Approximate Bayesian Inference in Spatial GLMMs with Skew Gaussian Latent Variables

Fatemeh Hosseini [1], Jo Eidsvik[2] and Mohsen Mohammadzadeh[1]

1. Department of Statistics, Tarbiat Modares University, Tehran, Iran

2. Department of Mathematical Science, Norwegian University of Science and Technology, Norway

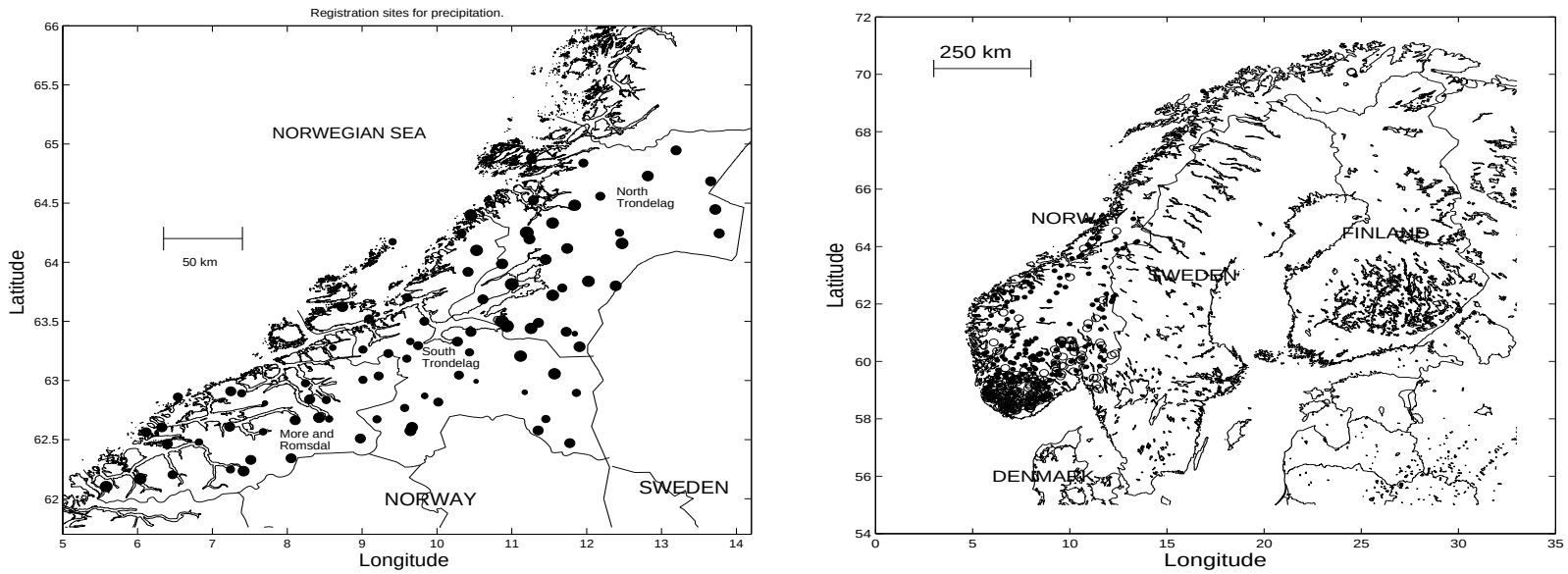
- Likelihood: $\pi(\boldsymbol{y}|\boldsymbol{x})$, exponential family. $\boldsymbol{y} = (y_1, \ldots, y_k)$.
- Prior for latent spatial variable: $\pi(\boldsymbol{x}|\boldsymbol{\eta})$. $\boldsymbol{x} = (x_1, \ldots, x_n)$, $n \geq k$.
- Prior for model parameters: $\pi(\boldsymbol{\eta})$.

GOALS:

- Present model with (closed) skew Normal $\pi(\boldsymbol{x}|\boldsymbol{\eta})$.
- Perform approximate Bayesian inference: $\hat{\pi}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\eta})$, $\hat{\pi}(\boldsymbol{\eta}|\mathbf{y})$, $\hat{\pi}(x_i|\boldsymbol{y})$.

(Closed) skew Normal generalization of Rue and Martino (2007), Eidsvik et al. (2009), Rue et al. (2009), who use Normal $\pi(\boldsymbol{x}|\boldsymbol{\eta})$.

Examples .....



(a) (b)

Figure 1: (a) Precipitation data. (b) Trout data.

$\boldsymbol{y}$: Counts/Proportions

$\boldsymbol{x}$: Latent variables/Logit field

$\boldsymbol{\eta}$: Cov/Mean/Skewness

GOALS:: $\hat{\pi}(x_i|\boldsymbol{y})$, $\hat{\pi}(\boldsymbol{\eta}|\boldsymbol{y})$

Examples .....

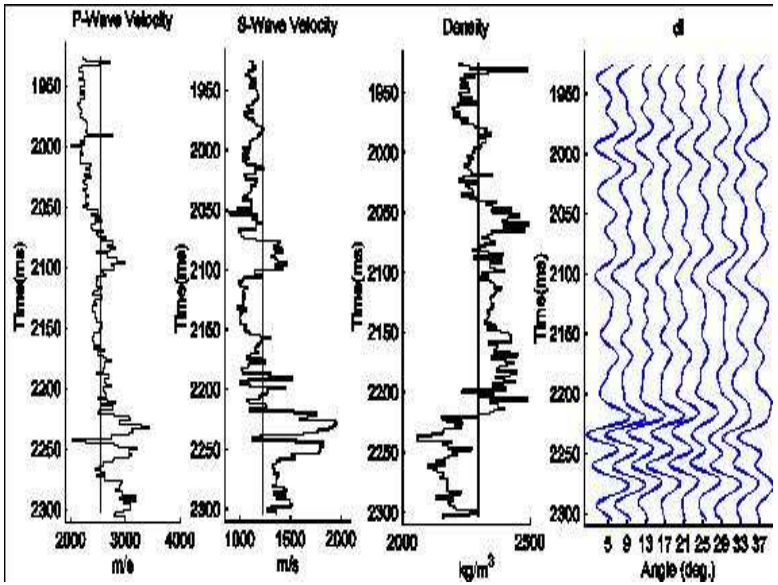


Figure 2: North Sea well log and synthetic seismic data (Karimi, Omre and Mohammadzadeh, 2009)

$\boldsymbol{y}$: Seismic data, Gaussian/linear

$\boldsymbol{x}$: Latent variables/elastic properties

$\boldsymbol{\eta}$: Cov/Mean/Skewness

GOALS:: $\hat{\pi}(x_i|\boldsymbol{y})$, $\hat{\pi}(\boldsymbol{\eta}|\boldsymbol{y})$

- Azzalini (1985) introduced the skew normal (SN) distribution.
- SN i) is larger class than Normal. ii) SN includes Normal with an extra skewness parameter
- Dominguez-Molina *et al.* (2003) introduced the closed skew normal (CSN) distribution.
- CSN i) is larger class than SN. ii) CSN includes SN with extra skewness parameters. iii) CSN is conjugate with Normal likelihood.

