Bayesian spatial prediction of weed intensities from exact count data and picture based indexes

Gilles Guillot
INRA, Applied Math. Dept, Paris France
Chalmers University

Joint work with
Niklas Lorén, SIK
Mats Rudemo, Chalmers University

Background on weed mapping

A precision agriculture context

An "operationnal context oriented" study

Previous approaches: weed counts only,
use of covariates (soil properties, remote sensing data),
ground pictures at small scale

A weed dataset

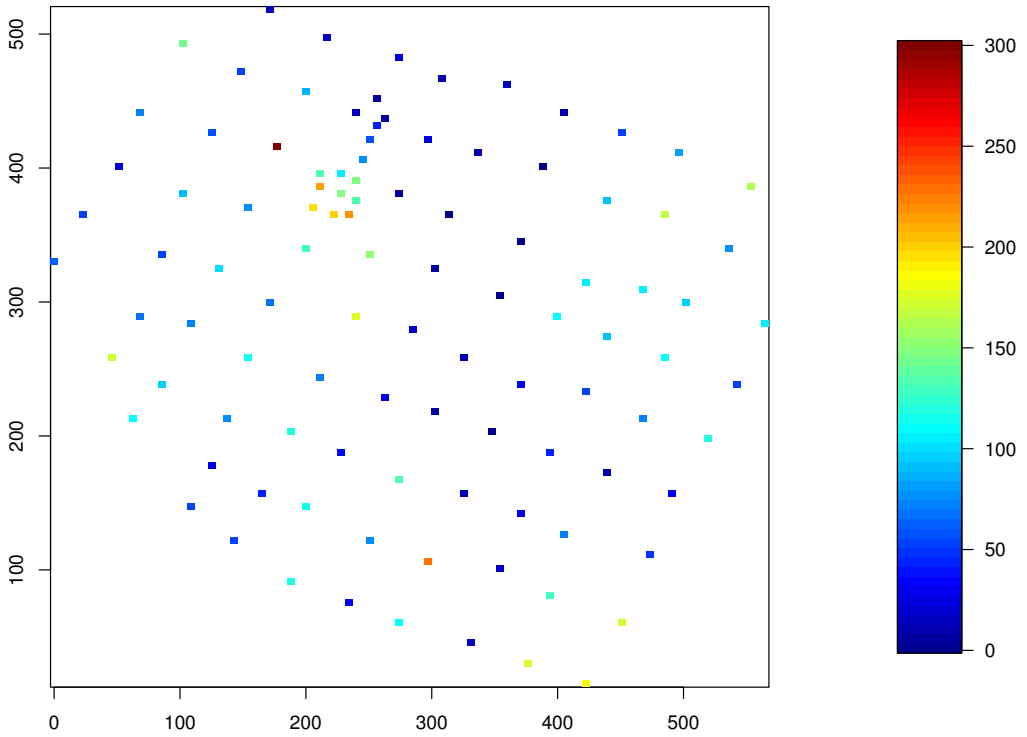
Data set collected on 15-16 may 2006 at the Bjertorp farm located 58.26°N-13.13°E

