The University of Auckland – Applied Mathematics

Bayesian Methods for Inverse Problems : why and how

Colin Fox

Tiangang Cui, Mike O'Sullivan (Auckland), Geoff Nicholls (Statistics, Oxford)

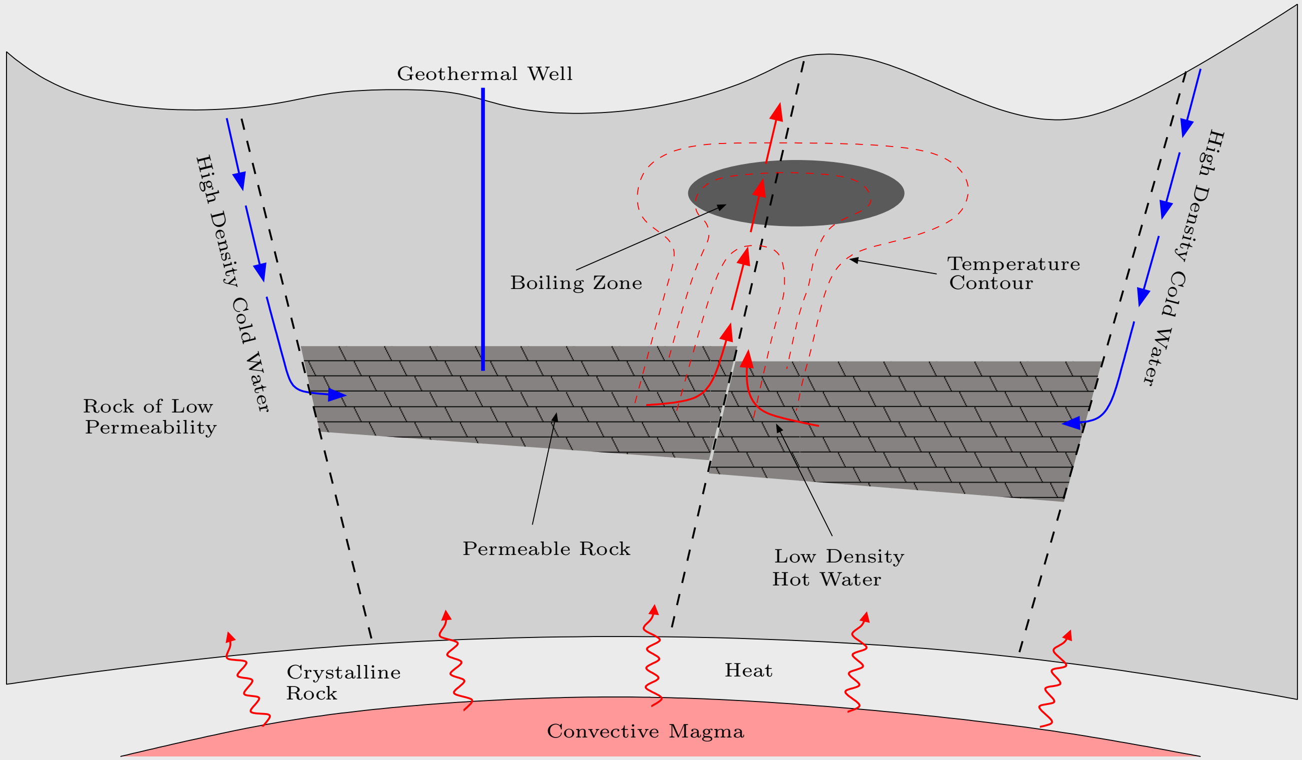
fox@math.auckland.ac.nz

In this talk

- An example in geothermal model calibration
- Inferential solutions to inverse problems, noise and all
- The need to integrate: pixel-wise degraded binary images
- Monte Carlo integration
- Computation (MCMC)
- Output analysis for geothermal problem
- Next talk :: better sampling algorithms

An inverse problem in geothermal fields

Schematic of a geothermal field



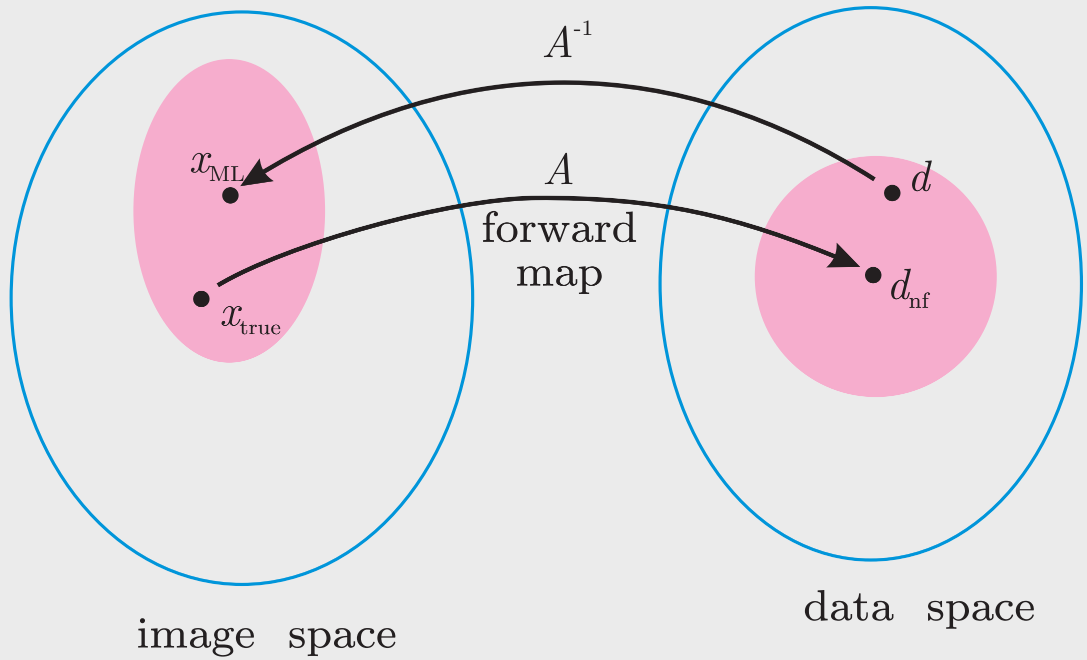
Given near-surface measurements of temperature, pressure, flow, want

- Model calibration :: rock type, porosity, fracture : heat sources
- Predict :: long term (50 year) power generation, land subsidence, etc
- Decide :: robust investment plan

