Some common tune(s)

- Real **inverse** problems with enormous interest
- Huge amounts of data and large dimensions
- Limited use of prior information (?)
- Modified MCMC algorithms
- Multiple levels of **[Gaussian, always Gaussian]** approximations
- Difficulty in recovering induced error [getting out of the model]

Ozone monitoring

Inverse Problem : Model defined in terms of untractable integrals of gas profiles

$$T^{\text{abs}}_{\lambda,l} = \exp\left[-\int_l \sum_{\text{gas}} \alpha^{\text{abs}}_{\lambda}\{z(s)\}\rho^{\text{abs}}\{z(s)\}\,\mathrm{d}s\right] \qquad (1)$$

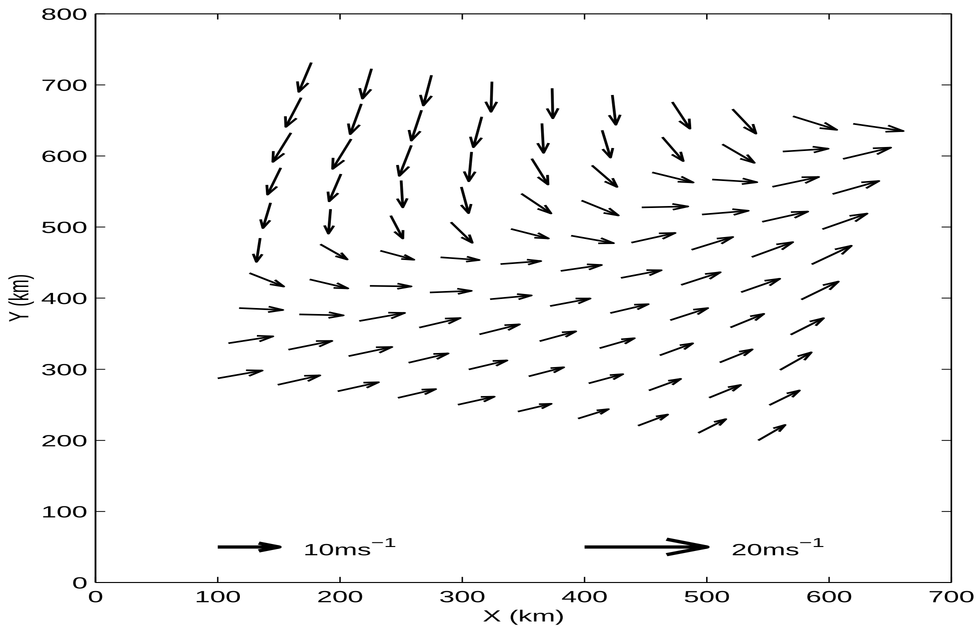
Gaussian observations

$$y_{\lambda,l} = T^{\text{abs}}_{\lambda,l} T^{\text{ref}}_{\lambda,l} + \epsilon_{\lambda,l}, \qquad \epsilon_{\lambda,l} \sim \mathcal{N}(0, \sigma^2_{\lambda,l})$$

O[450 occultations $\times$ 1400 wavelengths $\times$ 70 altitudes]

From waves to winds

Inverse Problem : Recover $\mathbf{V}$ (winds) from $\mathbf{S^0}$ (waves)



Forward model

$$s_i^0 = (\hat{a}_0(\mathbf{v}_i), \hat{a}_1(\mathbf{v}_i), \hat{a}_2(\mathbf{v}_i), \hat{a}_3(\mathbf{v}_i), \hat{a}_4(\mathbf{v}_i)) \begin{pmatrix} 1 \\ \cos\chi \\ \cos 2\chi \\ \cos 3\chi \\ \cos 4\chi \end{pmatrix} + \epsilon$$

O$[60,000]$

Approximations [Gaussian, so Gaussian]