A 30 hectare field

100 measurement sites with exact counts

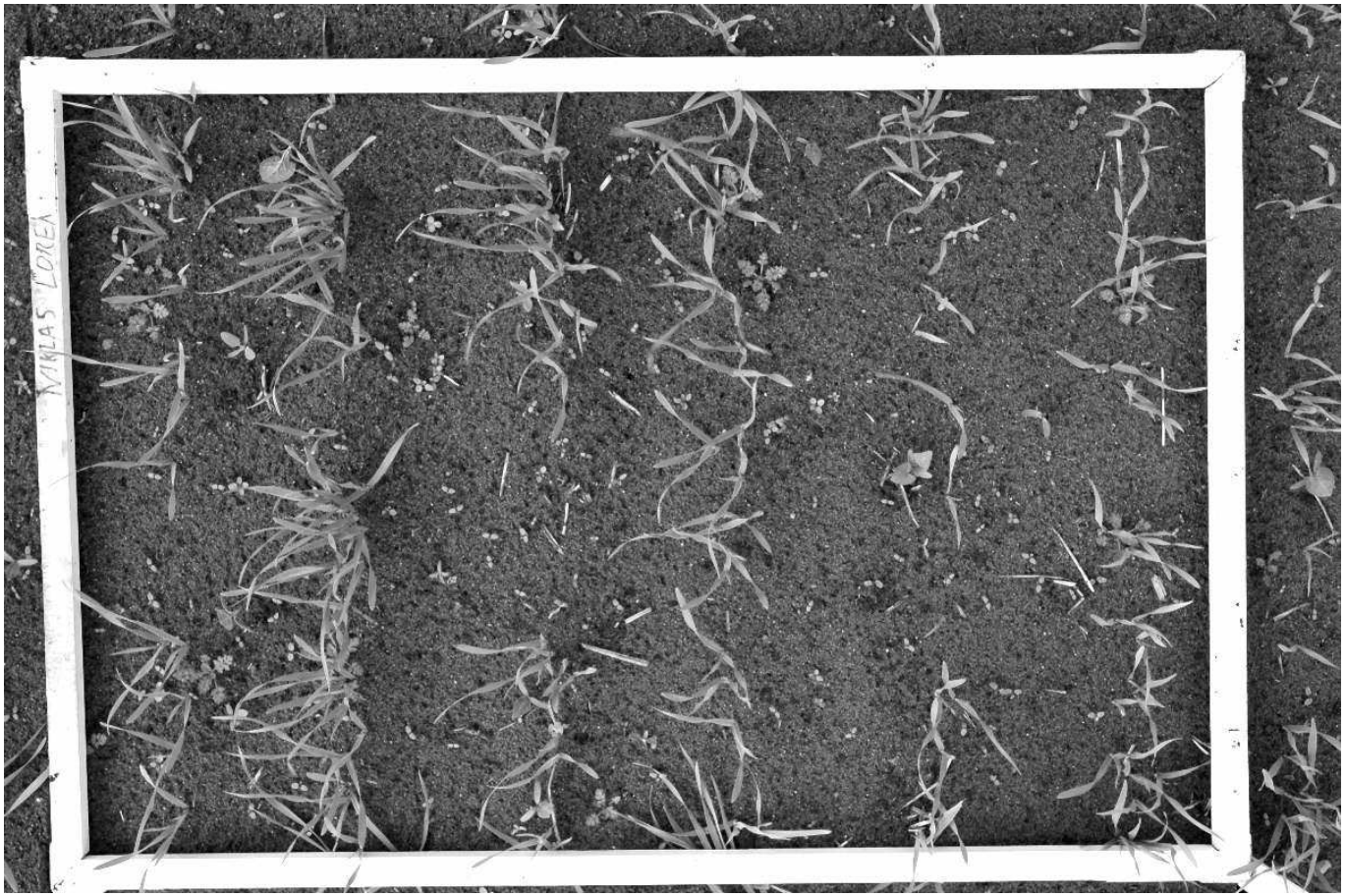
pictures over a frame of 0.5 m by 0.75 m

