A fast Bayesian model for latent radio signal prediction

Brett Houlding[1] Arnab Bhattacharya[1] **Simon Wilson**[1]
Tim Forde[2]

Centre for Telecommunications Value-Chain Research
[1]School of Computer Science and Statistics / [2]School of Engineering
Trinity College Dublin

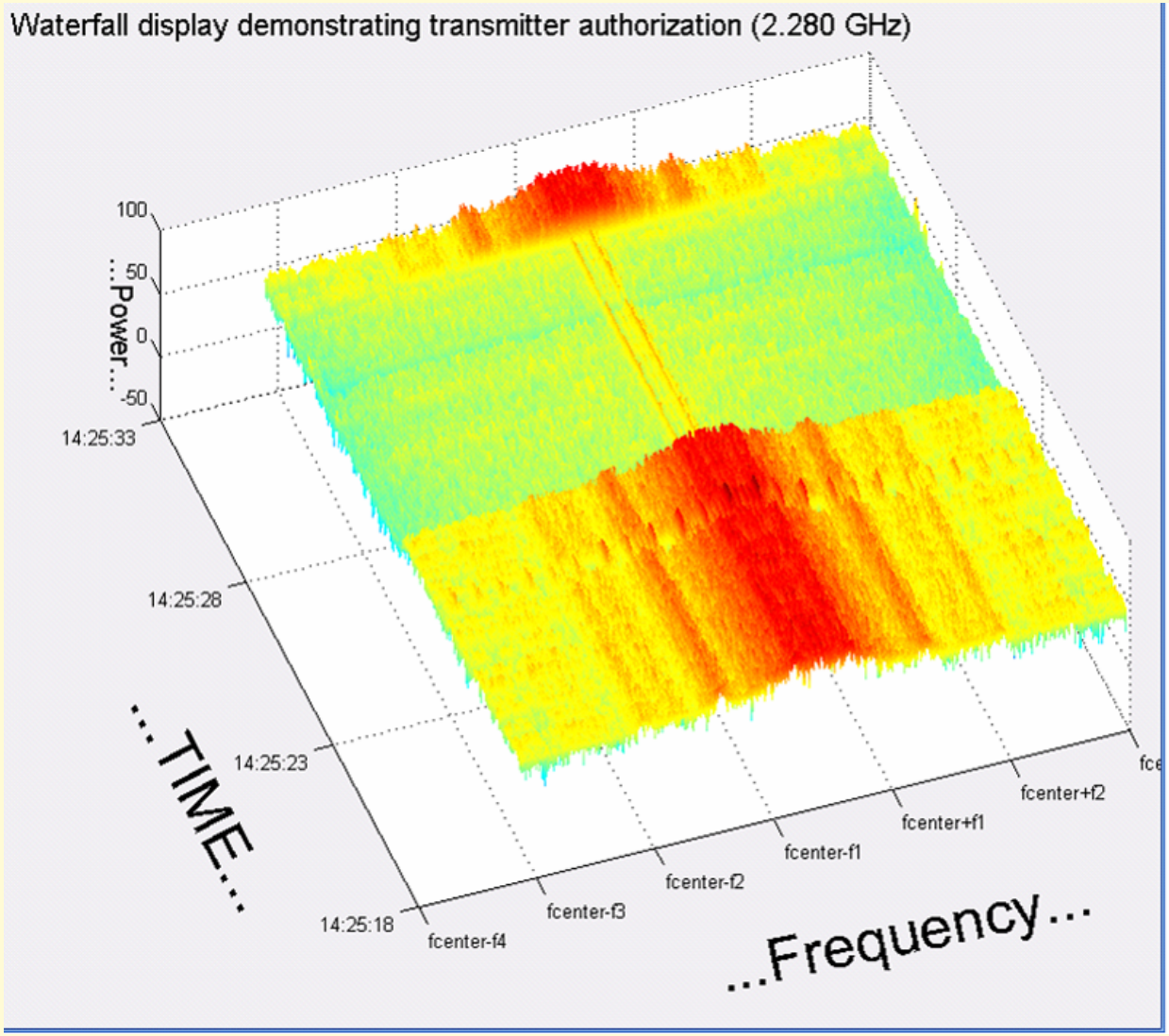
Context

- Interest in dynamic spectrum allocation for wireless networks to increase bandwidth:
  - Use under-utilised frequencies and abandon them when the "primary" user needs them;
  - Form ad-hoc networks with nearby devices;
- Some of the questions raised with a statistical element e.g.:
  - Optimisation/decision problems under uncertainty (frequency selection, choice of transmission parameters, feature selection for prediction, trade-off between prediction accuracy and transmission overhead);
  - Sufficient statistics for acceptable network performance:
    - Does the amount of information required by devices on the network make it impractical?
- $\Rightarrow$ need for fast online inference and prediction of spectrum usage

Context: why Bayesian methods?

