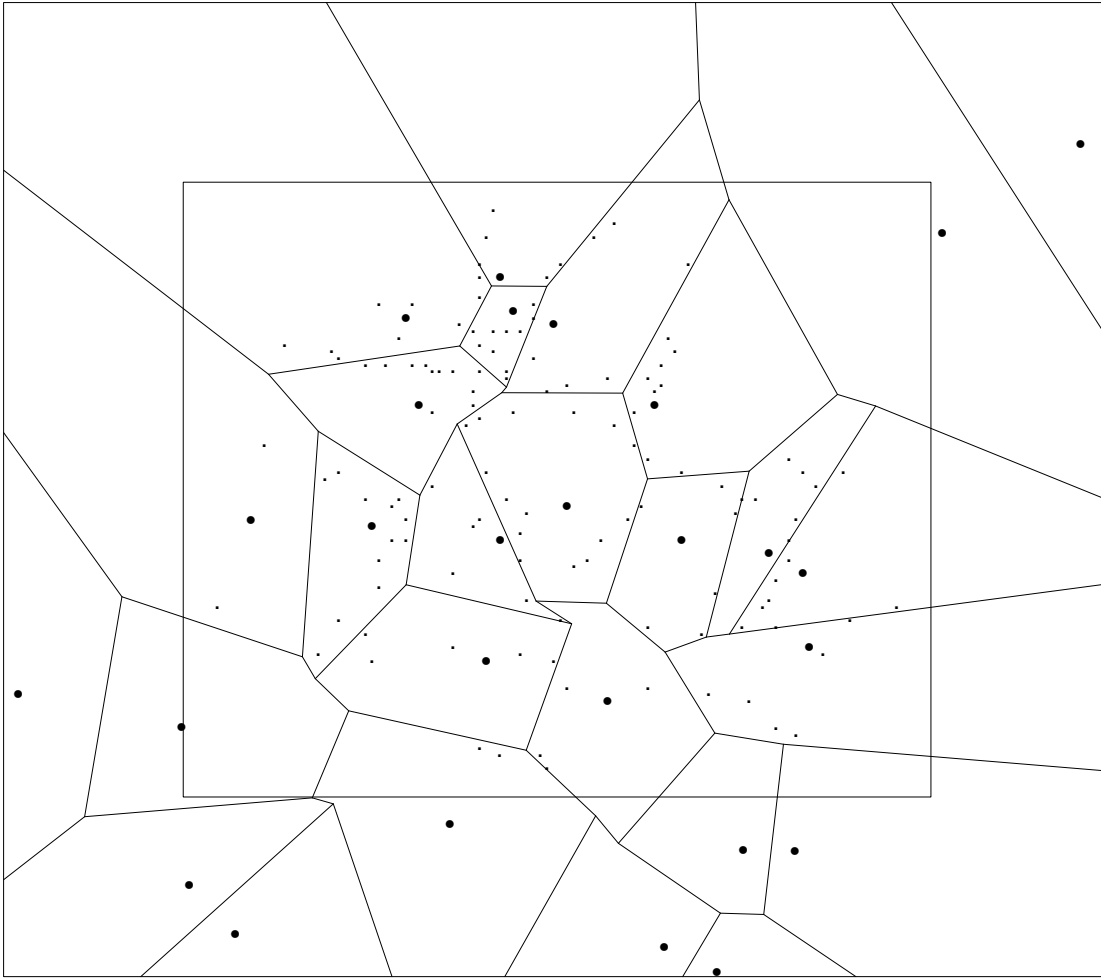
$\beta > 0$ represents the extra intensity at a boundary;

$\gamma > 0$ determines the scale over which the extra intensity decreases.

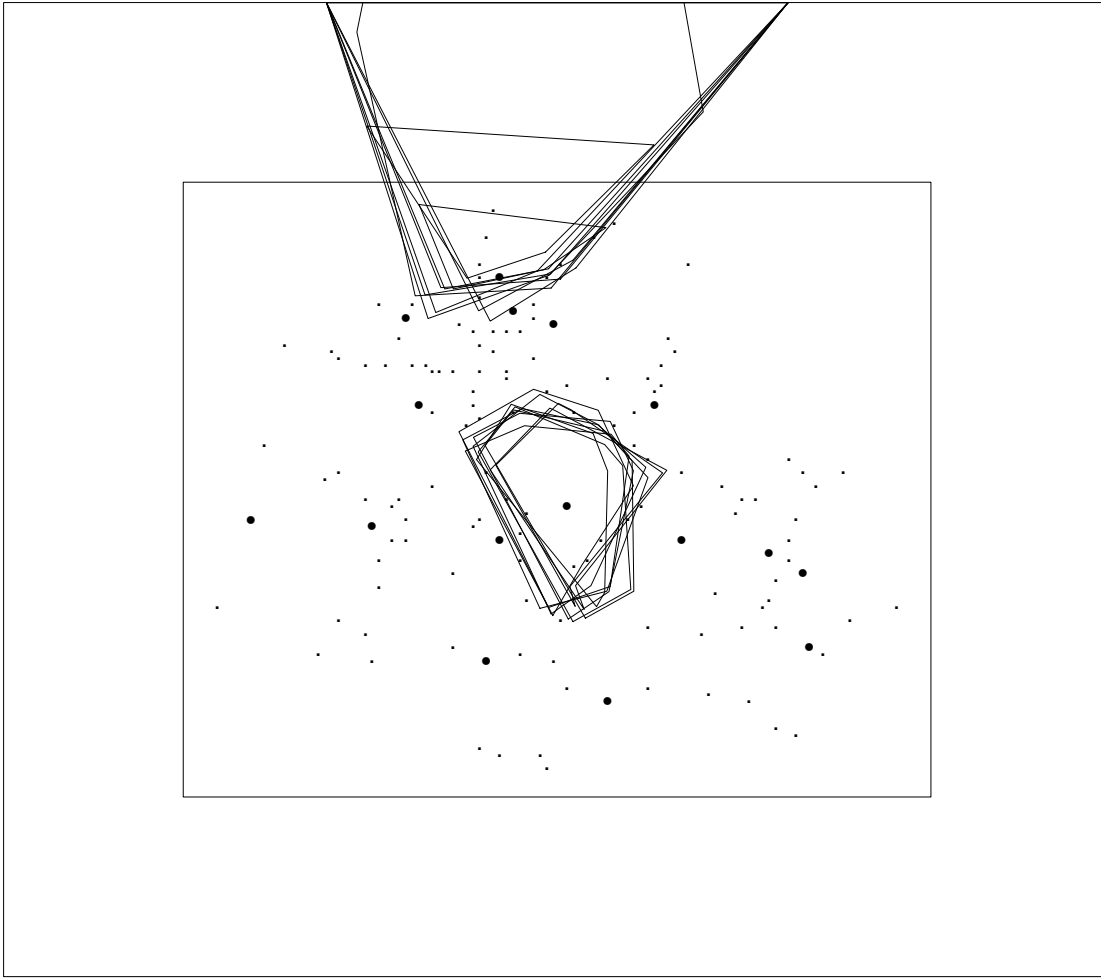
Badger setts and edge effects

- There is important additional information: badger setts are *also* observed, and for present purposes are taken to coincide with the nuclei.
- But there may also be unobserved nuclei (or badger setts) outside the observation window.
- We handle edge effects: *split region* into
 - an *inner part* (the observation window), with latrines observed and nuclei known,
 - and an *outer part*, with nuclei unknown and latrines unobserved.