The exact weed counts dataset

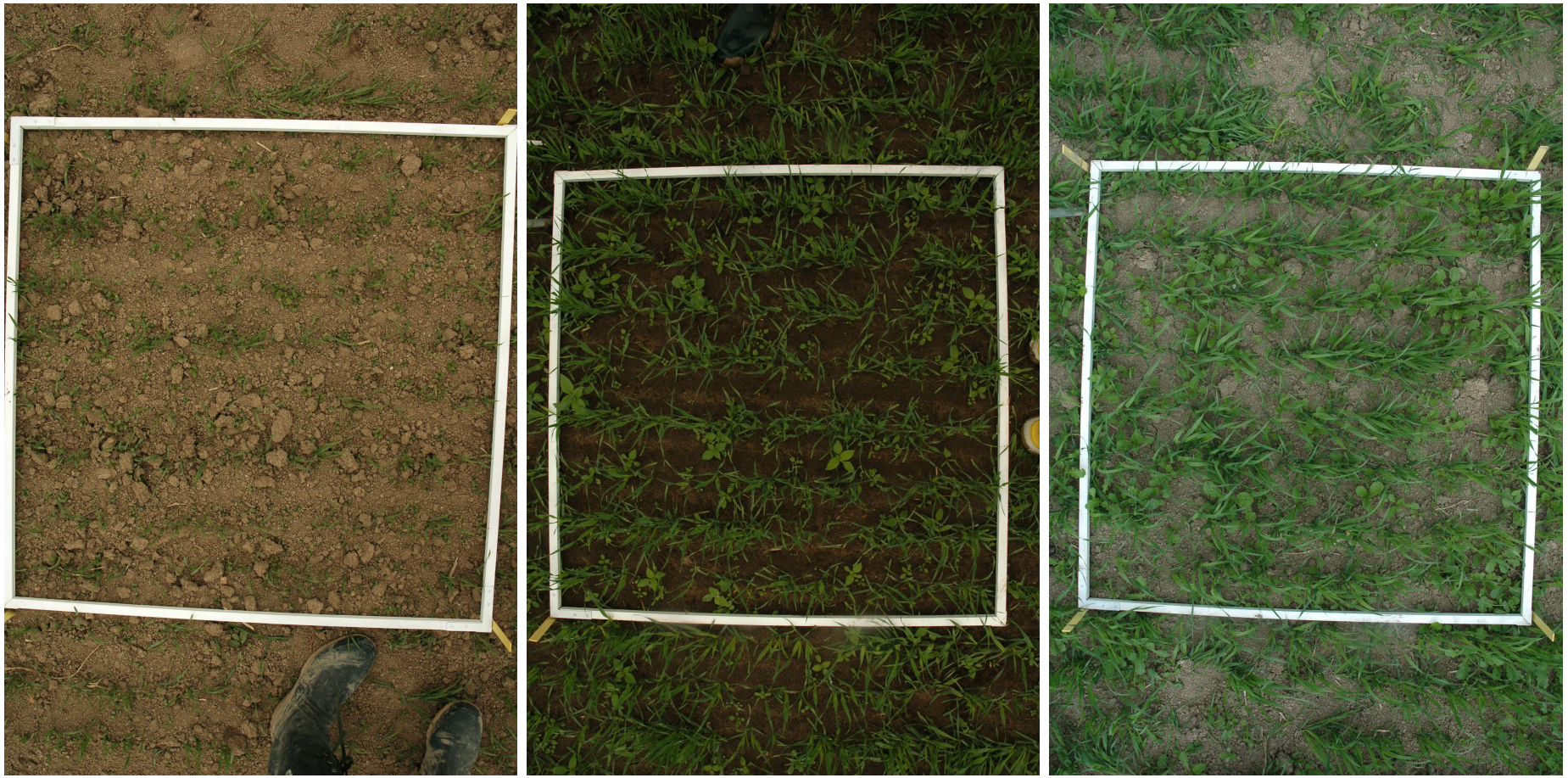


Display of exact weed count data. The color code stands for the number of observed plants.

Examples of picture from the 2006 dataset



Examples of pictures (from the 2005 dataset)



