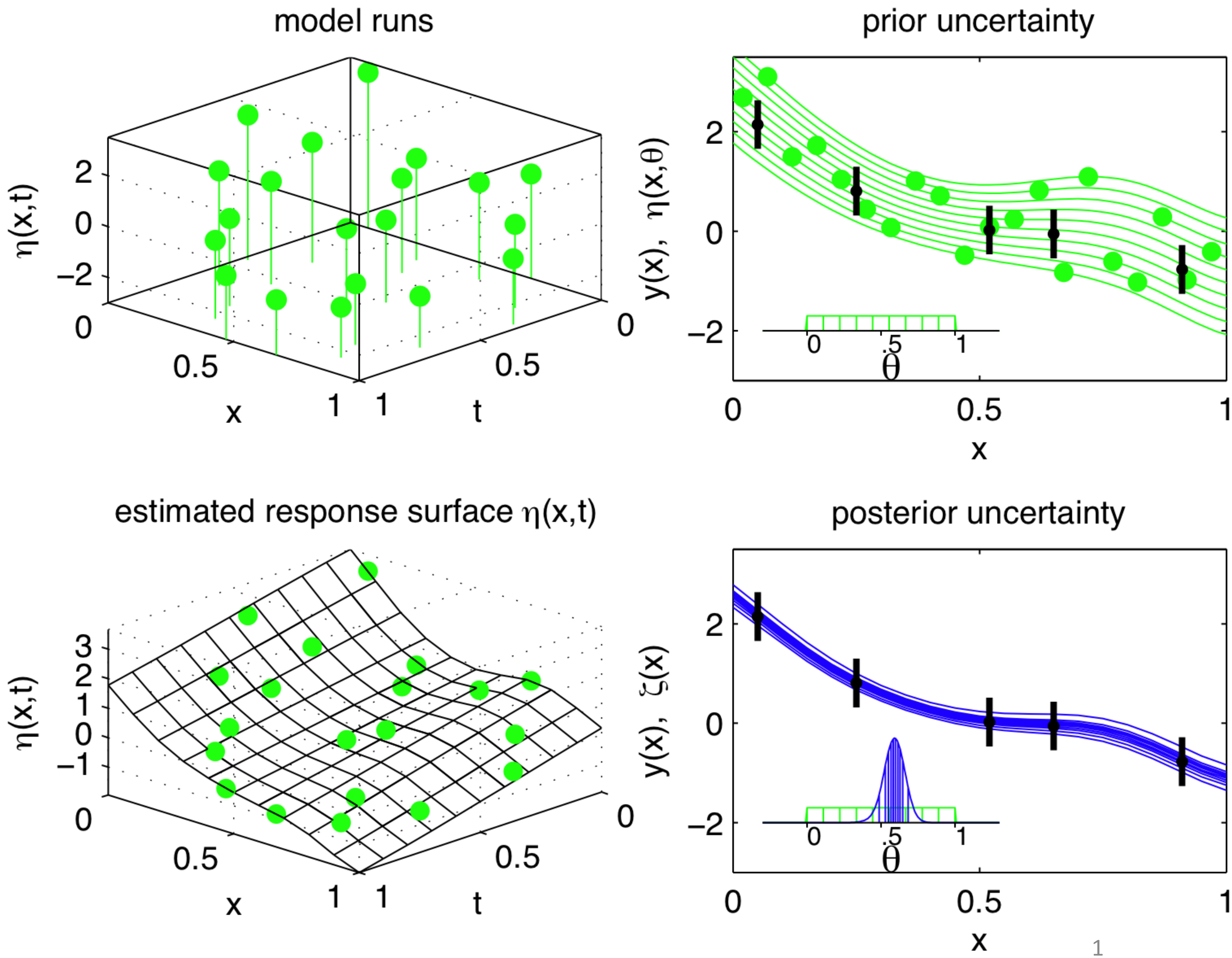
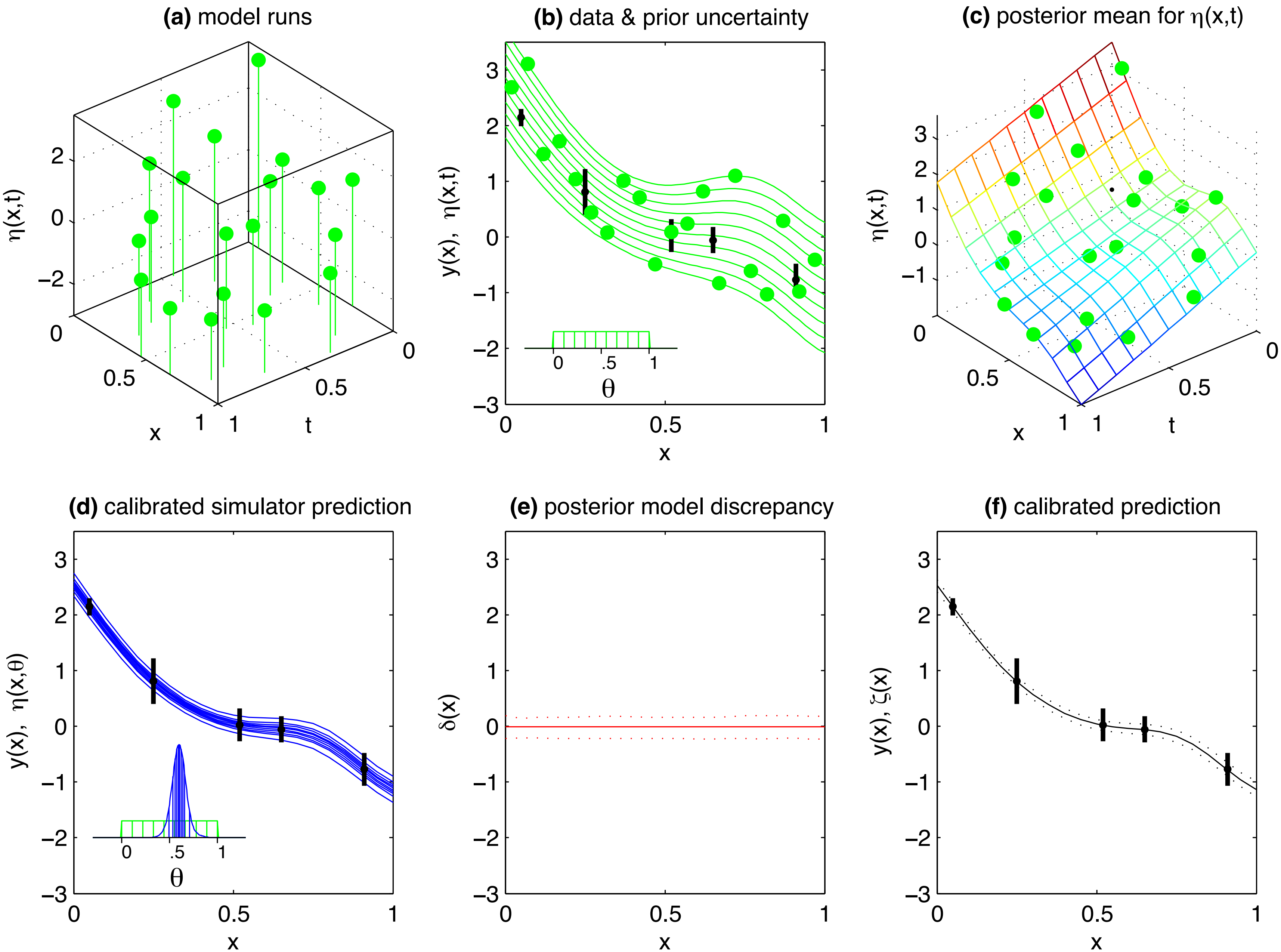
Modeling framework $y(x_i) = \eta(x_i, \theta) + \epsilon_i$

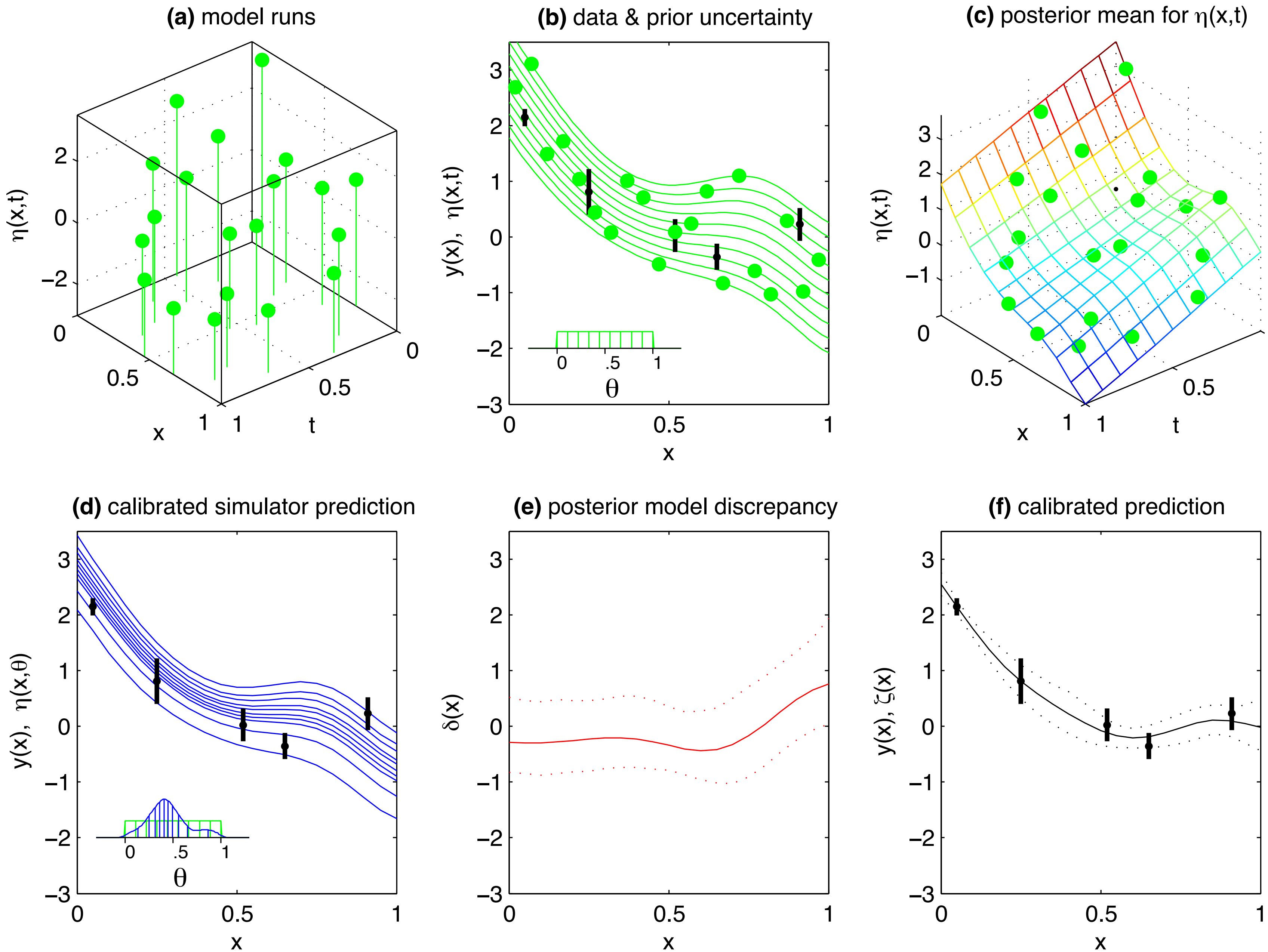


An Example

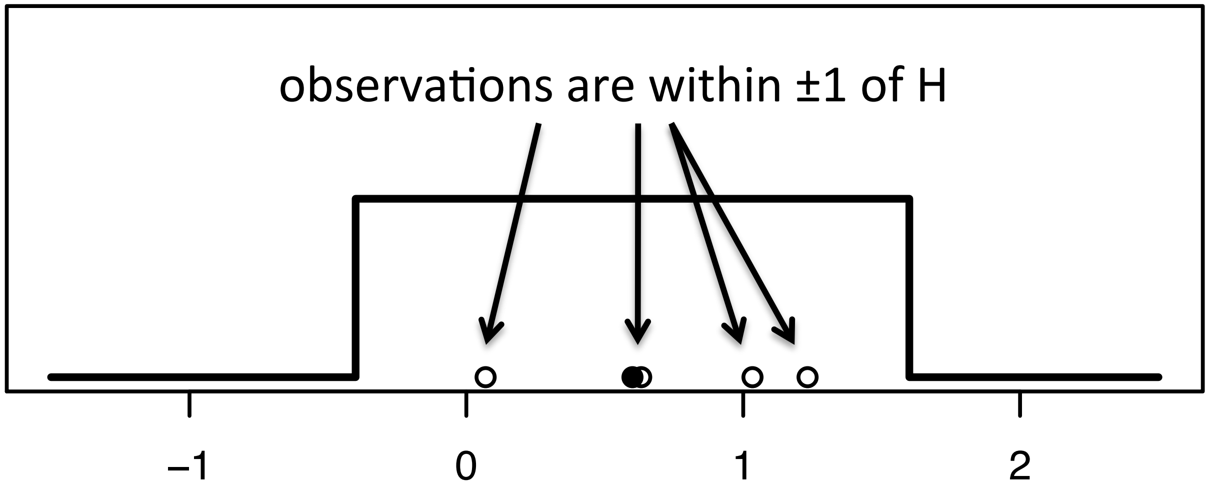


(a) model runs
η(x,t)
2
0
−2
0
0.5
x
1
1
t
0.5
0
(b) data & prior uncertainty
y(x), η(x,t)
3
2
1
0
−1
−2
−3
0
.5
1
θ
0
0.5
1
x
(c) posterior mean for η(x,t)
η(x,t)
3
2
1
0
−1
0
0.5
x
1
1
t
0.5
0
(d) calibrated simulator prediction
y(x), η(x,θ)
3
2
1
0
−1
−2
−3
0
.5
1
θ
0
0.5
1
x
(e) posterior model discrepancy
δ(x)
3
2
1
0
−1
−2
−3
0
0.5
1
x
(f) calibrated prediction
y(x), ζ(x)
3
2
1
0
−1
−2
−3
0
0.5
1
x