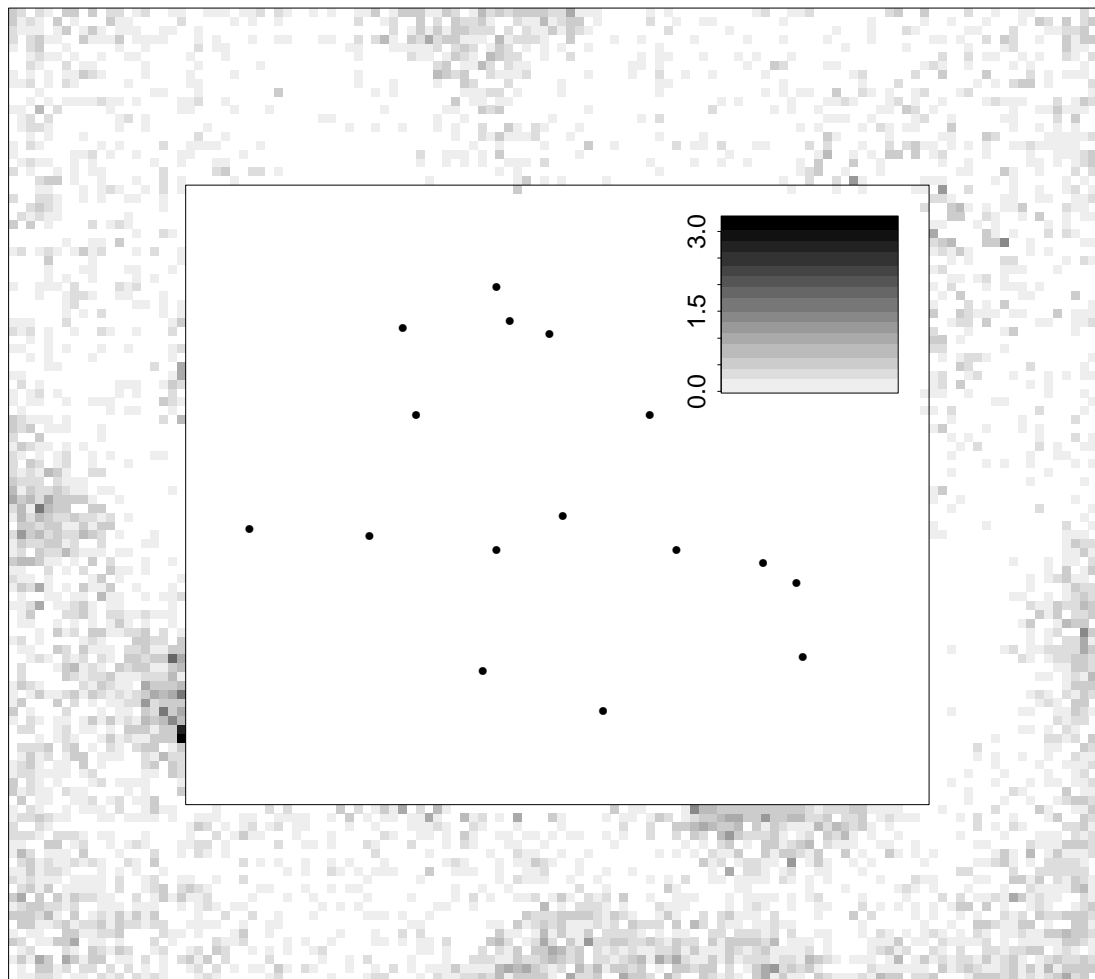
Typical reconstruction.



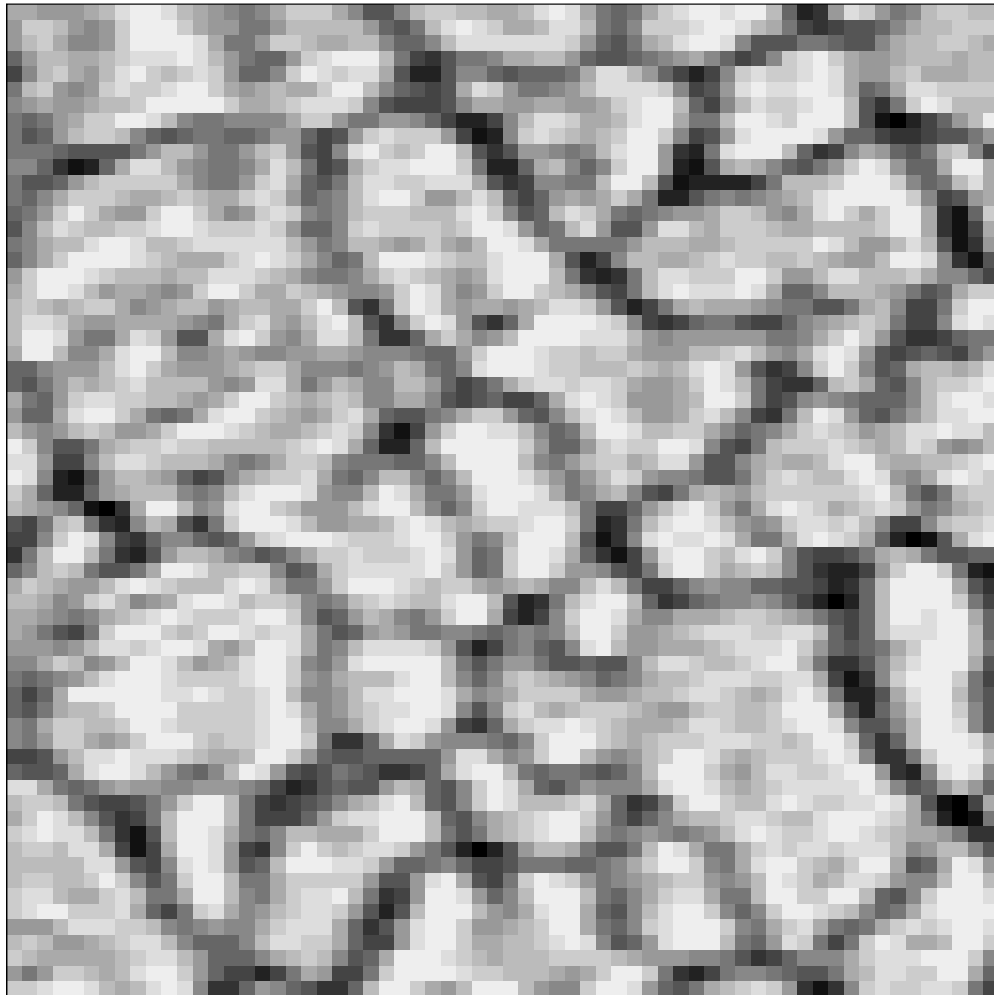
Posterior samples for two cells.



