ICOSSAR 19 June 2013

Bayesian System Identification and Response Predictions Robust to Modeling Uncertainty

James L. Beck

George W. Housner Professor of Engineering & Applied Science

Department of Mechanical & Civil Engineering
Department of Computing & Mathematical Sciences

California Institute of Technology

www.its.caltech.edu/~jimbeck/

Focus of Presentation

Performance prediction using a computational model for a dynamic system treating both excitation and modeling uncertainties

- Long history in structural reliability of treating excitation uncertainty under wind and earthquakes (random vibrations, stochastic dynamics)
- Rigorous treatment of modeling uncertainty is only recent

Modeling Uncertainty

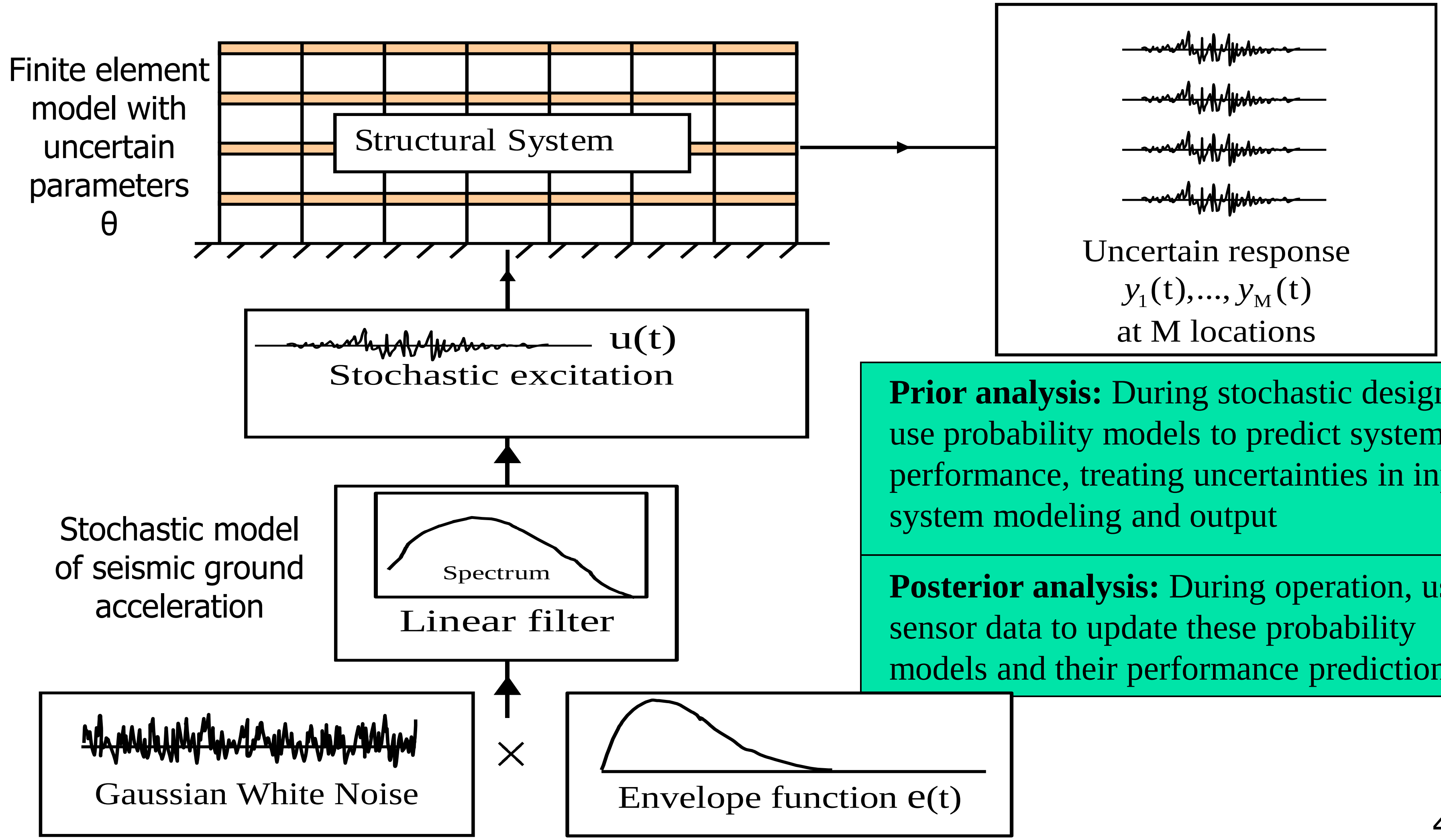
Requires a Bayesian probability approach where the probability of a model is a meaningful concept