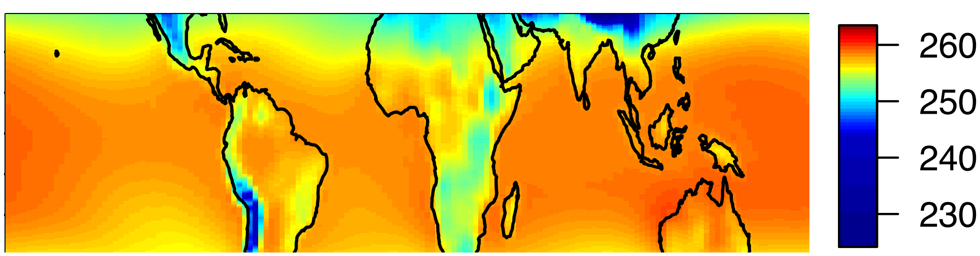
An Example - with discrepancy



Inference from observations

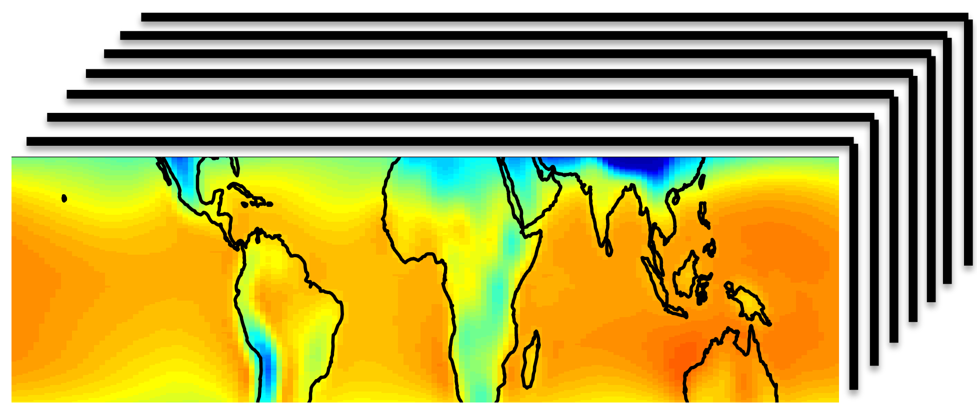


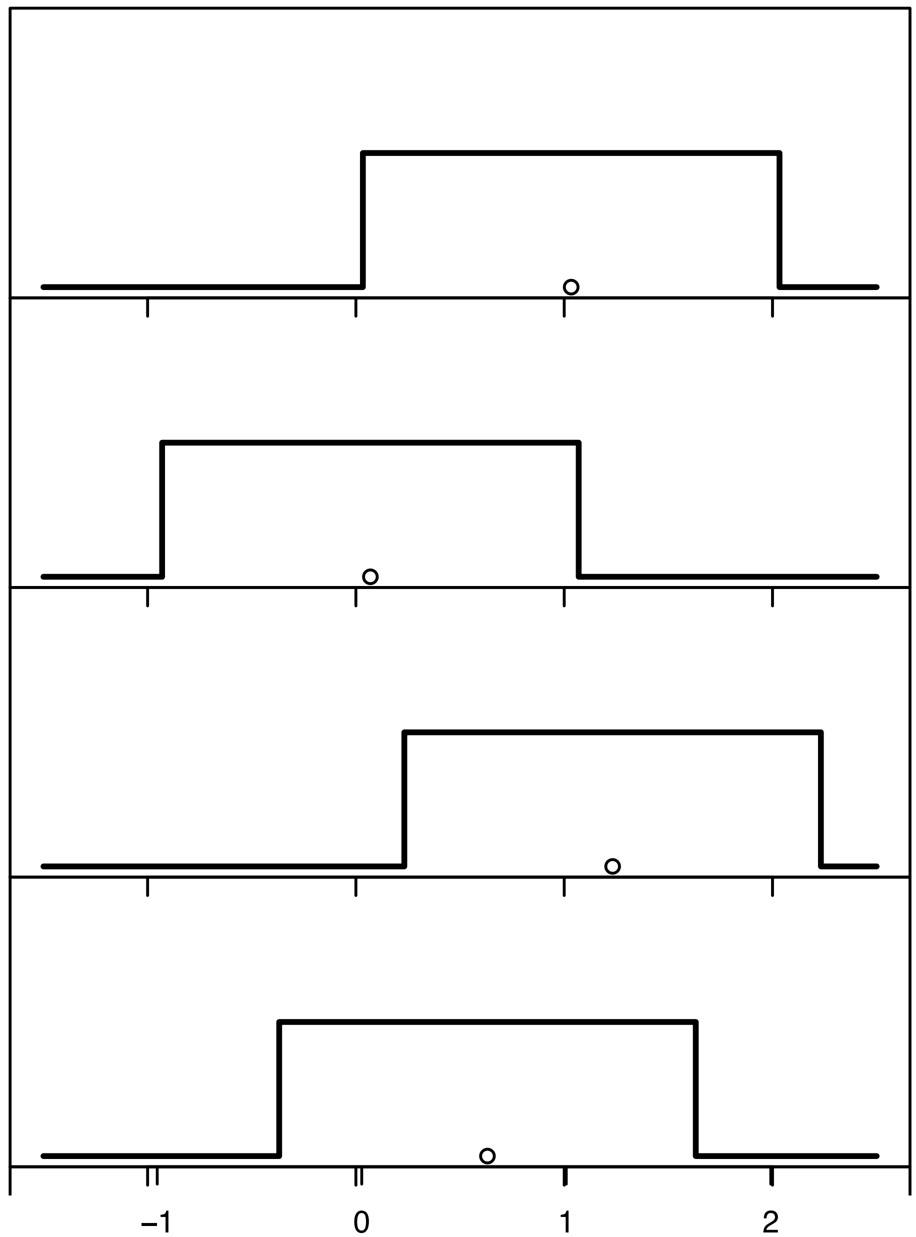
Physical observations



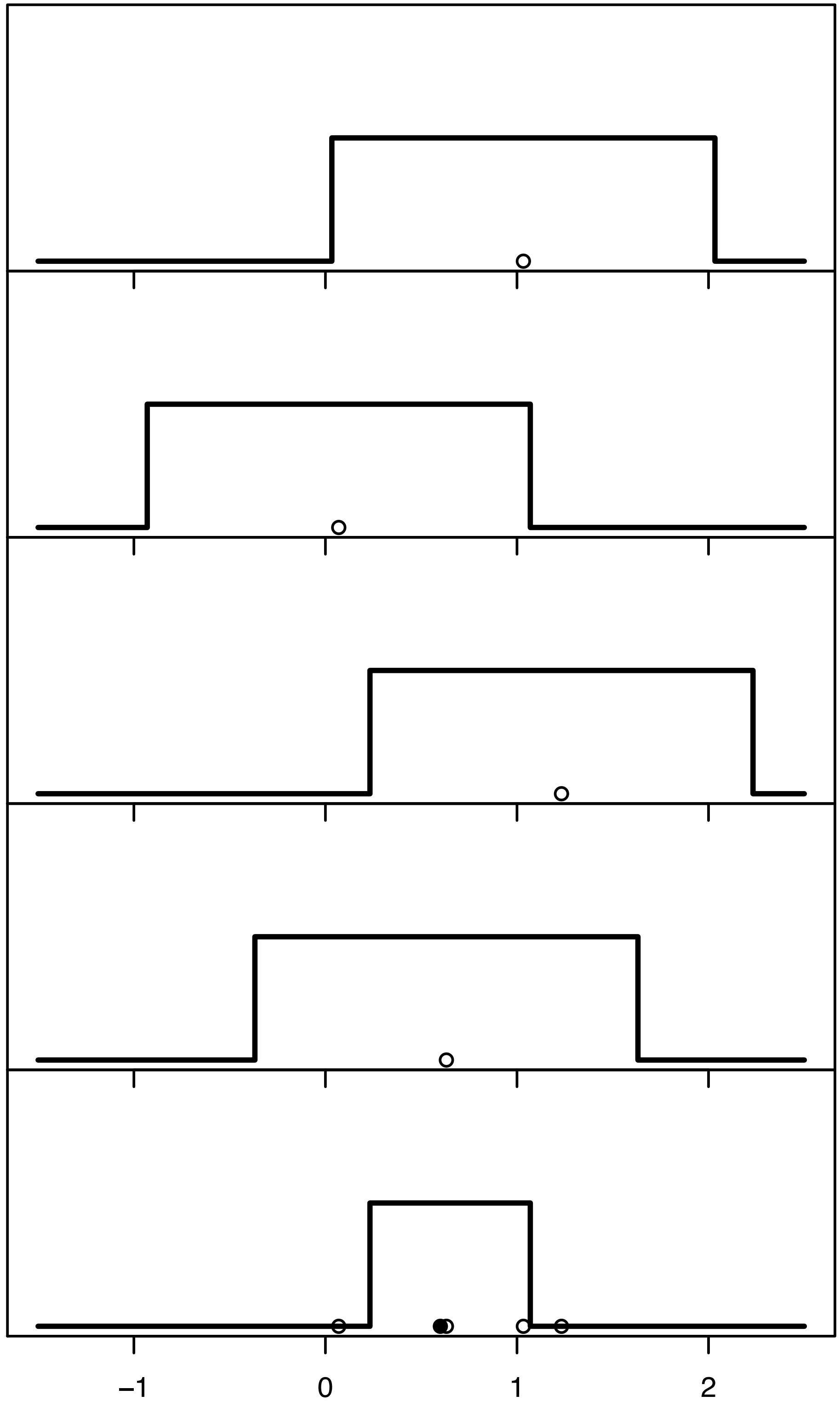
Surface temperature, winter (DJF)

CAM simulations

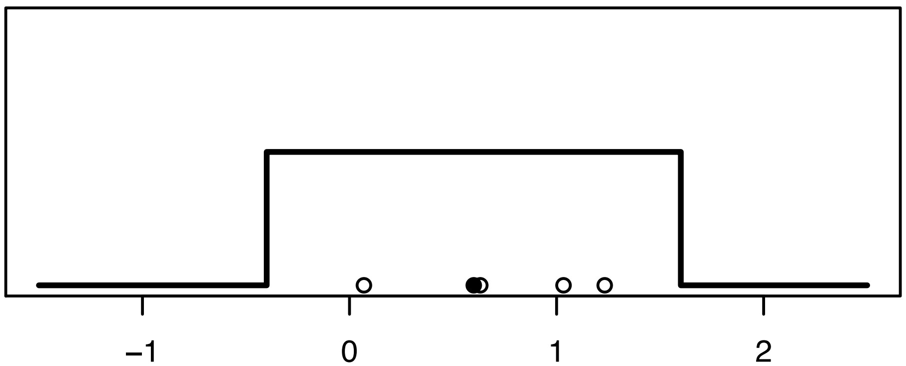


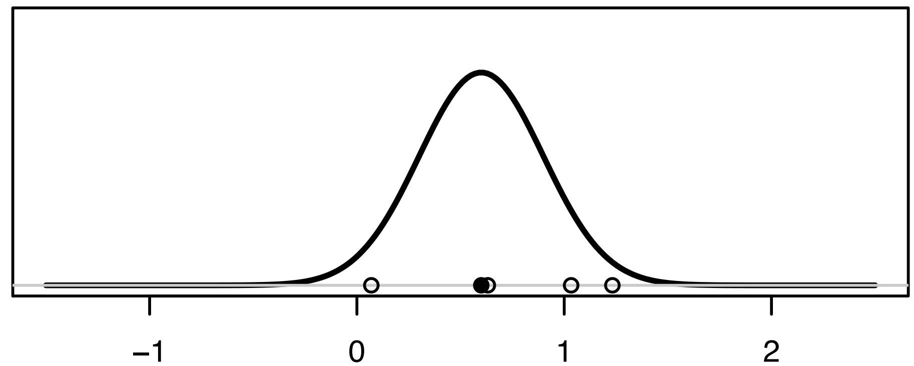


−1
0
1
2



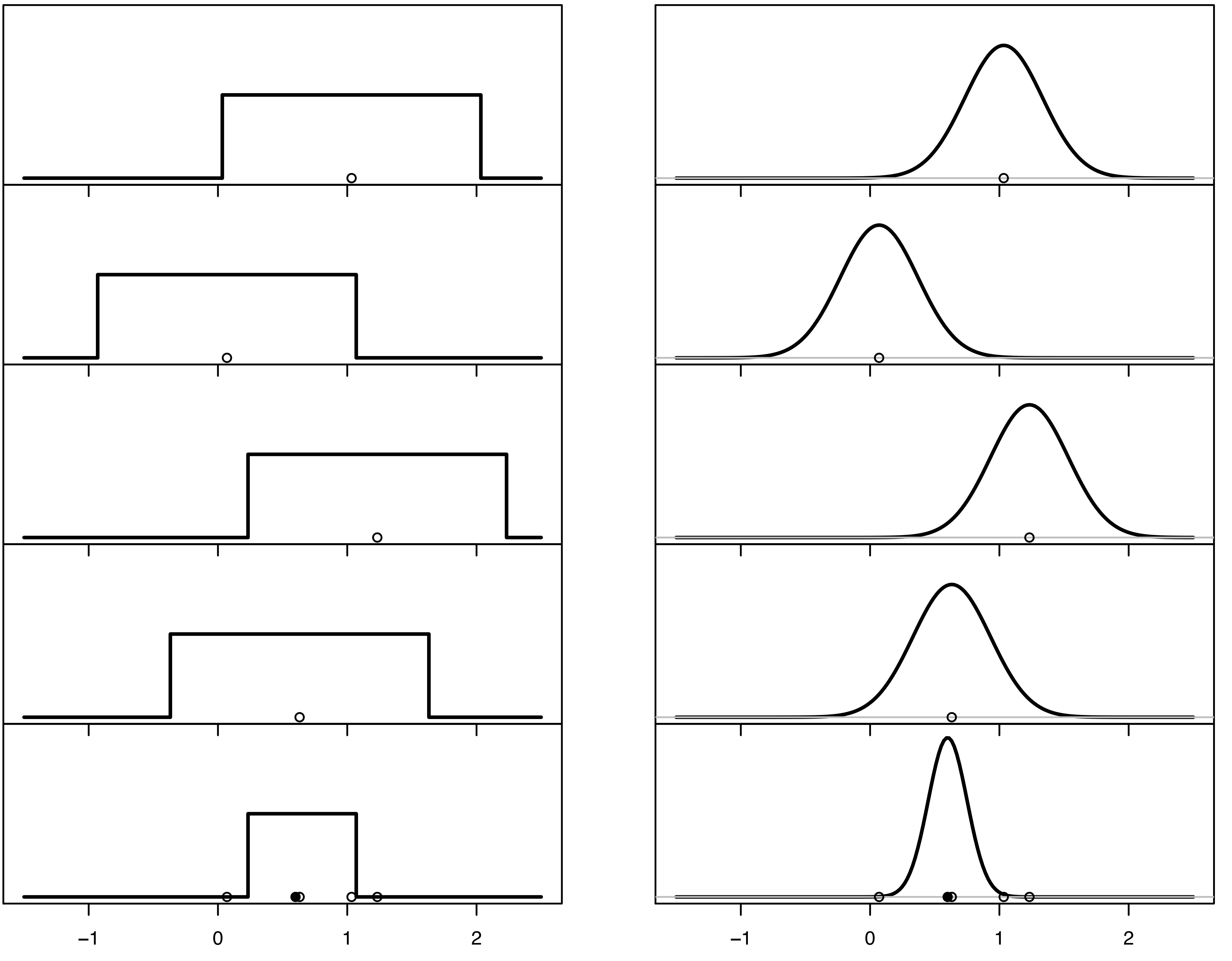
−1
0
1
2





$$y_i \text{ are iid } U[H-1, H+1] \qquad y_i \text{ are iid } N(H, \text{sd} = .3)$$

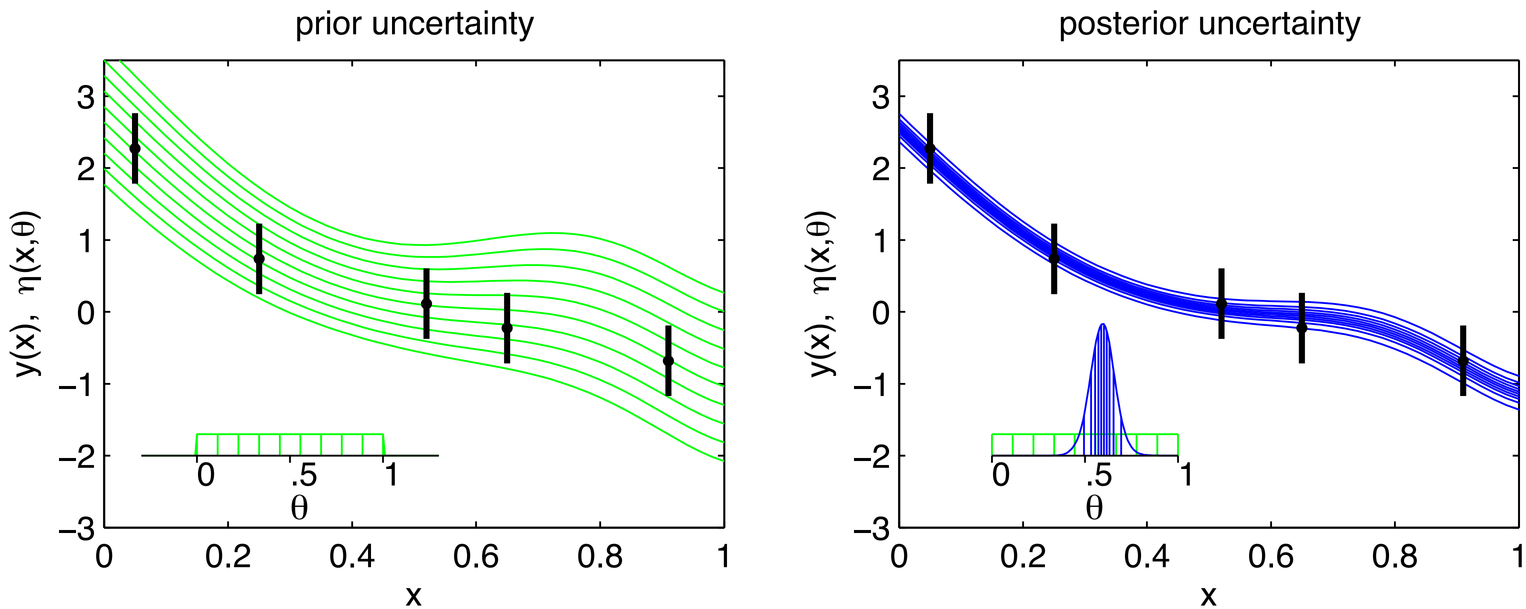
different models for how data are generated given truth



−1
0
1
2
−1
0
1
2

BRIEF INTRO TO BAYESIAN SOLUTIONS FOR INVERSE PROBLEMS

Bayesian analysis of an inverse problem



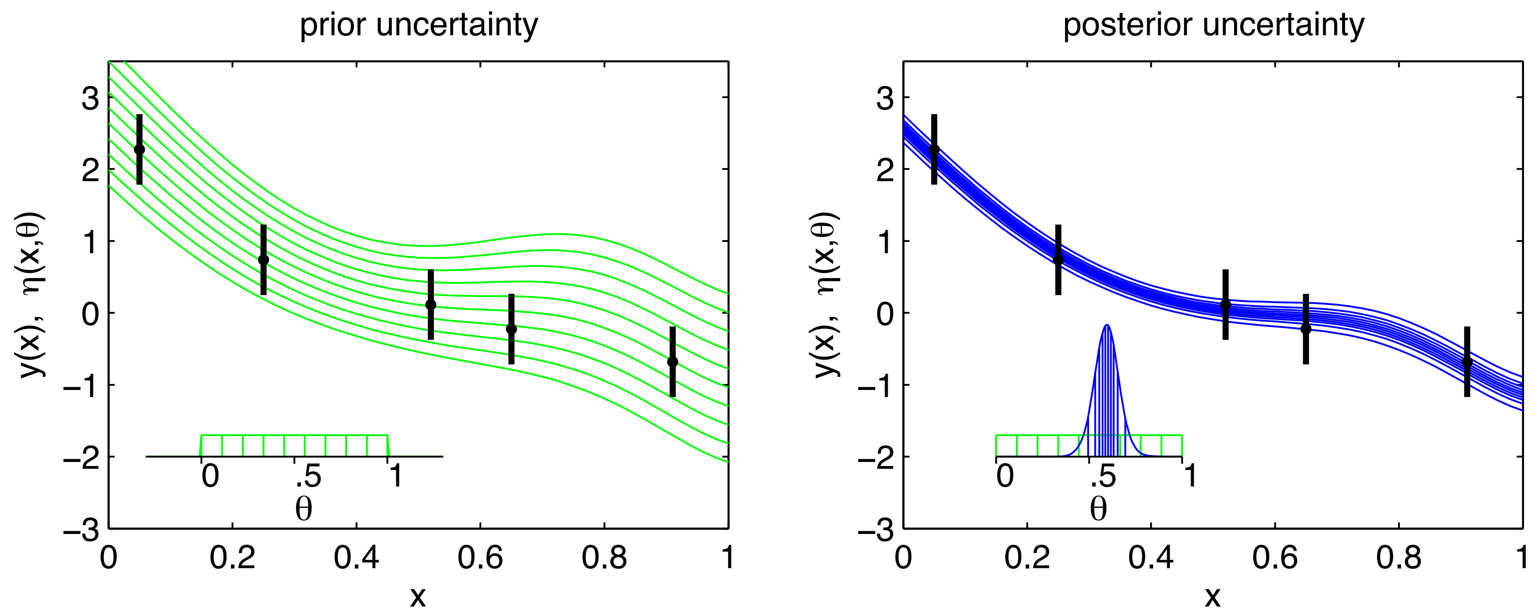
- A simple example...

x	experimental conditions
θ	model calibration parameters
$\zeta(x)$	true physical system response given inputs x
$\eta(x,\theta)$	forward simulator response at x and θ.
$y(x)$	experimental observation of the physical system
$e(x)$	observation error of the experimental data

Assume:

$$\begin{aligned} y(x) &= \zeta(x) + e(x) \\ &= \eta(x,\theta) + e(x) \end{aligned} \quad \theta \text{ unknown.}$$