The image analysis algorithm

Resolution of 3008 $\times$ 2000 pixels

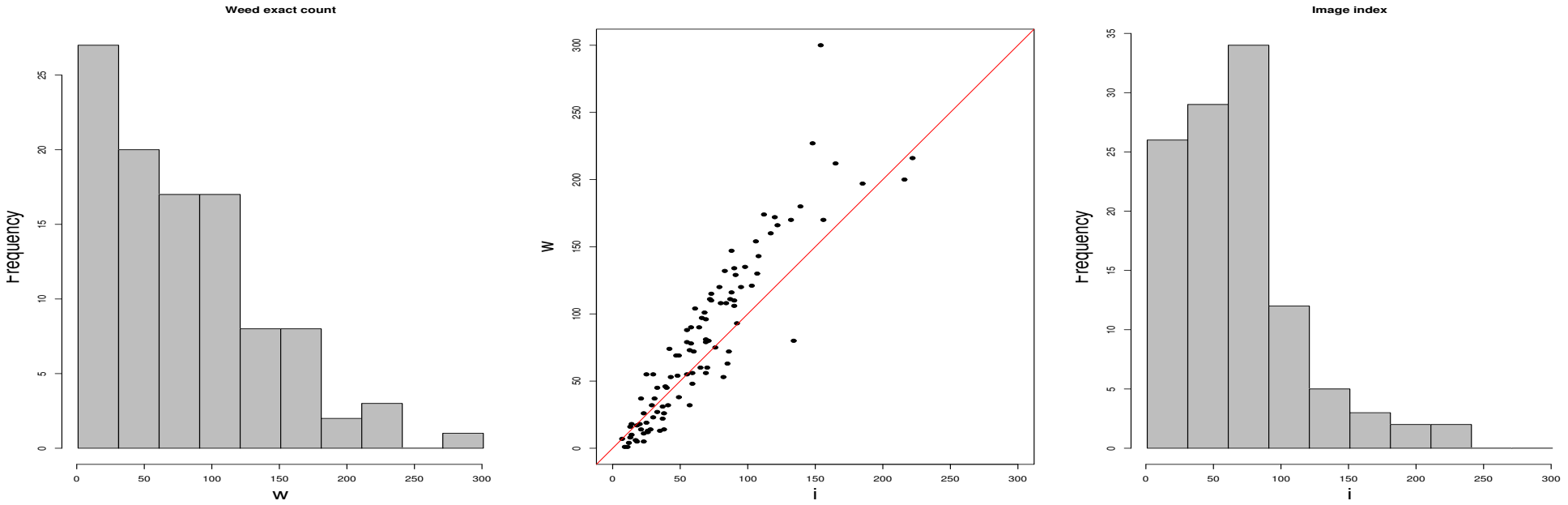
Segmentation of soil and plants using excess green transform