Skew Normal .....

- $\boldsymbol{x} \sim SN(\boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda})$

$$f(\boldsymbol{x}|\boldsymbol{\lambda}, \boldsymbol{\mu}, \Sigma) = 2\phi_n(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma)\Phi(\boldsymbol{\lambda}'\Sigma^{-\frac{1}{2}}(\boldsymbol{x} - \boldsymbol{\mu})), \tag{1}$$

  - $\boldsymbol{\lambda} \in R^n$ is a skewness parameter.

- If $\boldsymbol{\lambda} = \mathbf{0} \Rightarrow$ SN reduces to Normal: $\boldsymbol{x} \sim N_n(\boldsymbol{\mu}, \Sigma)$.

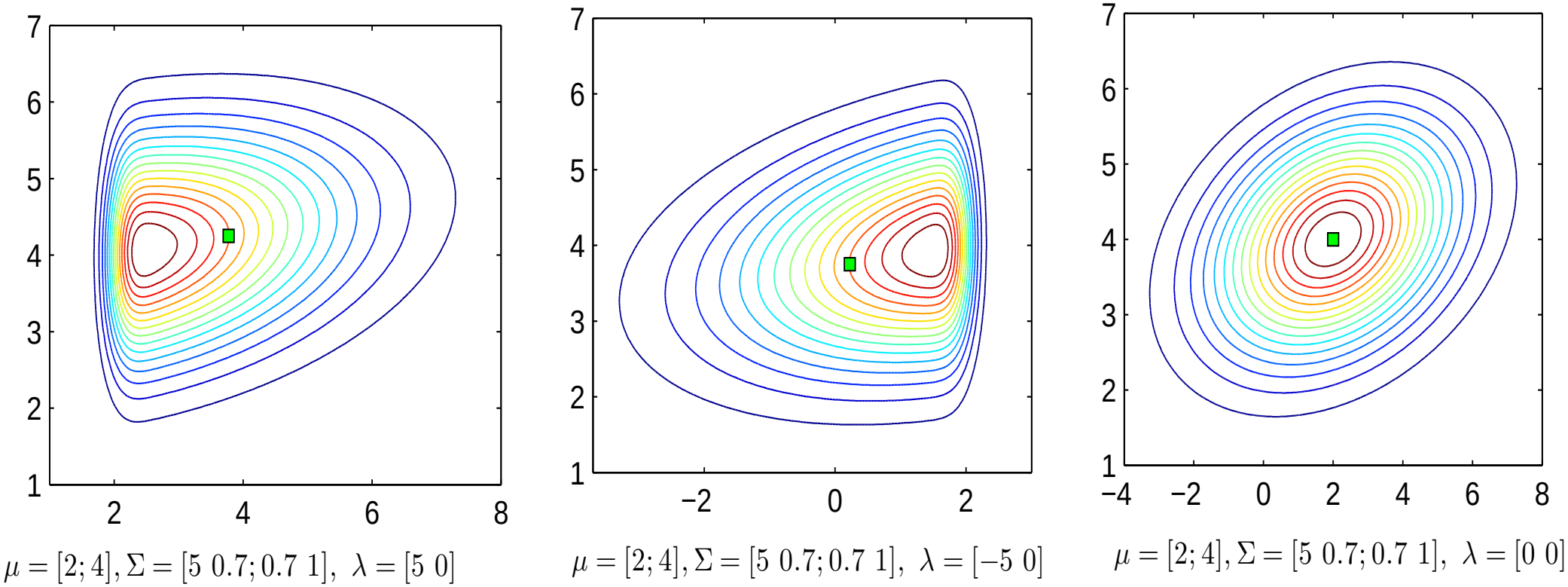


Figure 3: Contour plots of SN distribution.

Closed Skew Normal .....

- $\boldsymbol{x} \sim CSN_{n,q}(\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta)$

$$f_{n,q}(\boldsymbol{x}|\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta) = \Phi_q^{-1}(\mathbf{0}; \boldsymbol{\nu}, \Delta + D\Sigma D^{'}) \\ \phi_n(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma)\ \Phi_q(D(\boldsymbol{x} - \boldsymbol{\mu}); \boldsymbol{\nu}, \Delta). \qquad (2)$$

- If $\boldsymbol{\nu} = 0$, $\Delta = I$ and $q = 1$, CSN reduces to SN: $CSN_{n,1}(\boldsymbol{\mu}, \Sigma, D, 0, 1) = SN(\boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda})$, $D = \boldsymbol{\lambda}'\Sigma^{-1/2}$.
- If $D = \mathbf{0}$, CSN reduces to Normal, $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \Sigma)$.
- Expectation of CSN is $\mathrm{E}(\boldsymbol{X}) = \boldsymbol{\mu} + \Sigma D'\psi$,

$$\psi = \psi(0, \Sigma, D, \nu, \Delta) = \frac{\Phi_q^*(r, \nu, \Delta + D\Sigma D')}{\Phi_q(0, \nu, \Delta + D\Sigma D')}|_{r=0}, \quad \Phi_q^*(r; \nu, \Omega) = [\nabla\Phi_q(r; \nu, \Omega)]'$$