- Bayesian inference methods have several appealing properties:
  - Natural inference method to combine with decision theoretic approaches to decision making;
  - Ideally suited to making reasoned inference with data from many sources of different quality;
- Principal objection is lengthy computation — unsuited to fitting complex models in real time;
- This talk explores one new approach to speeding up implementation of Bayesian methods.

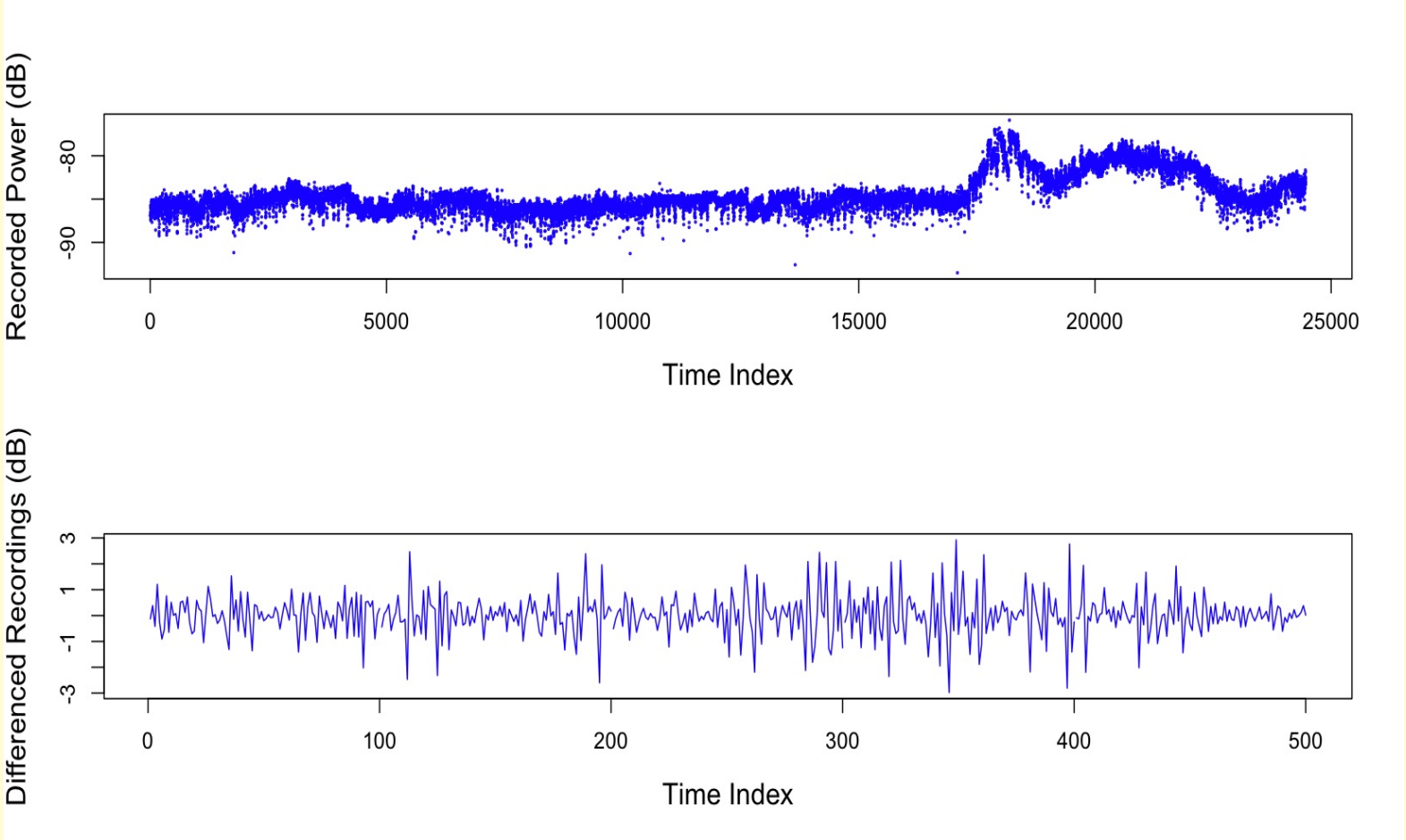
Typical Spectrum Usage Data over Time and Frequency



In this talk

- A model for the spatio-temporal evolution of ambient power at a particular frequency;
  - Model is based on analysis of data collected in Dublin;
  - It has a standard latent Gaussian form.
- Fast fitting of such models through Bayesian methods:
  - Introduce the integrated nested Laplace approximation (INLA);
  - Applicable to a wide class of potentially interesting models for these data.