Posterior intensity surface for nuclei in outer region.



A cross-section of a metal shows a micro-crystalline structure at a very fine scale, with considerable noise in the grey-scale image.



Likelihood

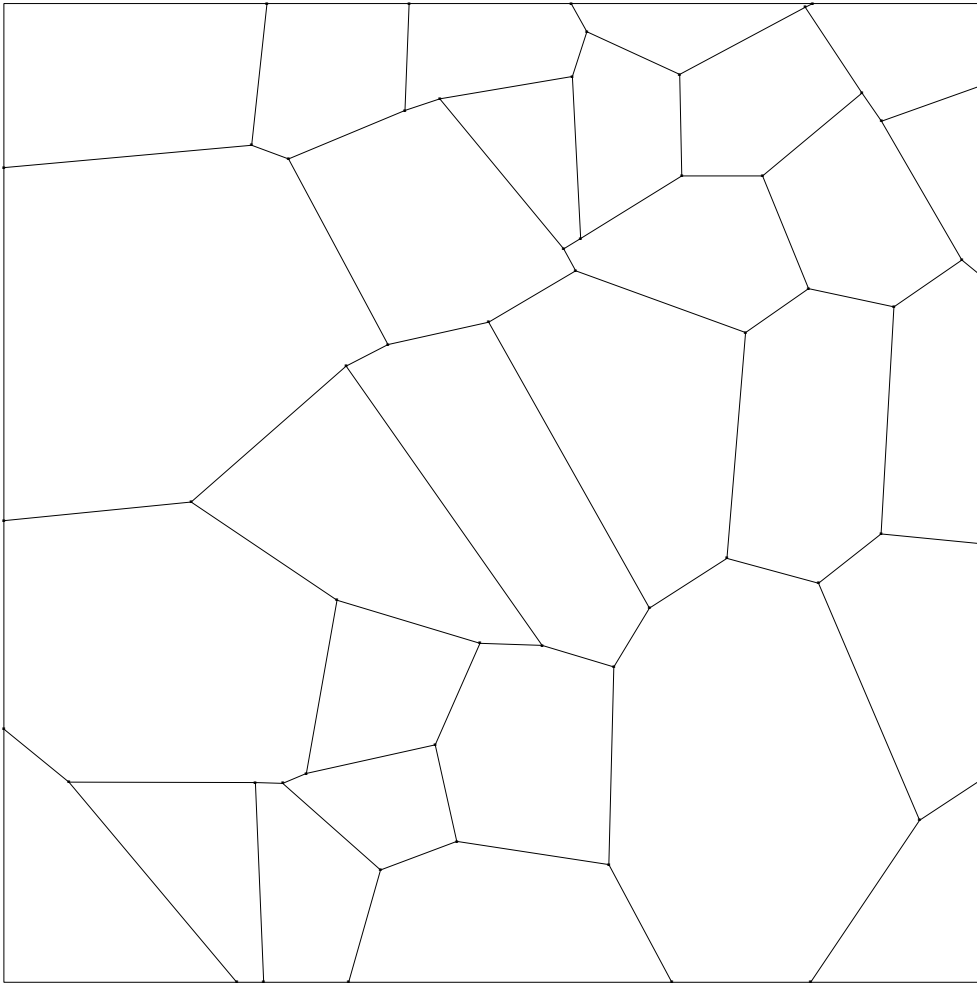
Letting Y_{ij} = grey-level at pixel (i, j) , the likelihood is

$$\text{prob}(y|E) = \prod_{ij} p_{I_{ij}(E)}(y_{ij})$$

where

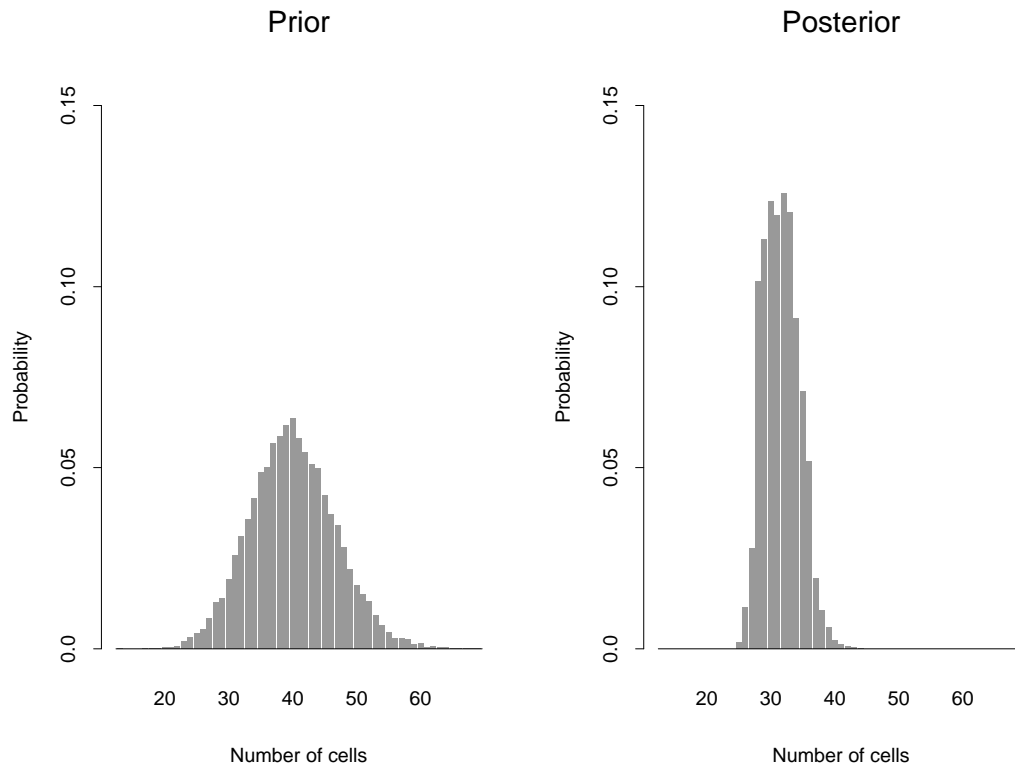
- $I_{ij}(E) \in \{0, 1\}$ is a discretized version of edges;
- $p_k(y_{ij})$ = probability of observing grey-level y_{ij} in a pixel which is ($k = 1$) or is not ($k = 0$) near a micro-crystal boundary;
- $p_k(\cdot)$ is obtained by smoothing with a locally linear ‘lowless’ algorithm (Cleveland, 1979).