- Huge increase in the development and use of Bayesian methods in the last decade or so
- Allows analysis that is robust to modeling uncertainties, both prior (e.g. design based on reliability or life-cycle cost optimization), & posterior (e.g. system ID, structural health monitoring, robust control, state &/or parameter estimation)

Outline of Presentation

- **Introduction**
- Probability logic as the foundation for Bayesian probability
- Stochastic model classes & stochastic embedding
- Bayesian system identification with Ex. 1
- Prior & posterior robust stochastic analyses with Ex. 2
- Posterior model class assessment & selection with Ex.3
- Concluding remarks

Predicting system performance under excitation and modeling uncertainty

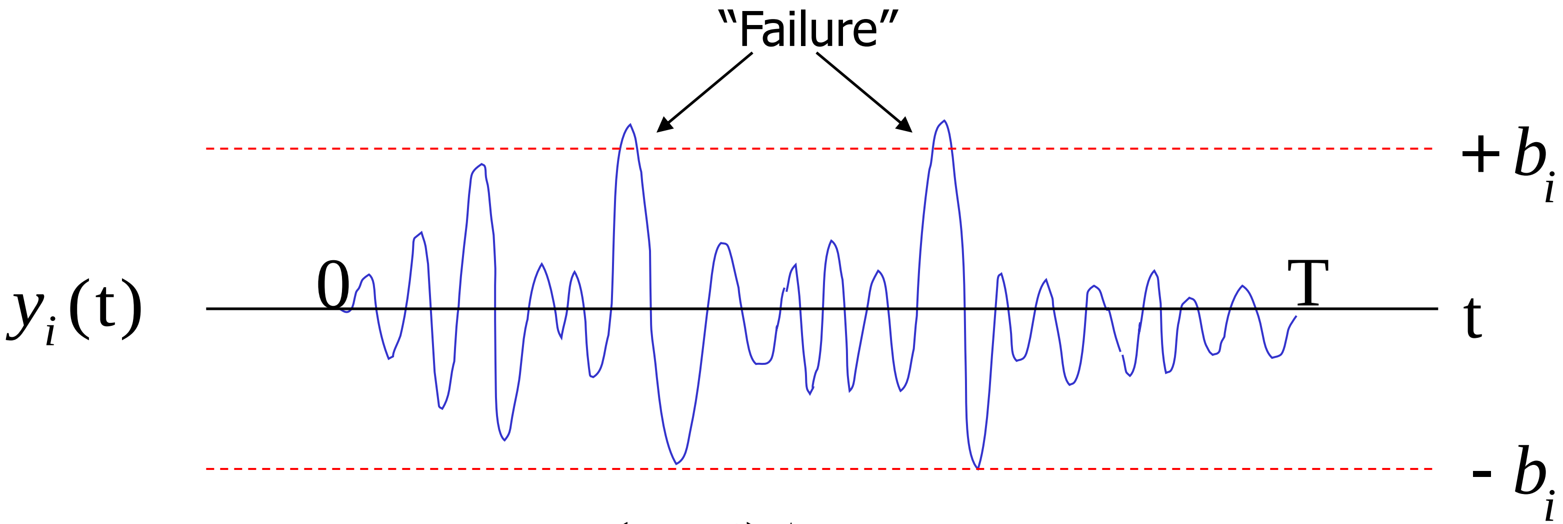


Prior analysis: During stochastic design, use probability models to predict system performance, treating uncertainties in input, system modeling and output

Posterior analysis: During operation, use sensor data to update these probability models and their performance predictions

System performance measure in the presence of uncertainty: Failure probability

- Failure = any one component of $y(\mathrm{t}) = [y_1(\mathrm{t}), \ldots, y_{\mathrm{M}}(\mathrm{t})] \in R^{\mathrm{M}}$ exceeds its specified threshold within a specified duration [0, T]:



- Discretize time so $\mathrm{t}_n = (n-1)\Delta\mathrm{t}$ and failure probability:

$$\mathrm{P}_{\mathrm{F}} = \int_{\mathrm{F}} p(Y_{\mathrm{N}})dY_{\mathrm{N}} = \int \mathrm{I}_{\mathrm{F}}(Y_{\mathrm{N}})\, p(Y_{\mathrm{N}})dY_{\mathrm{N}}, \quad Y_{\mathrm{N}} = [y(t_1), \ldots, y(t_{\mathrm{N}})] \in R^{\mathrm{MN}}$$

$$\mathrm{F} = \{Y_{\mathrm{N}} : |y_i(t_n)| > b_i \text{ for some } i = 1, \ldots, \mathrm{M} \;\&\; n = 1, \ldots, \mathrm{N} = \mathrm{Int}[\mathrm{T}/\Delta\mathrm{t}] + 1\}$$