Power observed from Three Rock Mountain transmitter, Dublin

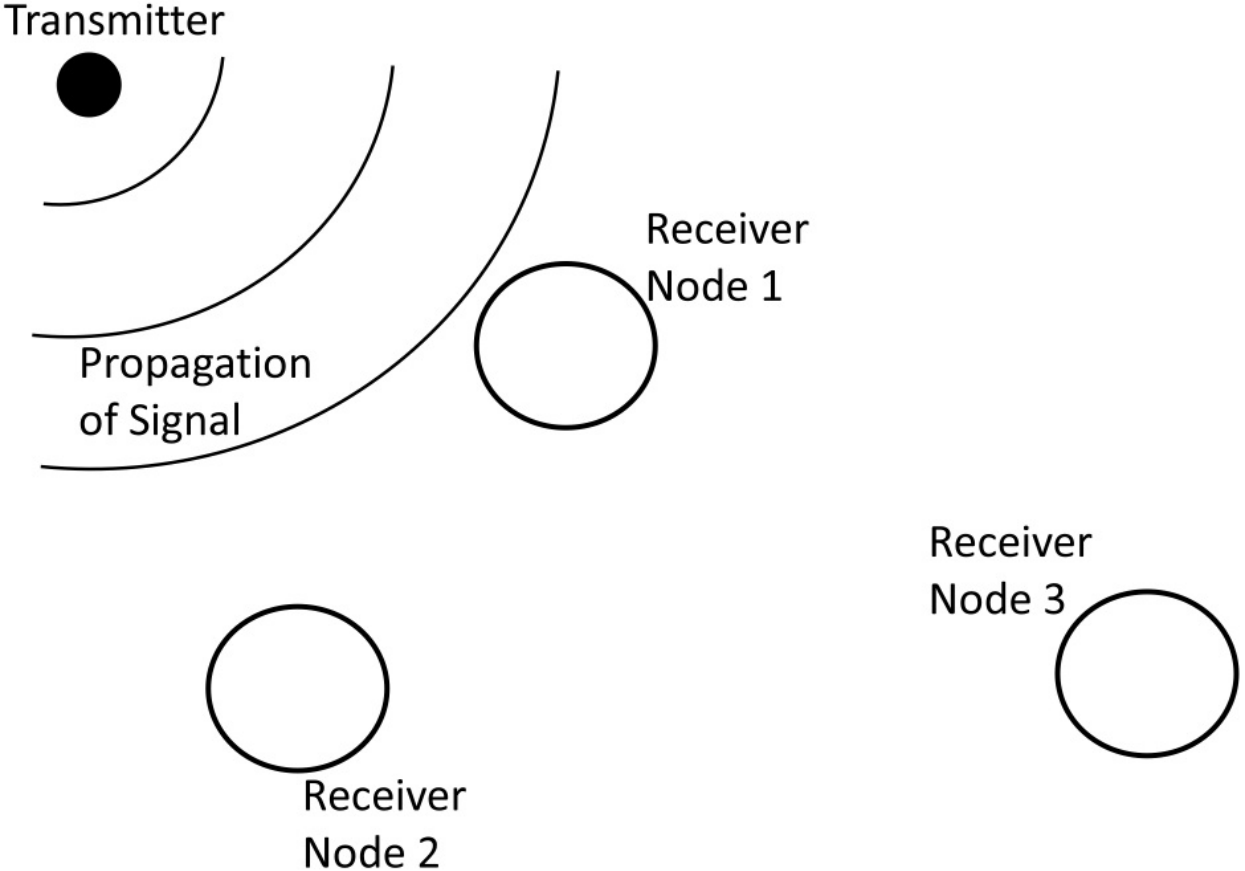


Model for transmitter signal

- Some initial data exploration showed the ARIMA(1,1,0) model fitting well:

$$
\begin{aligned}
X_t &= (1+\phi)X_{t-1} - \phi X_{t-2} + \epsilon_t, \\
\epsilon_t &\sim N(0, \sigma_\epsilon^2)
\end{aligned}
$$

Extending to a transmitter–receiver model for a laboratory cell



Extending to a transmitter–receiver model for a laboratory cell

- One transmitter, 3 receivers in our lab set-up;
- Assume transmitter signal X_t is ARIMA(1,1,0);
- Receiver signal is Y_t:

$$
\begin{aligned}
Y_t \,|\, X_t = \begin{pmatrix} Y_{t1} \\ Y_{t2} \\ Y_{t3} \end{pmatrix} \Bigg| X_t &\sim N\left(\begin{pmatrix} X_t \\ X_t \\ X_t \end{pmatrix}, \Sigma \right); \\
X_t &\sim N((1+\phi)X_{t-1} - \phi X_{t-2}, \sigma^2).
\end{aligned}
$$