Most likely reconstruction



- The results shown here are from an analysis which ignores boundary effects.
- Some obvious grains are identified, but there is considerable uncertainty in some parts of the image.
- There is also considerable uncertainty about the number of crystals making up the image, though the posterior is more concentrated than the prior.

Number of cells



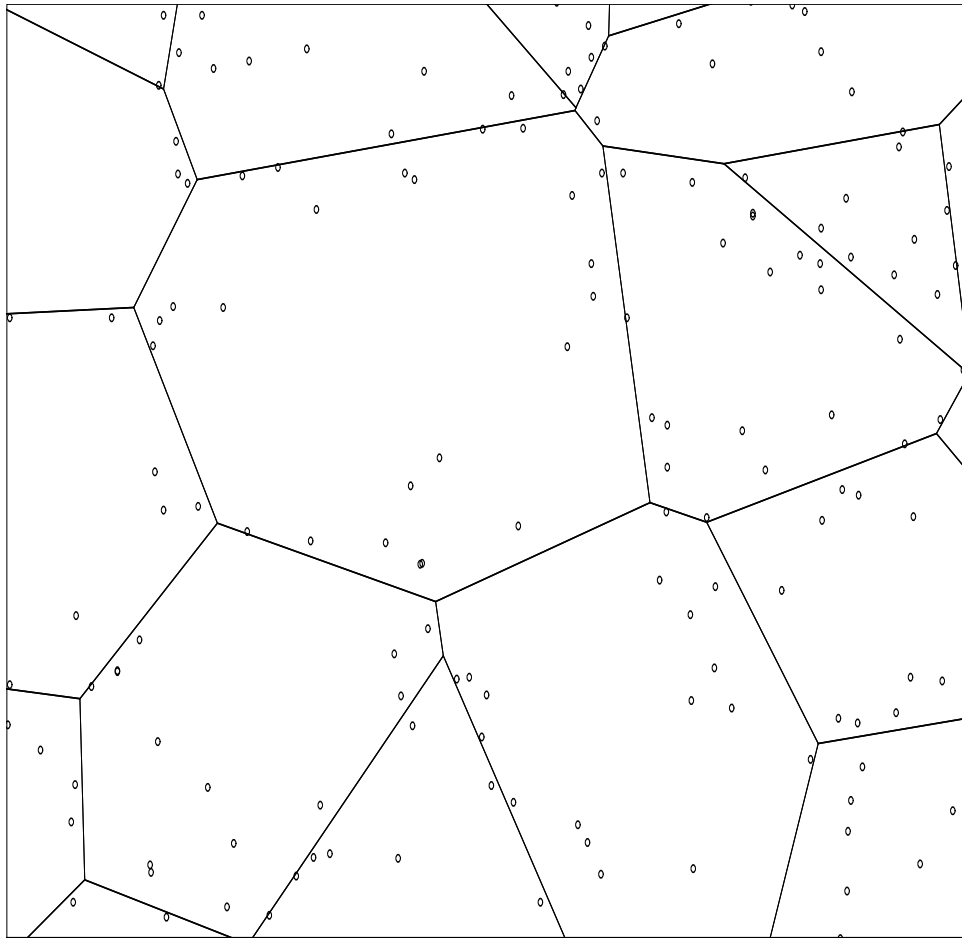
An empirical reconstruction (computational easy but no notion of uncertainty) has 40 cells, a value out in the upper tail of the posterior distribution.

Paper 2

Bayesian analysis of spatial point processes in the neighbourhood of Voronoi networks

- Joint with Øivind Skare and Eva B.V. Jensen.
- Submitted, 2006.

Final



Observation model

Given $E =$ union of edges of Voronoi tessellation on S generated by the nuclei in x ,

our data is $y_W = y \cap W$, where $W \subseteq S$ is an observation window and $y = \{y_i\}$ with

- $y_i = z_i + e_i$
- $z = \{z_i\}$ is a Poisson process on E with intensity $\rho > 0$
- iid errors $e_i \sim N_d(0, \sigma^2 I)$.

Likelihood

- $y|(E, \sigma^2, \rho)$ = Poisson process with an intensity function $\chi(s|E, \sigma^2, \rho)$ of closed form.
- If we could observe y (i.e. when $S = W$), the likelihood is

$$\text{prob}(y|E, \sigma^2, \rho) = \exp(-\rho|E|) \prod_{y_i \in y} \chi(y_i|E, \sigma^2, \rho)$$

where $|E|$ = length of edges.

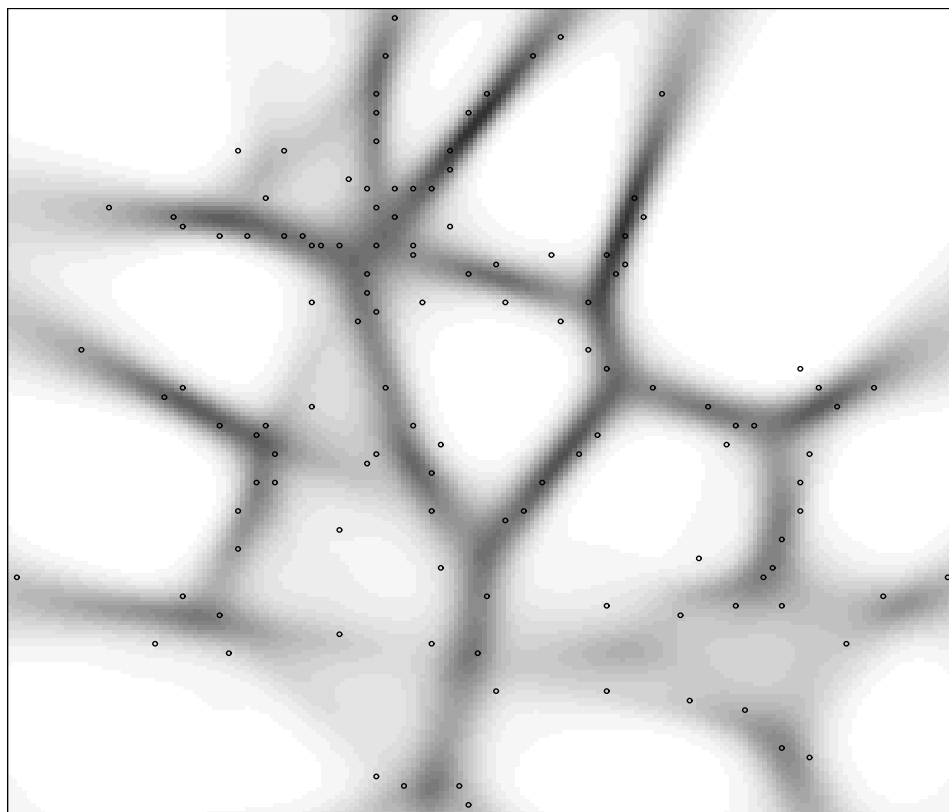
- However, as W is strictly included in S , the likelihood turns out to be intractable.
- Therefore, in the Bayesian analysis, we treat y on the outer region $S \setminus W$ as missing data (i.e. as an unknown parameter).