Ozone: Cross-sections

$$T_{\lambda,s}^{\text{abs}} = \exp\left[\sum_{\text{gas}} \alpha_{\lambda,l}^{\text{abs}} N_l^{\text{abs}}\right]$$

[Spectral inversion]

$$N_l^{\text{abs}} = \int_l \rho^{\text{abs}}\{z(s)\}\, \mathrm{d}s$$

[Vertical inversion]

Approximate likelihood

$$\exp\left\{-(T_l(N_l) - y_l)^{\mathsf{T}} C_l^{-1} (T_l(N_l) - y_l)/2\right\}$$

Waves/Winds: Inverse model

$$\mathbf{v}_i | s_i^0 \sim \sum_{k=1}^{4} \alpha_k(s_i^0) \mathscr{N}_2(\mu_k(s_i^0), \sigma_k^2(s_i^0) I_2)$$

[Neural network estimation of $\alpha_k, \mu_k, \sigma_k$'s]

- Why 4?!
- Which learning set?

Waves/Winds: Inverse model

$$\mathbf{v}_i | s_i^0 \sim \sum_{k=1}^{4} \alpha_k(s_i^0) \mathscr{N}_2(\mu_k(s_i^0), \sigma_k^2(s_i^0) I_2)$$

[Neural network estimation of $\alpha_k, \mu_k, \sigma_k$'s]

- Why 4?!
- Which learning set?

Note: A mixture model with noisy observations is still a mixture model

$$\sum_{k=1}^{4} \alpha_k \mathscr{N}_2(\mu_k, (\sigma_k^2 + \tau^2)\, I_2)$$

Where is the difficulty in

$$p(\mathbf{V}|\mathbf{S^0}) \propto \prod_i \frac{p(\mathbf{v}_i|s_i^0)}{p(\mathbf{v}_i)} p(\mathbf{V})$$

since

1. **the function can be computed analytically**
2. **a missing data structure can be introduced**
3. **modes need not be known in advance**

Besides,

1. how can modes be identified in $60,000$ dimensions?

2. is bimodality a simple consequence of direction lack of identifiability?!

Which prior modelling?!

Effect of the Gaussian assumption

Ozone: Vague and limited

Seems to only correspond to the discretization/regularization choice

$$x_i = \hat{x}_i \pm \varepsilon^{\text{reg}} \sqrt{(\Delta z^{-1})}$$

Which prior modelling?!

Effect of the Gaussian assumption

Ozone: Vague and limited

Seems to only correspond to the discretization/regularization choice

$$x_i = \hat{x}_i \pm \varepsilon^{\text{reg}} \sqrt{(\Delta z^{-1})}$$

Waves/Winds: Prior information = NWP forecasts or (?) Gaussian

Use of seasonal data only through the covariance matrix

Modified MCMC

Ozone: Adaptive covariance structure

$$C_t = s_d \text{cov}(X_1, \ldots, X_t) + s_d \epsilon I_d$$

[Haario, Saksman & Tamminen, 2003]

1. **For practical purposes, relevance of ergodic theorem?**

   Adaptativity can be embedded in a grand-scale scheme

$$(X^{(t)}, \Sigma^{(t)}) \overset{\mathscr{L}}{\rightsquigarrow} \pi(X) \times |\Sigma|^{-\alpha} \exp -\alpha \mathrm{tr}(\Sigma^{-1})$$

   and regular Markovian updating

$$\Sigma^{(t)} = \beta\, \Sigma^{(t-1)} + (1-\beta) \left\{ (X^{(t)})^{\mathsf{T}} X^{(t)} + \varepsilon \mathbf{I}_d \right\}$$

1. **For practical purposes, relevance of ergodic theorem?**

   Adaptativity can be embedded in a grand-scale scheme

$$(X^{(t)}, \Sigma^{(t)}) \overset{\mathscr{L}}{\rightsquigarrow} \pi(X) \times |\Sigma|^{-\alpha} \exp -\alpha \mathrm{tr}(\Sigma^{-1})$$

   and regular Markovian updating

$$\Sigma^{(t)} = \beta\, \Sigma^{(t-1)} + (1-\beta)\, \left\{ (X^{(t)})^{\mathsf{T}} X^{(t)} + \varepsilon \mathbf{I}_d \right\}$$

2. how can SCAM one-dimensional update be efficient in **high-D**?