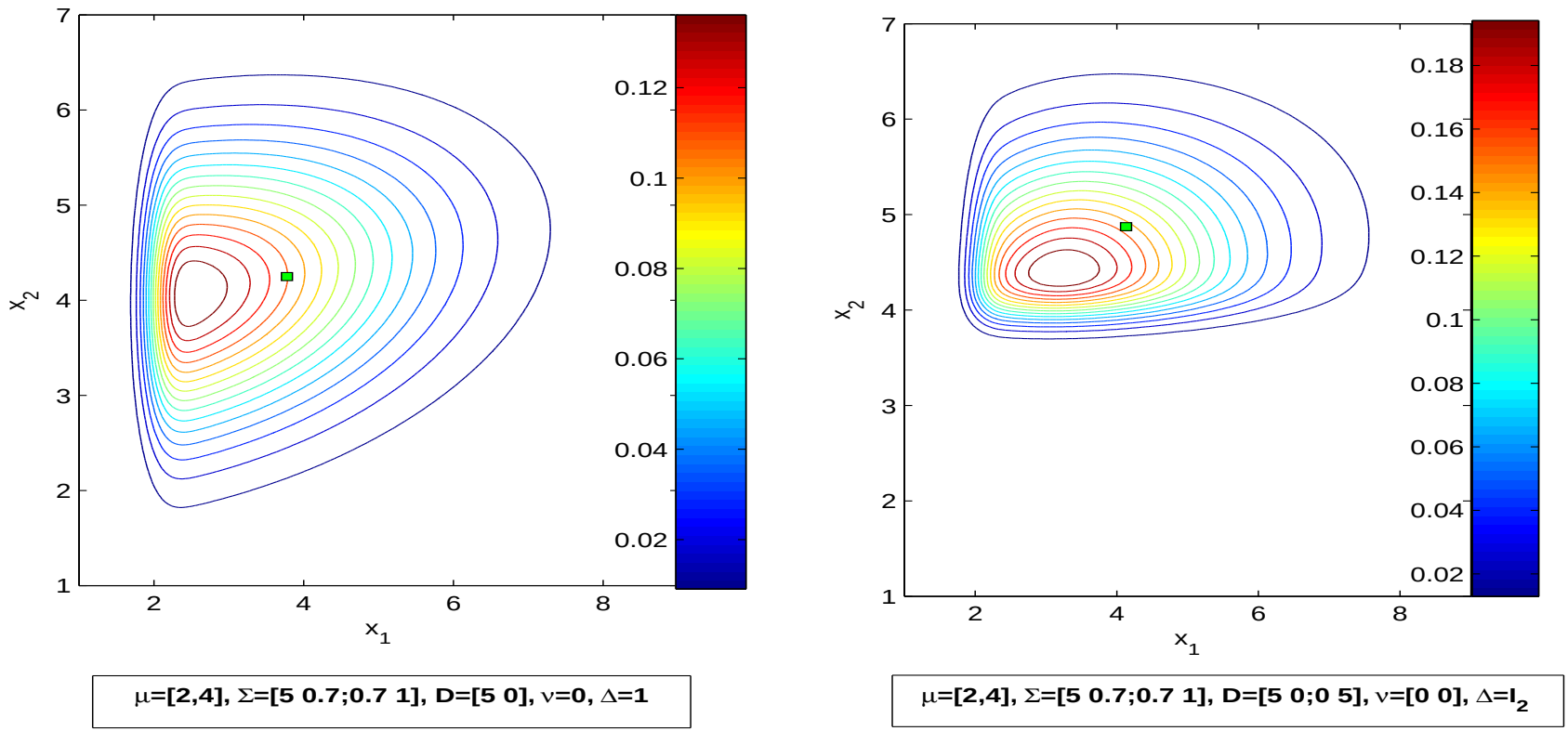


Figure 4: Contour plots of CSN distribution.

- The CSN distribution is closed under:
  - Linear transformations: Let $\boldsymbol{x}$ be CSN, then $A\boldsymbol{x} + b$ is

$$CSN_{k,q}(A\boldsymbol{\mu} + b, A\Sigma A', D\Sigma A'(A\Sigma A')^{-1}, \boldsymbol{\nu}, \Delta + D\Sigma D' - D\Sigma A'(A\Sigma A')^{-1}A\Sigma D'),$$

  - Marginalization: Let $\boldsymbol{x} = (\boldsymbol{x}_1', \boldsymbol{x}_2')'$, $\boldsymbol{x}_1$ size k, $\boldsymbol{x}_2$ size $n-k$, then

$$\boldsymbol{x}_1 \sim CSN_{k,q}(\boldsymbol{\mu}_1, \Sigma_{11}, D^*, \boldsymbol{\nu}, \Delta^*),$$

  - Conditioning: $\boldsymbol{x}_2 | \boldsymbol{x}_1$

$$CSN_{n-k,q}(\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1), \Sigma_{22\cdot 1}, D_2, \boldsymbol{\nu} - D^*(\boldsymbol{x}_1 - \boldsymbol{\mu}_1), \Delta).$$

where

$$D^* = D_1 + D_2\Sigma_{21}\Sigma_{11}^{-1}, \Delta^* = \Delta + D_2\Sigma_{22\cdot 1}D_2', \Sigma_{22\cdot 1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}.$$

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad D = (D_1, \ D_2)$$

.

Closed Skew Normal .....

A Normal / conditioning formulation of CSN:

$$\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{x}_0 \end{pmatrix} \sim N_{n+q}\left(\begin{pmatrix} \boldsymbol{\mu} \\ -\boldsymbol{\nu} \end{pmatrix}, \begin{pmatrix} \Sigma & \Sigma D' \\ D\Sigma & \Delta + D\Sigma D' \end{pmatrix} \right).$$

Then, $(\boldsymbol{x}|\boldsymbol{x}_0 > \mathbf{0}) \sim CSN_{n,q}(\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta)$.
Sample CSN by rejection sampling.

Spatial GLMM defined at three levels:

- $\pi(y_i|x_i) = \exp\{y_i x_i - b(x_i) + c(y_i)\}$, $i = 1, \ldots, k$. Conditionally independent.
- $\boldsymbol{x} = (\boldsymbol{x}^{obs}, \boldsymbol{x}^{pred})$, $\boldsymbol{x}^{obs} = A\boldsymbol{x}$ at k observation sites, $\boldsymbol{x}^{pred}$ at $(n-k)$ prediction sites. $\pi(\boldsymbol{x}|\boldsymbol{\eta}) = SN(H\boldsymbol{\beta}, \Sigma_\theta, \boldsymbol{\lambda})$, $\boldsymbol{\lambda} = \lambda_0 \mathbf{1}$.

$$\pi(\boldsymbol{x}|\boldsymbol{\eta}) = \frac{2}{(2\pi)^{n/2}|\Sigma_\theta|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - H\boldsymbol{\beta})'\Sigma_\theta{}^{-1}(\boldsymbol{x} - H\boldsymbol{\beta})\right) \cdot \Phi(\boldsymbol{\lambda}'\Sigma_\theta^{-\frac{1}{2}}(\boldsymbol{x} - H\boldsymbol{\beta}))$$

- $\Sigma_\theta(i,j) = \sigma^2 \exp(-\frac{|i-j|}{\varphi})$, $\boldsymbol{\theta} = (\sigma, \varphi)$.
- Model parameters $\boldsymbol{\eta} = (\boldsymbol{\beta}, \boldsymbol{\theta}, \lambda_0)$. Proper prior density $\pi(\boldsymbol{\eta})$.

Spatial GLMM with SN prior .....

- Joint:

$$
\begin{aligned}
\pi(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\eta}) &= \pi(\boldsymbol{y}|\boldsymbol{x})\pi(\boldsymbol{x}|\boldsymbol{\eta})\pi(\boldsymbol{\eta}) \\
&= \frac{2}{(2\pi)^{n/2}|\Sigma_\theta|^{1/2}} \exp\left(\sum_{i=1}^{k}[y_i x_i - b(x_i) + c(y_i)] - \frac{1}{2}(\boldsymbol{x} - H\boldsymbol{\beta})'{\Sigma_\theta}^{-1}(\boldsymbol{x} - H\boldsymbol{\beta})\right) \\
&\quad \times \Phi(\boldsymbol{\lambda}'\Sigma_\theta^{-\frac{1}{2}}(\boldsymbol{x} - H\boldsymbol{\beta}))\pi(\boldsymbol{\eta}),
\end{aligned} \tag{3}
$$

- Full conditional $\pi(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\eta})$ not generally available.
- If Gaussian likelihood $b(x_i) = b_0 x_i^2$, then $\pi(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\eta})$ is CSN.