Prior

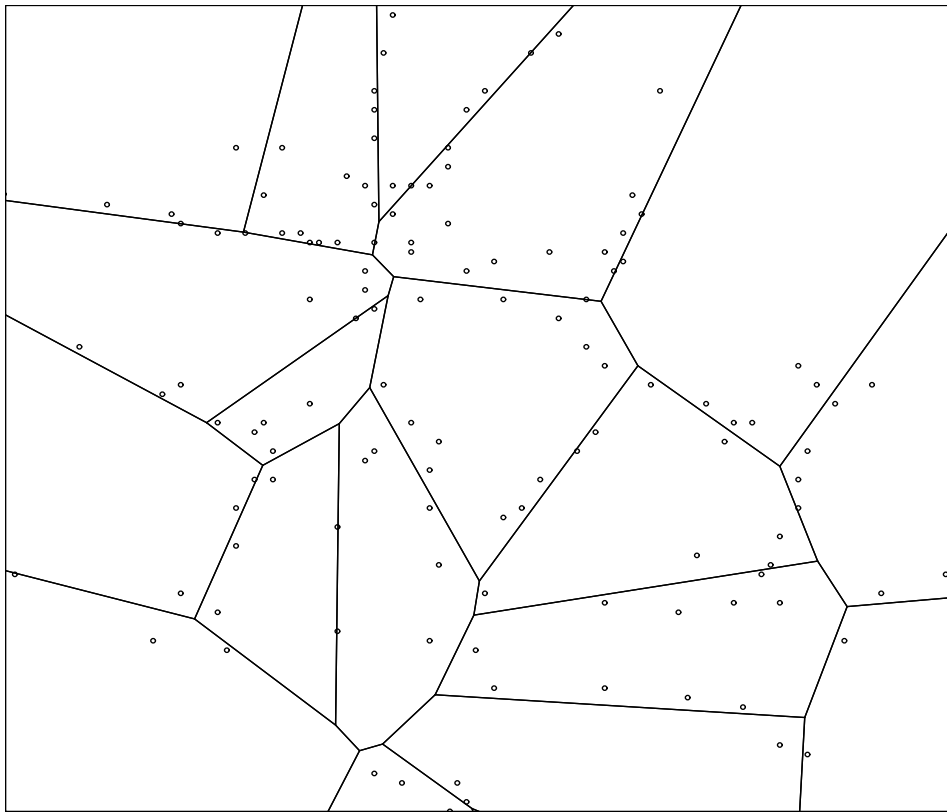
- Still $x|\lambda$ is a Poisson process of nuclei on some bounded region S in d -dimensional space.
- Prior assumptions on the intensity $\lambda > 0$ and other parameters...

Example: Badger latrines

Posterior edge intensity



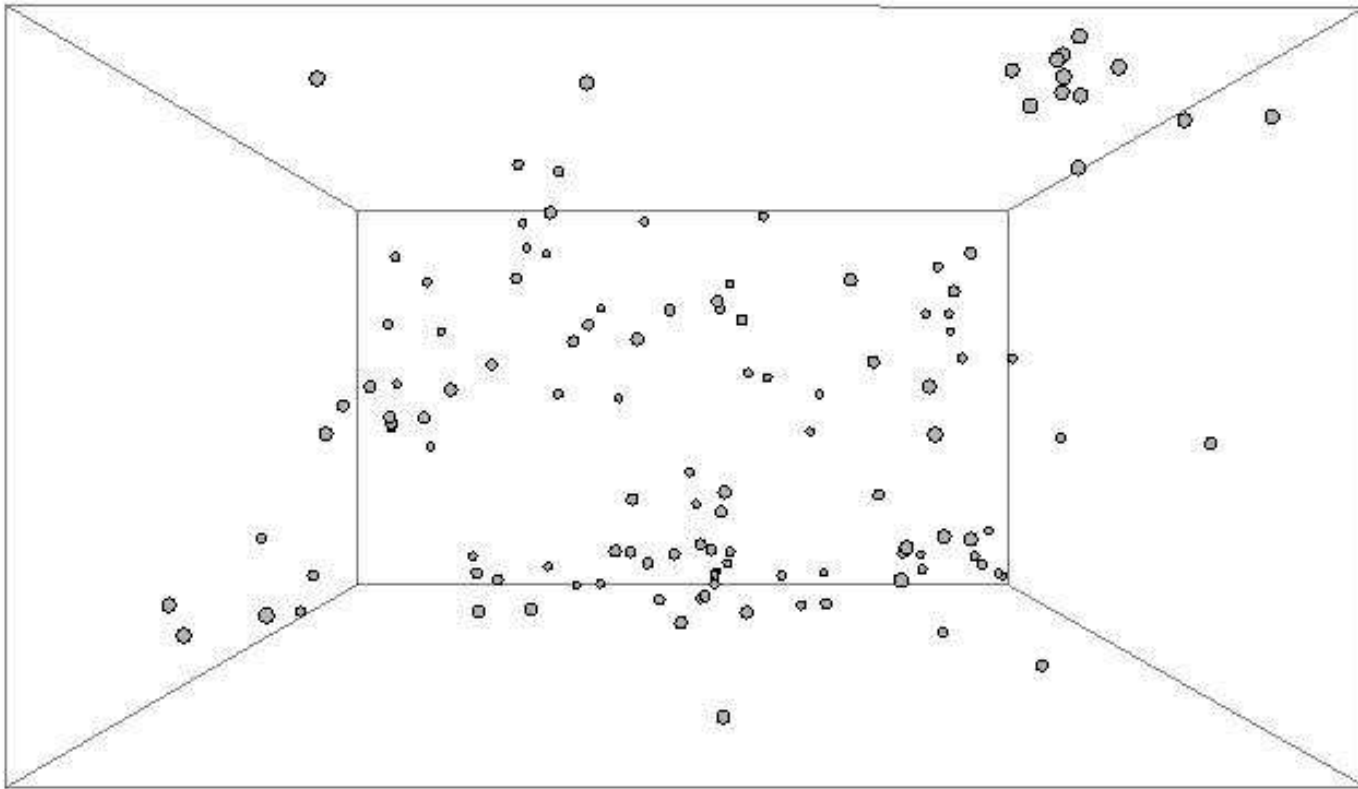
Most likely (Iteration 96112675)