Wairakei geothermal power generator

Bayesian formulation of inverse problems

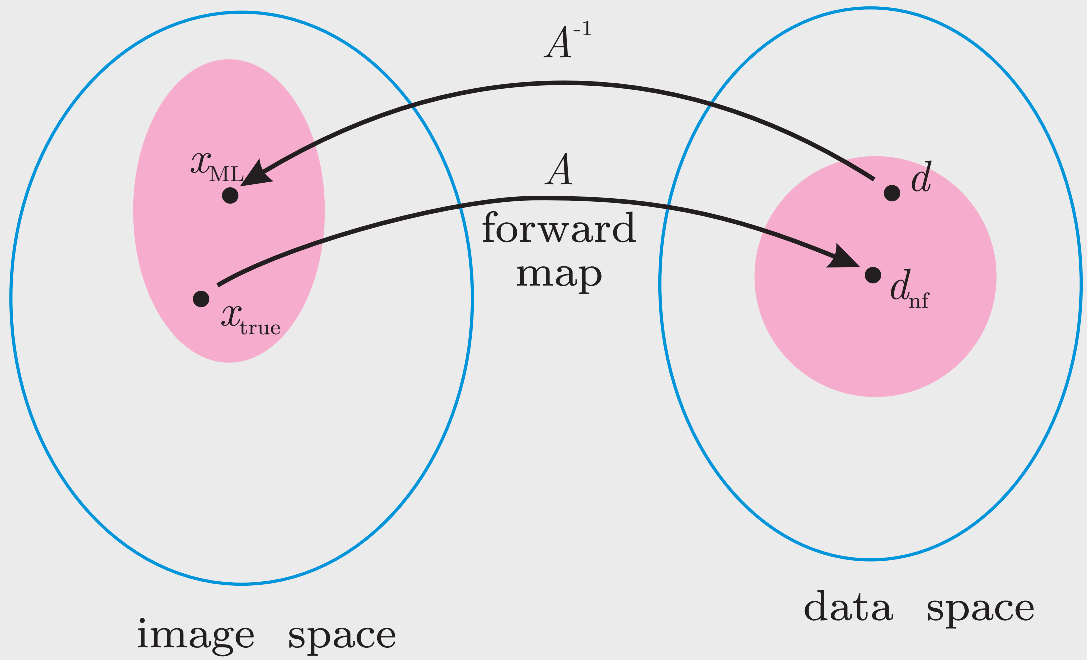
$d = Ax + n$: data d, image x, measurement noise n, forward map A



$$\pi\,(x|d,m) \propto \mathrm{Pr}\,(d|x,m)\mathrm{Pr}\,(x|m) \qquad \text{(Bayes' rule)}$$

Bayesian formulation of inverse problems

$d = Ax + n$: data d, image x, measurement noise n, forward map A

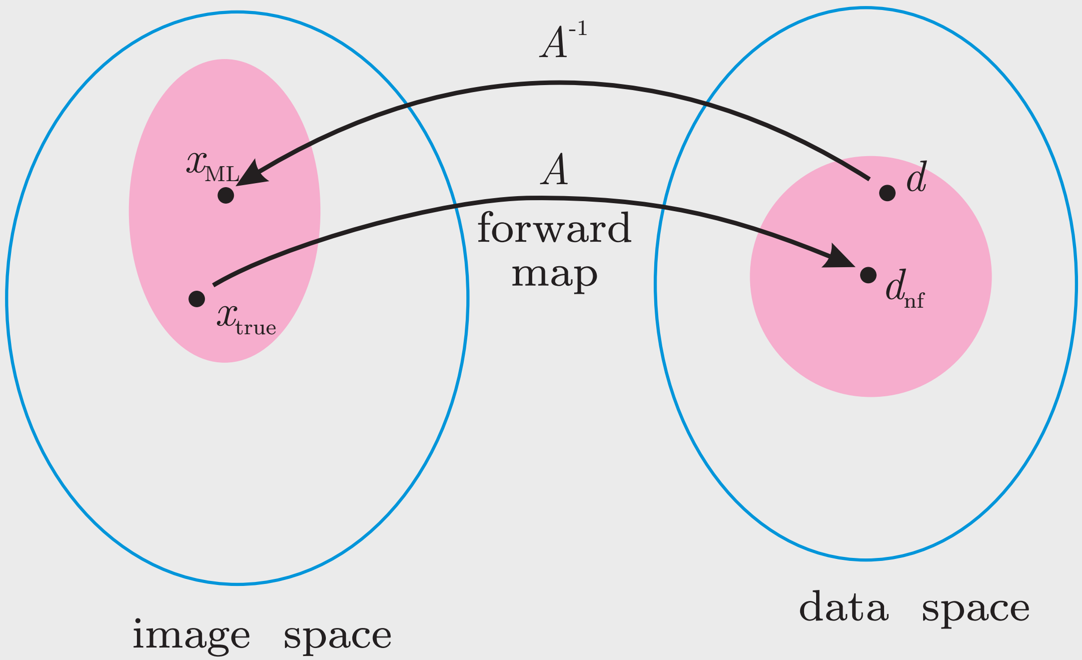


$$\pi(x|d,m) \propto \mathrm{Pr}(d|x,m)\,\mathrm{Pr}(x|m) \qquad \text{(Bayes' rule)}$$

Likelihood: measurement and noise (Physics, instrumentation, probability)

Bayesian formulation of inverse problems

$d = Ax + n$: data d, image x, measurement noise n, forward map A



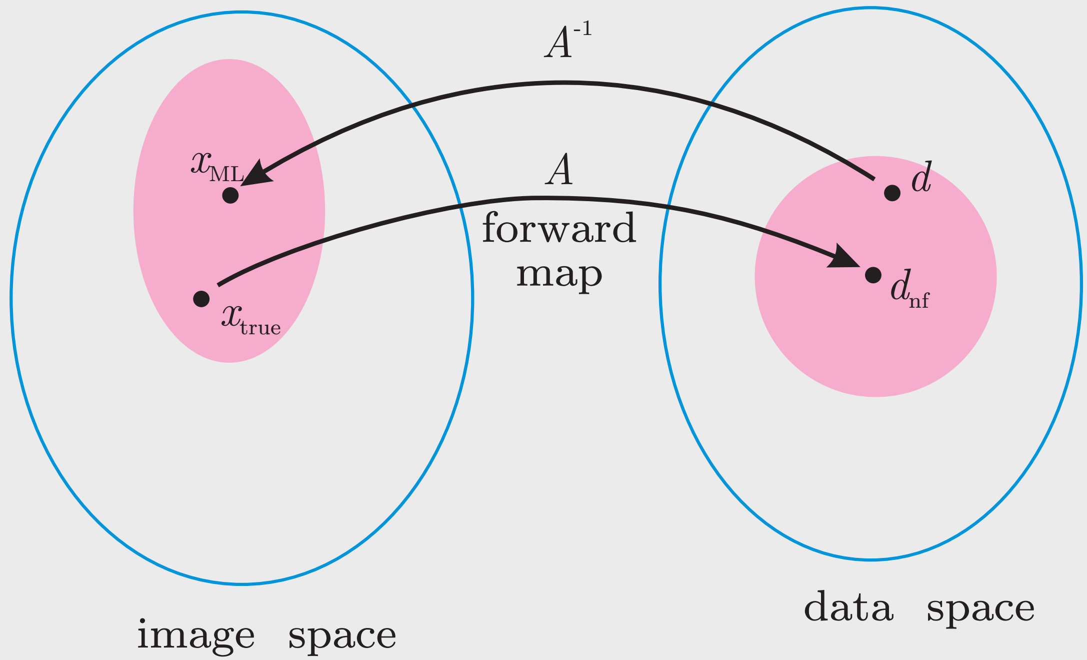
$$\pi\,(x|d,m) \propto \mathrm{Pr}\,(d|x,m)\mathrm{Pr}\,(x|m) \qquad \text{(Bayes' rule)}$$

Likelihood: measurement and noise (Physics, instrumentation, probability)

Prior / state space: stochastic modelling, physical laws, previous measurements, expert opinion