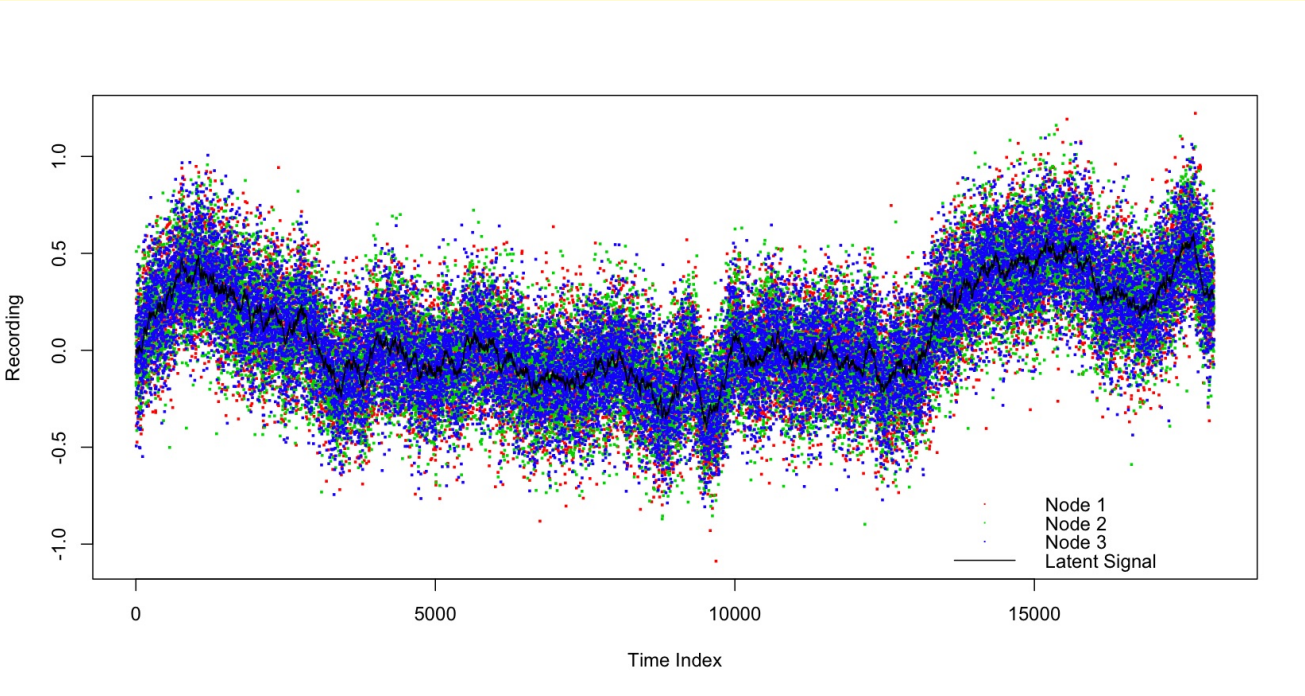
  - Generalise mean of Y_t via propagation model to $(g(X_t,\theta), g(X_t,\theta), g(X_t,\theta))^T$;
  - Σ reflects spatial correlation in errors: $\Sigma_{ij} = \sigma_i \sigma_j e^{-rd(i,j)}$;

Extending to a transmitter–receiver model for a laboratory cell

- X_t is a Gaussian Markov random field in 1d;
- Model is an example of a latent GMRF;
- It's an empirical model but good for our purposes;
- Fast Bayesian computation is possible with these models through a generalisation of the Laplace approximation.

Simulated data from latent GMRF model

Parameters $\phi = 0.35$, $\sigma_\epsilon = 0.0042$, $\sigma_i^2 = 0.04$;
Receiver distances $d(1,2) = 1$, $d(1,3) = 3$, $d(2,3) = \sqrt{10}$.



Model fitting

- Model parameters $\theta = (\phi, \sigma_\epsilon^2, r, \sigma^2)$;
- Observe $Y_{1:T} = y_{1:T} = (y_1, \ldots, y_T)$;
- Interest is in learning about:
  - θ, for predictive purposes;
  - ...and possibly $X_{1:T} = (X_1, \ldots, X_T)$, for smoothing purposes;
- Bayesian approach: compute

$$
\begin{aligned}
p(\theta, X_{1:T} \,|\, y_{1:T}) \;&\propto\; p(y_{1:T} \,|\, X_{1:T}, \theta)\, p(X_{1:T} \,|\, \theta)\, p(\theta) \\
&\propto \left(\prod_{t=1}^{T} p(y_t \,|\, X_t, \Sigma) \right) \left(\prod_{t=3}^{T} p(X_t \,|\, X_{t-1}, X_{t-2}, \phi, \sigma_\epsilon) \right) \\
&\qquad \times\, p(X_1, X_2 \,|\, \phi, \sigma_\epsilon)\, p(\theta).
\end{aligned}
$$

- Typically computed via a slow MCMC simulation.