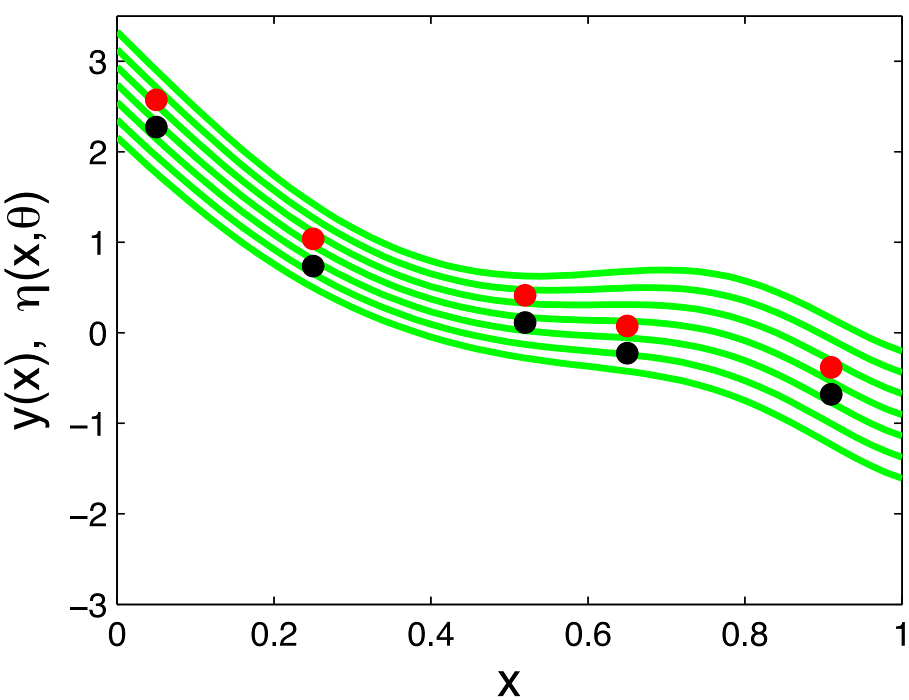
Data for the toy inverse problem

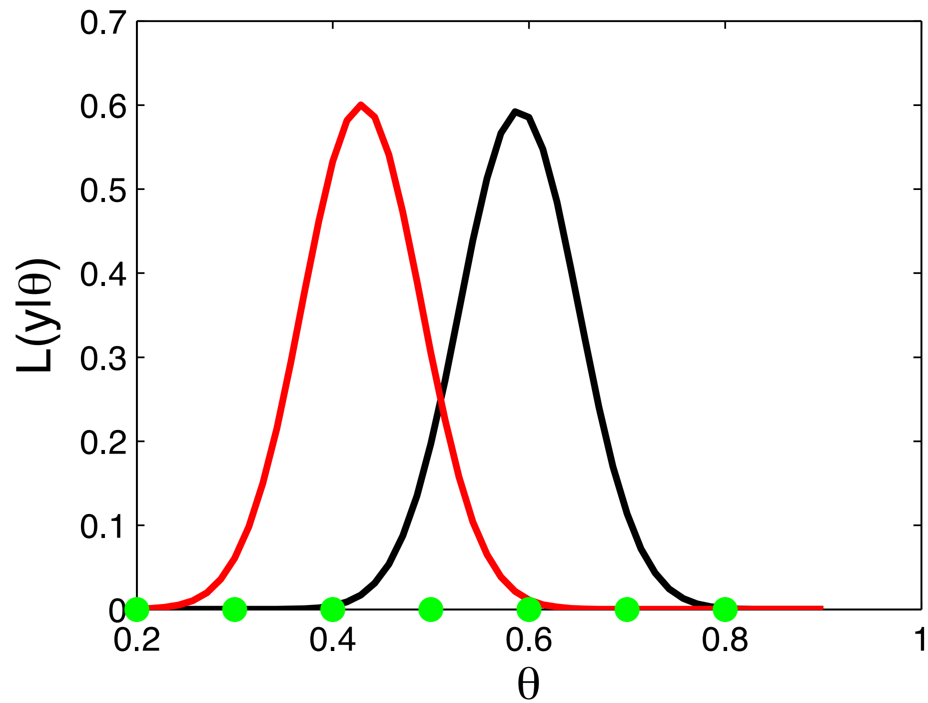


$n = 5$ physical observations

variable	data				
i	1	2	3	4	5
x	0.05	0.25	0.52	0.65	0.91
y	2.2731	0.7371	0.1138	−0.2254	−0.6807

Likelihood

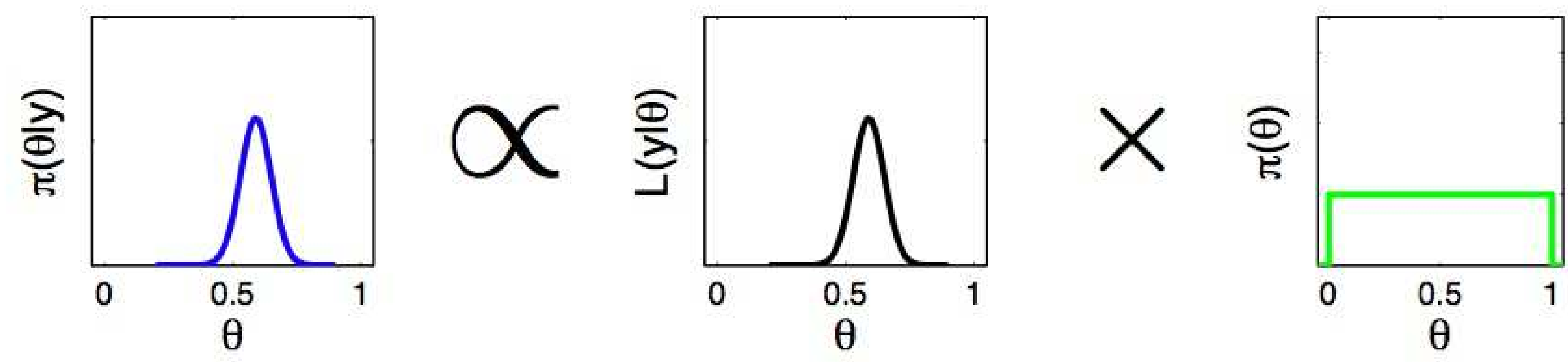




$$L(y|\eta(x,\theta)) \propto \lambda_y^{\frac{n}{2}} \exp\left\{-\frac{1}{2\cdot .25^2}\lambda_y (y-\eta(x,\theta))^T(y-\eta(x,\theta))\right\}$$

- $L(y|\theta)$ is the probability model for the data y given θ
- tells us which values of θ are likely given the observed data y
- can combine with the prior $\pi(\theta)$ to describe posterior uncertainty for θ

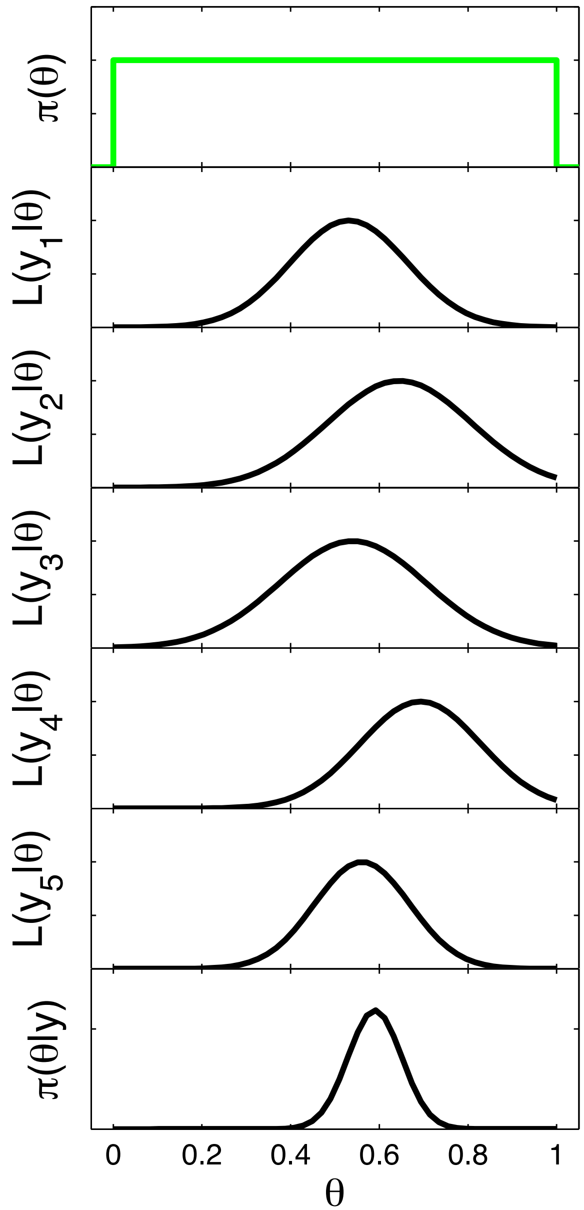
Bayes' Rule



$$\pi(\theta|y) \propto L(y|\theta) \times \pi(\theta)$$

- pointwise multiplication over the support of θ
- very general approach for inference
- prior pdf for θ is required
- normalizing $\pi(\theta|y)$ is generally difficult, but rarely necessary
- high dimensional θ can lead to computational challenges

Bayes' Rule (independent components)



- With independent data, the likelihood is a product of independent components:

$$
\begin{aligned}
L(y|\theta) &= \prod_{i=1}^{n} L(y_i|\theta) \\
&= \prod_{i=1}^{n} \exp\left\{-\tfrac{1}{2}\lambda_y (y_i - \eta(x_i,\theta))^2\right\}
\end{aligned}
$$

  (here we fix $\lambda_y = 4$).

- A central limit theorem:
  0. $y_i \sim L(y_i|\theta)$, $i = 1, \ldots, n$, independent
  1. regularity on $L(y_i|\theta)$'s
  2. prior support for $\pi(\theta)$ covers true θ

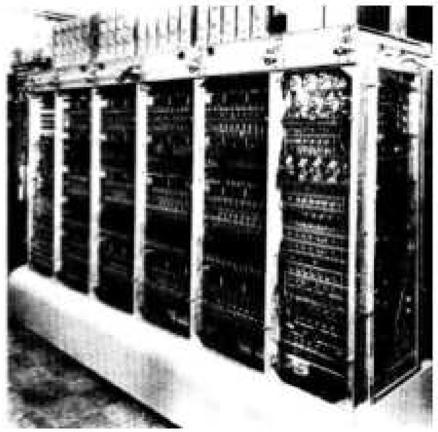
$$\pi(\theta|y) \rightarrow \text{dnorm}(\theta, \lambda_n^{-1})$$

  where $\lambda_n = \frac{d^2}{d\theta^2} \log \pi(\theta|y)$

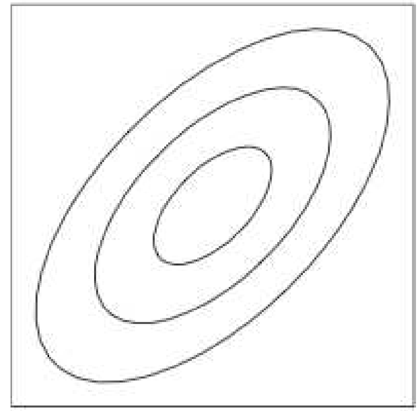
Exploring the posterior distribution

Nick Metropolis

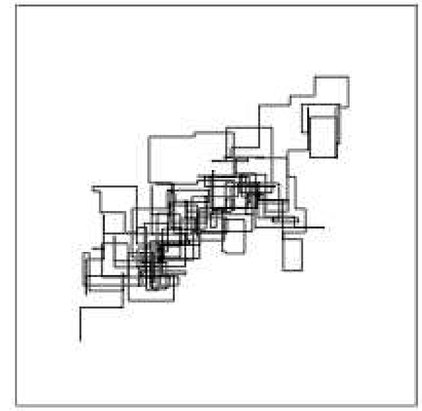
MANIAC I



2-d pdf



MCMC sample

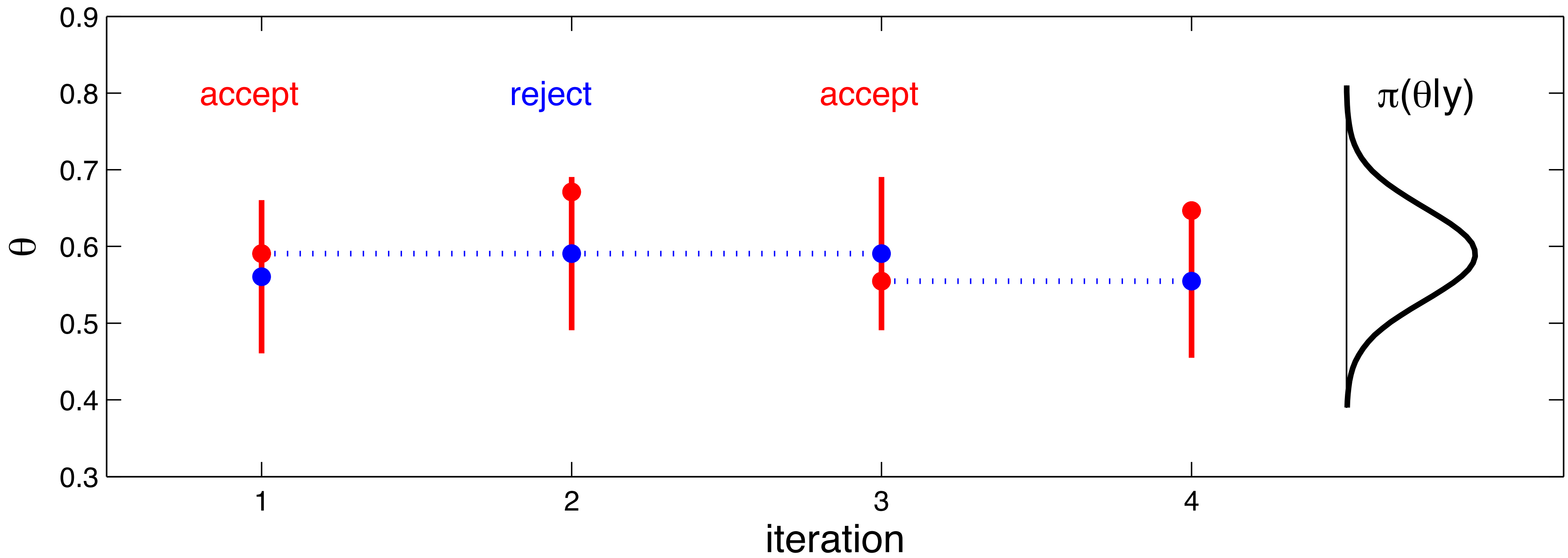


~1953

~1991

- Use Markov chain Monte Carlo to build a Markov chain with stationary distribution $\pi(\theta|y)$
- Realizations are a (correlated) sample from $\pi(\theta|y)$
- $\pi(\theta|y)$ need not be normalized

Metropolis recipe for MCMC



Initialize chain at θ^0

1. Given current realization θ^t, generate θ^* from a symmetric kernel $q(\theta^t \rightarrow \theta^*)$

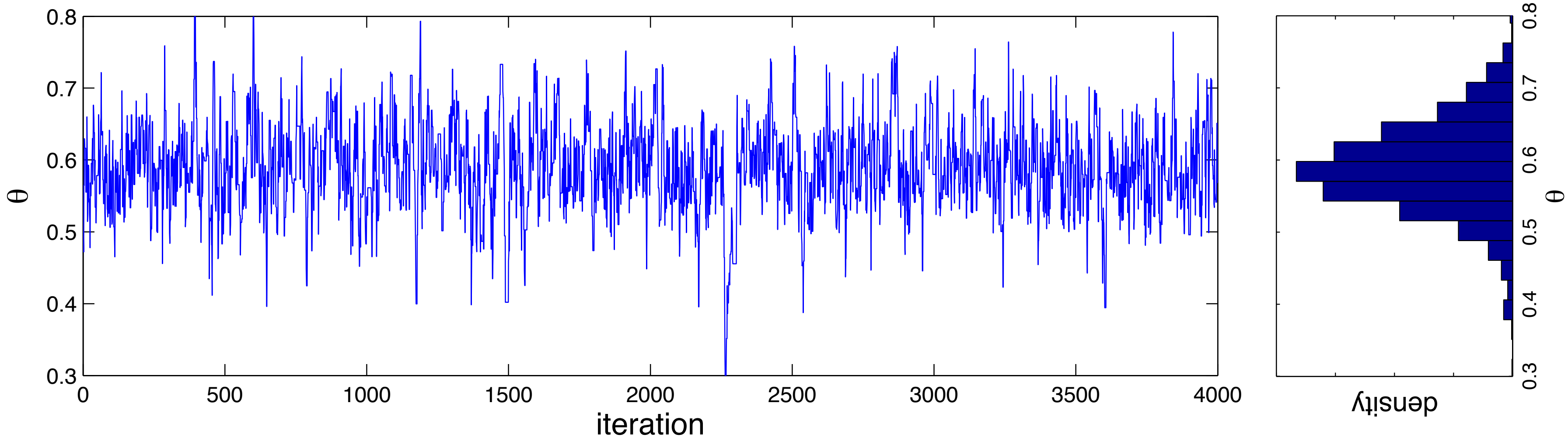
   i.e. $q(\theta^t \rightarrow \theta^*) = q(\theta^* \rightarrow \theta^t)$

2. Compute acceptance probability $\alpha = \min\left\{1, \frac{\pi(\theta^*|y)}{\pi(\theta^t|y)}\right\}$

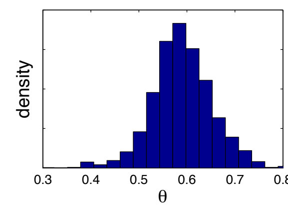
3. Set $\theta^{t+1} = \theta^*$ with probability α, otherwise $\theta^{t+1} = \theta^t$