- Linearize full conditional for $\boldsymbol{x}$:

$$\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta}) \approx CSN_{n,1}(\boldsymbol{\mu}_{x|y,\eta}, \Sigma_{x|y,\eta}, D_{x|y,\eta}, \boldsymbol{\nu}_{x|y,\eta}, 1),$$

▷ $\boldsymbol{\mu}_{x|y,\eta} = Z\boldsymbol{\beta} + \Sigma_\theta A' R^{-1}(z(\boldsymbol{y}, \boldsymbol{x}^{obs}) - AZ\boldsymbol{\beta}),$
$z_i(y_i, x_i) = [y_i - b'(x_i) + x_i b''(x_i)]/b''(x_i).$
$R = A\Sigma_\theta A' + P$ and $P_{ii} = 1/b''(x_i),\ i = 1, \cdots, k.$

▷ $\Sigma_{x|y,\eta} = \Sigma_\theta - \Sigma_\theta A' R^{-1} A \Sigma_\theta.$

▷ $D_{x|y,\eta} = \boldsymbol{\lambda}' \Sigma_\theta^{-\frac{1}{2}},$ ▷ $\boldsymbol{\nu}_{x|y,\eta} = \boldsymbol{\lambda}' \Sigma_\theta^{-\frac{1}{2}}(Z\boldsymbol{\beta} - \boldsymbol{\mu}_{x|y,\eta}).$

- Iterative algorithm for fitting a CSN approximation:

1. Choose starting value $\boldsymbol{x}^{(0)}$, set $m = 0$.
2. Calculate $\hat{\pi}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta}) = CSN_{n,1}(\boldsymbol{x}; \hat{\boldsymbol{\mu}}_{x|y,\eta}(\boldsymbol{x}^{(m)}), \hat{\Sigma}_{x|y,\eta}(\boldsymbol{x}^{(m)}), D_{x|y,\eta}, \hat{\boldsymbol{\nu}}_{x|y,\eta}(\boldsymbol{x}^{(m)}), 1)$.
3. Calculate
$$\boldsymbol{x}^{(m+1)} = \hat{\mathrm{E}}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta}) = \hat{\boldsymbol{\mu}}_{x|y,\eta}(\boldsymbol{x}^{(m)}) + \hat{\Sigma}_{x|y,\eta}(\boldsymbol{x}^{(m)}) D'_{x|y,\eta}\hat{\psi},$$
$$\hat{\psi} = \psi(0, \hat{\Sigma}_{x|y,\eta}(\boldsymbol{x}^{(m)}), D_{x|y,\eta}, \hat{\nu}_{x|y,\eta}(\boldsymbol{x}^{(m)}), 1)$$
4. Set $m = m + 1$.
5. Convergence is obtained after a few iterations.
$$\hat{\pi}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta}) = CSN_{n,1}(\boldsymbol{x}; \hat{\boldsymbol{\mu}}_{x|y,\eta}(\boldsymbol{x}^{(m+1)}), \hat{\Sigma}_{x|y,\eta}(\boldsymbol{x}^{(m+1)}), D_{x|y,\eta}, \hat{\boldsymbol{\nu}}_{x|y,\eta}(\boldsymbol{x}^{(m+1)}), 1).$$

$$\hat{\pi}(\boldsymbol{\eta}|\boldsymbol{y}) \propto \frac{\pi(\boldsymbol{y}|\boldsymbol{x})\pi(\boldsymbol{x}|\boldsymbol{\eta})\pi(\boldsymbol{\eta})}{\hat{\pi}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta})}\Big|_{\boldsymbol{x}=\hat{\mathrm{E}}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta})}.$$

Similar to the Laplace approximation of Tierney and Kadane (1986).

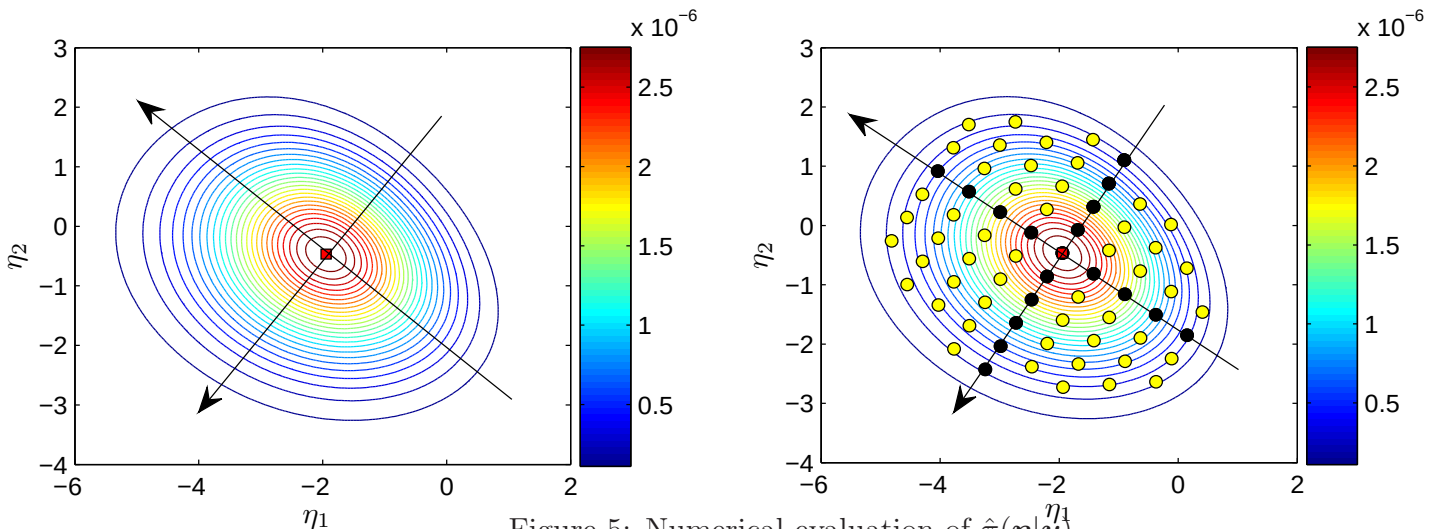


Figure 5: Numerical evaluation of $\hat{\pi}(\boldsymbol{\eta}|\boldsymbol{y})$

Approximate Spatial Prediction

- Predict the latent variables $\boldsymbol{x}^{pred}$.

- $\hat{\pi}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta})$ is CSN $\rightarrow$ marginal $\hat{\pi}(x_j|\boldsymbol{y},\boldsymbol{\eta})$ is also CSN.

- $$\hat{\pi}(x_j|\boldsymbol{y}) = \sum_{\ell} \hat{\pi}(x_j|\boldsymbol{y},\boldsymbol{\eta}_\ell) \cdot \hat{\pi}(\boldsymbol{\eta}_\ell|\boldsymbol{y}), \quad j = k+1, \cdots, n.$$

- Mixture of CSN distributions.