Remove small objects considered as noise

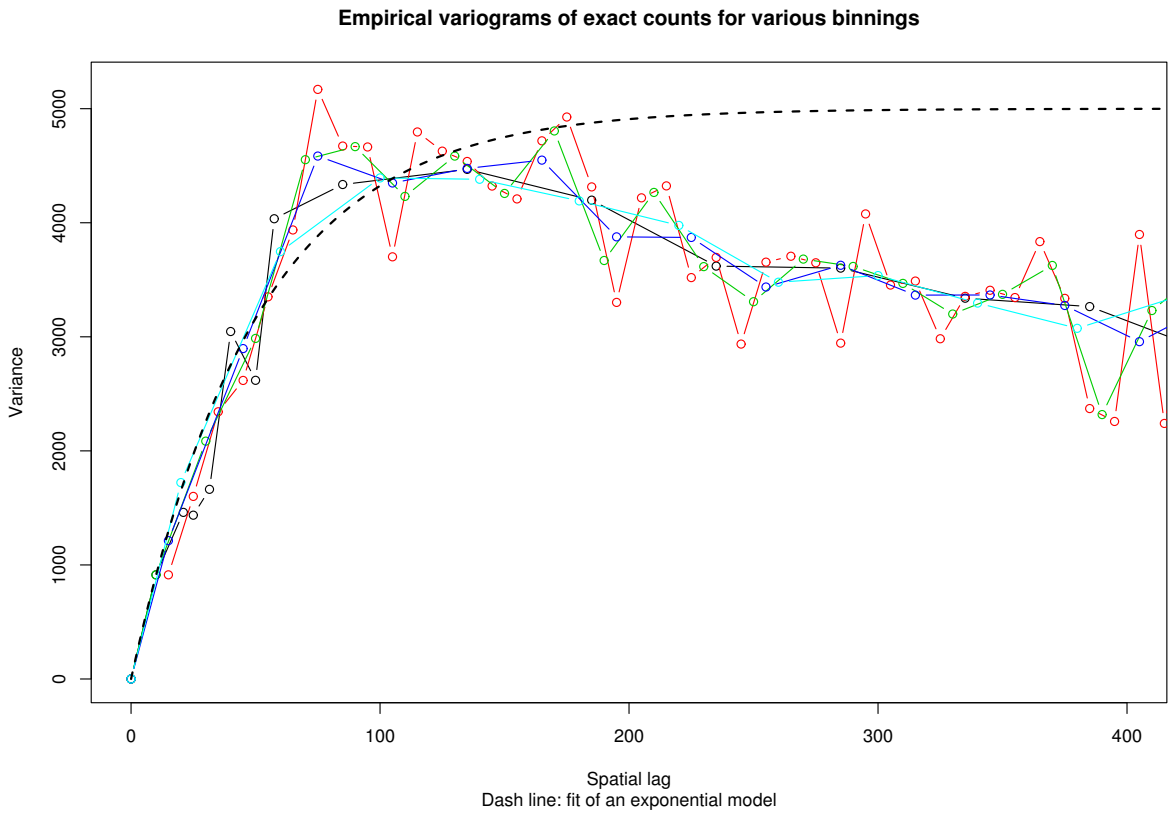
Hough transform to identify crop rows

Extract large weeds covered by crop straws using combinations of morphological operations

Statistical features



Empirical histogram of exact counts pair plot and histogram of and image indexes



Empirical variograms of the exact count data $w(s)$ for various binnings of the spatial lag, and theoretical fit with an exponential model (dashed line) with scale parameter equal to 50 m.

Desired features of the model

Goal: making spatial prediction from a small sample of exact weed counts + a large sample of images indexes

Take into account spatial correlation
Comply with the observed marginal distributions
Take into account the close relationship between w and i

Model (1): spatial variation of weed counts

We assume a transformed Gaussian model, i.e. there is a zero mean, centered Gaussian random field $y(s), s \in \mathbb{R}^2$ and a function $\phi : \mathbb{R} \longrightarrow \mathbb{R}$ such that $w(s) = \phi(y(s))$.

Accounts for non-Gaussian marginal distributions, while keeping the parsimony and flexibility of Gaussian random fields

We assume that y has a stationary and isotropic correlation function $\rho(h)$ with exponential decay, namely

$$Cor[y(s), y(s+h)] = \exp(-||h||/\kappa) \tag{1}$$

for some unknown spatial correlation parameter κ.

Model (2): marginal distribution of weed counts

The weed counts being positive and displaying an asymmetric histogram, we assume that the marginal distribution of w is Gamma, with shape parameter α and scale parameter β, namely: $\pi(w(s)|\alpha,\beta) \propto w(s)^{\alpha-1}e^{-\beta w(s)}$.

Assuming a Gamma distribution, $w(s)$ writes

$$w(s) = \phi(y(s)) = F^{-1}_{\alpha,\beta} \circ G(y(s)) \tag{2}$$

where $F_{\alpha,\beta}$ is the c.d.f of a Gamma(α,β) distribution and G is the c.d.f of a Normal$(0,1)$ distribution.

Poorer results with exponential, log-normal or Poisson distribution

Model (3): linking exact counts and picture indexes

Exact counts and image indexes are equal up to the blurring by a certain noise.
- Image values have to be positive whatever the values of the noise,
- Images covered by a large number of weeds more difficult to handle than those with a few weeds only.

Hence the multiplicative model:

$$i(s) = \lambda w(s)\varepsilon(s) \tag{3}$$

λ is a scaling factor

ε error with uncorrelated Gamma distribution with unit mean and variance τ.

Bayesian specification

Likelihood level

$w(s) = F^{-1}_{\alpha,\beta} \circ G(y(s)),$ $\qquad Cor[y(s), y(s+h)] = \exp(-||h||/\kappa)$

$y \sim GRF(0, \rho)$

$i(s) = \lambda w(s) \varepsilon(s)$ $\qquad \varepsilon(s) \overset{\text{i.i.d}}{\sim} \text{Gamma}(1/\tau, 1/\tau)$

5 unknown scalar parameters: $\theta = (\kappa, \alpha, \beta, \lambda, \tau)$

Prior: Independent flat Gamma prior on the components of θ

Connection with other approaches

Point process based models (Møller et al. 1998, Brix and Chadœuf 2001, Brix and Møller 2001): appealing from a theoretical and an ecological point of view, but interest not assessed for a prediction purpose.