Interpretation of probability

- The axioms of probability are well-established but after three centuries, the meaning of probability is still in dispute
- The interpretation is important in applications to real systems and phenomenon – it governs:
  - perceived domain of its applicability;

    e.g. is the probability of a model meaningful?
  - understanding of the results of stochastic analysis;

    e.g. what does the failure probability mean?

    (Is it an inherent property of the system, or

    a property of what we know about the system and its future excitation?)

Two prevailing interpretations of probability: Frequentist & Bayesian

Frequentist

<u>Defn</u>: Probability is the relative frequency of occurrence of an "inherently random" event in the "long run"

1) Probability distributions are <u>inherent</u> properties of "random" phenomena
2) Limited scope, e.g. no meaning for the probability of a model
3) "Inherent randomness" is assumed but cannot be proved

Bayesian

<u>Defn</u>: Probability is a measure of the plausibility of a statement based on specified information

1) Probability distributions represent states of <u>plausible knowledge</u> about systems and phenomena, not inherent properties of them
2) Probability of a model is a measure of its <u>plausibility</u> relative to other models in a set
3) Pragmatically quantifies uncertainty due to missing information without any claim that this is due to nature's "inherent randomness"

Frequentist vs Bayesian

Is it "inherent randomness " or does it just "look random"?

(a) Data-stream from a random number generator "looks random" but it is deterministic if the algorithm and initial condition ("seed") are known;

(b) The outcome of coin tosses "looks random" but it is a deterministic mechanics problem if the initial conditions are known

Philosophical question

Is uncertainty about predicting any phenomenon because of:

(a) an <u>inherent</u> "randomness" property of the phenomenon (producing aleatory/aleatoric uncertainty), or

(b) due to our limited capacity to collect or understand the relevant information (producing epistemic uncertainty due to missing information)?

- E.T. Jaynes' answer (2003): (a) is an example of the <u>Mind-Projection Fallacy</u>:
  Our models of reality are confused with reality,
  or more specifically here:
  <u>Our</u> uncertainty is projected onto nature as its inherent property

Early quotes regarding probability theory

- **Pierre Simon Laplace (1814):**

  Probability theory is nothing but common sense reduced to calculation [Supplement to his Analytical Theory of Probabilities, 1812]

- **James Clerk Maxwell (1850):**

  The true logic of this world is the calculus of probabilities, which takes account of the magnitude of the probability which is, or ought to be, in a reasonable man's mind

A connection between Bayesian probability and relative frequency

Take a binomial probability model for an uncertain event A with model parameter θ = probability of event A occurring.

If all values of θ are equally plausible a priori, then the most probable value a posteriori of parameter θ is:

$\hat{\theta}$ = relative frequency of event A in any finite number of trials

[The full posterior distribution is a beta PDF]

Outline of Presentation

- Introduction
- **Probability logic as the foundation for Bayesian probability**
- Stochastic model classes & stochastic embedding
- Bayesian system identification with Ex. 1
- Prior & posterior robust stochastic analyses with Ex. 2
- Posterior model class assessment & selection with Ex.3
- Concluding remarks

Probability logic: Rigorous foundation for Bayesian probability

- **Extends binary Boolean logic** to a multi-valued logic for quantification of plausible reasoning under incomplete information

- **Key idea**: **Probability** $\mathbf{P}[\boldsymbol{b}|\boldsymbol{a}]$ = measure of **plausibility** of statement $\boldsymbol{b}$ based on the information in statement $\boldsymbol{a}$

  [$\mathbf{P}[\boldsymbol{b}|\boldsymbol{a}]$=1 if $\boldsymbol{a}$ is true implies $\boldsymbol{b}$ is true; =0 if $\boldsymbol{a}$ implies $\boldsymbol{b}$ is false]

- **Seminal work on foundations by R.T. Cox:**

  "Probability, Frequency and Reasonable Expectation", Amer. J. Physics 1946

  The Algebra of Probable Inference, Johns Hopkins Press 1961

- **Treatise on theory and applications by E.T. Jaynes:**

  Probability Theory – The Logic of Science, Cambridge U. Press 2003

Probability logic axioms

- By extending Boolean logic to incomplete information, R.T. Cox derived a minimal set of axioms for probability logic (1946, 1961):