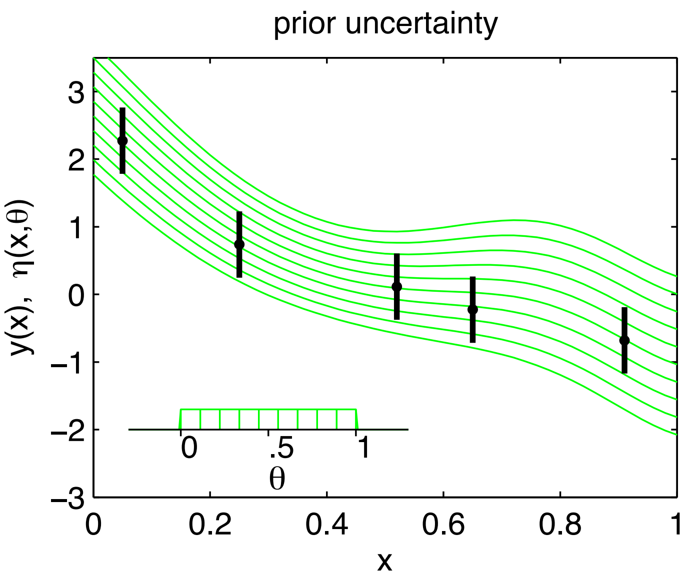
4. Iterate steps 1 – 3

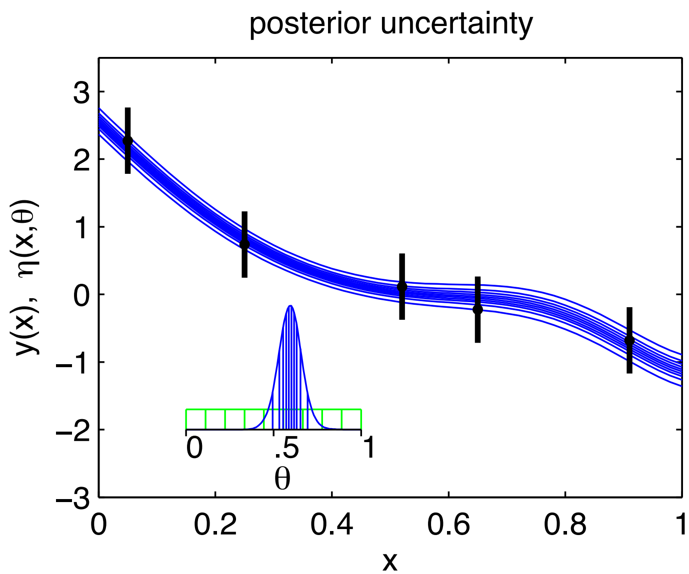
Metropolis sampling for the inverse problem



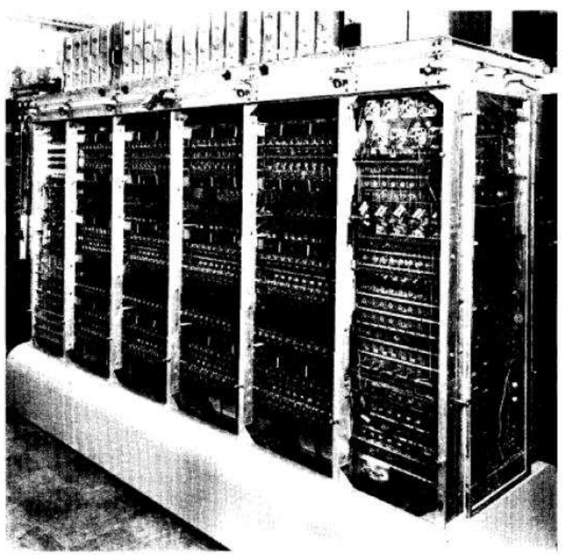
- chain $\theta^0, \theta^1, \ldots, \theta^{4000}$ is a draw from $\pi(\theta|y)$
- use Monte Carlo sample to estimate expectations, variances, probabilities, etc.







Sampling from non-standard multivariate distributions



Nick Metropolis – Computing pioneer at Los Alamos National Laboratory

– inventor of the Monte Carlo method

– inventor of Markov chain Monte Carlo:

Equation of State Calculations by Fast Computing Machines (1953) by N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller and E. Teller, *Journal of Chemical Physics*.

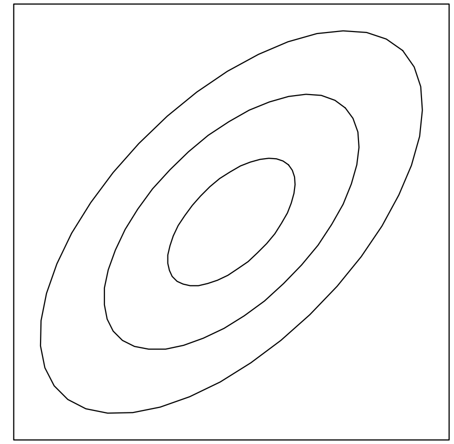
Originally implemented on the MANIAC1 computer at LANL

Algorithm constructs a Markov chain whose realizations are draws from the target (posterior) distribution.

Constructs steps that maintain detailed balance.

Gibbs Sampling and Metropolis for a bivariate normal density

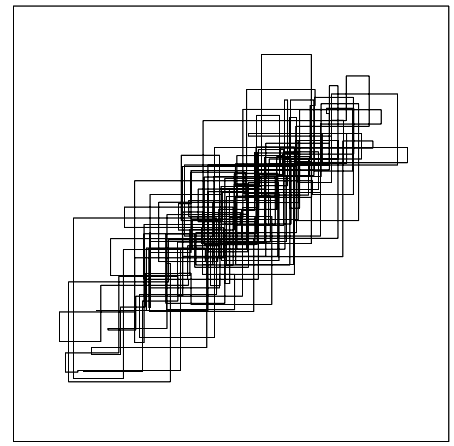
$$\pi(z_1, z_2) \propto \begin{vmatrix} 1 & \rho \\ \rho & 1 \end{vmatrix}^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \begin{pmatrix} z_1 & z_2 \end{pmatrix} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}^{-1} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \right\}$$



sampling from the full conditionals

$$\begin{aligned} z_1 | z_2 &\sim N(\rho z_2, 1 - \rho^2) \\ z_2 | z_1 &\sim N(\rho z_1, 1 - \rho^2) \end{aligned}$$

also called heat bath

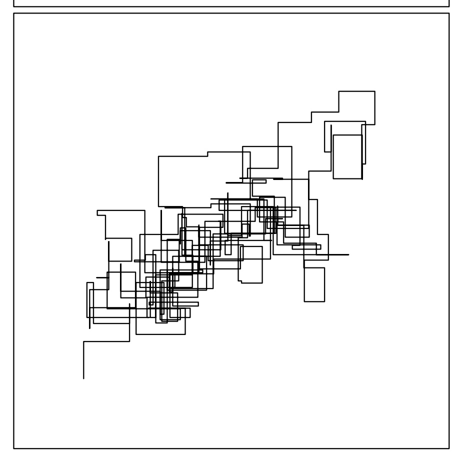


Metropolis updating:

generate $z_1^* \sim U[z_1 - r, z_1 + r]$

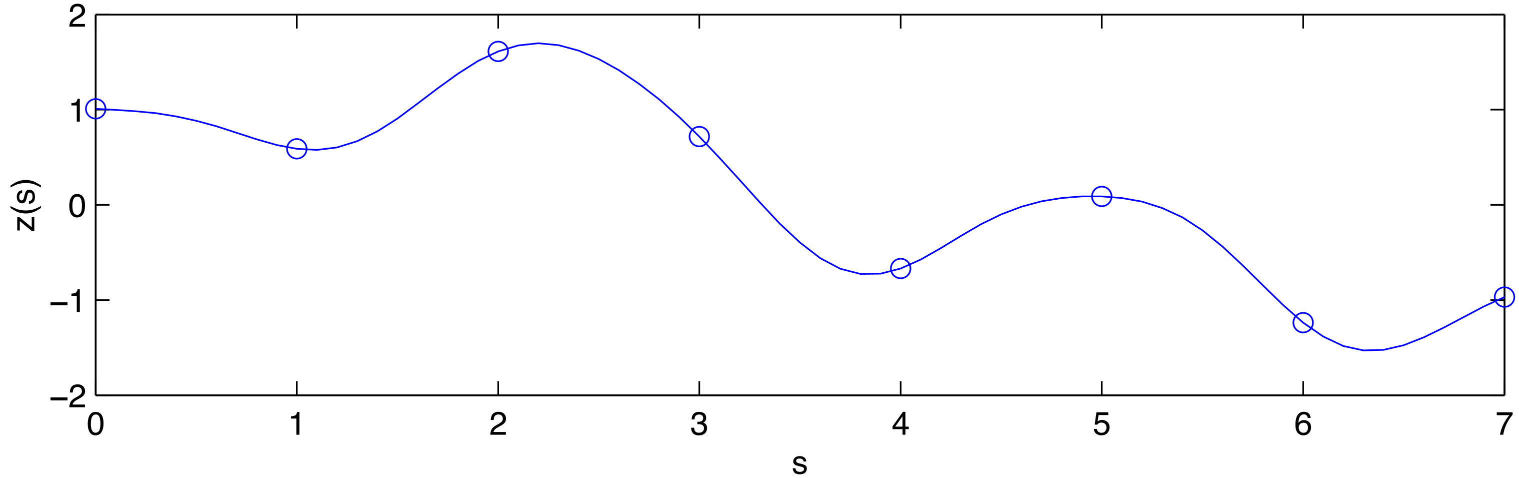
calculate $\alpha = \min\{1, \frac{\pi(z_1^*, z_2)}{\pi(z_1, z_2)} = \frac{\pi(z_1^* | z_2)}{\pi(z_1 | z_2)}\}$

set $z_1^{\text{new}} = \begin{cases} z_1^* \text{ with probability } \alpha \\ z_1 \text{ with probability } 1 - \alpha \end{cases}$



GAUSSIAN PROCESSES 1

Gaussian process models for spatial phenomena



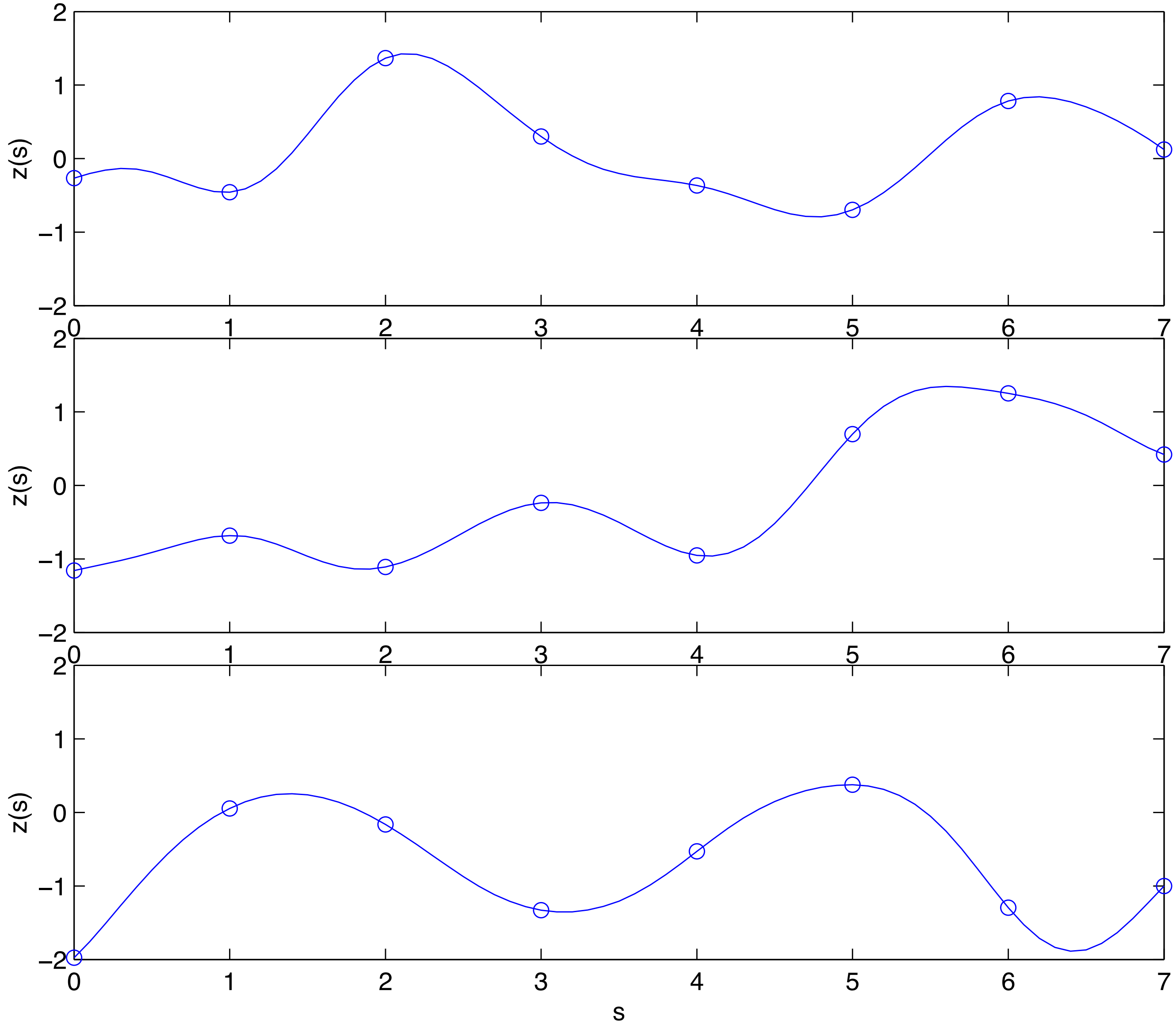
An example of $z(s)$ of a Gaussian process model on $s_1, \ldots, s_n$

$$z = \begin{pmatrix} z(s_1) \\ \vdots \\ z(s_n) \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} & \Sigma & \end{pmatrix} \right), \quad \text{with } \Sigma_{ij} = \exp\{-||s_i - s_j||^2\},$$

where $||s_i - s_j||$ denotes the distance between locations s_i and s_j.

z has density $\pi(z) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\{-\frac{1}{2} z^T \Sigma^{-1} z\}$.

Realizations from $\pi(z) = (2\pi)^{-\frac{n}{2}}|\Sigma|^{-\frac{1}{2}} \exp\{-\frac{1}{2}z^T\Sigma^{-1}z\}$



model for $z(s)$ can be extended to continuous s

Generating multivariate normal realizations

Independent normals are standard for any computer package

$$u \sim N(0, I_n)$$

Well known property of normals:

$$\text{if } u \sim N(\mu, \Sigma), \quad \text{then } z = Ku \sim N(K\mu, K\Sigma K^T)$$

Use this to construct correlated realizations from iid ones.

Want $z \sim N(0, \Sigma)$

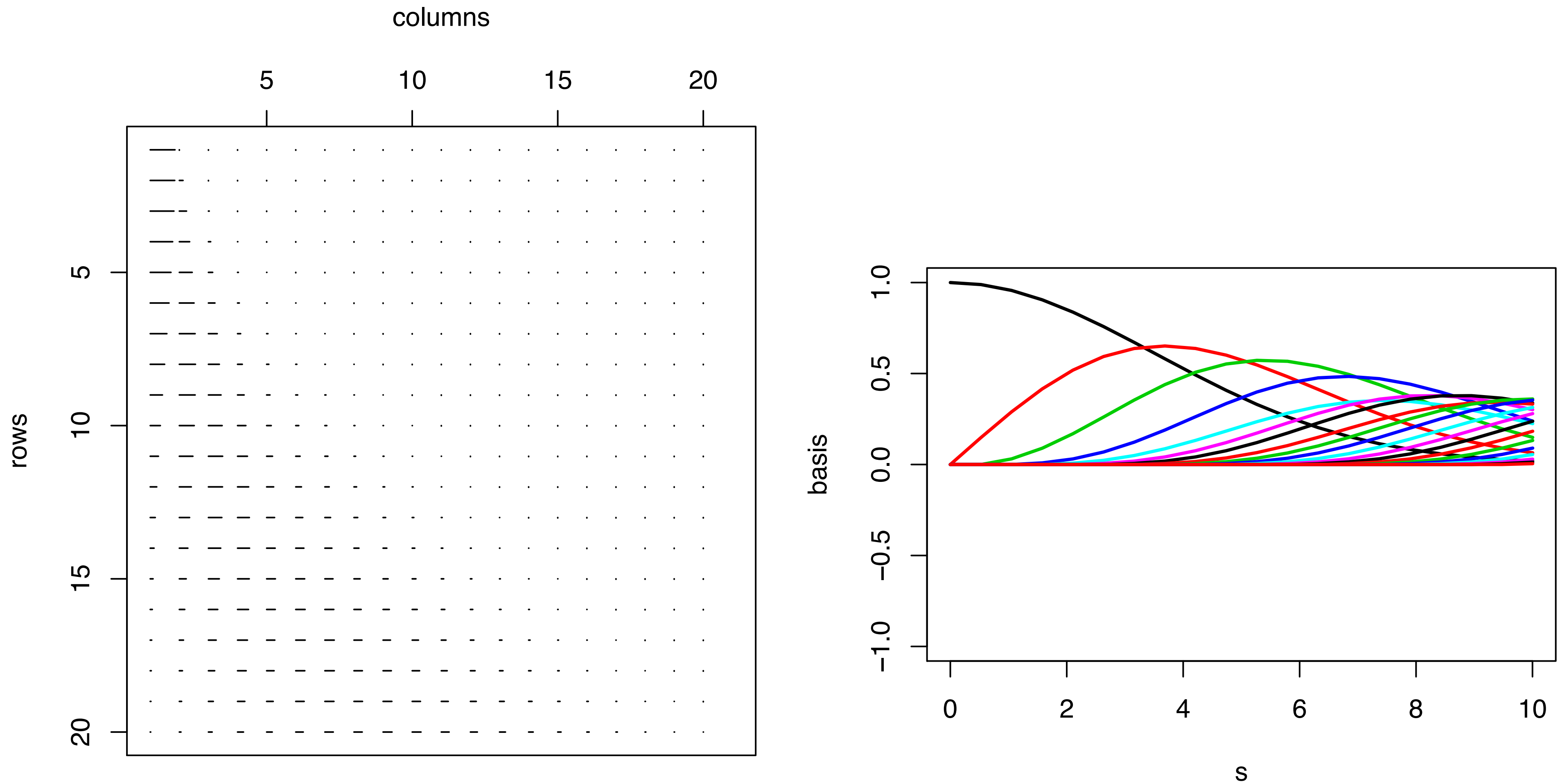
1. compute square root matrix L such that $LL^T = \Sigma$;
2. generate $u \sim N(0, I_n)$;
3. Set $z = Lu \sim N(0, LI_nL^T = \Sigma)$

- Any square root matrix L will do here.
- Columns of L are basis functions for representing realizations z.
- L need not be square – see over or under specified bases.