Bayesian formulation of inverse problems

$d = Ax + n$: data d, image x, measurement noise n, forward map A



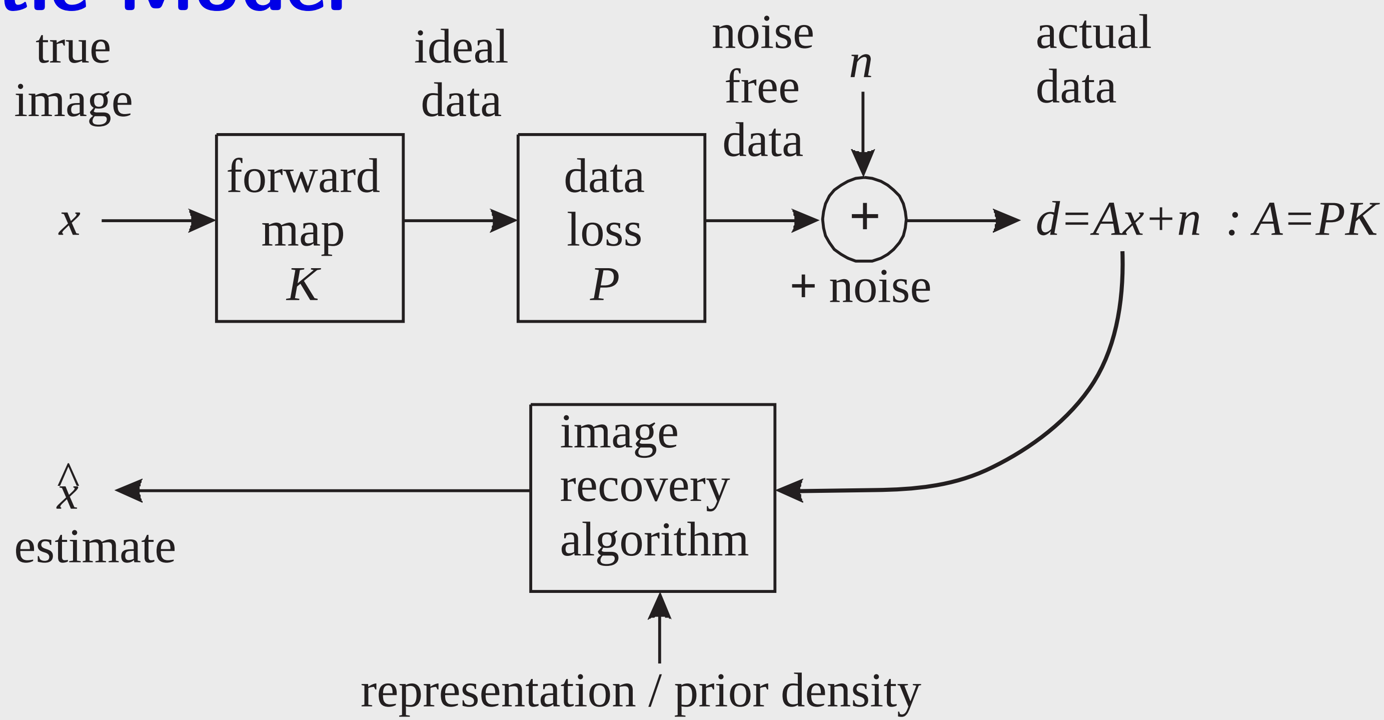
$$\pi\,(x|d,m) \propto \Pr\,(d|x,m)\Pr\,(x|m) \qquad \text{(Bayes' rule)}$$

Likelihood: measurement and noise (Physics, instrumentation, probability)

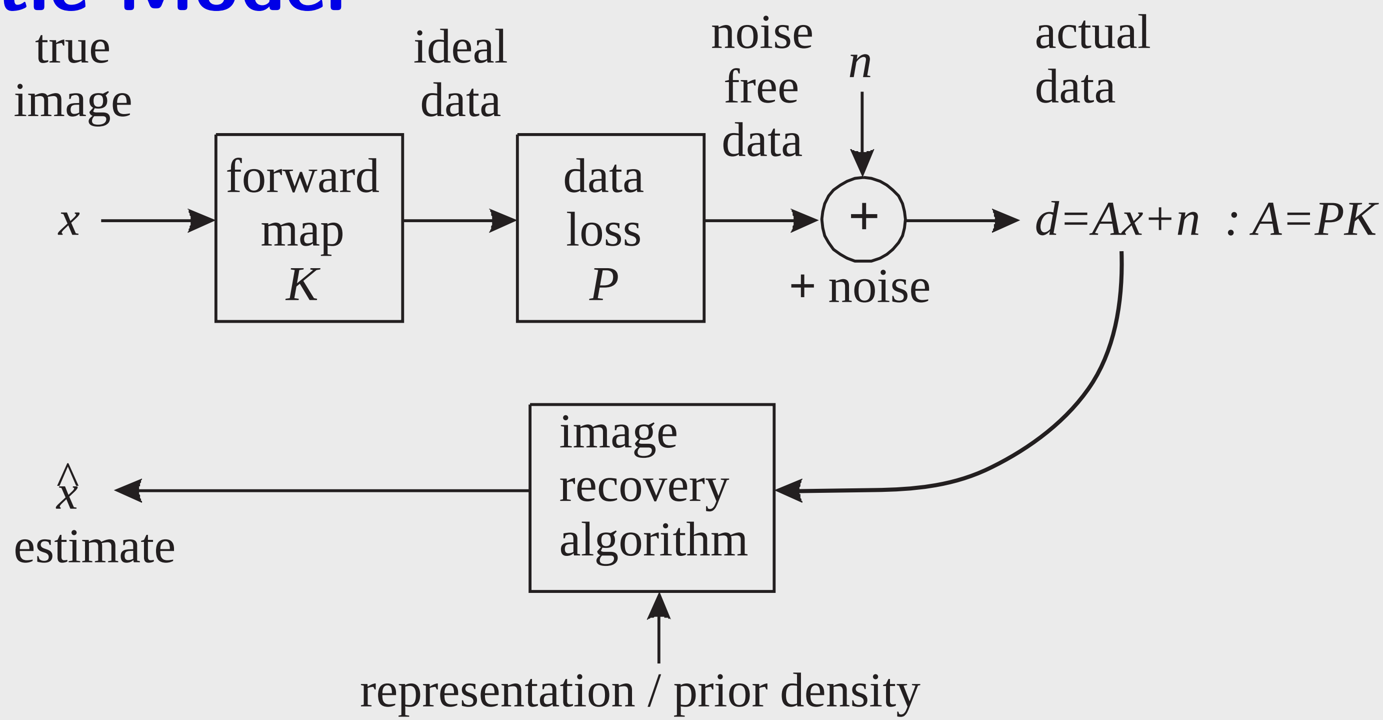
Prior / state space: stochastic modelling, physical laws, previous measurements, expert opinion

Inference based on posterior distribution conditioned on data (computational statistics)