  For any statements a, b and c,

  P1: $\mathrm{P}[b|a] \geq 0$ [By convention]

  P2: $\mathrm{P}[\sim b|a] = 1 - \mathrm{P}[b|a]$ [Negation Function]

  P3: $\mathrm{P}[c\&b|a] = \mathrm{P}[c|b\&a]\mathrm{P}[b|a]$ [Conjunction Function]

  - These axioms and De Morgan's Law of Logic imply Disjunction Function:

    $$\mathrm{P}[c \text{ or } b|a] = \mathrm{P}[c|a] + \mathrm{P}[b|a] - \mathrm{P}[c\&b|a]$$

- They imply Kolmogorov's statement of probability axioms (1933) for probability measure $P(A)$, which has no built-in meaning, by using:

  If x = uncertain-valued variable with possible values in set X, then:

  $P(A) = \mathrm{P}[x \in A \,|\, M],\ A \subset X$ (M specifies probability model for x)

Outline of Presentation

- Introduction
- Probability logic as the foundation for Bayesian probability
- **Stochastic model classes & stochastic embedding**
- Bayesian system identification with Ex. 1
- Prior & posterior robust stochastic analyses with Ex. 2
- Posterior model class assessment & selection with Ex.3
- Concluding remarks

Fundamental concept for treating system modeling uncertainty

- **Stochastic model class** $\mathcal{M}$ defined by two steps:

  (1) Choose set of stochastic I/O (forward) models for prediction:

  $$\{\, p(Y_n \mid U_n, \theta, \mathcal{M}) : \theta \in \Theta \subset R^{N_p} \,\}$$

  U_n, Y_n= input & output time histories up to any time t_n

  θ=uncertain model parameters

  e.g. use stochastic embedding of a parameterized deterministic I/O model

  (2) Choose PDF $p(\theta|\mathcal{M})$ (prior) over this set that expresses the initial relative plausibility of each stochastic model in (1)

- Model class $\mathcal{M}$ treats modeling uncertainty, both "parametric" and "non-parametric" (from "unmodeled dynamics")
- If system data $\mathcal{D}_N = \{\hat{Y}_N, \hat{U}_N\}$ is available, then Bayes' Theorem using (1) and (2) gives updated PDF (posterior):

$$p(\theta \mid \mathcal{D}_N, \mathcal{M}) \propto p(\hat{Y}_N \mid \hat{U}_N, \theta, \mathcal{M})\, p(\theta \mid \mathcal{M})$$

Developing stochastic I/O models for dynamic systems

- **General strategy:** Develop a probability model for stochastic predictions of the system output <u>conditional</u> on its input and model parameters θ:

$$\boxed{p(Y_n \mid U_n, \theta, \mathcal{M})}$$

$$u_k \in R^{N_I} \xrightarrow[\text{Inputs}]{\text{Given}} \boxed{\text{Actual System}} \xrightarrow[\text{Outputs}]{\text{Uncertain}} y_k \in R^{N_o}$$

where <u>discrete time histories</u> for N_o predicted <u>outputs</u> and N_I corresponding <u>inputs</u> up to any time t_n are denoted by:

$$Y_n = [y_k : k = 1, \ldots, n] \in R^{nN_o}$$

$$U_n = [u_k : k = 1, \ldots, n] \in R^{nN_I}$$

(possibly derived from continuous-time stochastic models)

Developing stochastic I/O models by stochastic embedding: Method 1

- **Stochastic embedding:** Start with <u>deterministic</u> I/O model of a dynamic system giving $q_n(U_n,\theta),\ \theta \in \Theta \subset R^{N_p}$

$$u_n \in R^{N_I} \xrightarrow{\text{Input}} \boxed{\begin{array}{l}\text{System}\\ \text{Model } \theta\end{array}} \xrightarrow{\text{Output}} q_n \in R^{N_o}$$

  - **Uncertain output prediction error** introduced by:

$$v_n = y_n - q_n(U_n,\theta)$$

Model output

System output

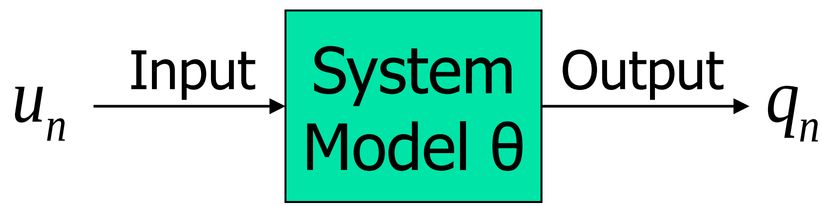
  - $p(Y_n \mid U_n,\theta,\sigma)$ from prob. model for prediction-error time history e.g. take prediction errors as zero-mean Gaussian white-noise (<u>maximum</u> <u>entropy</u> <u>probability</u> <u>model</u>) so system output is i.i.d.