   [Lack of covariance structure]

1. **For practical purposes, relevance of ergodic theorem?**

   Adaptativity can be embedded in a grand-scale scheme

$$(X^{(t)}, \Sigma^{(t)}) \stackrel{\mathscr{L}}{\rightsquigarrow} \pi(X) \times |\Sigma|^{-\alpha} \exp -\alpha \text{tr}(\Sigma^{-1})$$

   and regular Markovian updating

$$\Sigma^{(t)} = \beta\,\Sigma^{(t-1)} + (1-\beta)\,\left\{(X^{(t)})^{\mathsf{T}} X^{(t)} + \varepsilon \mathbf{I}_d\right\}$$

2. how can SCAM one-dimensional update be efficient in **high-D**?

   [Lack of covariance structure]

3. Does it really work better?!

   [Is SCAM a scam?!]

$\mathscr{T}$-posterior example

For a t-distribution posterior

$$\prod_{j=1}^{n} \left[\nu + (x_j - \theta^{(t)})^2\right]^{-(\nu+1)/2},$$

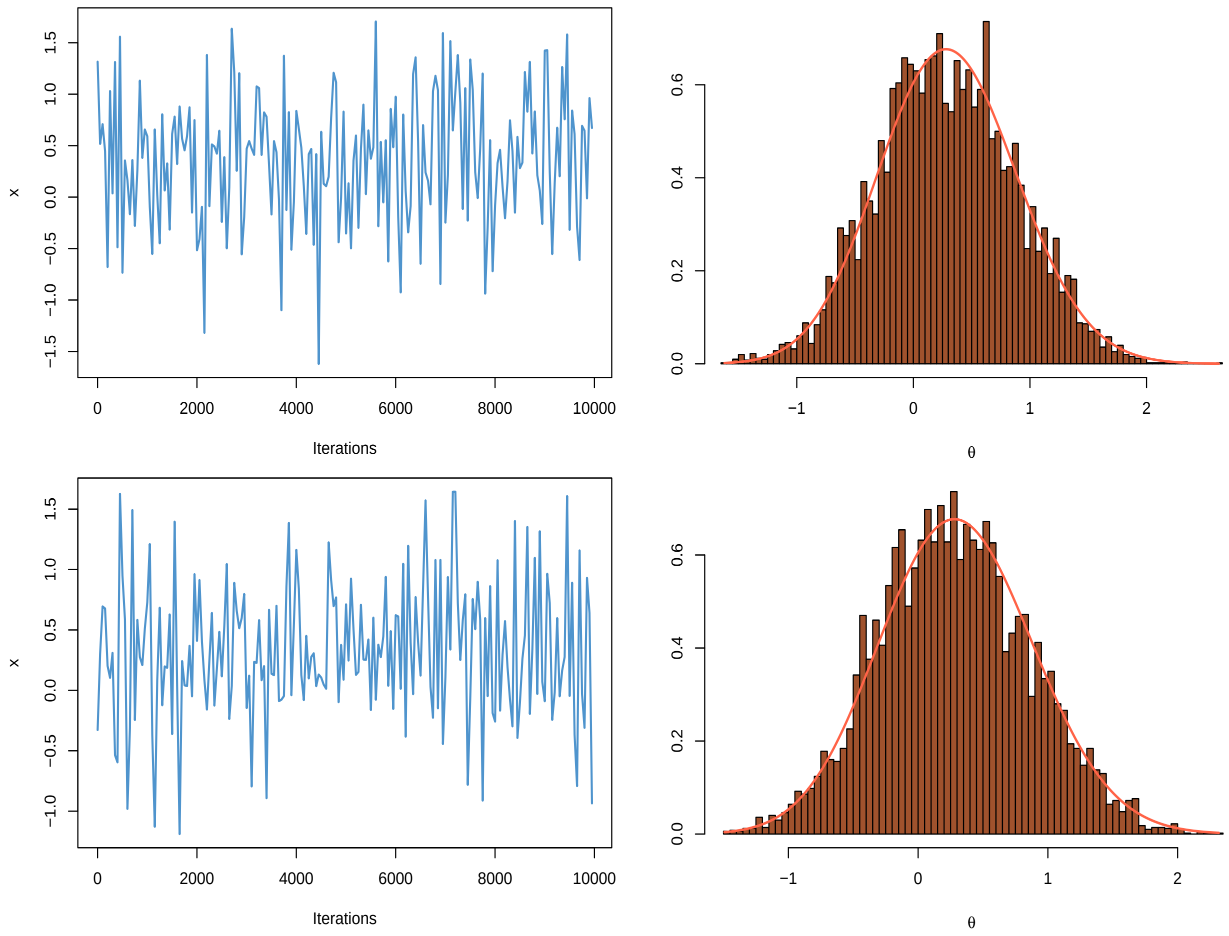
proposal based on variance of the previous values of the chain,