Stochastic Model

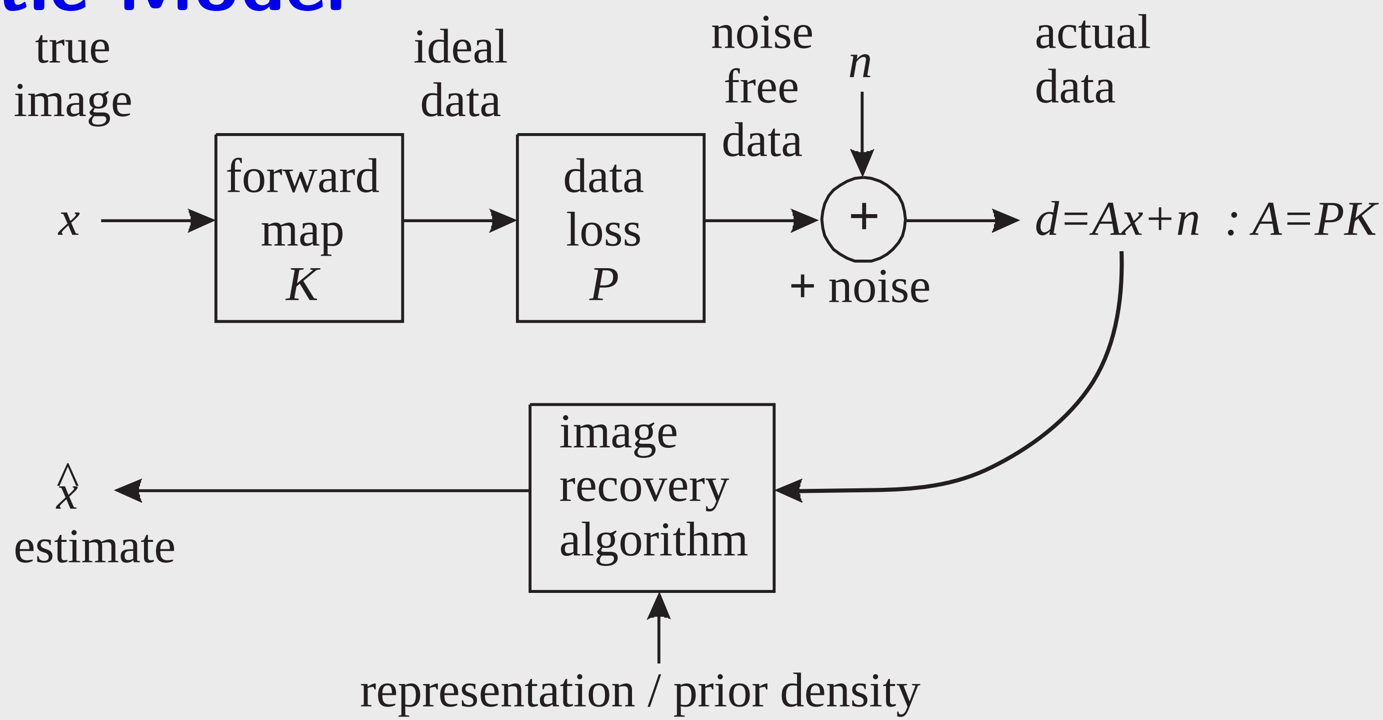


Stochastic Model



if $n \sim f_N(\cdot)$ $\quad\quad$ $\Pr(d|x,m) = f_N\left(d - Ax\right)$

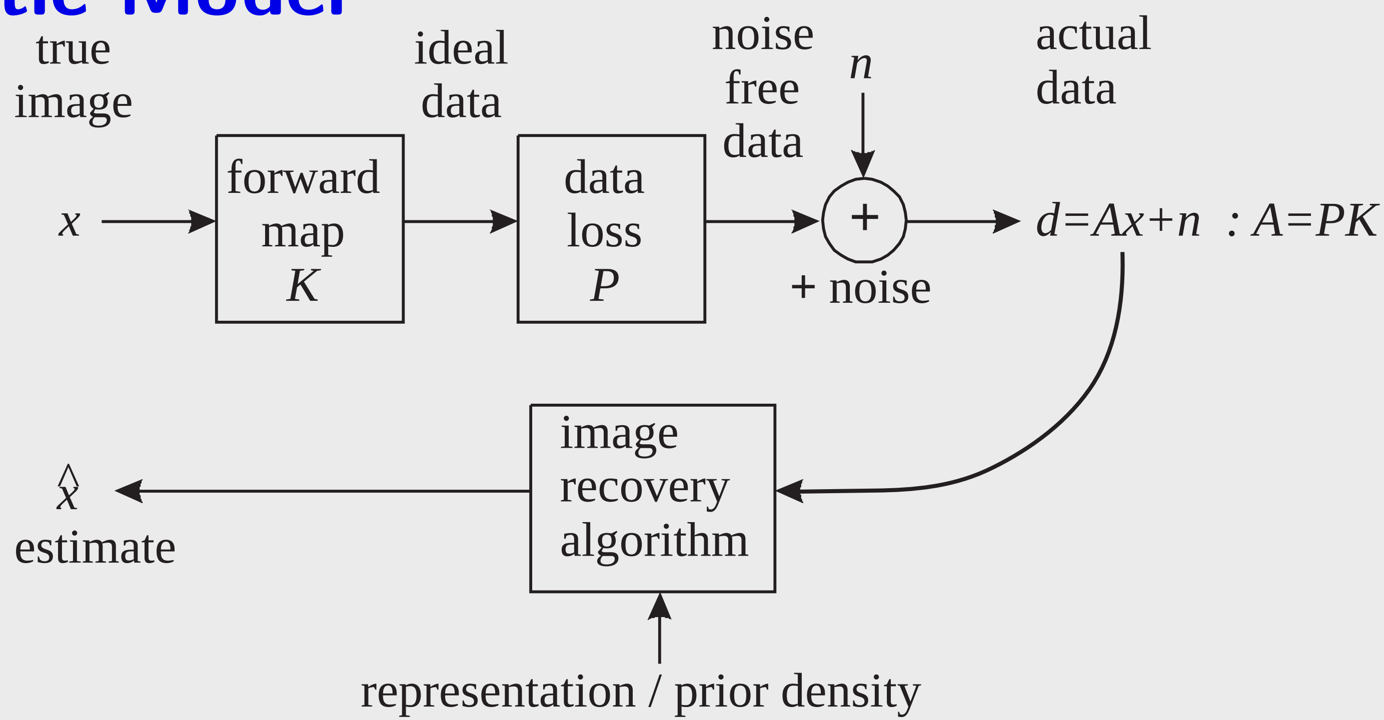
Stochastic Model



if $n \sim f_N(\cdot)$ $\Pr(d|x, m) = f_N\,(d - Ax)$

Prior distribution $\pi_{\text{pr}}(x)$ uses physical laws, expert knowledge, stochastic modelling

Stochastic Model



if $n \sim f_N(\cdot)$ $$\Pr(d|x,m) = f_N\left(d - Ax\right)$$

Prior distribution $\pi_{\text{pr}}(x)$ uses physical laws, expert knowledge, stochastic modelling

Focus of inference is posterior distribution for x given d

$$\pi\left(x|d\right) \propto \Pr\left(d|x\right) \pi_{\text{pr}}\left(x\right)$$

Solutions to Inverse Problem = Summary Statistics

Posterior distribution $\pi(x|d)$ encapsulates all information about x

Regularized Inversion - Modes

$$\hat{x}_{\mathsf{MLE}} = \arg\max \Pr(d|x) \qquad \hat{x}_{\mathsf{MAP}} = \arg\max \pi(x|d)$$

(least-squares, Moore-Penrose inverse, Tikhonov regularization, Kalman filtering, Backus-Gilbert, Prussian-hat cleaning)