- Improved approximation for predictive marginals.

- $$\tilde{\pi}(x_j|\boldsymbol{y},\boldsymbol{\eta}) \propto \frac{\pi(\boldsymbol{y}|\boldsymbol{x})\pi(\boldsymbol{x}|\boldsymbol{\eta})}{\tilde{\pi}(\boldsymbol{x}_{-j}|x_j,\boldsymbol{y},\boldsymbol{\eta})}\Big|_{\boldsymbol{x}=E(\boldsymbol{x}_{-j}|x_j,\boldsymbol{y},\boldsymbol{\eta})}, \tag{4}$$

- $\tilde{\pi}(\boldsymbol{x}_{-j}|x_j,\boldsymbol{y},\boldsymbol{\eta})$ is CSN, approximated and evaluated for fixed x_j.

- $\tilde{\pi}(x_j|\boldsymbol{y}) = \sum_\ell \tilde{\pi}(x_j|\boldsymbol{y},\boldsymbol{\eta}_\ell)\cdot\hat{\pi}(\boldsymbol{\eta}_\ell|\boldsymbol{y}), \quad j = k+1,\cdots,n.$

Example 1 : Precipitation in Middle Norway .....

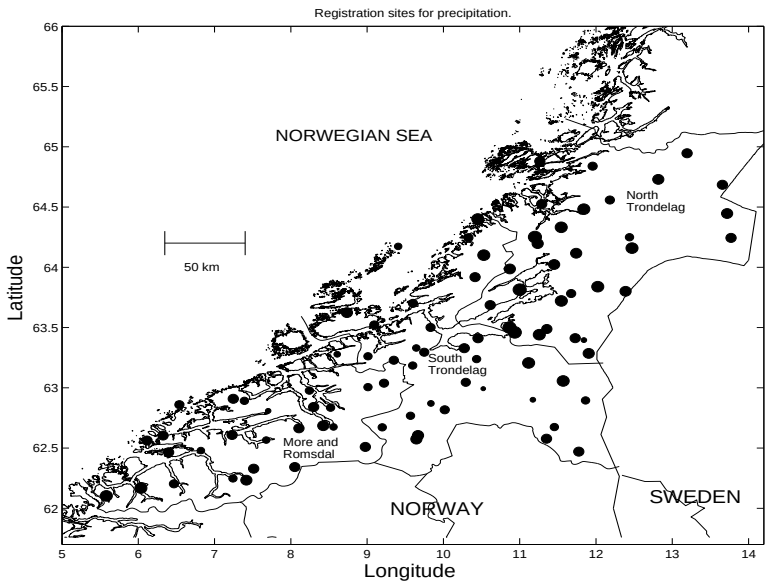


Figure 6: Map of observation sites

- The data are number of rainy days and the number of days in operation for $i = 1, \cdots, 92$ registration sites.

- Binomial data, $\pi(y_i|x_i) = Bin(\frac{e^{x_i}}{1+e^{x_i}}, 14)$.

Example 1 : Precipitation in Middle Norway .....

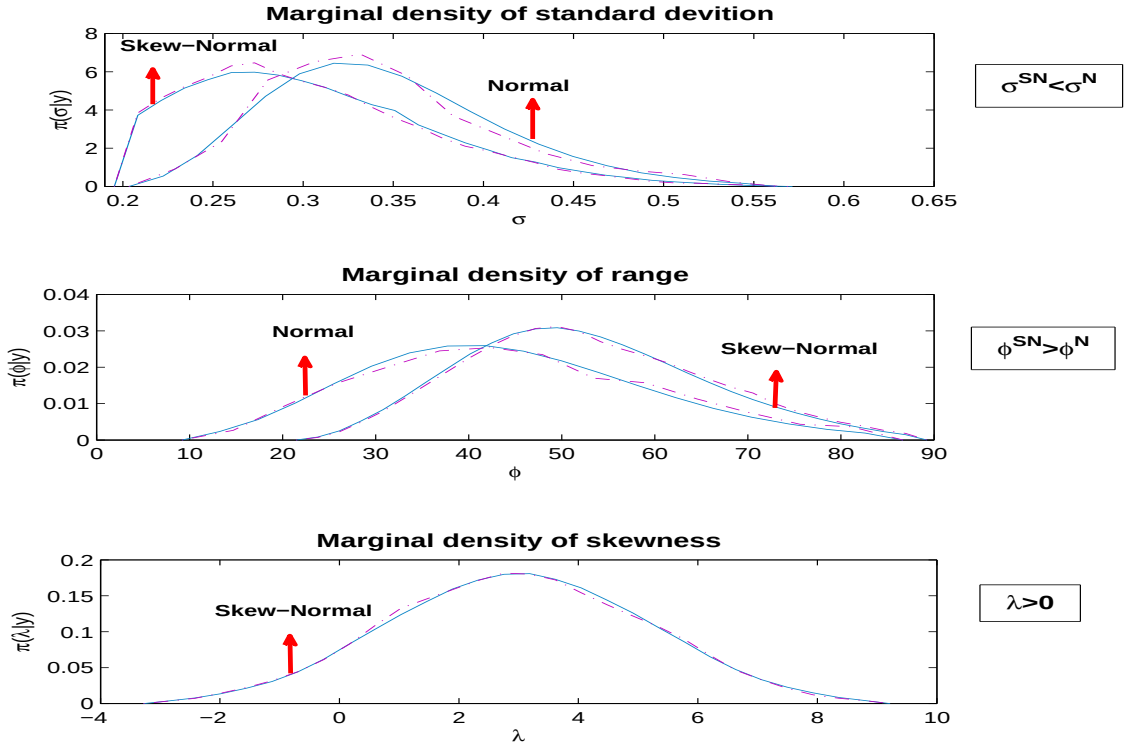


Figure 7: Posterior marginals for parameters

Solid is approximate Bayesian inference.
Dashed is independent proposal MCMC $\hat{\pi}(\boldsymbol{\eta}|\boldsymbol{y}) \cdot \hat{\pi}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\eta})$.
Acceptance rates $70 - 80\%$.

Example 1 : Precipitation in Middle Norway .....