$$y_n \sim \mathrm{N}(q_n(U_n,\theta),\Sigma(\sigma)) \quad \text{at each time } t_n$$

Developing stochastic I/O models by stochastic embedding: Method 2

- **Stochastic embedding of state-space models:** Start with <u>deterministic</u> state-space model of a dynamic system giving $q_n(U_n,\theta)$ for $\theta \in \Theta \subset R^{N_p}$ in terms of a state evolution eqn:

$$\tilde{x}_n = F(\tilde{x}_{n-1}, u_{n-1}, \theta)$$
$$q_n = H(\tilde{x}_n, u_n, \theta)$$



- **Uncertain state and output prediction errors** (from "unmodeled dynamics") introduced by:

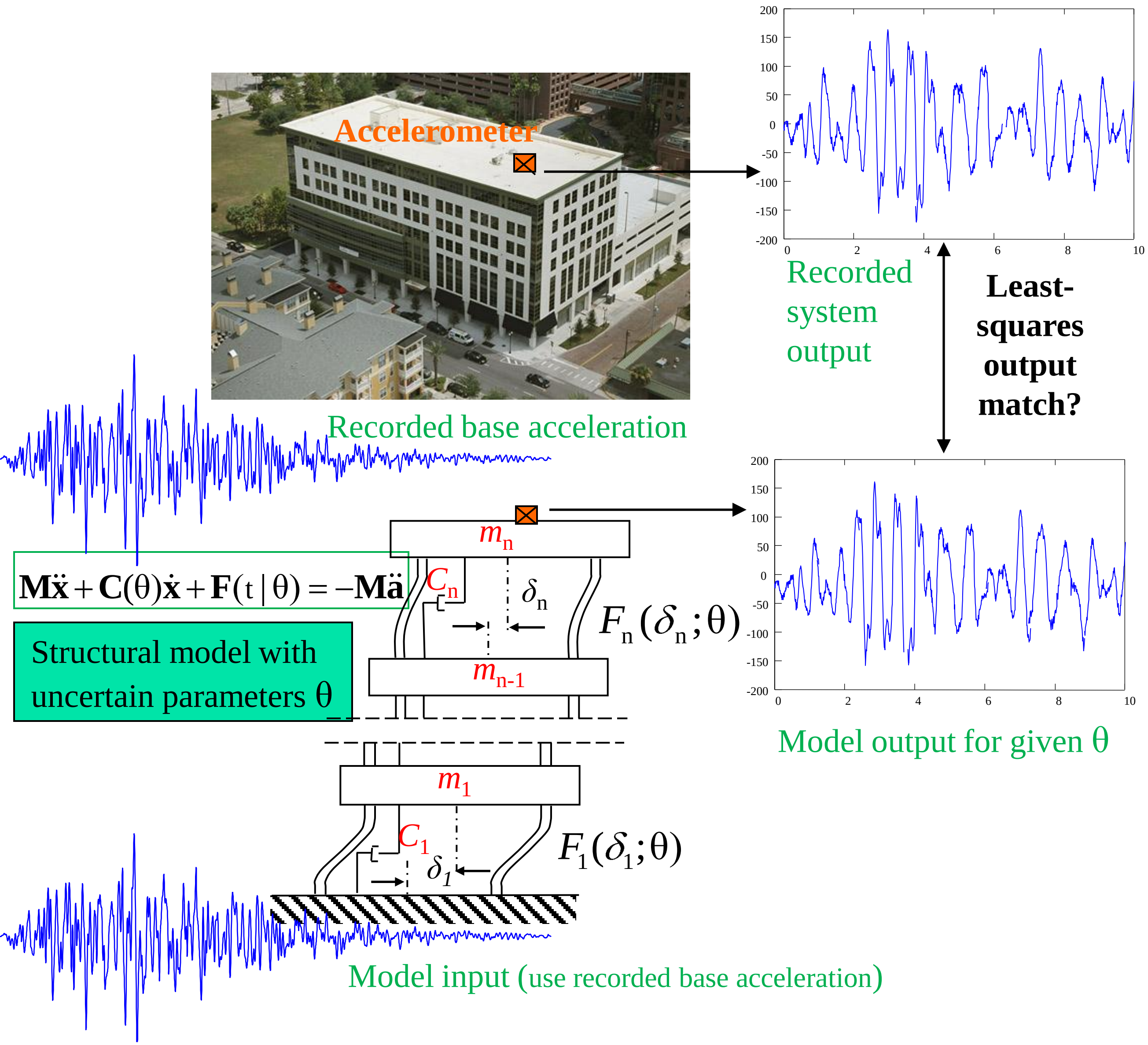
$$x_n = F(x_{n-1}, u_{n-1}, \theta) + w_n \quad \& \quad y_n = H(x_n, u_n, \theta) + v_n$$