Approximation of a log-Gaussian Cox process by a SGLMM (Christensen and Waagepetersen, 2002): same appealing feature as above but still complies better with discontinuous trajectory

Simulation based inference and prediction (1): data augmentation

If $D = (w_s, i_t)$ and $P = w_u$, simulation and inference carried out jointly by sampling from $\pi(P, \theta|D)$

Involves evaluating

$$\pi(i_t|w_s, \theta) = \int_{w_t \in \mathbb{R}_+^{n_t}} \pi(i_t|w_t, w_s, \theta) d\pi(w_t|\theta) \quad (4)$$

Data augmentation: make prediction of w also at sites t
P becomes $= (w_t, w_u)$
and we sample from $\pi(P, \theta|D) = \pi(w_t, w_u, \theta|D)$ instead of $\pi(w_u, \theta|D)$

Simulation based inference and prediction (2): reparameterisation

Single site Metropolis Hastings update of w_t do not work

Metropolis Adapted Langevin Algorithm (MALA) often recommended

Proposed alternative:
In the our model, define $\mathbf{w} = (w_s, w_t, w_u)$

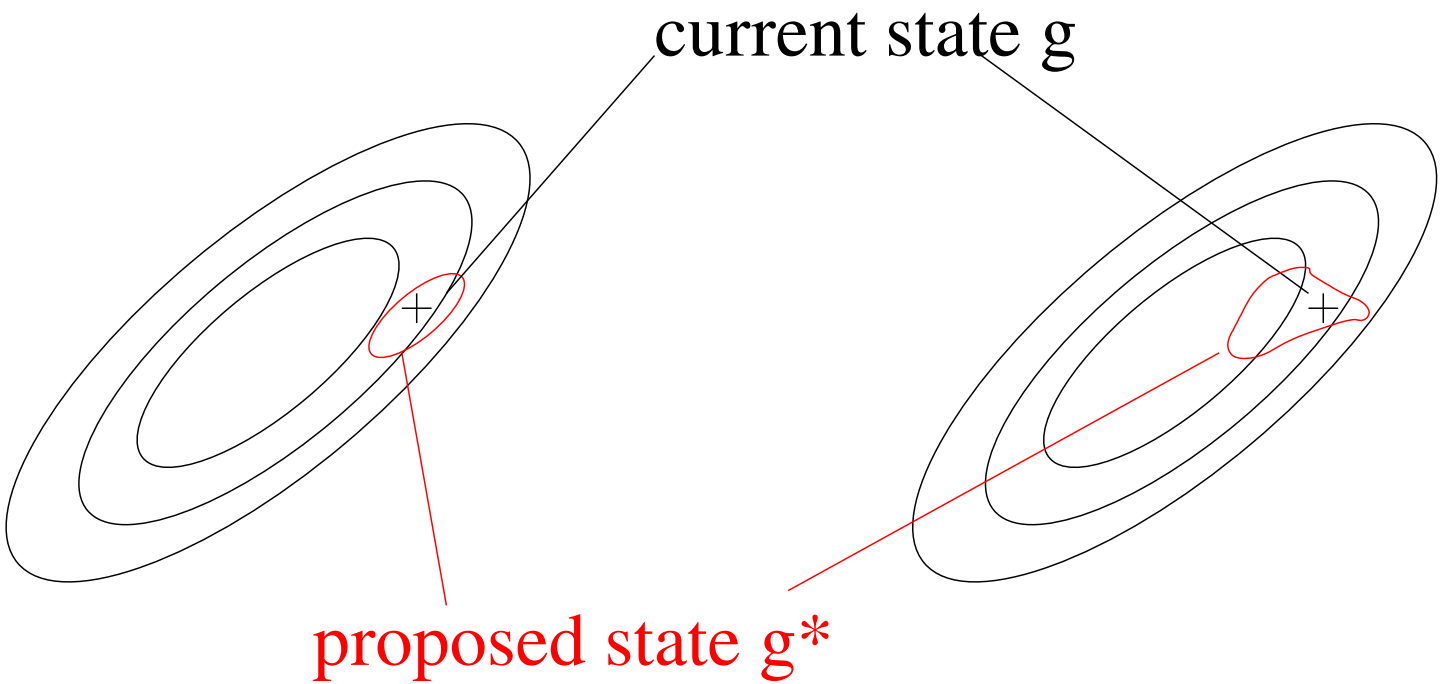
$\mathbf{y} = (y_s, y_t, y_u)$

L_κ lower triang. Choleski decomposition of $Var[\mathbf{y}]$

$\mathbf{h} = L_\kappa^{-1}\mathbf{y}$

Propose update of $\mathbf{w}$ of the form $\mathbf{w}^* = \phi(L_\kappa \mathbf{h}^*)$