Integrated nested Laplace approximation (INLA)

- Appropriate for latent Gaussian MRF models of the general form

$$\begin{aligned} Y_t \,|\, X_t, \theta &\sim \text{some distribution;} \\ X_{1:T} \,|\, \theta &\sim \textit{GMRF}; \\ \theta &\sim \text{some distribution} \end{aligned}$$

  where $\dim(\theta) \leq 5$ or so;
- Computes a functional approximation to:

$$p(\theta \,|\, Y_{1:T} = y_{1:T})$$

  and

$$p(X_t \,|\, Y_{1:T} = y_{1:T}), t = 1, \ldots, T.$$

- Is **not** a Monte Carlo method $\Rightarrow$ fast;

INLA: approximating $p(\theta \mid y_{1:T})$

- A Gaussian approximation $p_G(X_{1:T} \mid y_{1:T}, \theta)$ to $p(X_{1:T} \mid y_{1:T}, \theta)$ is found:
  - Find $x^*_{1:T}(\theta) = \arg\max p(X_{1:T} \mid y_{1:T}, \theta)$;
  - The mean of p_G is $x^*_{1:T}(\theta)$;
  - The precision matrix of p_G is the Hessian of $\log(p(X_{1:T} \mid y_{1:T}, \theta))$ at $x^*_{1:T}(\theta)$;
- Note that for any $X_{1:T}$:

$$p(\theta \mid y_{1:T}) \propto \frac{p(y_{1:T} \mid X_{1:T}, \theta)\, p(X_{1:T} \mid \theta)\, p(\theta)}{p(X_{1:T} \mid y_{1:T}, \theta)}$$

INLA: approximating $p(\theta \mid Y_{1:T})$

- So up to a constant

$$
\begin{aligned}
p(\theta|y_{1:T}) &\approx p_{\text{INLA}}(\theta \mid y_{1:T}) \\
&\propto g_{\text{INLA}}(\theta \mid y_{1:T}) \\
&= \left. \frac{p(y_{1:T} \mid X_{1:T}, \theta)\, p(X_{1:T} \mid \theta)\, p(\theta)}{p_G(X_{1:T} \mid y_{1:T}, \theta)} \right|_{X_{1:T}=x^*_{1:T}(\theta)}
\end{aligned}
$$

- Find the mode of $g_{\text{INLA}}(\theta \mid y_{1:T})$ by some search algorithm;
- Note that each evaluation of $g_{\text{INLA}}(\theta \mid y_{1:T})$ requires an optimisation over $X_{1:T}$;
- Explore space around mode to find high density region;
- Compute $g_{\text{INLA}}(\theta \mid y_{1:T})$ on a finite grid $\Theta = \{\theta_i \mid i \in I\}$ over the high density region;

INLA: approximating $p(\theta \mid Y_{1:T})$

- Finally, normalise over elements of Θ to get $p_{\text{INLA}}(\theta \mid y_{1:T})$on Θ:

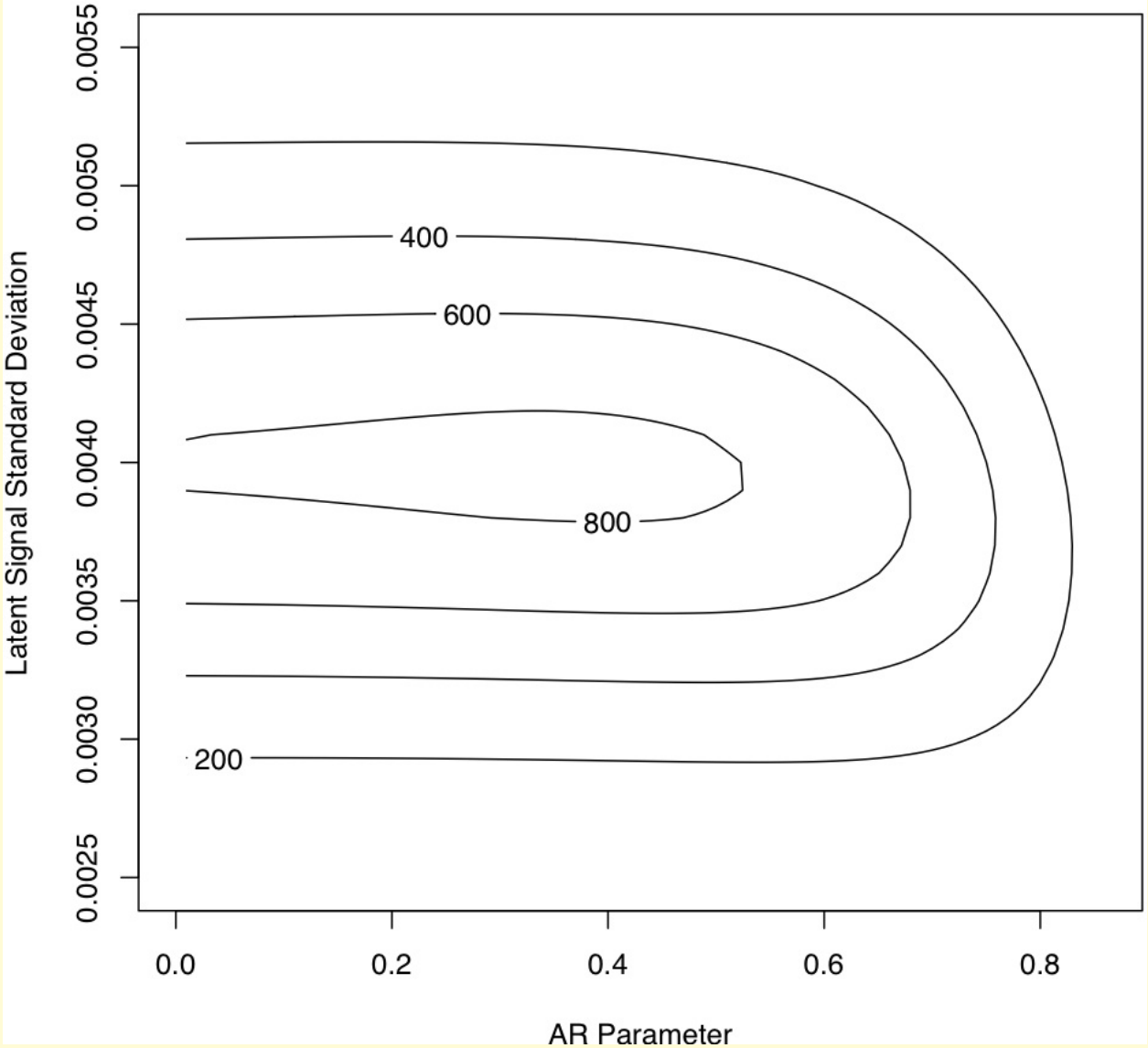
$$\begin{aligned} p(\theta_i | y_{1:T}) &\approx p_{\text{INLA}}(\theta_i \mid y_{1:T}) \\ &\approx \frac{g_{\text{INLA}}(\theta_i \mid y_{1:T})}{\sum_{j \in I} g_{\text{INLA}}(\theta_j \mid y_{1:T})}, \ i \in I. \end{aligned}$$

- $\dim(\theta)$ needs to be small — $p_{\text{INLA}}(\theta_i \mid y_{1:T})$ needs to be computed and stored over I;
- GMRF assumption on $X_{1:T}$ needed in 2 places:
  - Precision matrix of $p_G(X_{1:T} \mid y_{1:T}, \theta)$ is sparse;
  - Needed in approximation of $p(X_t \mid y_{1:T})$.