Standard Cholesky decomposition

$z = N(0, \Sigma)$, $\Sigma = LL^T$, $z = Lu$ where $u \sim N(0, I_n)$, L lower triangular

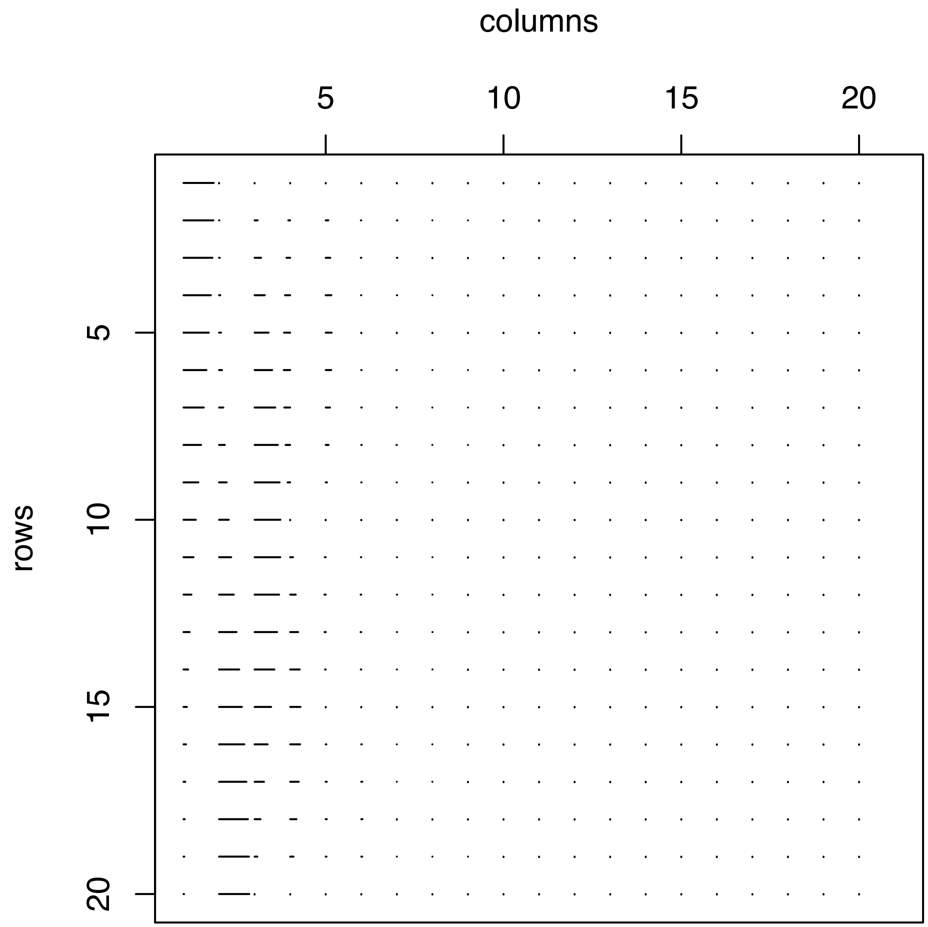
$\Sigma_{ij} = \exp\{-||s_i - s_j||^2\}$, $s_1, \ldots, s_{20}$ equally spaced between 0 and 10 :

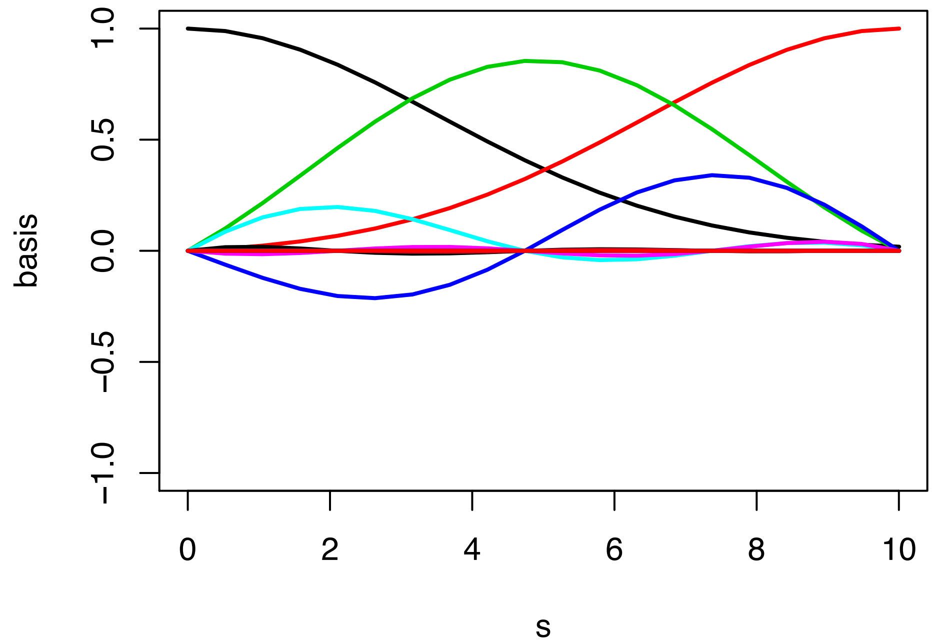


Cholesky decomposition with pivoting

$z = N(0, \Sigma), \;\; \Sigma = LL^T, \;\; z = Lu$ where $u \sim N(0, I_n), \;\; L$ permuted lower triangular

$\Sigma_{ij} = \exp\{-||s_i - s_j||^2\}, \; s_1, \ldots, s_{20}$ equally spaced between 0 and 10 :

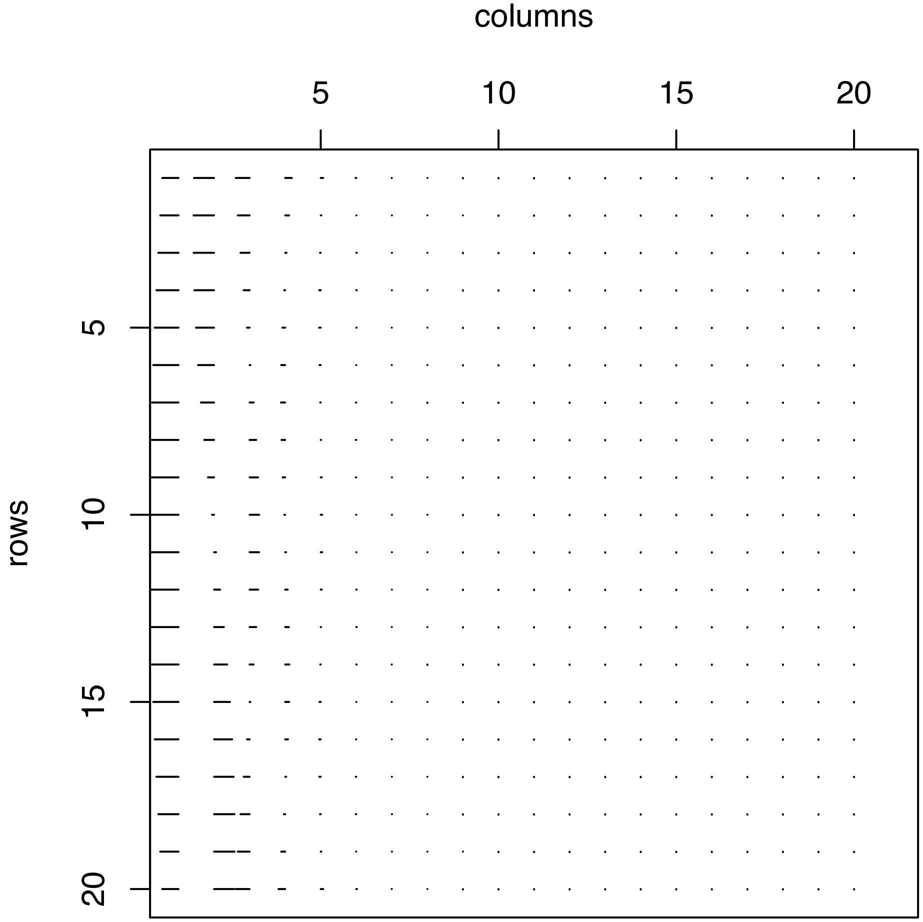


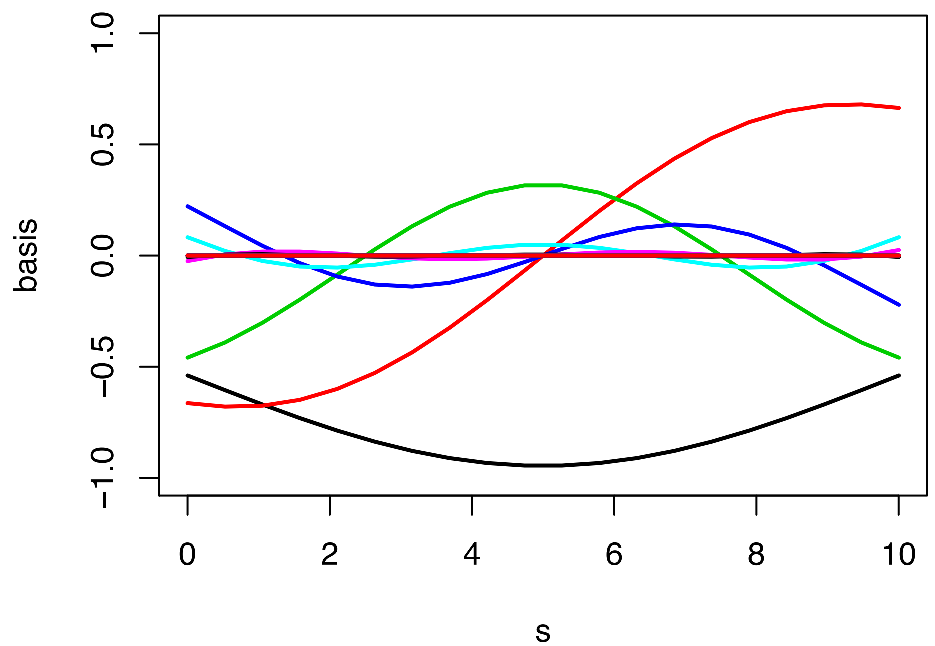


Singular value decomposition

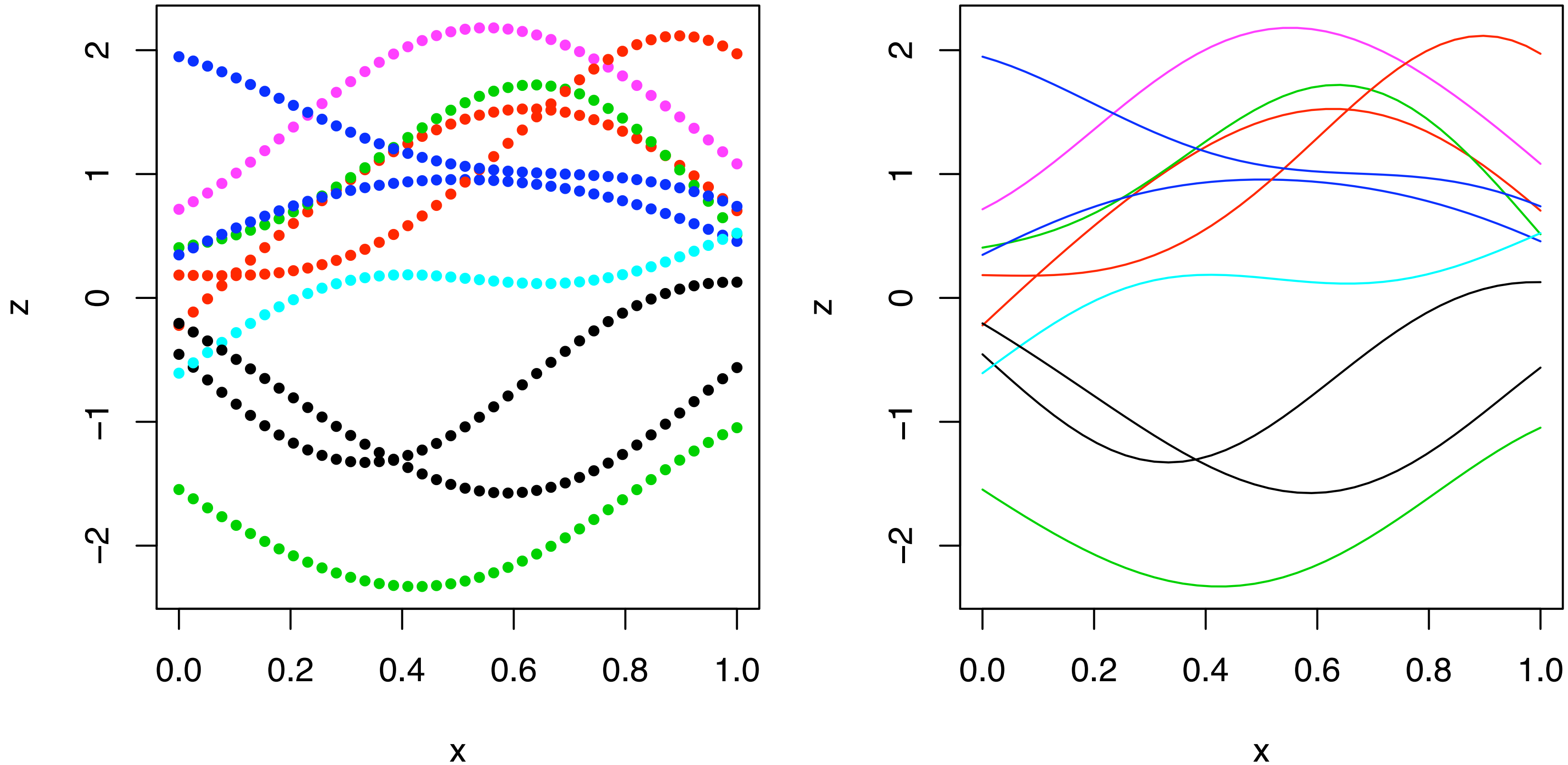
$$z = N(0,\Sigma),\ \ \Sigma = U\Lambda U^T = LL^T,\ \ z = Lu \text{ where } u \sim N(0, I_n)$$

$$\Sigma_{ij} = \exp\{-||s_i - s_j||^2\},\ s_1,\ldots,s_{20} \text{ equally spaced between 0 and 10}:$$





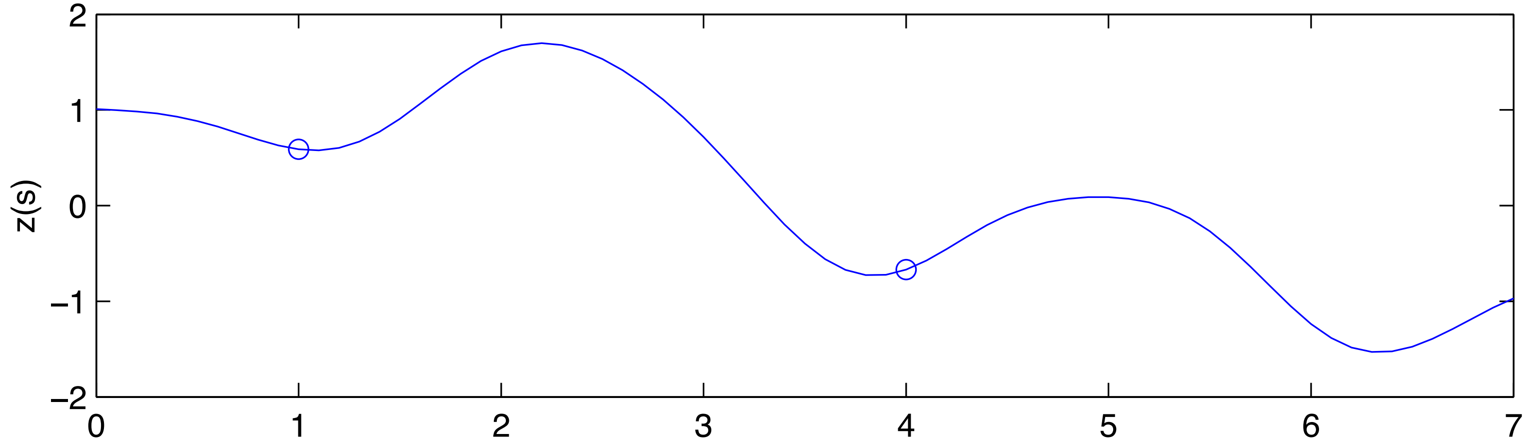
Gaussian Process Models



$$z = \begin{pmatrix} z(x_1) \\ \vdots \\ z(x_{40}) \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} & \Sigma & \end{pmatrix} \right) \text{ with } \Sigma_{ij} = \rho^{||x_i - x_j||^2}$$

$$\pi(z) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\{-.5 z^T \Sigma^{-1} z\}$$