Inferential Solutions

"Answers" are expectations over the posterior $\pi(x|d)$

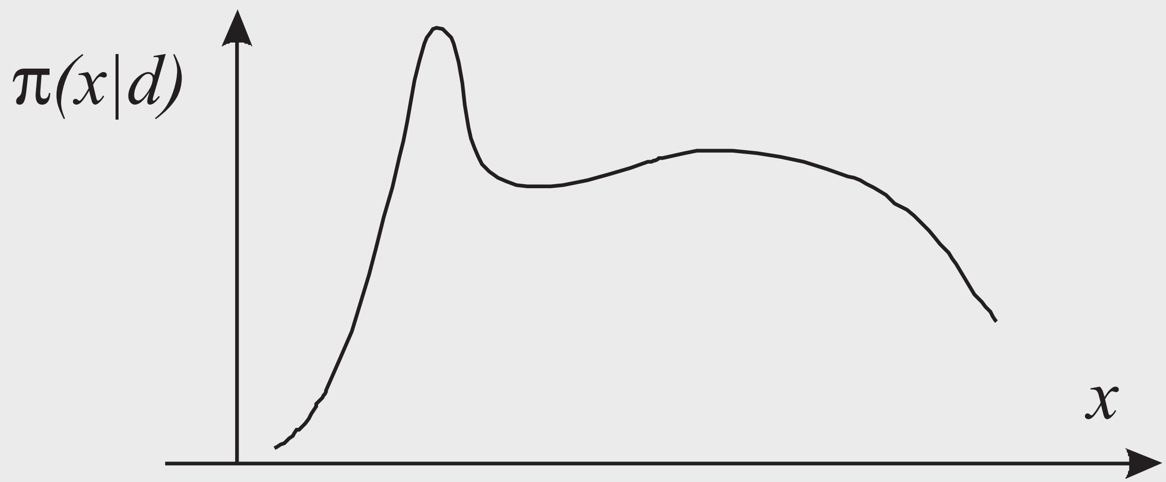
$$\mathsf{E}_\pi[f(x)] = \int_X f(x)\,\pi(x|d)\,dx$$

If

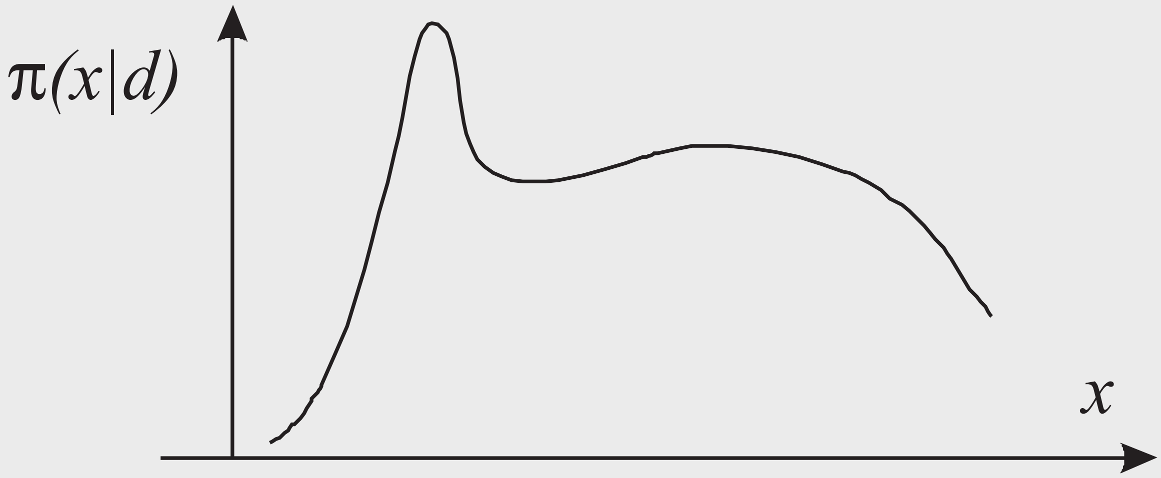
$$f(x) = \text{indicator function that image shows cancer}$$

$\mathsf{E}[f(x)]$ is posterior probability (based on measurements, prior) that patient has cancer.