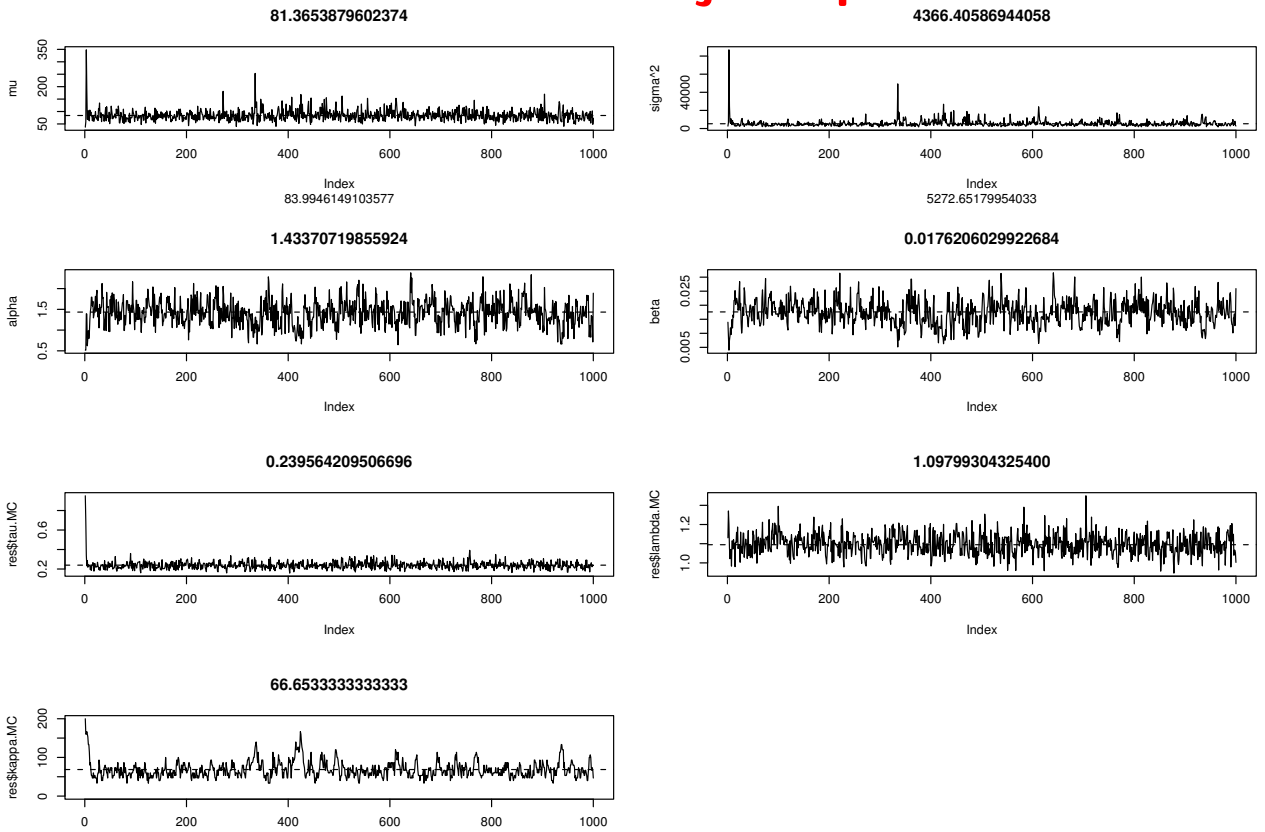
Illustration of the proposed proposal for a bivariate normal distribution



Our proposal: $y^* = y + L_\kappa \delta h$
small increment with same distribution as $\pi(y)$