$$\sigma_t^2 = \frac{5}{t} \sum_{i=1}^{t} (\theta^{(i)} - \hat{\mu}_t)^2$$

or on ridge-stabilised version

$$\tilde{\sigma}_t^2 = \frac{5}{t} \sum_{i=1}^{t} (\theta^{(i)} - \hat{\mu}_t)^2 + \epsilon$$

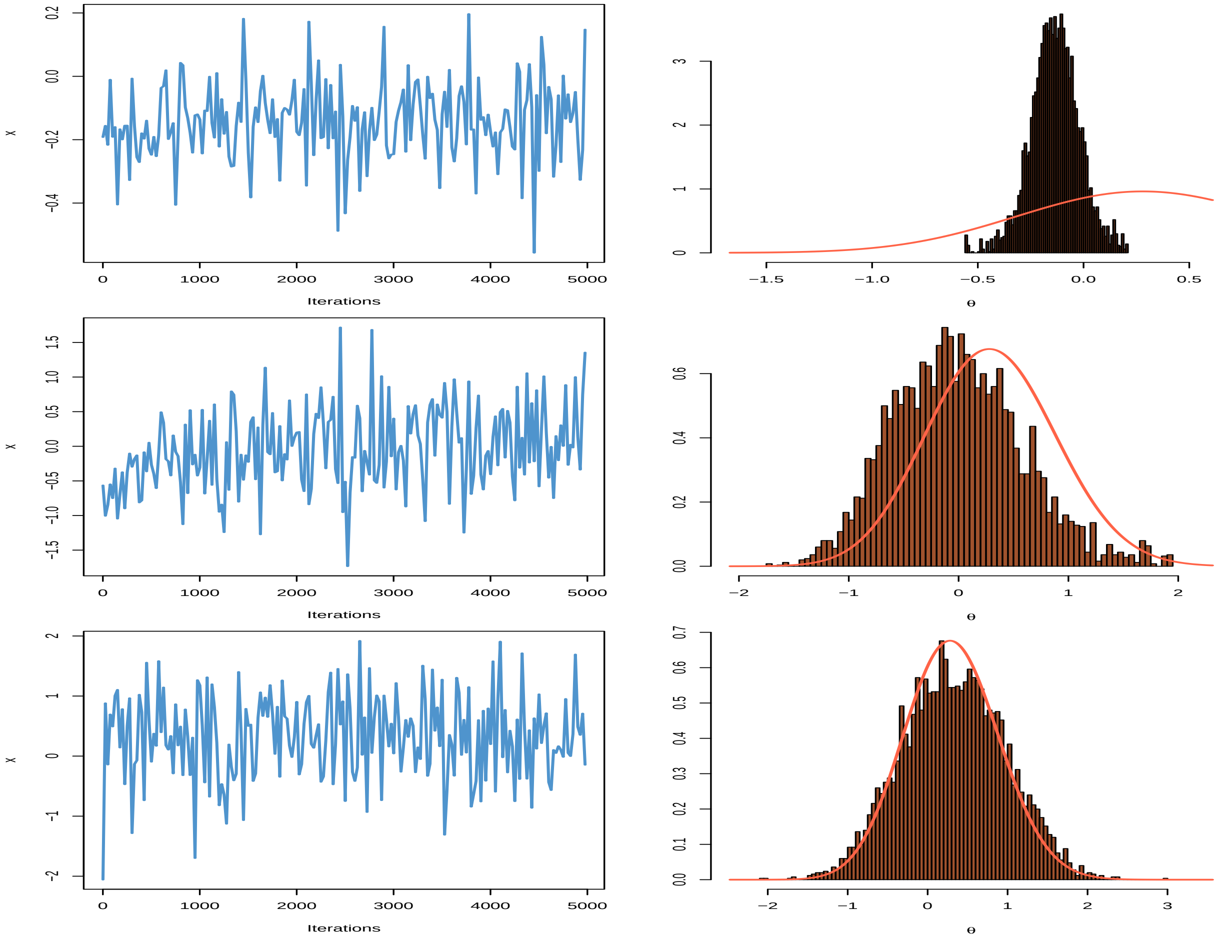
Similar behaviour of the output


x
Iterations
θ
x
Iterations
θ

Less accurate if the mean is also updated from the empirical mean

$$\hat{\mu}_t = \frac{1}{t} \sum_{i=1}^{t} \theta^{(i)}$$

[Haario et al. stabilisation = ?]



Influence of the original variance

Waves/Wind:

1. Mode-jump $(J = 1)$ with probability $.05$
2. Random-walk proposal $\mathbf{V}^*$ with drift μ_J $(\mu_0 = 0)$
3. Acceptance probability

$$\alpha = \frac{p(\mathbf{V}^*|\mathbf{S}^0)}{p(\mathbf{V}_t|\mathbf{S}^0)} \frac{\varphi(\mathbf{V}_t - \mathbf{V}^* \pm \mu_J)}{\varphi(\mathbf{V}^* - \mathbf{V}_t - \mu_J)}$$

Inefficient and useless:

1. correct $\varphi(\mathbf{V}_t - \mathbf{V}^* \pm \mu_J)\big/\varphi(\mathbf{V}^* - \mathbf{V}_t - \mu_J)$ ratio?
2. detailed balance uncalled for
3. multiscale proposals (**high-D** troubles of same magnitude)
4. Rao–Blackwellisation $.5\mathscr{N}(\mathbf{V}_t, \Sigma) + .5\mathscr{N}(\mathbf{V}_t + \mu_J, \Sigma)$
5. delayed rejection
6. particle systems
7. population Monte Carlo

Iterated sequential processing reminiscent of Gibbs sampling

More questions

Ozone:

1. Influence of the prior distribution on the $N_l^{(i)}$'s ?
2. Assessment that convergence is not a problem ?
3. How is the problem in its original form (1) solved ?

More questions

Ozone:

1. Influence of the prior distribution on the $N_l^{(i)}$'s ?
2. Assessment that convergence is not a problem ?
3. How is the problem in its original form (1) solved ?

Waves/Wind:

1. Multiple levels of Gaussian approximations, but no clear assessment of the effect of these approximations
2. No clear quantitative assessment of the computational difficulties

Conclusion

- Very real and very interesting problems
- Worthy illustration of Statistical inference in huge dimensions
- Adapted and adaptive MCMC algorithms with more universal bearings

I propose the vote of thanks!