- $p(Y_n \mid U_n, \theta)$ defined by probability models for missing information, i.e. initial state x_0 and time histories of uncertain state and output prediction errors, w_n and v_n (use <u>maximum</u> <u>entropy</u> PDFs)

Outline of Presentation

- Introduction
- Probability logic as the foundation for Bayesian probability
- Stochastic model classes & stochastic embedding
- **Bayesian system identification with Ex. 1**
- Prior & posterior robust stochastic analyses with Ex. 2
- Posterior model class assessment & selection with Ex.3
- Concluding remarks

System identification: Typical approach



Goal: Use system I/O data $\mathcal{D}$ to develop a model to represent I/O behavior of a system

Typical Approach: Propose a deterministic model with uncertain parameter vector $\theta \in \Theta$ then **estimate** its value by using data $\mathcal{D}$

e.g. least-squares output-error, maximum likelihood or maximum a posteriori (MAP) estimates

System identification: Typical approach

- **Problem #1**: No model is expected to exactly represent system I/O behavior – so no true parameter values to estimate!
- **Problem #2**: Parameter estimates are often **non-unique** (model is **unidentifiable** based on data $\mathcal{D}$) – fixing some parameter values to make unique estimates for the others may produce biased predictions
- **Problem #3**: Every model will have uncertain **prediction errors** (e.g. "unmodeled dynamics") – how can we quantify this uncertainty?
- **Resolution:** Use **Bayesian system identification**

Bayesian system identification

- System ID for dynamic systems is viewed as inference about plausible system models based on data - not a quest for the "true" model or parameter values

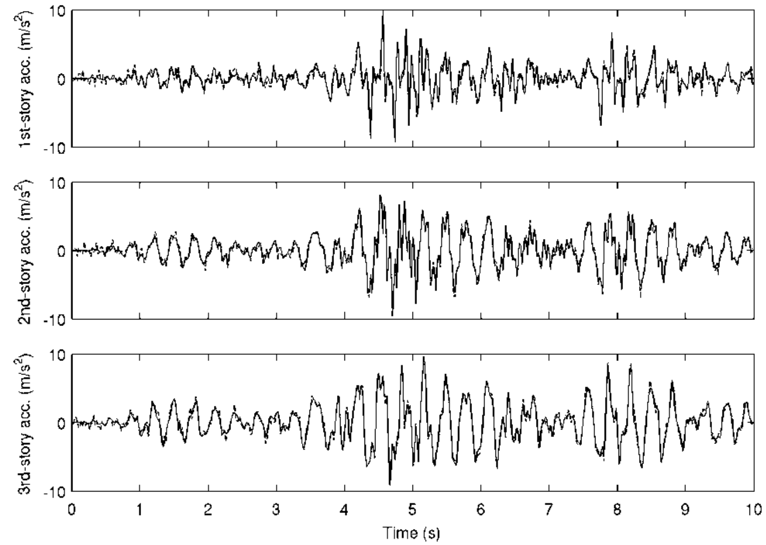
- Provides a rigorous stochastic framework for quantifying modeling uncertainty based only on probability axioms and probability models defined by stochastic model classes

- Instead of parameter estimation, system data is used to do Bayesian updating of the probability of each stochastic I/O model in the parameterized set