Two modes over 100×100 pixel image

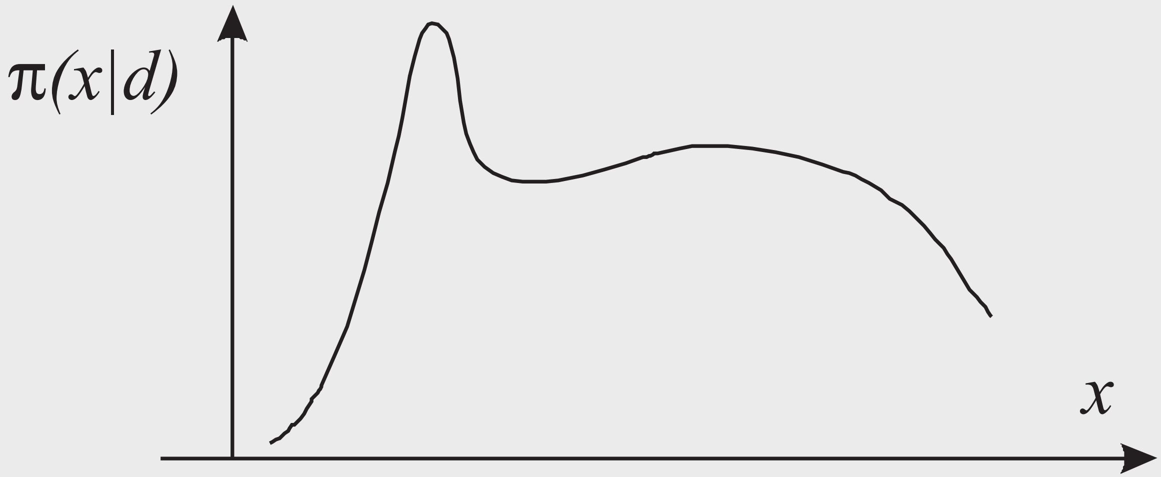


Two modes over 100×100 pixel image



if maximum value (at mode) is twice the value of the lower local maximum

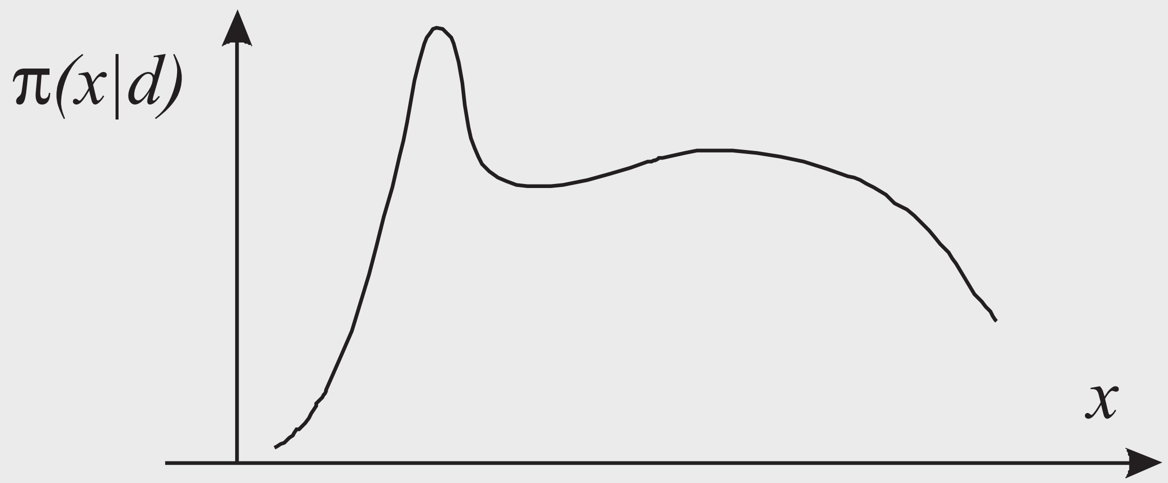
Two modes over 100×100 pixel image



if maximum value (at mode) is twice the value of the lower local maximum

width of global mode is half the width of the lower local mode

Two modes over 100×100 pixel image

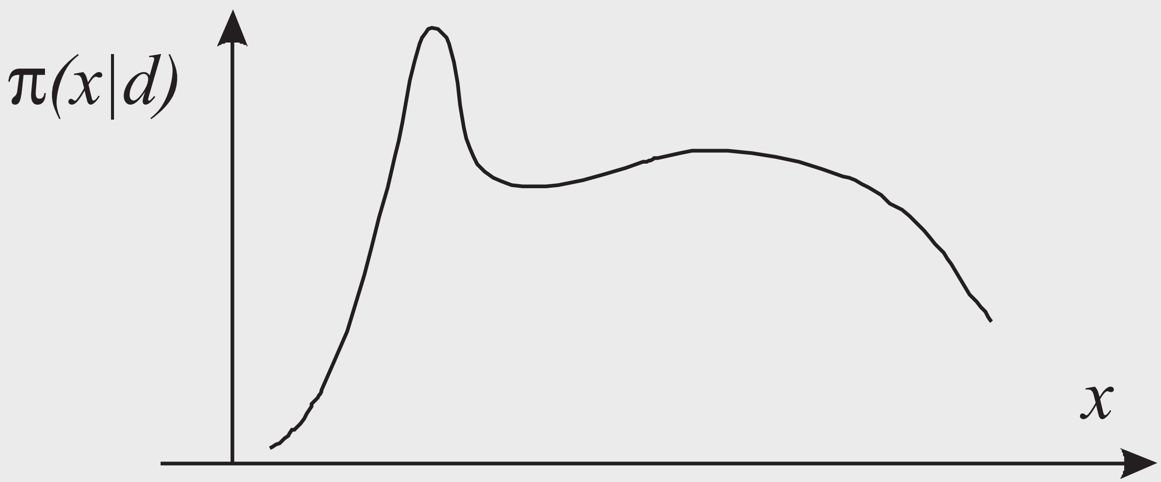


if maximum value (at mode) is twice the value of the lower local maximum

width of global mode is half the width of the lower local mode

in each of the coordinate directions $x_1, x_2, \cdots, x_{100\times 100}$

Two modes over 100×100 pixel image



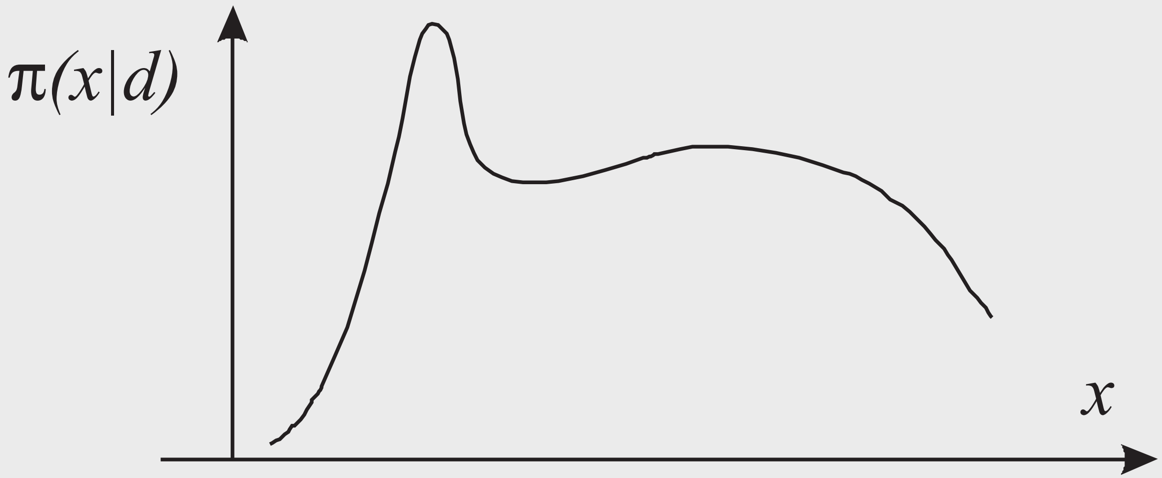
if maximum value (at mode) is twice the value of the lower local maximum

width of global mode is half the width of the lower local mode

in each of the coordinate directions $x_1, x_2, \cdots, x_{100\times100}$

there is $2^{10000-1} \approx 10^{3010}$ times *more* probability mass in the lower, broader, peak than around the mode

Two modes over 100×100 pixel image



if maximum value (at mode) is twice the value of the lower local maximum

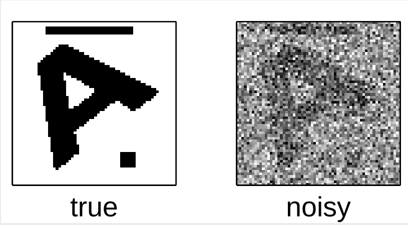
width of global mode is half the width of the lower local mode

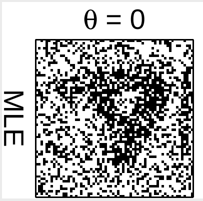
in each of the coordinate directions $x_1, x_2, \cdots, x_{100\times 100}$

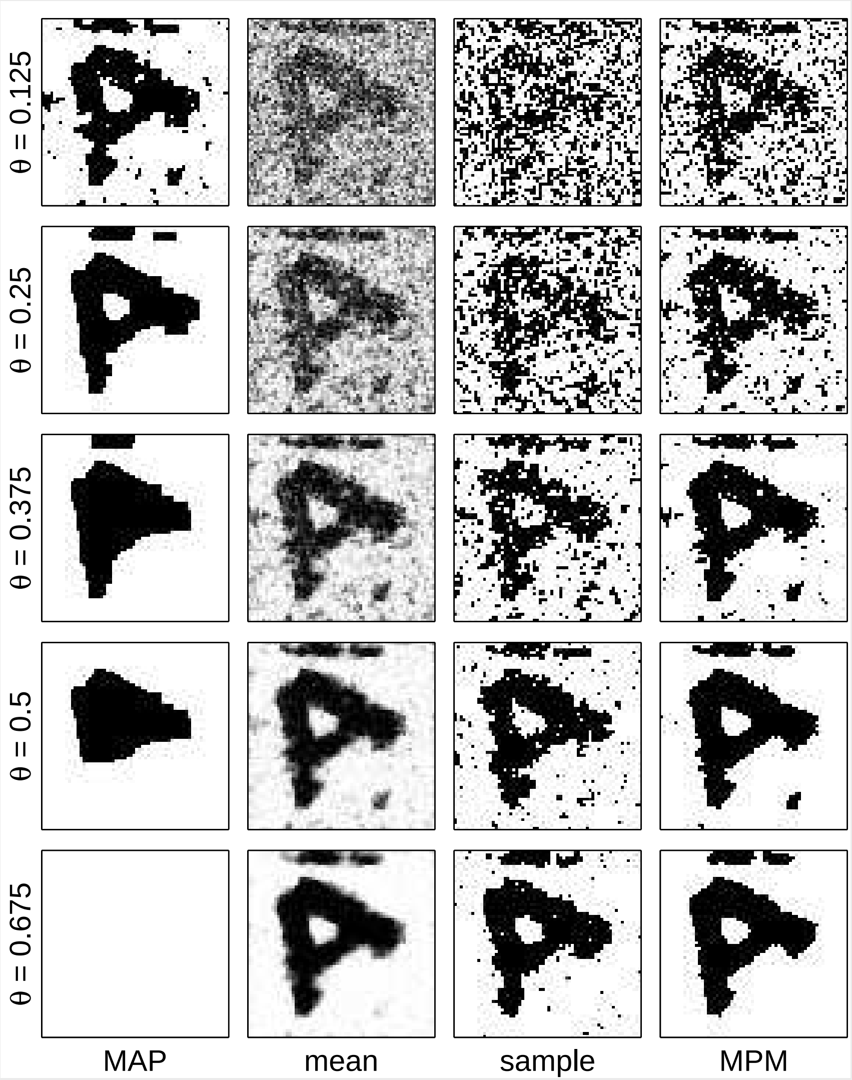
there is $2^{10000-1} \approx 10^{3010}$ times *more* probability mass in the lower, broader, peak than around the mode