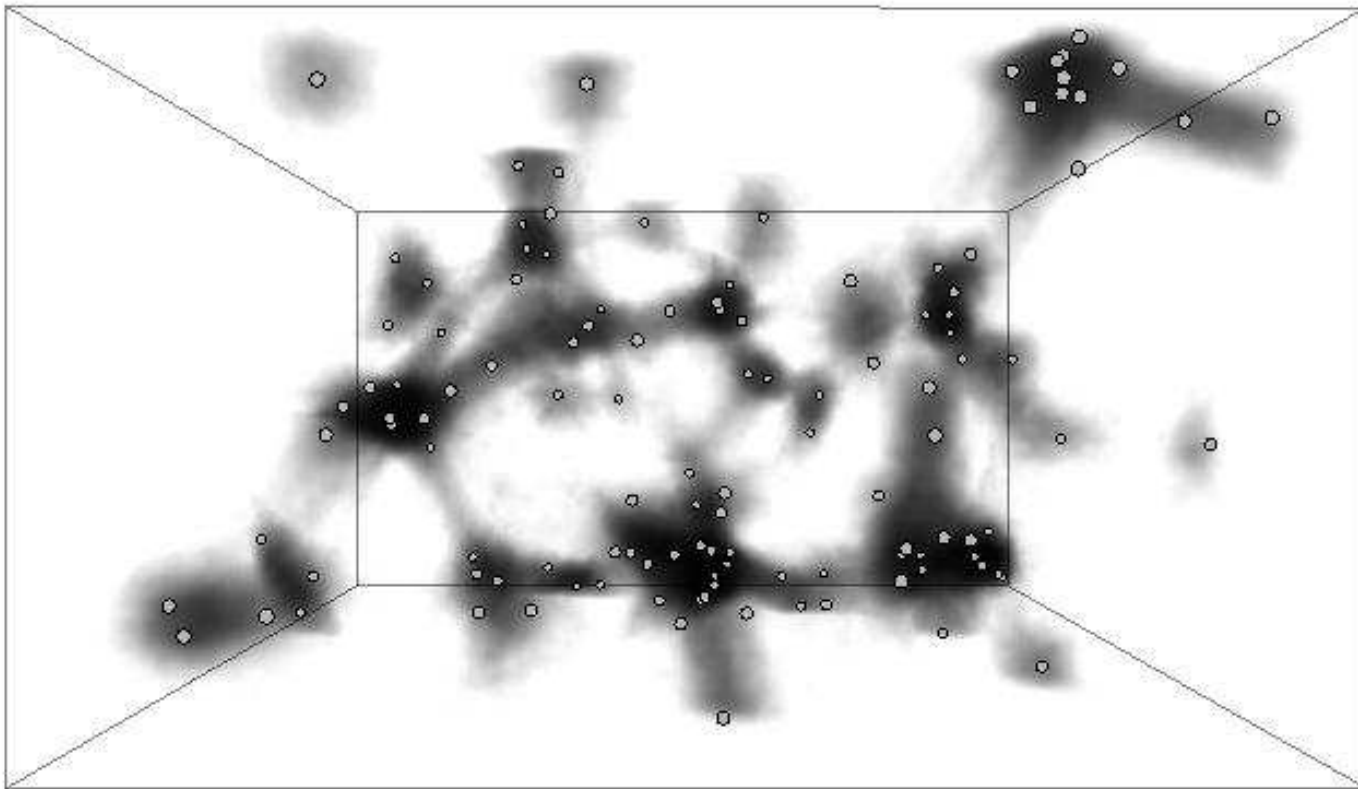
Example: Aluminum grain structure

Position of pores in aluminum (3D).

Supposed to be at the edges of aluminum grains.



Posterior pore intensity and observed pores.



Another application area: Cosmology

- Work in progress together with Rien van de Weygaert, Vicent Martinez, Øivind Skare and Eva B.V. Jensen.
- Data = various huge point patterns of galaxies.
- Similar model except for the intensity ρ of the Poisson process on Voronoi edges E : now, it is relevant to let $\rho = \rho(s|E)$ depend on the distance to nuclei defining the edges...
- Still the likelihood is tractable and we expect a Bayesian MCMC reconstruction to be feasible...

Paper 3

Bayesian image analysis with coloured Voronoi tessellations and a view to applications in reservoir modelling