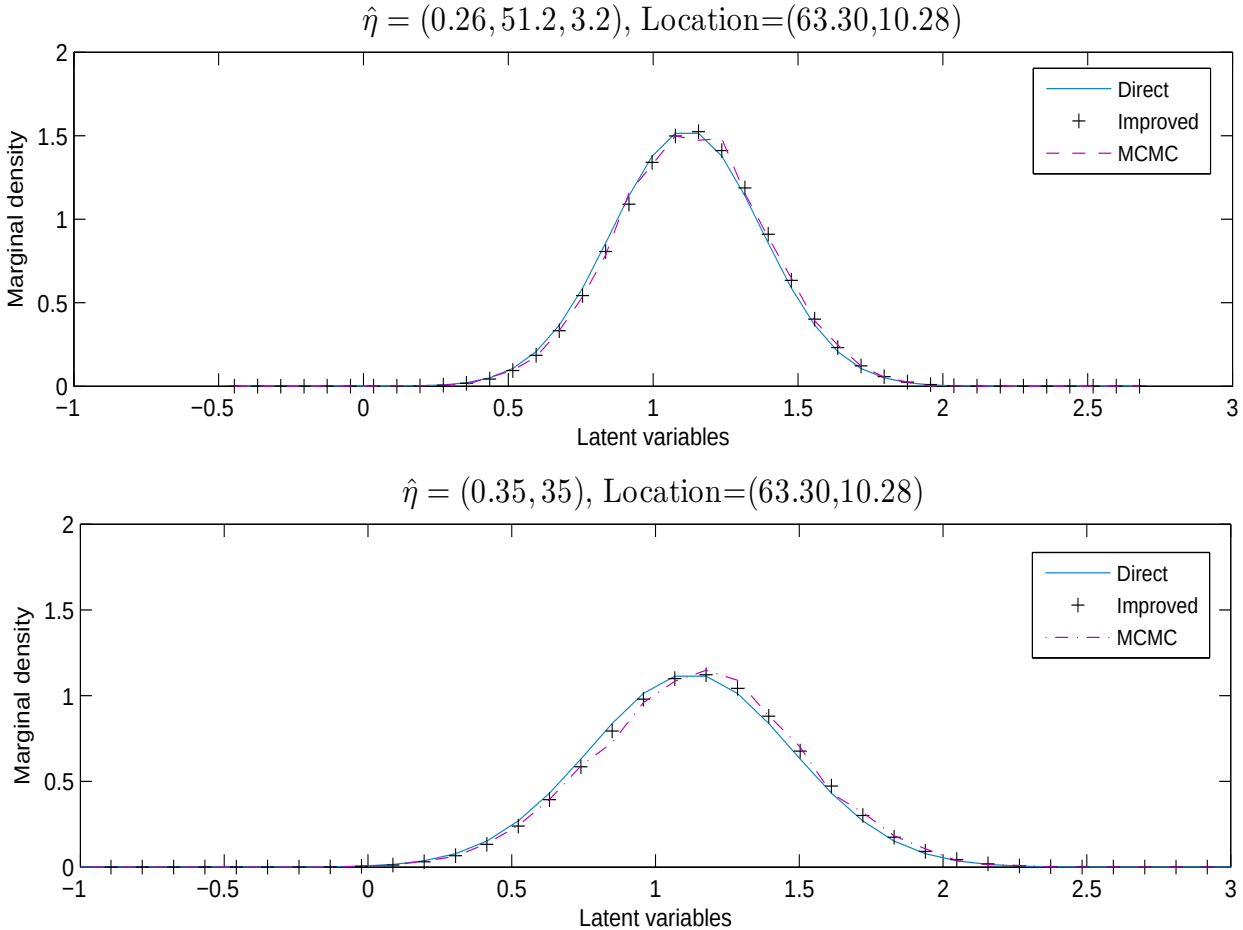


Figure 8: Predictions at unobserved site. Top is SN, below is Normal.

Three approximations: $\hat{\pi}(x_j|\boldsymbol{y}, \hat{\boldsymbol{\eta}}), \quad \tilde{\pi}(x_j|\boldsymbol{y}, \hat{\boldsymbol{\eta}}), \quad \pi^{mcmc}(x_j|\boldsymbol{y}, \hat{\boldsymbol{\eta}})$

Example 2: Norwegian Lakes Acidification

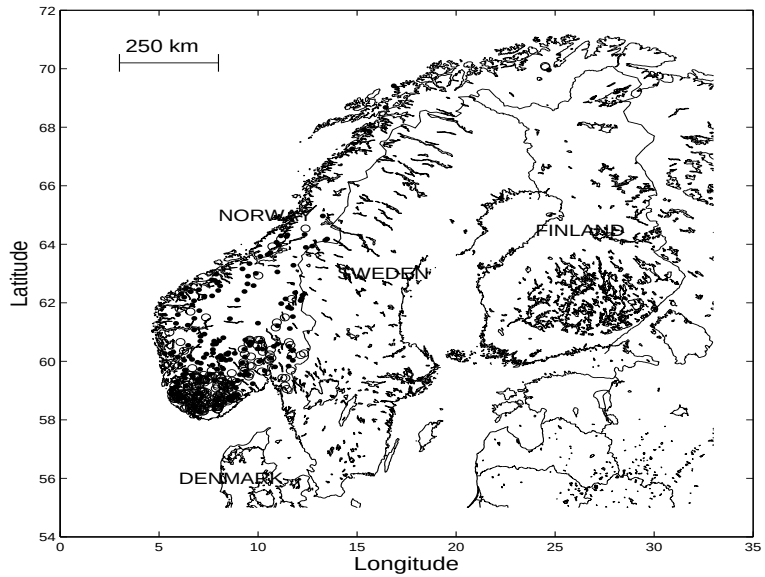


Figure 9: Map of observation sites.

- Data are high/low fish quality in 542 lakes.
- Bernoulli data, $\pi(y_i|x_i) = Bin(\frac{e^{x_i}}{1+e^{x_i}}, 1)$.
- Explanatory variable is acid capacity index (pH).