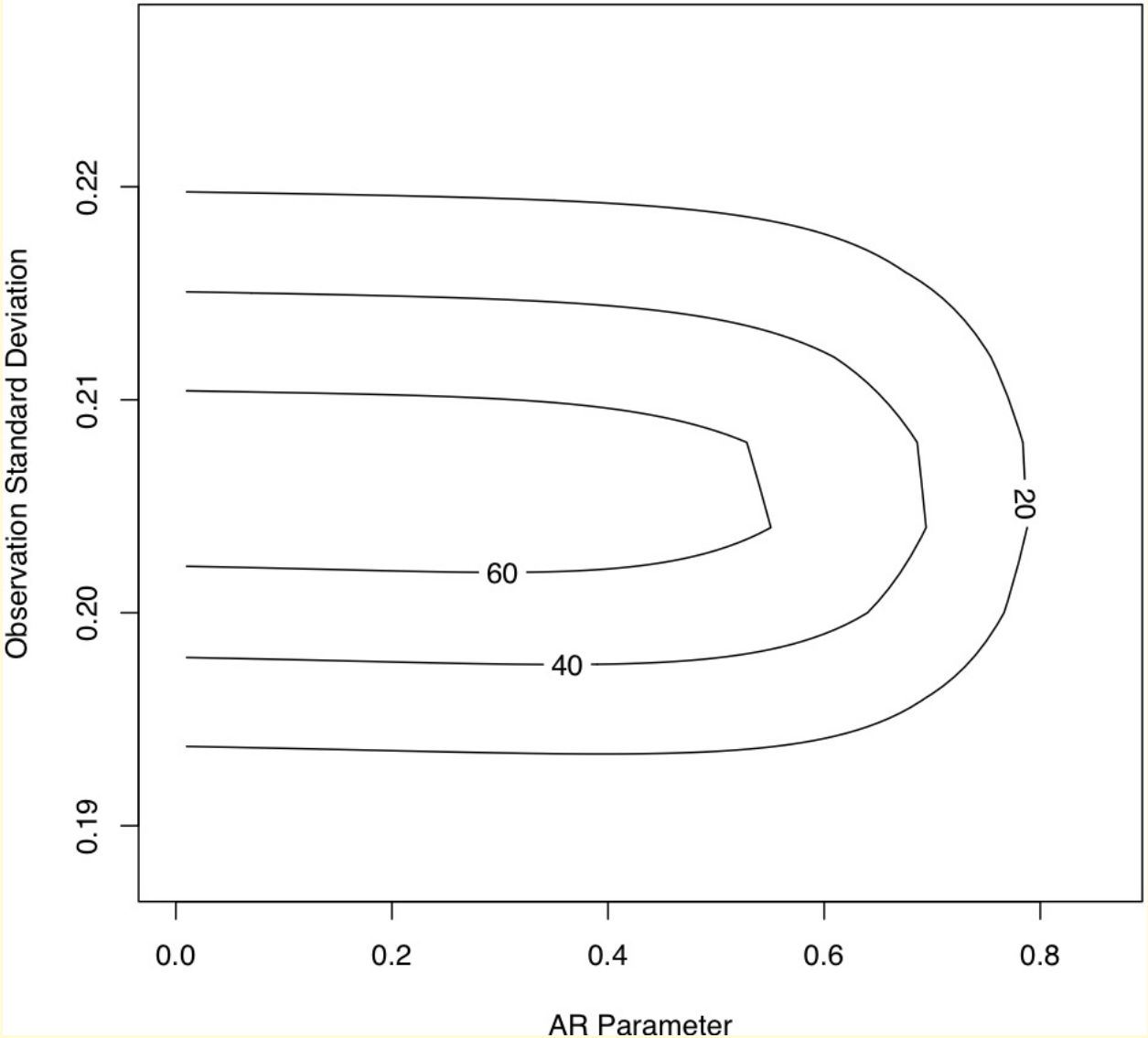
Results

- For the proposed model, p_G is exact (likelihood for Y_t is Gaussian);

Results


Latent Signal Standard Deviation
0.0025
0.0030
0.0035
0.0040
0.0045
0.0050
0.0055
400
600
800
200
0.0
0.2
0.4
0.6
0.8
AR Parameter

Results


Observation Standard Deviation
0.19
0.20
0.21
0.22
0.0
0.2
0.4
0.6
0.8
AR Parameter
60
40
20

Computation Details

- 10 minutes in badly written, very slow, R code on a 100×100 grid for ϕ and σ_ϵ^2.

Sequential INLA

- Computation time is $O(T^2)$;
- Options:
  - Fit to a data window $Y_{T-W}, \ldots, Y_T$;
  - Sequentially update INLA approximation quickly;

Sequential INLA

- Sequential INLA: quick way to update from $p_{\text{INLA}}(\theta \mid y_{1:T})$ to $p_{\text{INLA}}(\theta \mid y_{1:T+1})$?
- Look at updates of the form:

$$p_{\text{INLA}}(\theta \mid y_{1:T+1}) \;=\; g(\theta, y_{1:T+1})\, p_{\text{INLA}}(\theta \mid y_{1:T});$$

- Exact update has:

$$g(\theta, y_{1:T+1}) = \frac{p_{\text{INLA}}(\theta \mid y_{1:T+1})}{p_{\text{INLA}}(\theta \mid y_{1:T})}$$
$$= \frac{p_G(X_{1:T} \mid \theta, y_{1:T})}{p_G(X_{1:T+1} \mid \theta, y_{1:T+1})} \frac{p(X_{1:T+1} \mid \theta)}{p(X_{1:T} \mid \theta)} p(y_{T+1} \mid X_{T+1}, \theta);$$

- As much work to compute as doing $p_{\text{INLA}}(\theta \mid y_{1:T+1})$ directly.