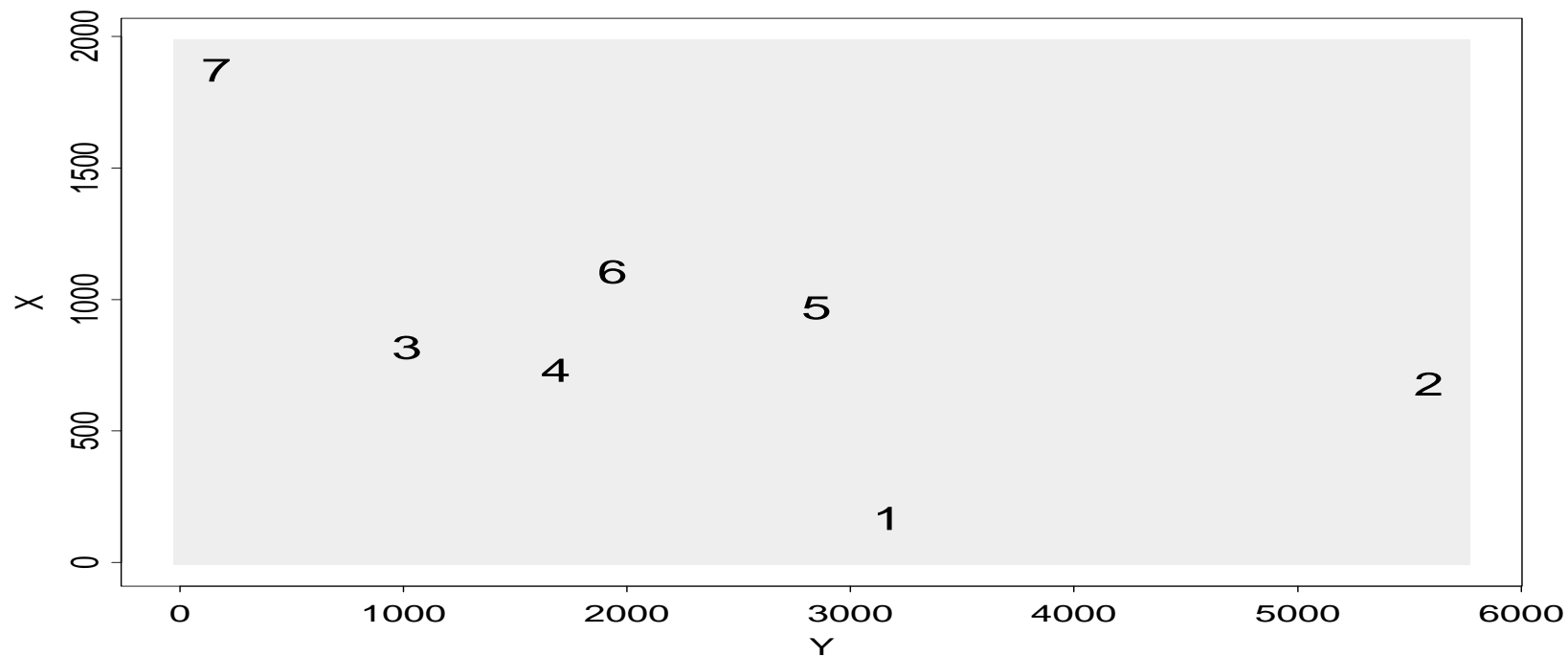
- Joint with Øivind Skare.
- *Statistical Modelling*, 2001.

Motivation and data

- Motivation: construct simple and flexible 2D and 3D priors in Bayesian image analysis and reservoir characterization.
- Data: typically very sparse or obtained by indirect observations.

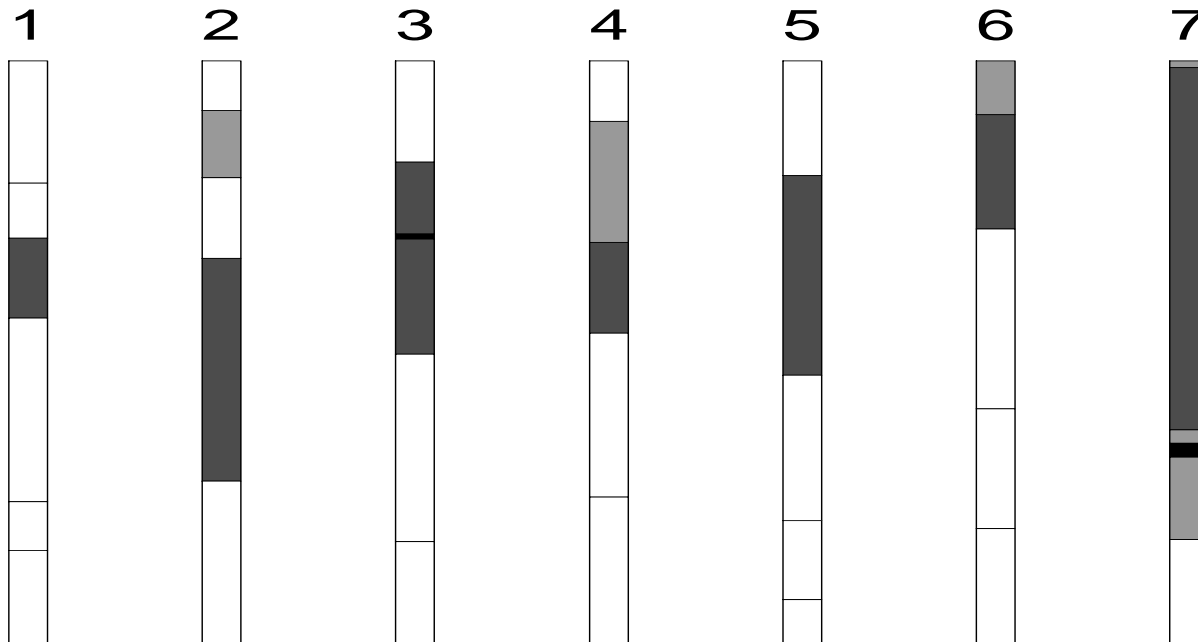
A reservoir modelling example

Reservoir of size $2000 \times 5800 \times 15 \text{ m}^3$ (somewhere in the North Sea). Observations at seven drilled wells.



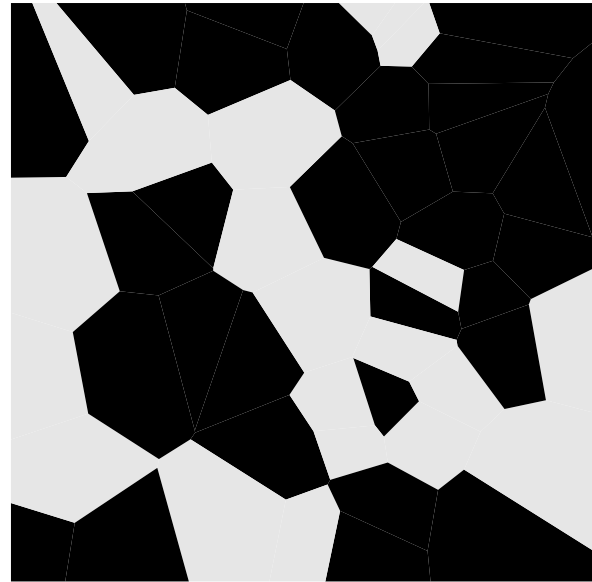
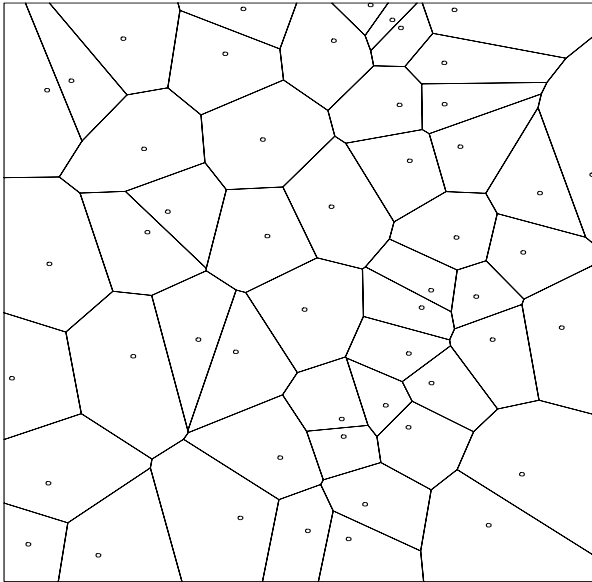
Data

Four different facies: shoal (dark shade), tidal channel sand (light shade), cemented facies (black) and background facies (white).