Example 2: Norwegian Lakes Acidification

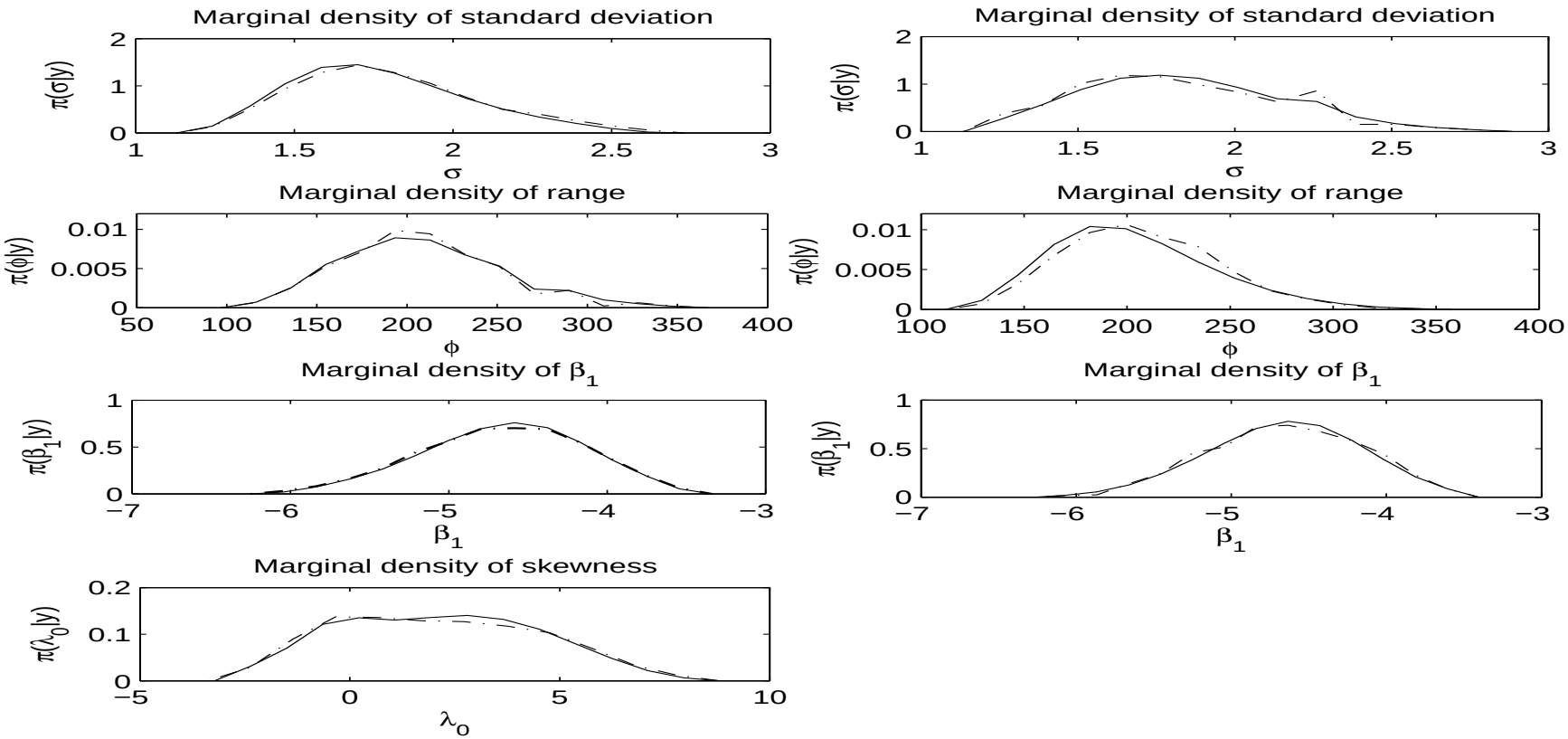


Figure 10: Posterior marginals for model parameters.

Solid is approximate Bayesian inference.
Dashed is MCMC (Acceptance rates about 80 percent)

Example 2: Norwegian Lakes Acidification .....

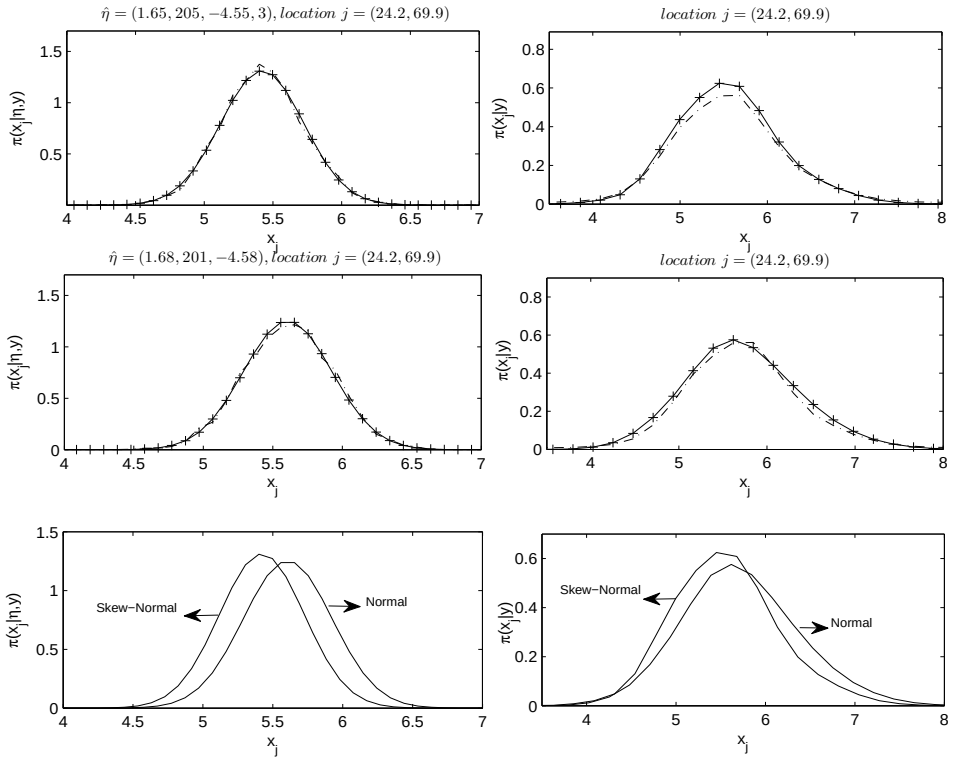


Figure 11: Prediction at unobserved site. Left is $\hat{\pi}(x_i|\boldsymbol{y}, \hat{\boldsymbol{\eta}})$. Right is $\hat{\pi}(x_i|\boldsymbol{y})$.