Sequential INLA

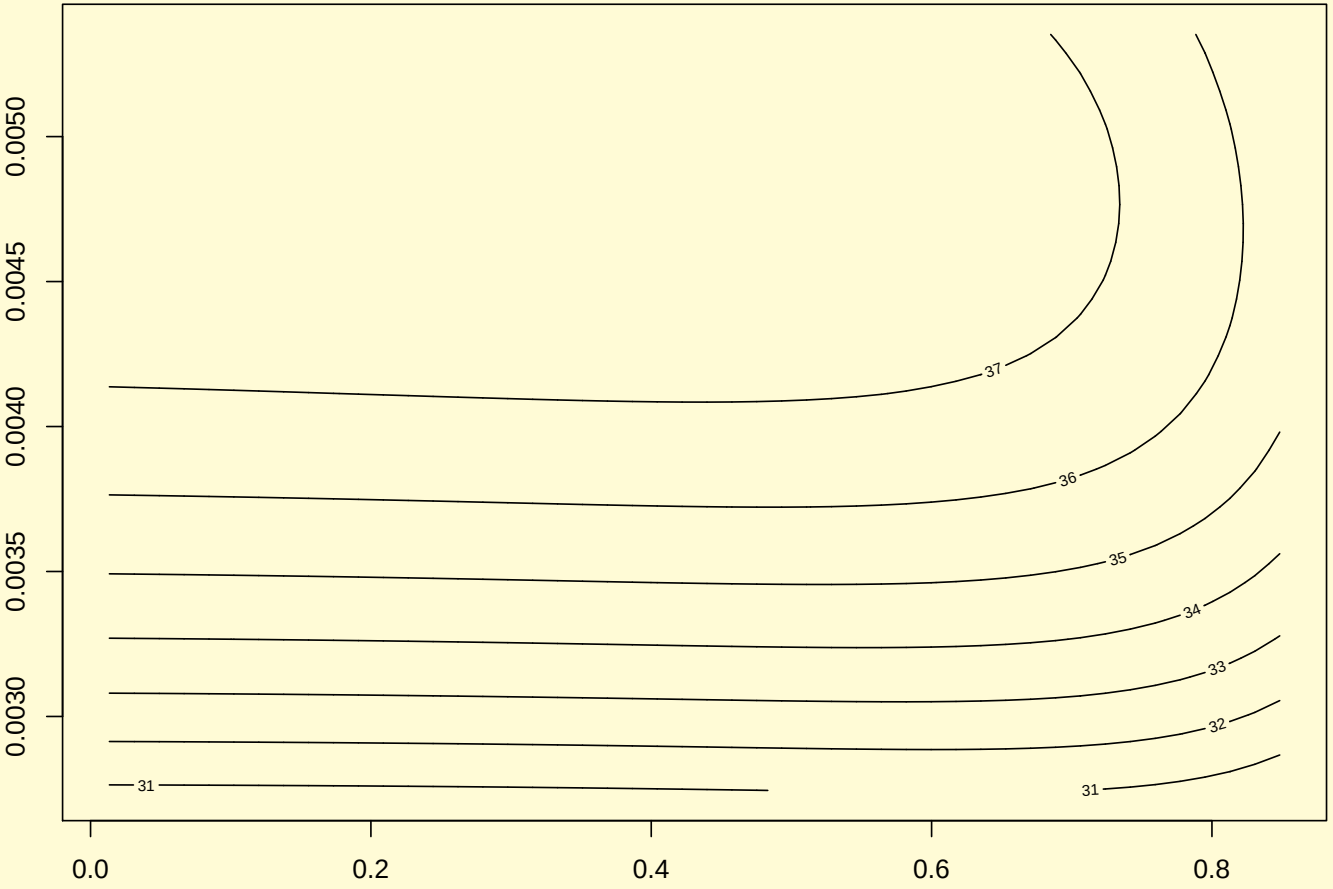
- Crude approximation:

$$g(\theta, y_{1:T+1}) \quad = \quad \frac{p(X_{T+1} \,|\, \theta)}{p_G(X_{T+1} \,|\, \theta, y_{t+1})} p(y_{T+1} \,|\, X_{T+1}, \theta);$$

- Compare to the exact g:

$$g(\theta, y_{1:T+1}) = \frac{p_G(X_{1:T} \,|\, \theta, y_{1:T})}{p_G(X_{1:T+1} \,|\, \theta, y_{1:T+1})} \frac{p(X_{1:T+1} \,|\, \theta)}{p(X_{1:T} \,|\, \theta)} \times p(y_{T+1} \,|\, X_{T+1}, \theta);$$

Sequential INLA: results



Conclusion

- INLA is a promising approach to fast Bayesian inference;
  - Works on a restricted but flexible set of models;
  - Possibility of sequential approximations to it for faster updating;
- Model proposed here is illustrative!
- Competing approaches:
  - Variational Bayes;
  - Expectation-propagation;
  - Sequential Monte Carlo;
- Exploration of sequential methods.