Conditioning on some observations of $z(s)$

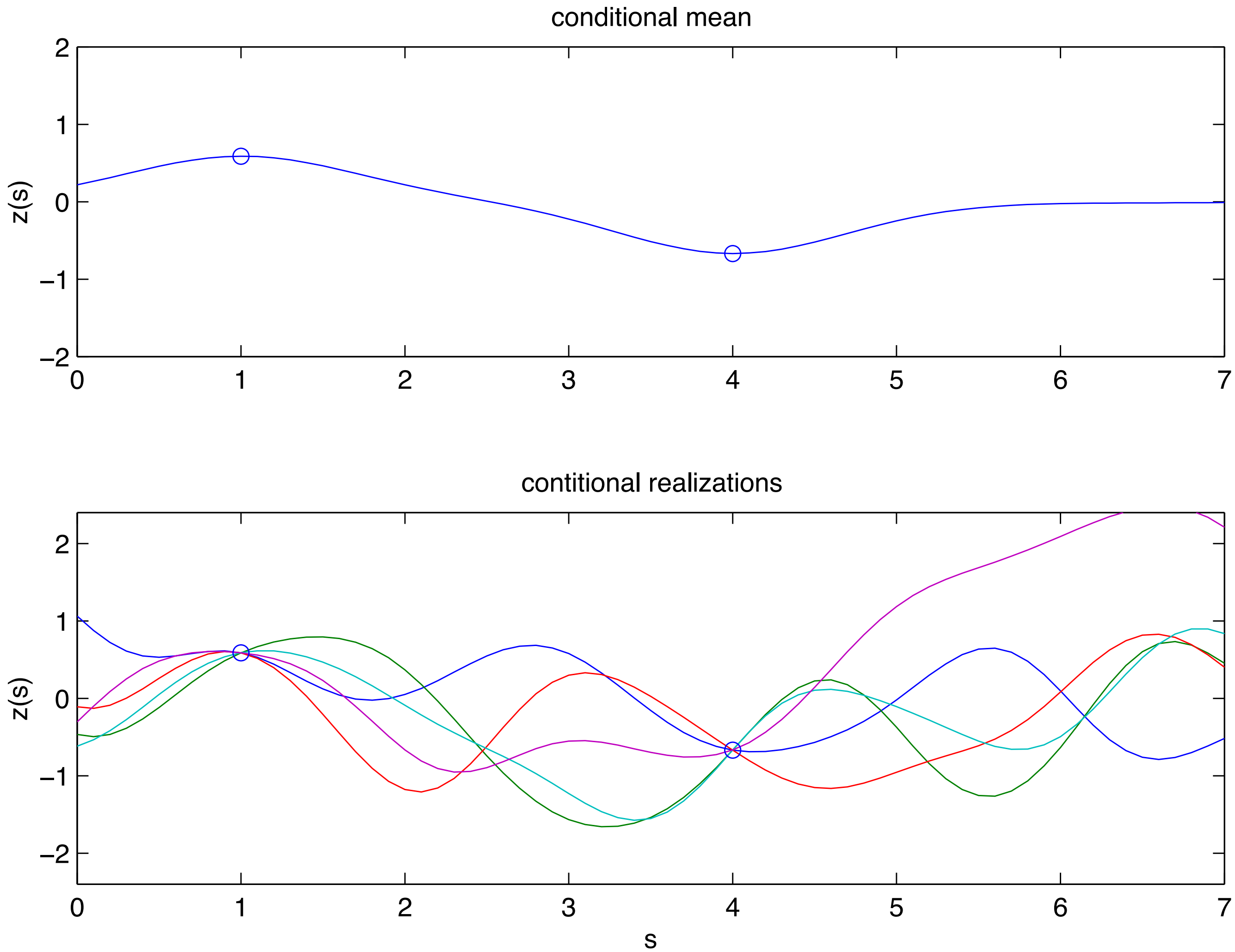


We observe $z(s_2)$ and $z(s_5)$ – what do we now know about $\{z(s_1), z(s_3), z(s_4), z(s_6), z(s_7), z(s_8)\}$?

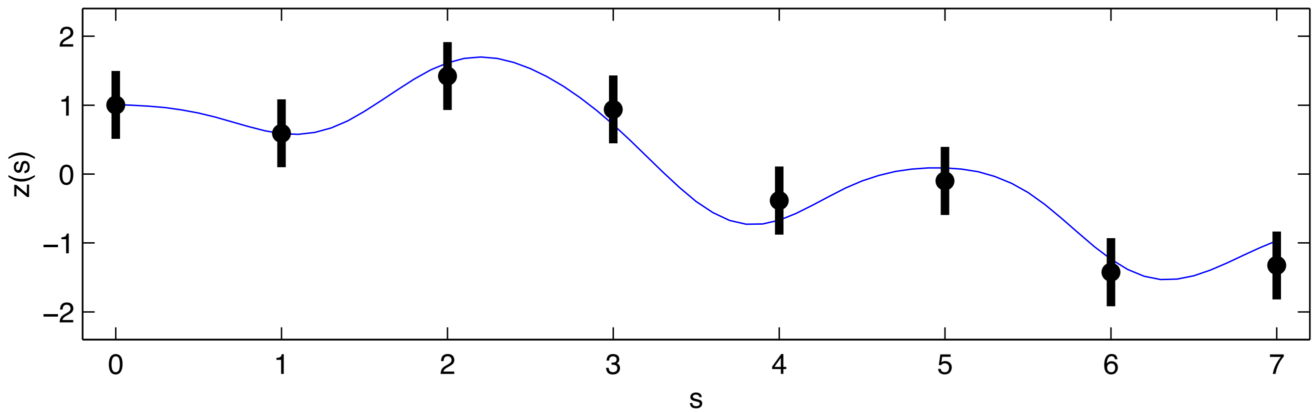
$$\begin{pmatrix} z(s_2) \\ z(s_5) \\ \hline z(s_1) \\ z(s_3) \\ z(s_4) \\ z(s_6) \\ z(s_7) \\ z(s_8) \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ \hline 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \left(\begin{array}{cc|ccc} 1 & .0001 & .3679 & \cdots & 0 \\ .0001 & 1 & 0 & \cdots & .0001 \\ \hline .3679 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \vdots & \ddots & \vdots \\ 0 & .0001 & 0 & \cdots & 1 \end{array} \right) \right)$$

Conditioning on some observations of $z(s)$

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right), \quad z_2 | z_1 \sim N(\Sigma_{21}\Sigma_{11}^{-1} z_1, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$$



Soft Conditioning (Bayes Rule)



Observed data y are a noisy version of z

$$y(s_i) = z(s_i) + \epsilon(s_i) \text{ with } \epsilon(s_k) \overset{iid}{\sim} N(0, \sigma_y^2),\ k = 1, \ldots, n$$

Data

$$y \qquad \Sigma_y = \sigma_y^2 I_n$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \begin{pmatrix} \sigma_y^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_y^2 \end{pmatrix}$$

spatial process prior for $z(s)$

$$\mu_z \qquad \Sigma_z$$

$$\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \quad \begin{pmatrix} & & \\ & \Sigma_z & \\ & & \end{pmatrix}$$

$$L(y|z) \propto |\Sigma_y|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(y-z)^T \Sigma_y^{-1} (y-z)\} \quad \pi(z) \propto |\Sigma_z|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2} z^T \Sigma_z^{-1} z\}$$

Soft Conditioning (Bayes Rule) ... continued

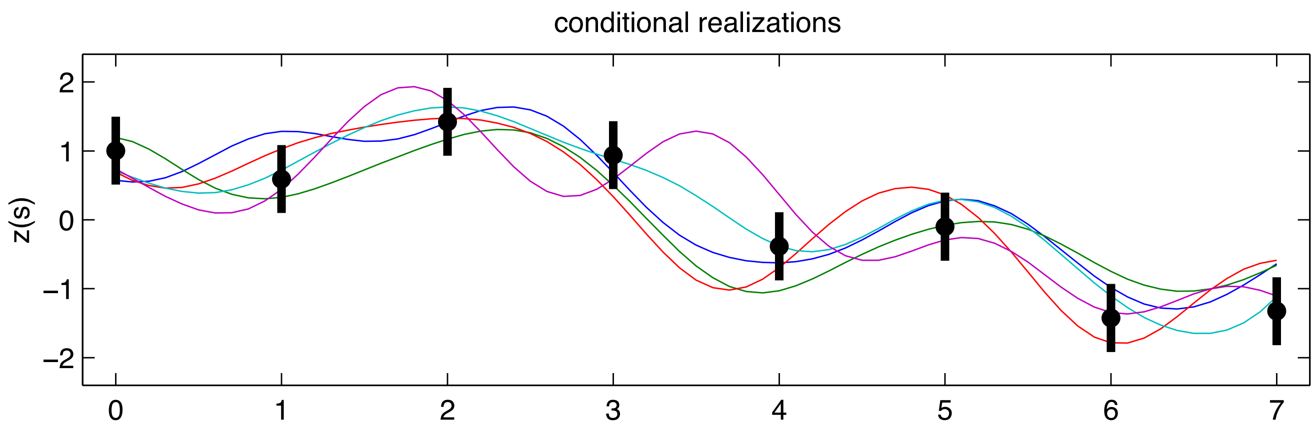
sampling model — spatial prior

$$L(y|z) \propto |\Sigma_y|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(y-z)^T \Sigma_y^{-1}(y-z)\} \quad \pi(z) \propto |\Sigma_z|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2} z^T \Sigma_z^{-1} z\}$$

$$\Rightarrow \pi(z|y) \;\propto\; L(y|z) \times \pi(z)$$

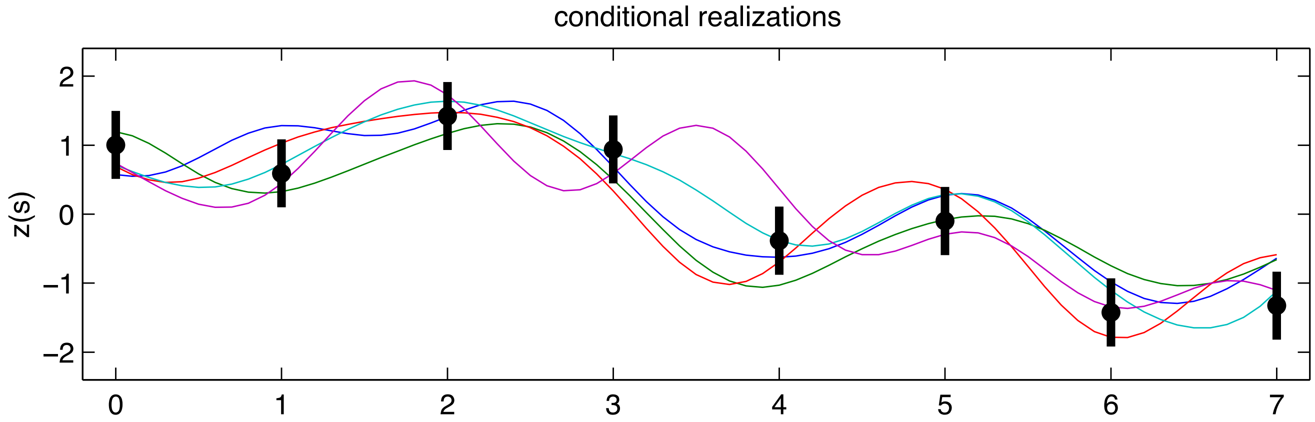
$$\Rightarrow \pi(z|y) \;\propto\; \exp\{-\tfrac{1}{2}[z^T(\Sigma_y^{-1} + \Sigma_z^{-1})z + z^T \Sigma_y^{-1} y + f(y)]\}$$

$$\Rightarrow \quad z|y \;\sim\; N(V\Sigma_y^{-1} y, V), \text{ where } V = (\Sigma_y^{-1} + \Sigma_z^{-1})^{-1}$$



$\pi(z|y)$ describes the updated uncertainty about z given the observations.

Updated predictions for unobserved $z(s)$'s



data locations	$y^d = (y(s_1), \ldots, y(s_n))^T$	$z^d = (z(s_1), \ldots, z(s_n))^T$
prediction locations	$y^* = (y(s_1^*), \ldots, y(s_m^*))^T$	$z^* = (z(s_1^*), \ldots, z(s_m^*))^T$
define	$y = (y^d; y^*)$	$z = (z^d; z^*)$

Data

$$y = \begin{pmatrix} y^d \\ y^* \end{pmatrix} = \begin{pmatrix} y^d \\ 0_m \end{pmatrix} \quad \Sigma_y = \begin{pmatrix} \sigma_y^2 I_n & 0 \\ 0 & \infty I_m \end{pmatrix}$$

spatial process prior for $z(s)$

$$\mu_z = \begin{pmatrix} 0_n \\ 0_m \end{pmatrix} \quad \Sigma_z = \begin{pmatrix} \text{cov rule applied} \\ \text{to } (s, s^*) \end{pmatrix}$$

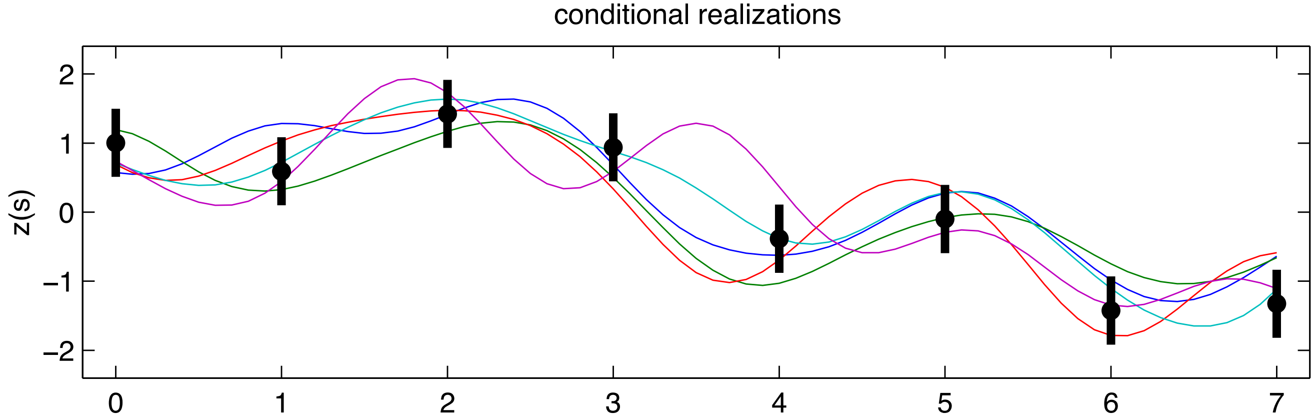
define $\Sigma_y^- = \begin{pmatrix} \frac{1}{\sigma_y^2} I_n & 0 \\ 0 & 0 \end{pmatrix}$

Now the posterior distribution for $z = (z^d, z^*)$ is

$$z|y \sim N(V\Sigma_y^- y, V), \quad \text{where } V = (\Sigma_y^- + \Sigma_z^{-1})^{-1}$$

Updated predictions for unobserved $z(s)$'s,

Alternative: use the conditional normal rules:



data locations $y = (y(s_1), \ldots, y(s_n))^T = (z(s_1) + \epsilon(s_1), \ldots, z(s_n) + \epsilon(s_n))^T$

prediction locations $z^* = (z(s_1^*), \ldots, z(s_m^*))^T$

$$\text{Jointly } \begin{pmatrix} y \\ z^* \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_y^2 I_n & 0 \\ 0 & 0 \end{pmatrix} + \Sigma_z \right)$$

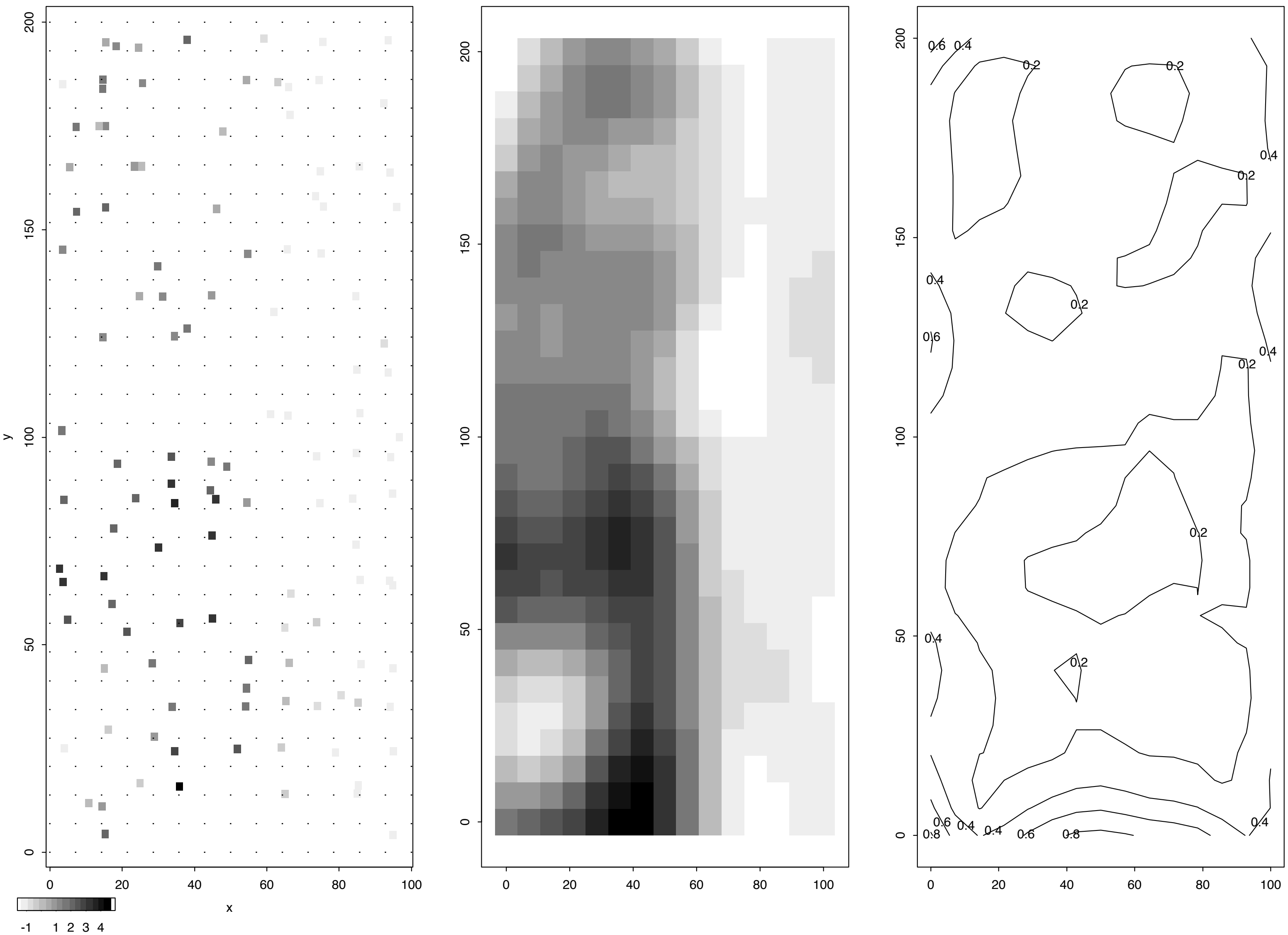
where

$$\Sigma_z = \begin{pmatrix} \Sigma_z(s,s) & \Sigma_z(s,s^*) \\ \Sigma_z(s^*,s) & \Sigma_z(s^*,s^*) \end{pmatrix} = \begin{pmatrix} \text{cov rule applied} \\ \text{to } (s, s^*) \end{pmatrix}_{(n+m)\times(n+m)}$$