Example 3: Seismic inversion

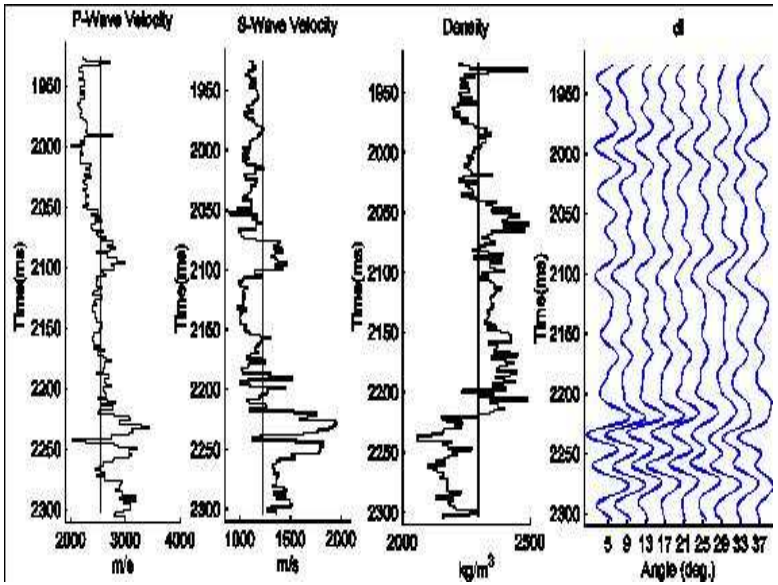
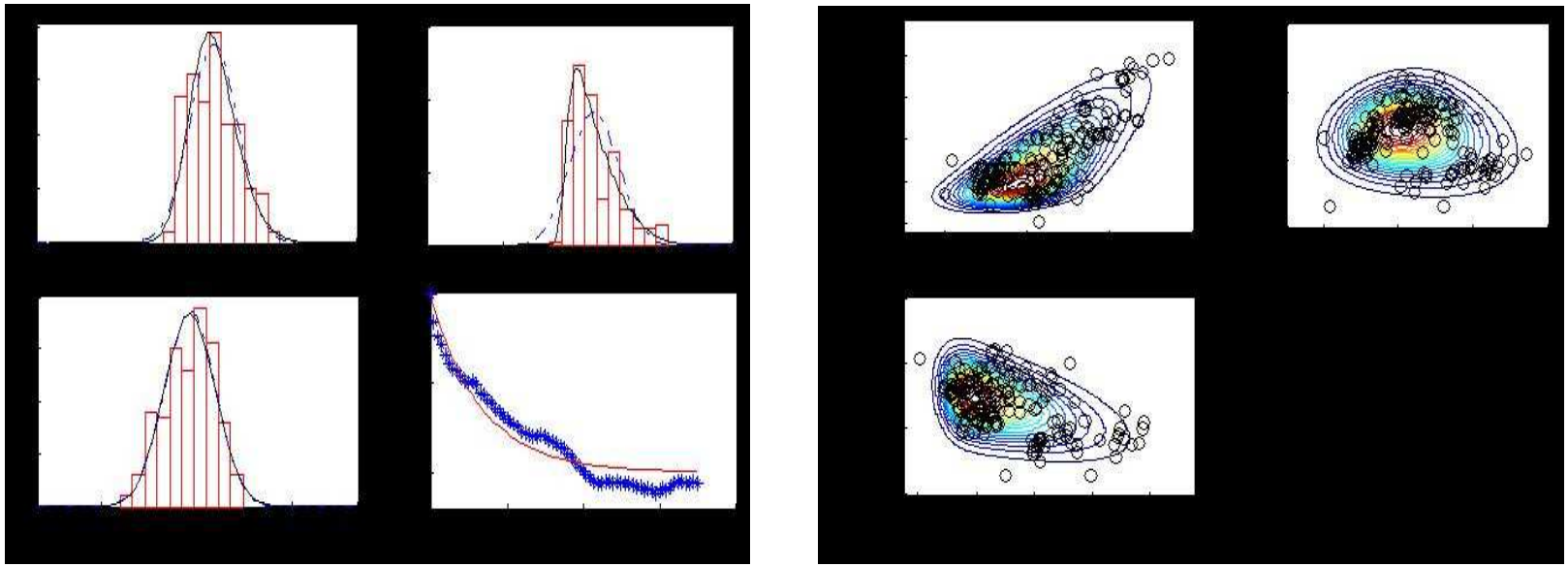


Figure 12: North Sea well log and synthetic seismic data (Karimi, Omre and Mohammadzadeh, 2009)

- Data are seismic reflection amplitudes for different traveltimes and several incidence angles.
- $\pi(\boldsymbol{y}|\boldsymbol{x}) = Normal(G\boldsymbol{x}, T)$.
- In this case $\pi(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\eta})$ is CSN without approximations.

Example 3: Seismic inversion

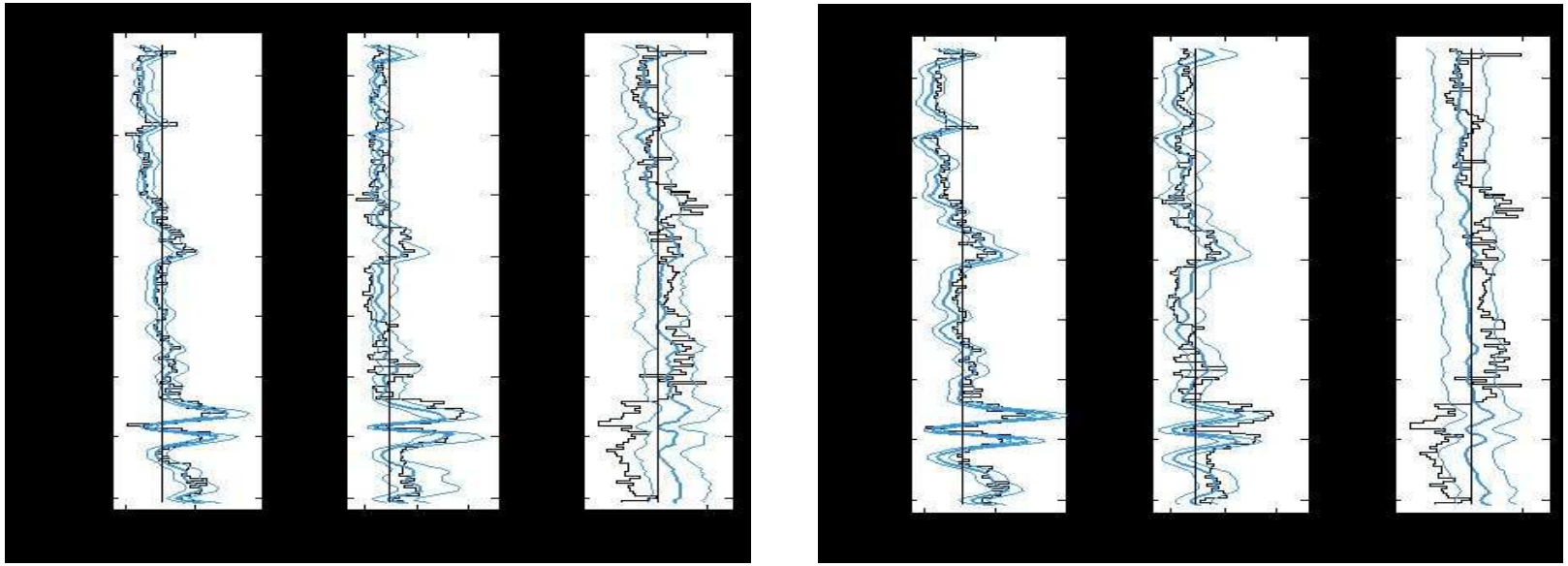


(a) (b)

Figure 13: CSN priors. From Karimi, Omre and Mohammadzadeh (2009)

Fitted tri-variate CSN prior from well logs.
Same for all depth, but correlated.

Example 3: Seismic inversion



(a) (b)

Figure 14: Inversion result. From Karimi, Omre and Mohammadzadeh (2009)

CSN result (left) and Normal result (right).

Conclusion

- Skew Normal prior for latent variables in Spatial GLMMs.
- Closed skew Normal for approximate Bayesian inference $\rightarrow$ approximations comparable with MCMC results. Direct inference (seconds), improved inference (minutes), MCMC (hours).
- Closed skew Normal results for parameter estimates differ from Normal: Larger 'correlation', less 'standard deviation'. Skewness is hard to estimate from our data.
- Closed skew Normal results for prediction differ from Normal only at some sites, where the skewness 'kicks in'. More skewness 'directions' could change this (larger q).
- Computational / numerical challenges: $\Phi_q(\boldsymbol{x})$, $E(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta})$, $\hat{\pi}(x_i|\boldsymbol{y},\boldsymbol{\eta})$, when n or k are very large.