$$\int_X \pi\left(\mathbf{x}|d\right)\, d\mathbf{x} = \int_X \pi\left(\mathbf{x}|d\right)\, dx_1\, dx_2 \cdots dx_{10000}$$

Mode vs Mean







Pixel-wise noise on binary image, Ising prior with "lumping" constant θ

F and Nicholls "Exact MAP states and expectations from perfect sampling: Greig, Porteous and Seheult revisited" 2001

Monte Carlo integration + importance sampling

If $x^{(1)}, \ldots, x^{(n)} \sim \pi\left(\cdot|d\right)$

$$\mathsf{E}_{\pi}\left[f\left(x\right)\right] \approx \frac{1}{n}\sum_{i=1}^{n} f\left(x^{(i)}\right)$$

Construct $x^{(1)}, \ldots, x^{(n)}$ as iterates of an ergodic map

Monte Carlo integration + importance sampling

If $x^{(1)}, \dots, x^{(n)} \sim \pi\left(\cdot|d\right)$

$$\mathsf{E}_{\pi}\left[f\left(x\right)\right] \approx \frac{1}{n}\sum_{i=1}^{n} f\left(x^{(i)}\right)$$

Construct $x^{(1)}, \dots, x^{(n)}$ as iterates of an ergodic map

Deterministic iteration: $x^{(n+1)} = M\left(x^{(n)}\right)$

$$\pi(x) = \sum_{M(y)=x} \frac{\pi(y)}{|M'(y)|} \qquad \text{(Frobenius-Perron)}$$

Stochastic Iteration: Markov chain with transition kernel $p\left(x, y\right)$

$$\int_X \pi\left(x\right) p\left(x, y\right)\, dx = \int_X \pi\left(y\right) p\left(y, x\right)\, dx \qquad \text{(global balance)}$$

$$\pi\left(x\right) p\left(x, y\right) = \pi\left(y\right) p\left(y, x\right) \qquad \text{(detailed balance)}$$

Sampling Algorithms in Physics

Hamiltonian Dynamics

Equate unknowns x to ‘position’ variables q with potential energy

$$E(q) = -T\log\left(\pi(x)\right) - T\log\left(Z\right)$$

Auxiliary 'momentum' variables p with kinetic energy $K(p) = \frac{1}{2}\|p\|^2$.

Hamiltonian $H(q,p) = E(q) + K(p)$ has canonical distribution

$$P(q,p) = \frac{1}{Z_P}\exp\left(-E(q)\right)\frac{1}{Z_K}\exp\left(-K(p)\right)$$

Hamiltonian dynamics leave $P(q,p)$ invariant $+$ stochastic transitions for ergodicity

Langevin diffusion

$$dx = s\,dB + \frac{s^2}{2}\nabla\log\left(\pi(x)\right)\,dt$$

Metropolis-Hastings algorithm

1. given state x_t at time t generate candidate state x' from a proposal distribution $q\left(.|x_t\right)$
2. With probability $\alpha\left(x_t \to x'\right) = \min\left(1, \frac{\pi(x')q\left(x_t|x'\right)}{\pi(x_t)q\left(x'|x_t\right)}\right)$ set $X_{t+1} = x'$
   otherwise $X_{t+1} = x_t$
3. Repeat

$q\left(.|x_t\right)$ can be any distribution that ensures the chain is irreducible and aperiodic.

Metropolis-Hastings algorithm

1. given state x_t at time t generate candidate state x' from a proposal distribution $q\left(.|x_t\right)$
2. With probability $\alpha\left(x_t \rightarrow x'\right) = \min\left(1, \frac{\pi(x')q\left(x_t|x'\right)}{\pi(x_t)q\left(x'|x_t\right)}\right)$ set $X_{t+1} = x'$ otherwise $X_{t+1} = x_t$
3. Repeat

$q\left(.|x_t\right)$ can be any distribution that ensures the chain is irreducible and aperiodic.

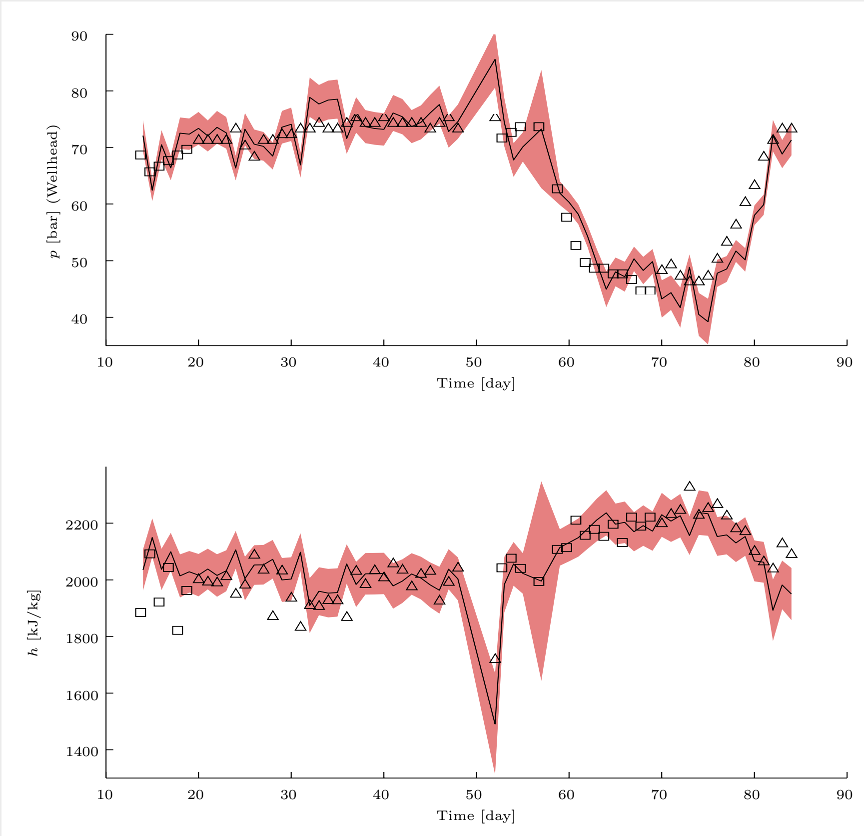
Pros

- Provably convergent under mild requirements
- State space can be continuous, discrete, stochastic, or variable dimension