Therefore $z^*|y \sim N(\mu^*, \Sigma^*)$ where

$$\begin{aligned} \mu^* &= \Sigma_z(s^*, s)[\sigma_y^2 I_n + \Sigma_z(s,s)]^{-1} y \\ \Sigma^* &= \Sigma_z(s^*, s^*) - \Sigma_z(s^*, s)[\sigma_z^2 I_n + \Sigma_z(s,s)]^{-1} \Sigma_z(s, s^*) \end{aligned}$$

Example: Dioxin concentration at Piazza Road Superfund Site

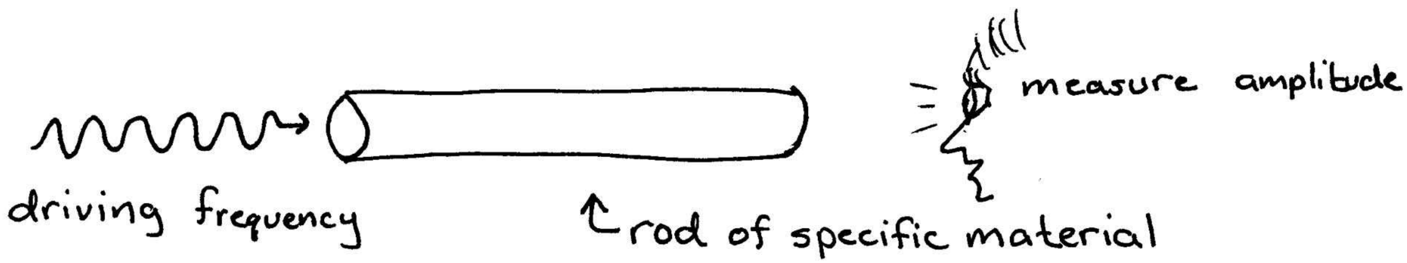


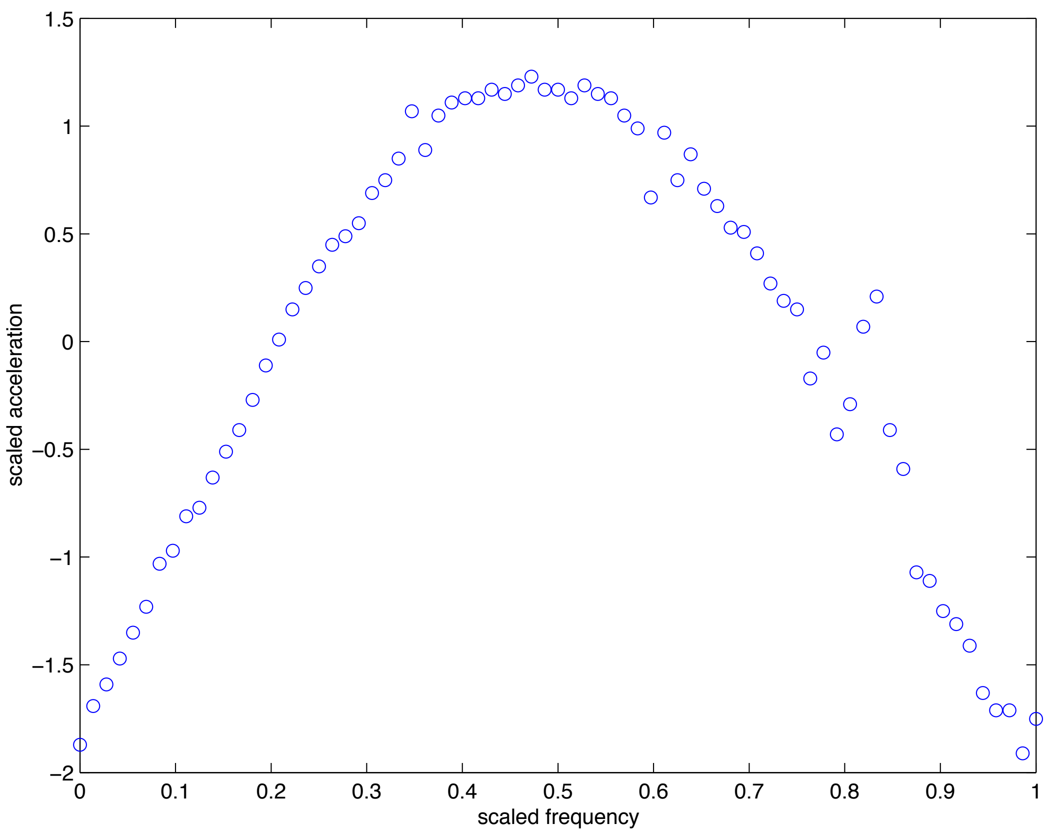
data Posterior mean of z^* pointwise posterior sd

GAUSSIAN PROCESSES 2

Gaussian process models revisited

Application: finding in a rod of material

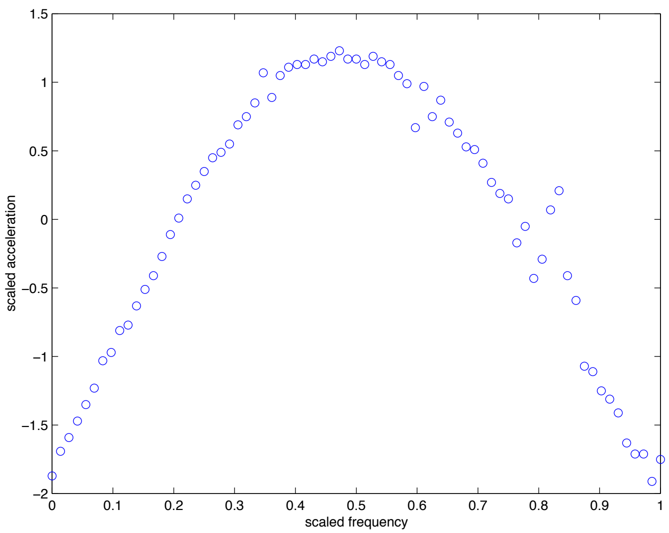




- for various driving frequencies, acceleration of rod recorded
- the true frequency-acceleration curve is smooth.
- we have noisy measurements of acceleration.
- estimate resonance frequency.
- use GP model for frequency-accel curve.
- smoothness of GP model important here.

Gaussian process models formulation

Take response y to be acceleration and spatial value s to be frequency.



data: $y = (y_1, \ldots, y_n)^T$ at spatial locations $s_1, \ldots, s_n$.

$z(s)$ is a mean 0 Gaussian process with covariance function

$$\text{Cov}(z(s), z(s')) = \frac{1}{\lambda_z} \exp\{-\beta(s - s')^2\}$$

β controls strength of dependence.

Take $z = (z(s_1), \ldots, z(s_n))^T$ to be $z(s)$ restricted to the data observations.

Model the data as:

$$y = z + \epsilon, \quad \text{where } \epsilon \sim N(0, \frac{1}{\lambda_y} I_n)$$

We want to find the posterior distribution for the frequency $s^\star$ where $z(s)$ is maximal.

Reparameterizing the spatial dependence parameter β

It is convenient to reparameterize β as:

$$\rho = \exp\{-\beta(1/2)^2\} \;\Leftrightarrow\; \beta = -4\log(\rho)$$

So ρ is the correlation between two points on $z(s)$ separated by $\frac{1}{2}$.

Hence z has spatial prior

$$z|\rho, \lambda_z \sim N(0, \frac{1}{\lambda_z} R(\rho; s))$$

where $R(\rho; s)$ is the correlation matrix with ij elements

$$R_{ij} = \rho^{4(s_i - s_j)^2}$$

Prior specification for $z(s)$ is completed by specfying priors for λ_z and ρ.

$\pi(\lambda_z) \;\propto\; \lambda_z^{a_z - 1}\exp\{-b_z\lambda_z\}$ if y is standardized, encourage λ_z to be close to 1 – eg.$a_z = b_z = 5$.

$\pi(\rho) \;\propto\; (1-\rho)^{-.5}$ encourages ρ to be large if possible

Bayesian model formulation

Likelihood

$$L(y|z,\lambda_y) \propto \lambda_y^{\frac{n}{2}} \exp\{-\tfrac{1}{2}\lambda_y (y-z)^T (y-z)\}$$

Priors

$$\pi(z|\lambda_z,\rho) \propto \lambda_z^{\frac{n}{2}} |R(\rho;s)|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}\lambda_z z^T R(\rho;s)^{-1} z\}$$

$$\pi(\lambda_y) \propto \lambda_y^{a_y-1} e^{-b_y\lambda_y}, \text{ uninformative here – } a_y = 1,\ b_y = .005$$

$$\pi(\lambda_z) \propto \lambda_z^{a_z-1} e^{-b_z\lambda_z}, \text{ fairly informative – } a_z = 5,\ b_z = 5$$

$$\pi(\rho) \propto (1-\rho)^{-.5}$$

Marginal likelihood (integrating out z)

$$L(y|\lambda_\epsilon,\lambda_z,\rho) \propto |\Lambda|^{\frac{1}{2}} \exp\{-\tfrac{1}{2} y^T \Lambda y\}$$

where $\Lambda^{-1} = \frac{1}{\lambda_y} I_n + \frac{1}{\lambda_z} R(\rho;s)$

Posterior

$$\pi(\lambda_y,\lambda_z,\rho|y) \propto |\Lambda|^{\frac{1}{2}} \exp\{-\tfrac{1}{2} y^T \Lambda y\} \times \lambda_y^{a_y-1} e^{-b_y\lambda_y} \times \lambda_z^{a_z-1} e^{-b_z\lambda_z} \times (1-\rho)^{-.5}$$

Posterior Simulation

Use Metropolis to simulate from the posterior

$$\pi(\lambda_y, \lambda_z, \rho|y) \propto |\Lambda|^{\frac{1}{2}} \exp\{-\tfrac{1}{2} y^T \Lambda y\} \times \lambda_y^{a_y-1} e^{-b_y \lambda_y} \times \lambda_z^{a_z-1} e^{-b_z \lambda_z} \times (1-\rho)^{-.5}$$

giving (after burn-in) $(\lambda_y, \lambda_z, \rho)^1, \ldots, (\lambda_y, \lambda_z, \rho)^T$

For any given realization $(\lambda_y, \lambda_z, \rho)^t$, one can generate $z^* = (z(s_1^*), \ldots, z(s_m^*))^T$ for any set of prediction locations $s_1^*, \ldots, s_m^*$.

From previous GP stuff, we know

$$\begin{pmatrix} z \\ z^* \end{pmatrix} | \cdots \sim N\left(V \Sigma_y^- \begin{pmatrix} y \\ 0_m \end{pmatrix}, V\right)$$

where

$$\Sigma_y^- = \begin{pmatrix} \lambda_\epsilon I_n & 0 \\ 0 & 0 \end{pmatrix} \text{ and } V^{-1} = \Sigma_y^- + \lambda_z R(\rho, (s, s^*))^{-1}$$

Hence, one can generate corresponding z^*'s for each posterior realization at a fine grid around the apparent resonance frequency $z^\star$.

Or use conditional normal formula with

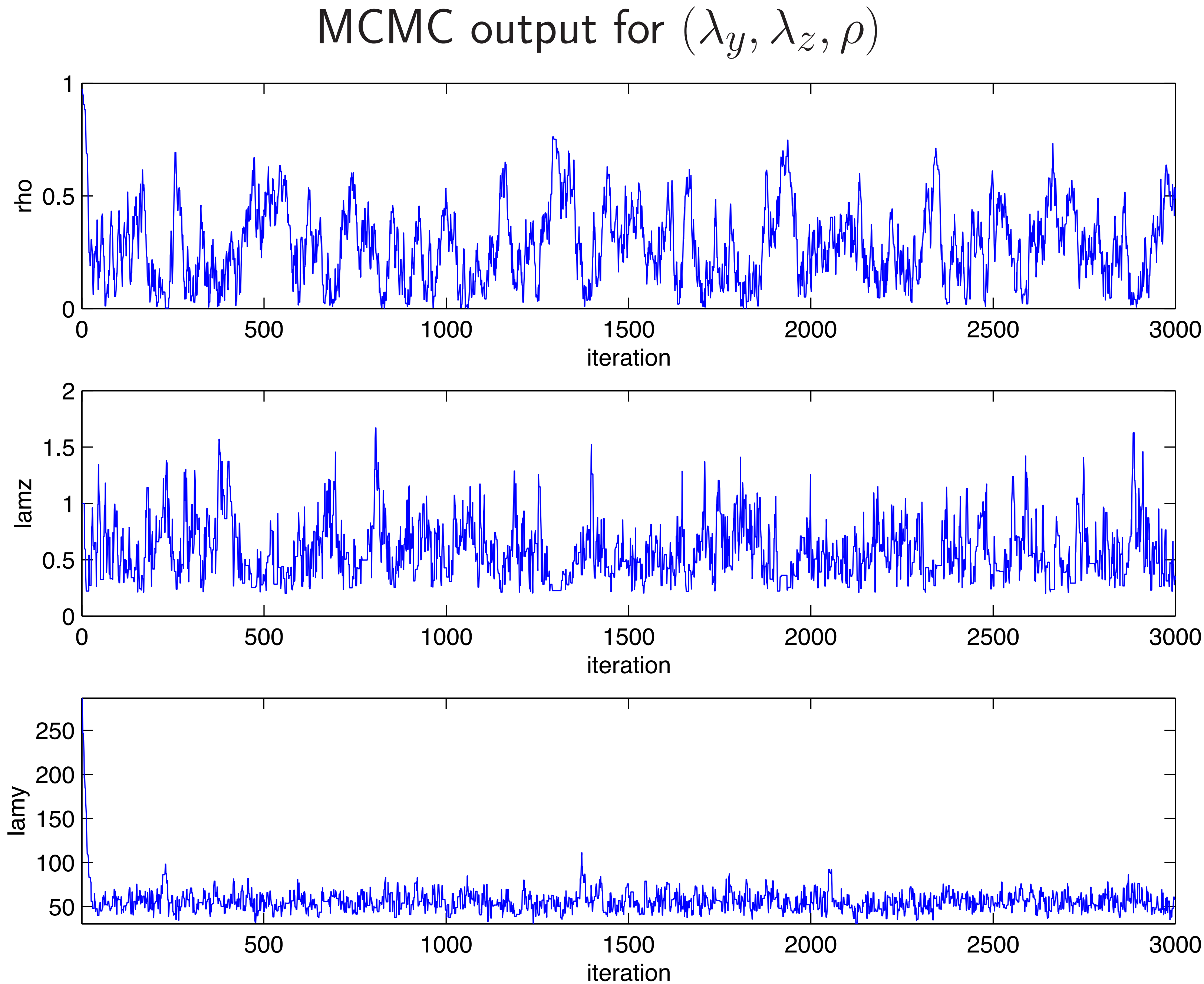
$$\begin{pmatrix} y \\ z^* \end{pmatrix} | \cdots \sim N\left(\begin{pmatrix} 0_n \\ 0_m \end{pmatrix}, \begin{pmatrix} \lambda_\epsilon^{-1} I_n & 0 \\ 0 & 0 \end{pmatrix} + \lambda_z^{-1} R(\rho, (s, s^*))\right)$$

where

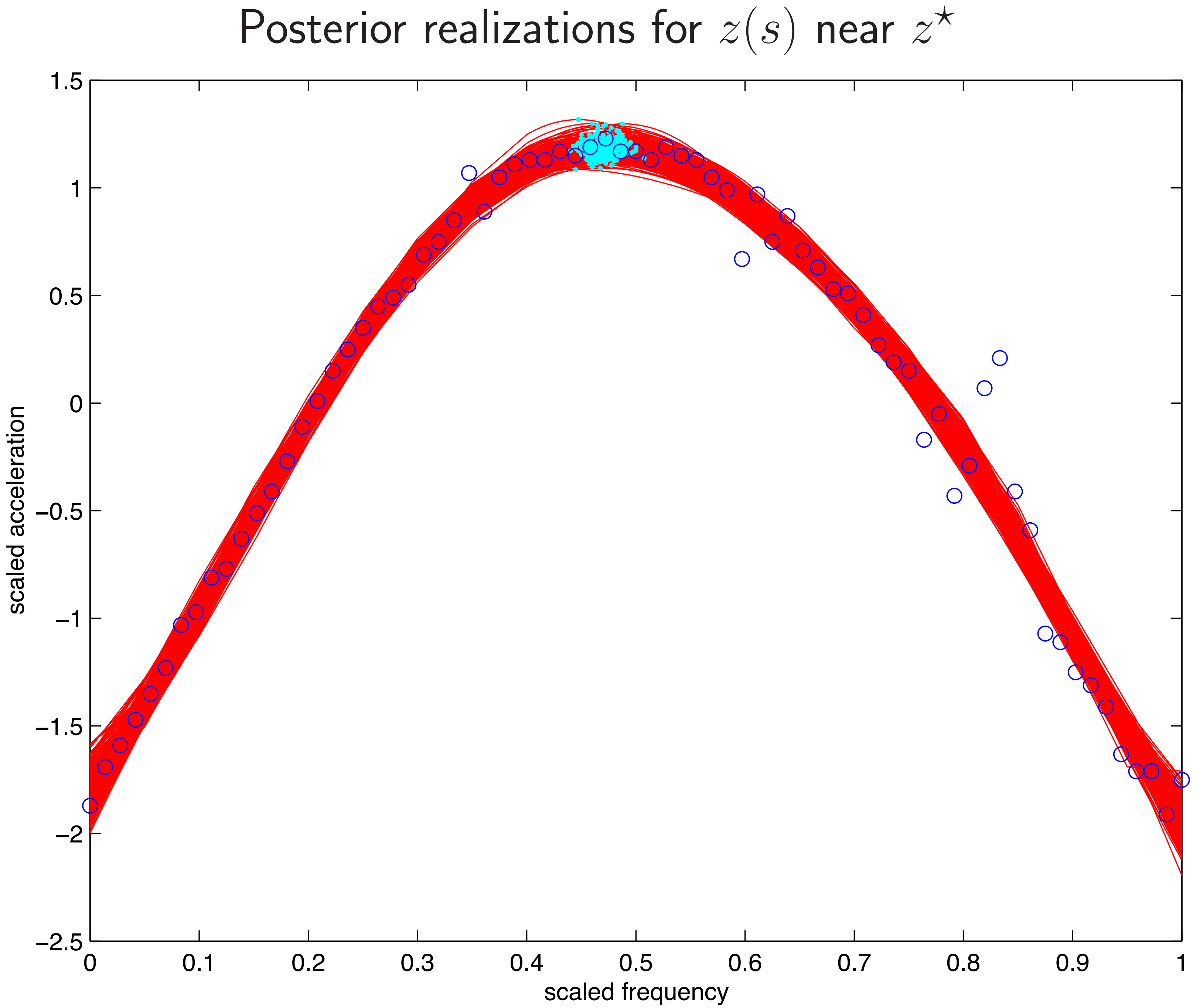
$$R(\rho, (s, s^*)) = \begin{pmatrix} R(\rho, (s, s)) & R(\rho, (s, s^*)) \\ R(\rho, (s^*, s)) & R(\rho, (s^*, s^*)) \end{pmatrix} = \begin{pmatrix} \text{cor rule applied} \\ \text{to } (s, s^*) \end{pmatrix}_{(n+m)\times(n+m)}$$