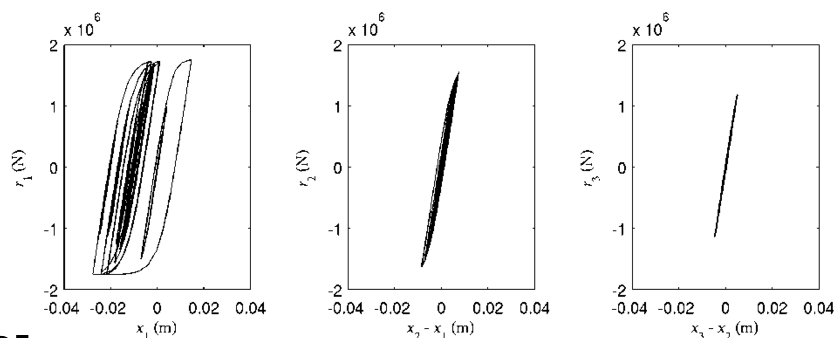
Example 1: Bayesian system ID using Masing hysteretic structural models

- **Three-story building**:
  - Input is strong ground motion record from 1994 Northridge Earthquake, Los Angeles (10 s with time-step=0.02 s)
  - Masing hysteretic model for inter-story force versus deformation used to generate synthetic sensor data
  - Output is noisy accelerations at each floor (20% RMS noise added to "simulate" modeling errors & measurement noise)

Floor accelerations (used in updating)



Hysteresis loops (not used in updating)



1st story **2nd story** **3rd story**

[From Muto & Beck: J. Vibration & Control 2008]

Stochastic model class $\mathcal{M}$ for system ID using inelastic seismic response

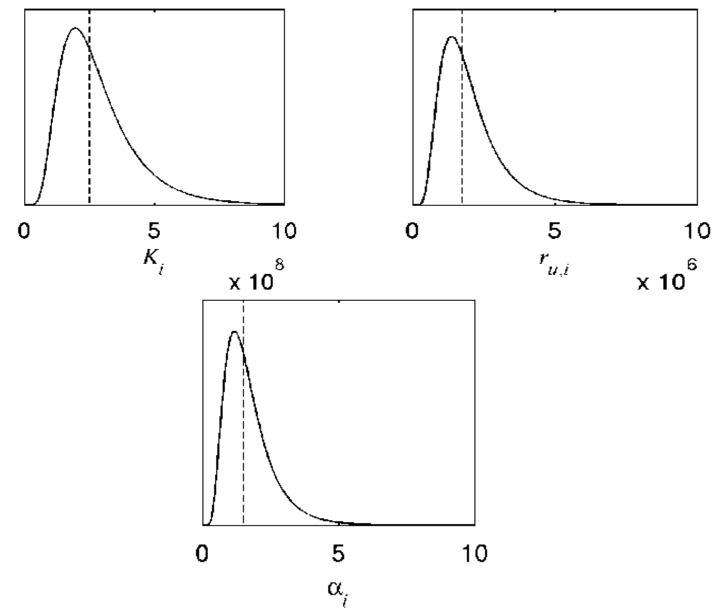
- **Stochastic I/O model:** From <u>stochastic embedding</u> of Masing shear building model (Method 1):

$$\mathbf{M}\ddot{\boldsymbol{x}}(t)+\mathbf{C}(\theta)\dot{\boldsymbol{x}}(t)+\mathbf{F}(t,\boldsymbol{x}\,|\,\theta)=-\mathbf{Mb}u(t)$$

$$\text{Output, } \boldsymbol{y}_n=\ddot{\boldsymbol{x}}(t_n)+\boldsymbol{e}_n \quad \leftarrow N(0,\sigma^2 I_3)$$

Model class $\mathcal{M}$ has 12 parameters (3 per story for force-deformation, 2 Rayleigh damping & σ^2, a common acceleration prediction-error variance)

Prior PDFs



- **Prior PDFs:** Plots show independent lognormal prior PDFs on hysteretic parameters (c.o.v. of 0.5) and exact values (dashed) used to generate data. Uniform priors on Rayleigh damping parameters & prediction-error variance (not shown)

Bayesian updating for model class $\mathcal{M}$

[T.Bayes 1763; P.S.Laplace 1774,1812; H.Jeffreys 1939]

- Given data $\mathcal{D}$, use Bayes' Theorem to update initial plausibility of each model:

Likelihood

Prior (Initial) PDF

$$p(\theta\,|\,\mathcal{D},\mathcal{M}) = \frac{p(\mathcal{D}\,|\,\theta,\mathcal{M})\,p(\theta\,|\,\mathcal{M})}{p(\mathcal{D}\,|\,\mathcal{M})}$$