Cons

- Can be (very) slow

Results for geothermal field

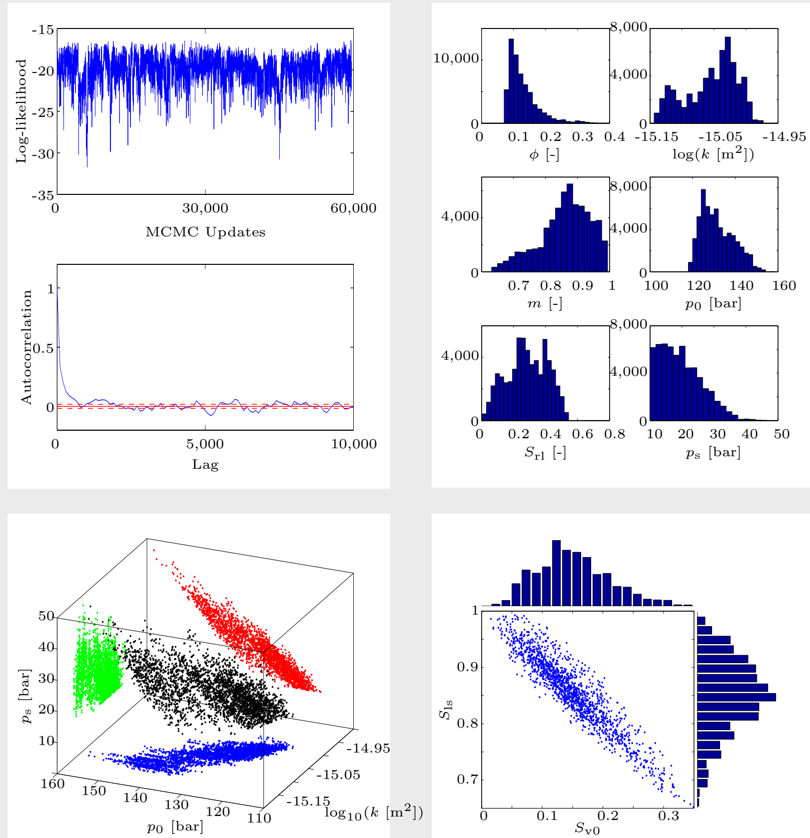


Wellhead pressure (top) and flowing enthalpy (bottom).

Black line is estimated result, triangles and squares are observed data.

Tiangang Cui, *Bayesian Inference for Geothermal Model Calibration* MEng Auckland, 2005

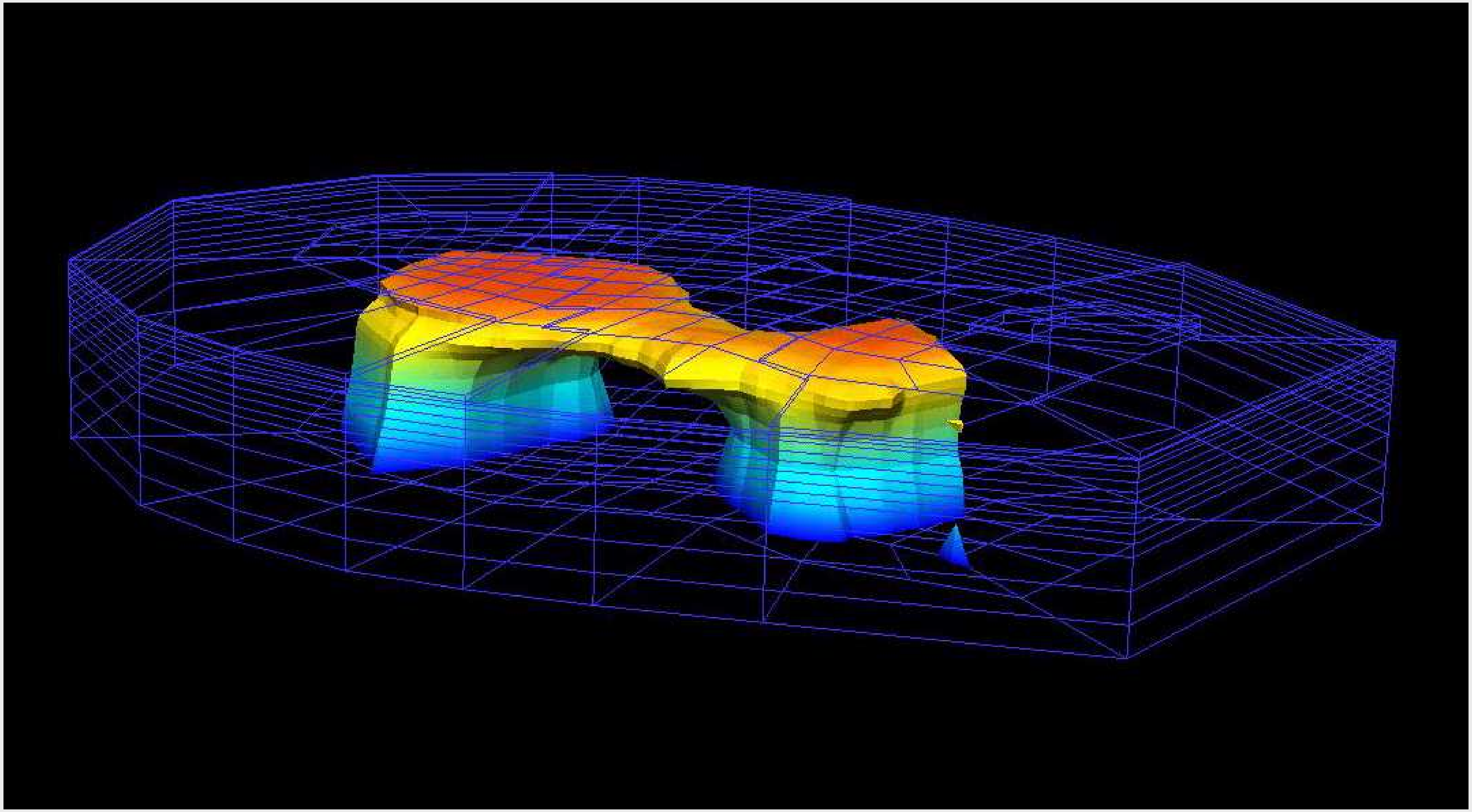
Output analysis for geothermal field



Autocorrelation and MCMC output trace. Histogram for the marginal distribution of parameters. Scatter plot of the joint marginal distribution for $\log(k)$, p_0 and p_s. Histogram for the marginal distribution of S_{v0} and S_{ls} and scatter plot of their joint marginal distribution.

Tiangang Cui, *Bayesian Inference for Geothermal Model Calibration* MEng Auckland, 2005

Inferred iso-temperature surface



Conclusions

1. Bayesian methods can solve substantial real-world inverse problems
2. Time to convergence depends on efficient simulation of the forward map, fast MCMC algorithms, good proposal distributions
3. Computation required is feasible