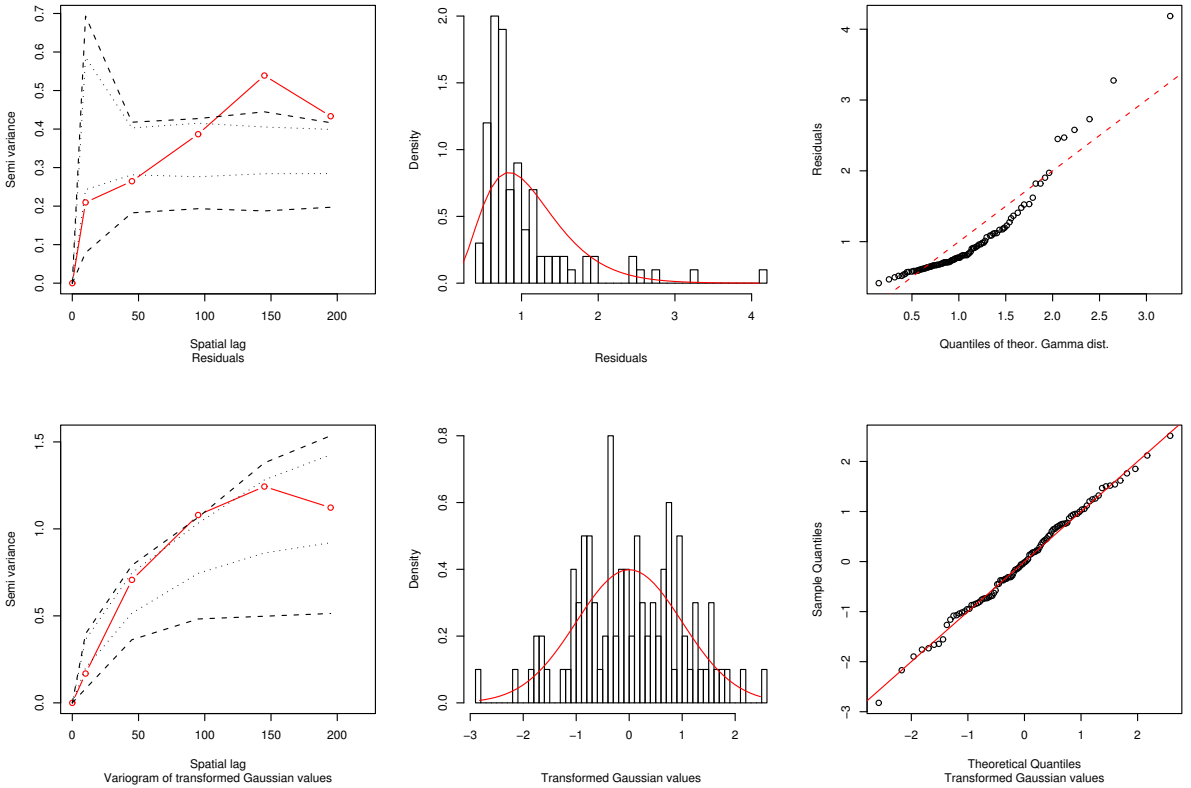
MALA proposal: $y^* = y + \eta U_\kappa^{-1} h + \delta h$
small increment approximately centered on the gradient at y

Inference for the Bjertorp dataset



Trace of the Markov Chains for the inference of θ on the full data set. First row: mean α/β and variance α/β^2 of w. Second row α and β. Third row: τ and λ. Last row: κ.

Goodness of fit assessment for the Bjertorp dataset



Goodness of fit diagnostic. Top, from left to right: variogram with confidence interval envelopes, histogram and fitted parametric density, quantile-quantile plot of predicted residuals ε^*. Bottom: variogram of predicted Gaussian values y^*, histogram and quantile-quantile plot of predicted uncorrelated components h^*.

Empirical cross-validation study

	Bayesian predictor	Kriging
10 counts, 30 images	54.5	58.1
10 counts, 50 images	53.7	59.6
50 counts, 90 images	51.4	52.4

Root mean square error on sub-sampled datasets for various sampling designs.

Simulation study

		Bayesian predictor		Kriging	
		τ		τ	
Counts	Images	0.23	0.5	0.23	0.5
50	100	51.1	53.2	55.1	54.7
30	100	54.4	54.6	59.9	58.0
10	100	59.9	60.9	66.6	67.2
10	300				
0	300				

Root mean square error on simulated datasets for various sampling designs and two levels of noise.

Conclusion

Simple model

Good fit

Improves on schemes based on weed counts only

THANKS!