Therefore $z^*|y \sim N(\mu^*, \Sigma^*)$ where

$$\begin{aligned} \mu^* &= \lambda_z^{-1} R(\rho, (s^*, s))[\lambda_\epsilon^{-1} I_n + \lambda_z^{-1} R(\rho, (s, s))]^{-1} y \\ \Sigma^* &= \lambda_z^{-1} R(\rho, (s^*, s^*)) - \\ &\quad \lambda_z^{-1} R(\rho, (s^*, s))[\lambda_\epsilon^{-1} I_n + \lambda_z^{-1} R(\rho, (s, s))]^{-1} \lambda_z^{-1} R(\rho, (s, s^*)) \end{aligned}$$

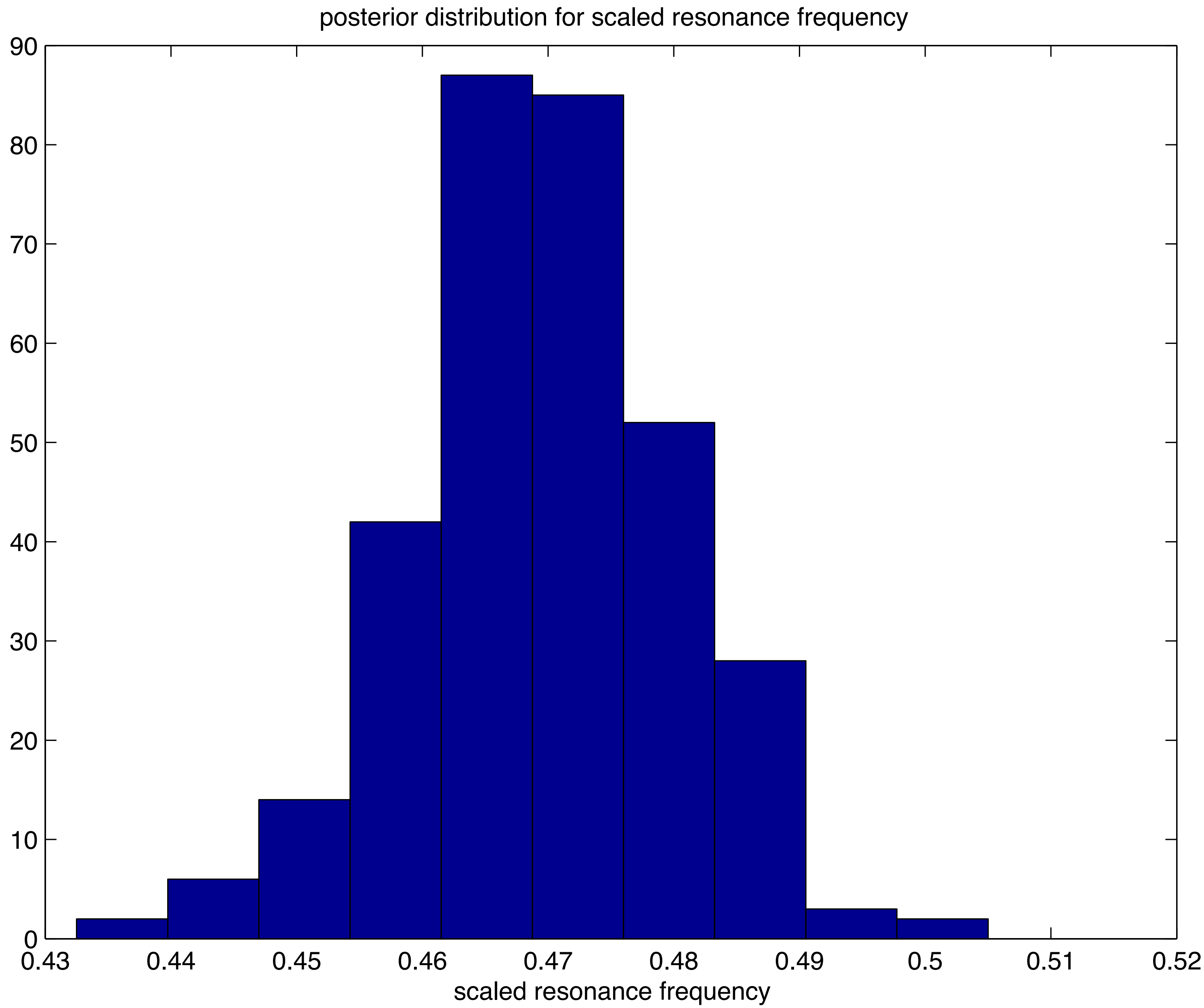


MCMC output for $(\lambda_y, \lambda_z, \rho)$
rho
0
0.5
1
0
500
1000
1500
2000
2500
3000
iteration
lamz
0
0.5
1
1.5
2
0
500
1000
1500
2000
2500
3000
iteration
lamy
50
100
150
200
250
500
1000
1500
2000
2500
3000
iteration



Posterior realizations for $z(s)$ near $z^\star$
scaled acceleration
scaled frequency
1.5
1
0.5
0
−0.5
−1
−1.5
−2
−2.5
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Posterior for resonance frequency $z^\star$



Gaussian Processes for modeling complex computer simulators

data input settings (spatial locations)

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \qquad S = \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix} = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ \vdots & \vdots & \vdots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{np} \end{pmatrix}$$

Model responses y as a (stochastic) function of s

$$y(s) = z(s) + \epsilon(s)$$

Vector form – restricting to the n data points

$$y = z + \epsilon$$

Model response as a Gaussian processes

$$y(s) = z(s) + \epsilon$$

Likelihood

$$L(y|z, \lambda_\epsilon) \propto \lambda_\epsilon^{\frac{n}{2}} \exp\{-\tfrac{1}{2}\lambda_\epsilon (y-z)^T (y-z)\}$$

Priors

$$\pi(z|\lambda_z, \beta) \propto \lambda_z^{\frac{n}{2}} |R(\beta)|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}\lambda_z z^T R(\beta)^{-1} z\}$$

$$\pi(\lambda_\epsilon) \propto \lambda_\epsilon^{a_\epsilon - 1} e^{-b_\epsilon \lambda_\epsilon}, \text{ perhaps quite informative}$$

$$\pi(\lambda_z) \propto \lambda_z^{a_z - 1} e^{-b_z \lambda_z}, \text{ fairly informative if data have been standardized}$$

$$\pi(\rho) \propto \prod_{k=1}^{p} (1-\rho_k)^{-.5}$$

Marginal likelihood (integrating out z)

$$L(y|\lambda_\epsilon, \lambda_z, \beta) \propto |\Lambda|^{\frac{1}{2}} \exp\{-\tfrac{1}{2} y^T \Lambda y\}$$

where $\Lambda^{-1} = \frac{1}{\lambda_\epsilon} I_n + \frac{1}{\lambda_z} R(\beta)$

GASP Covariance model for $z(s)$

$$\text{Cov}(z(s_i), z(s_j)) = \frac{1}{\lambda_z} R(\beta) = \frac{1}{\lambda_z} \prod_{k=1}^{p} \exp\{-\beta_k (s_{ik} - s_{jk})^\alpha\}$$

- Typically $\alpha = 2 \Rightarrow z(s)$ is smooth.
- Separable covariance – a product of componentwise covariances.
- Can handle large number of covariates/inputs p.
- Can allow for multiway interactions.
- $\beta_k = 0 \Rightarrow$ input k is "inactive" $\Rightarrow$ variable selection
- reparameterize: $\rho_k = \exp\{-\beta_k d_0^\alpha\}$ – typically d_0 is a halfwidth.

Posterior Distribution and MCMC

$$\pi(\lambda_\epsilon, \lambda_z, \rho|y) \propto |\Lambda_{\lambda,\rho}|^{\frac{1}{2}} \exp\{-\tfrac{1}{2} y^T \Lambda_{\lambda,\rho} y\} \times \lambda_\epsilon^{a_\epsilon - 1} e^{-b_\epsilon \lambda_\epsilon} \times \lambda_z^{a_z - 1} e^{-b_z \lambda_z} \times \prod_{k=1}^{p} (1-\rho_k)^{-.5}$$

- MCMC implementation requires Metropolis updates.

- Realizations of $z(s)|\lambda, \rho, y$ can be obtained post-hoc:

$-$ define $z^* = (z(s_1^*), \ldots, z(s_m^*))^T$ to be predictions at locations $s_1^*, \ldots, s_m^*$, then

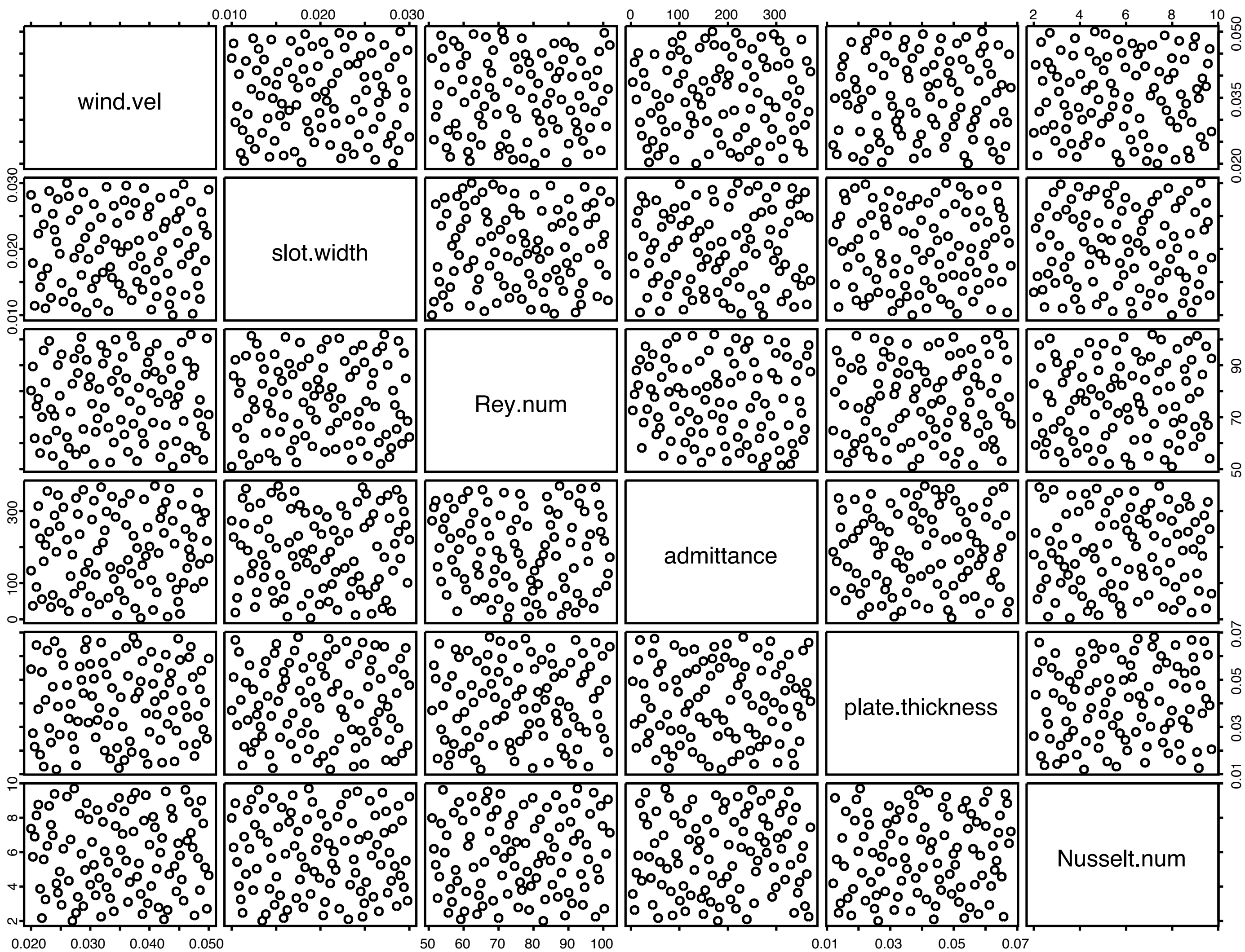
$$\begin{pmatrix} z \\ z^* \end{pmatrix} | \cdots \sim N\left(V \Sigma_y^- \begin{pmatrix} y \\ 0_m \end{pmatrix}, V\right)$$

where

$$\Sigma_y^- = \begin{pmatrix} \lambda_\epsilon I_n & 0 \\ 0 & 0 \end{pmatrix} \text{ and } V^{-1} = \Sigma_y^- + \lambda_z R(\rho, (s, s^*))^{-1}$$

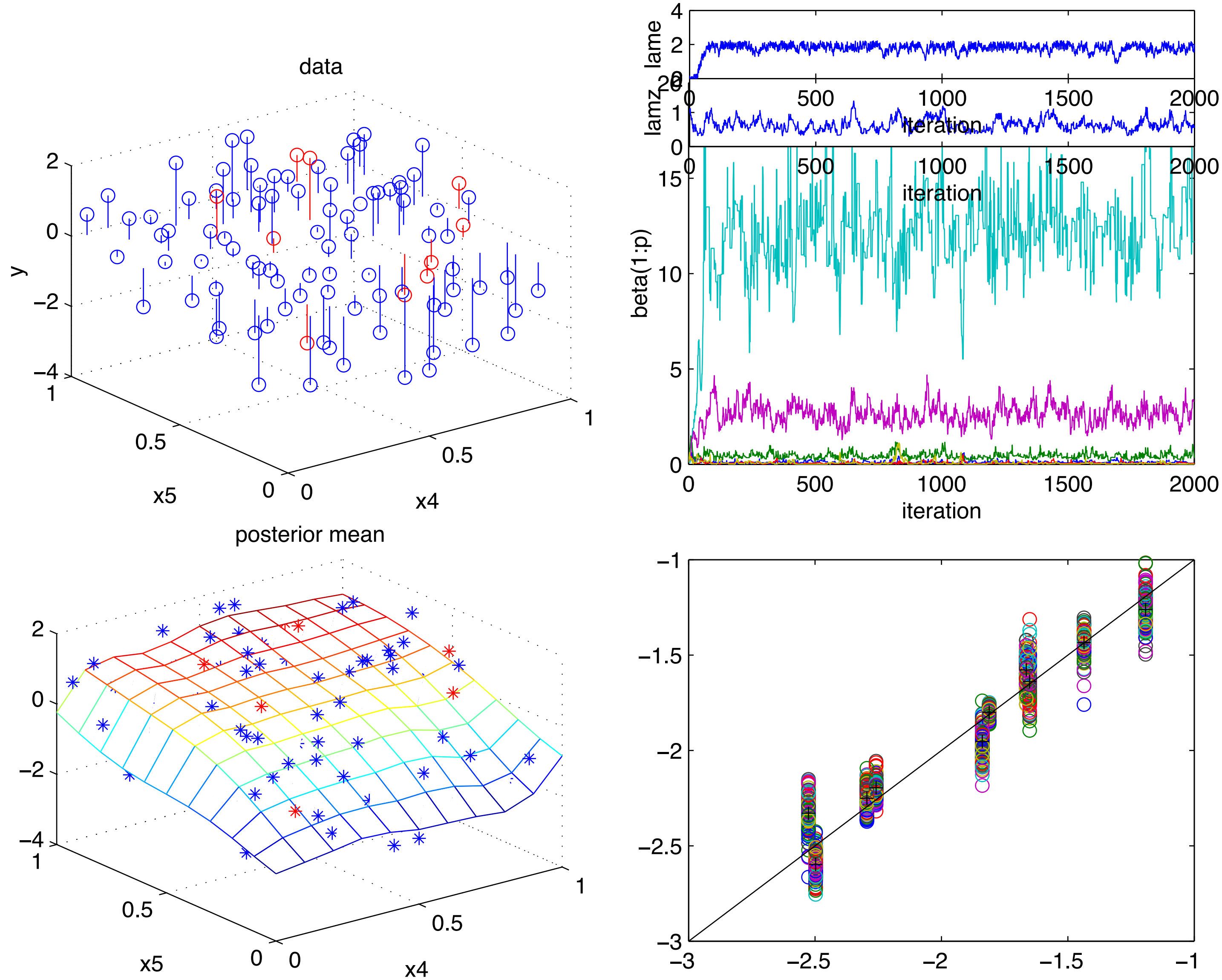
Example: Solar collector Code (Schonlau, Hamada and Welch, 1995)

- $n = 98$ model runs, varying 6 independent variables.
- Response is the increase in heat exchange effectiveness.
- A latin hypercube (LHC) design was used with 2-d space filling.



Example: Solar collector Code

- Fit of GASP model and predictions of 10 holdout points
- Two most active covariates are shown here.



Example: Solar collector Code

- Visualizing a 6-d response surface is difficult
- 1-d marginal effects shown here.