Coloured Voronoi tessellation

To each Voronoi cell with nucleus x_i we associate a mark/colour m_i .



Prior

$z = \{(x_1, m_1), \dots, (x_n, m_n)\}$ is a marked point process with density

$$f(z|\beta, \gamma) = \frac{1}{c(\beta, \gamma)} \beta^n e^{-\gamma s(z)}$$

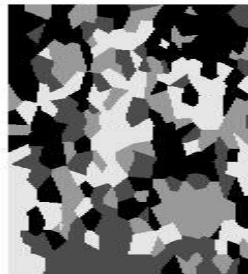
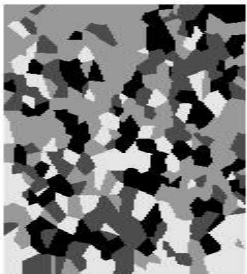
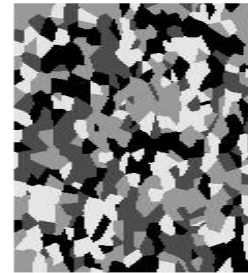
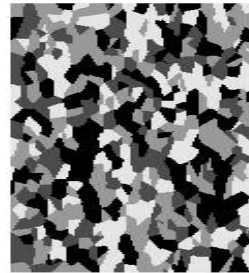
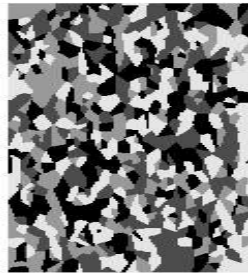
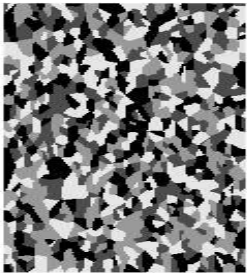
where

- $\beta > 0$ and $\gamma > 0$ are parameters,
- $c(\beta, \gamma)$ is a normalizing constant,
- $s(z) =$ number of pairs of neighbouring Voronoi cells of different colours.

Realizations from the prior

Space = unit square; 4 colours; $\beta = 1000$;

$\gamma = 0, .1, .2, .3, .4, .5, .6, .7$.

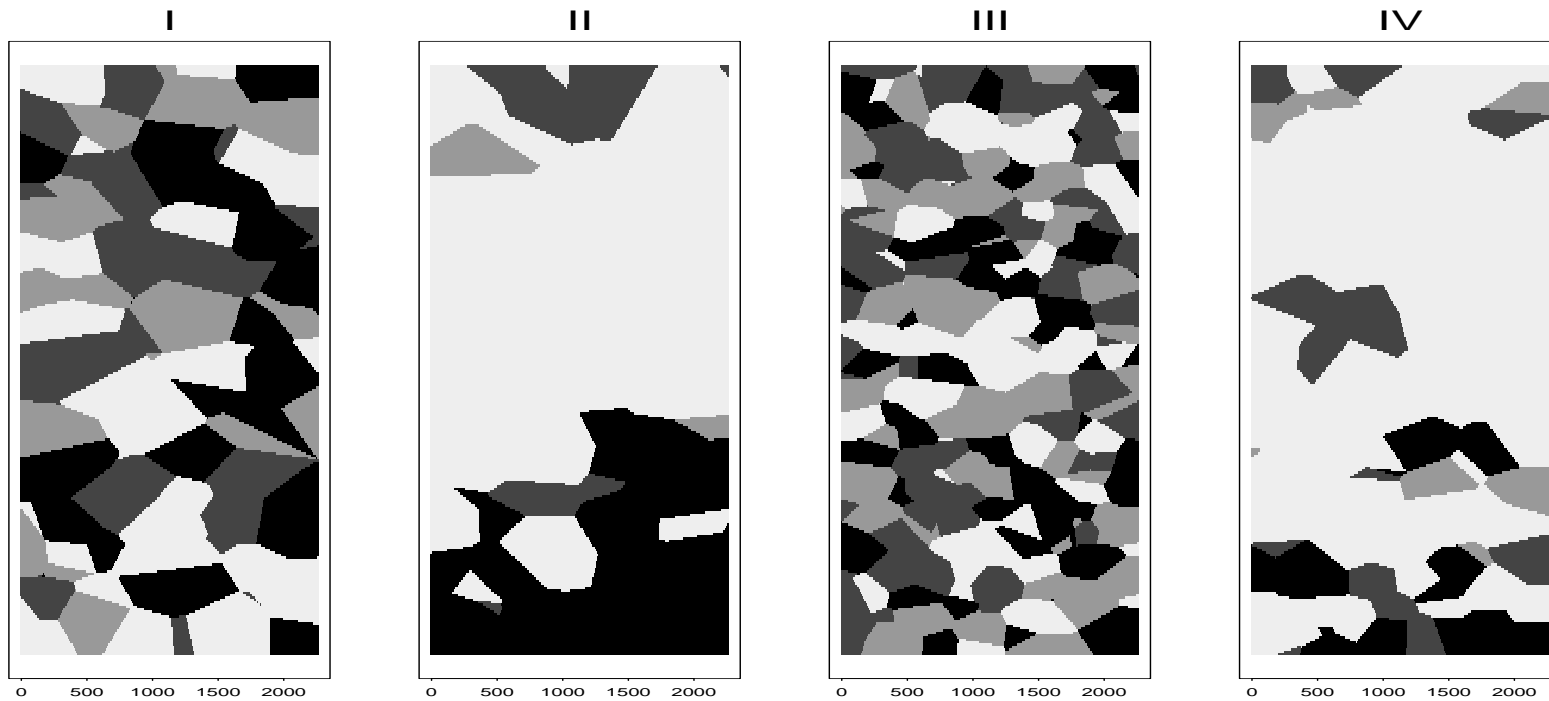


Likelihood

Terrible long story...

Posterior simulations = predictions

Central horizontal cuts through the reservoir for 4 different values of (β, γ) .



In conclusion

Voronoi tessellations provide

- flexible/realistic/illuminating models for many kinds of ‘noisy’ data (e.g. badger latrines, metal micro-crystalline structure) and images (e.g. reservoirs)
- using complex Bayesian model constructions
- where MCMC algorithms can be designed such that
 - Voronoi tessellations can be reconstructed
 - uncertainty can be quantified
 - and questions of scientific relevance can be answered.