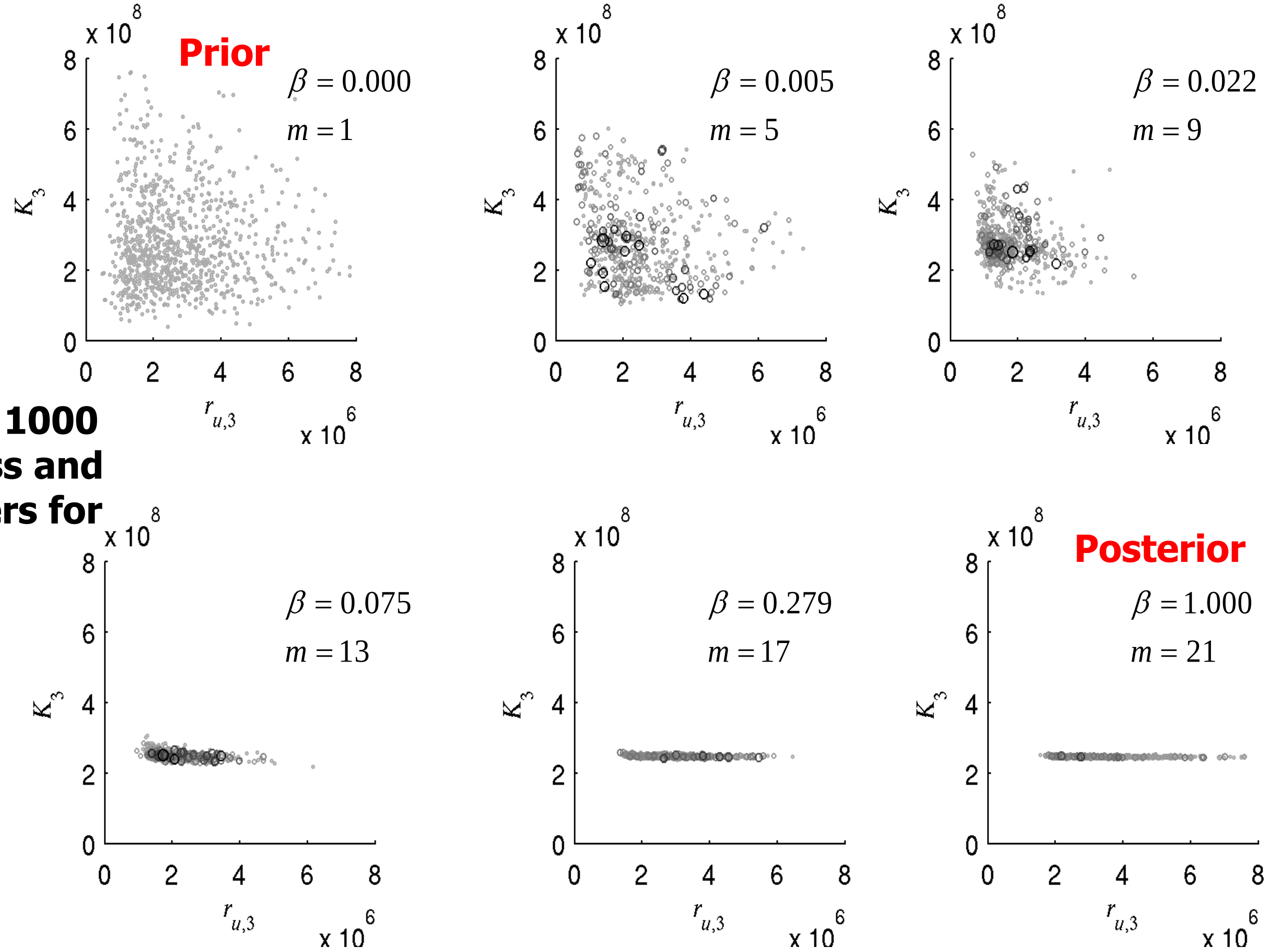
Normalizing Constant
= Evidence for $\mathcal{M}$

Posterior (Updated) PDF

- Laplace's method for large amount of data:
  Gaussian approximation of posterior PDF about MAP or MLE: $\mathrm{N}(\hat{\theta},\Sigma(\hat{\theta}))$
  - challenging high-D optimization to find MAP/MLE parameter values
  - requires $\mathcal{M}$ to be globally identifiable based on data $\mathcal{D}$ (then MAP≈MLE)
  - covariance matrix $\Sigma(\hat{\theta})$ is the inverse of the negative of Hessian matrix of log posterior (for MAP) or log likelihood (for MLE)
- Markov Chain Monte Carlo methods now dominant: draw posterior samples using either Gibbs Sampler, Metropolis-Hastings algorithm, Hamiltonian/Hybrid MC, etc.

Multi-level MCMC sampling by annealing

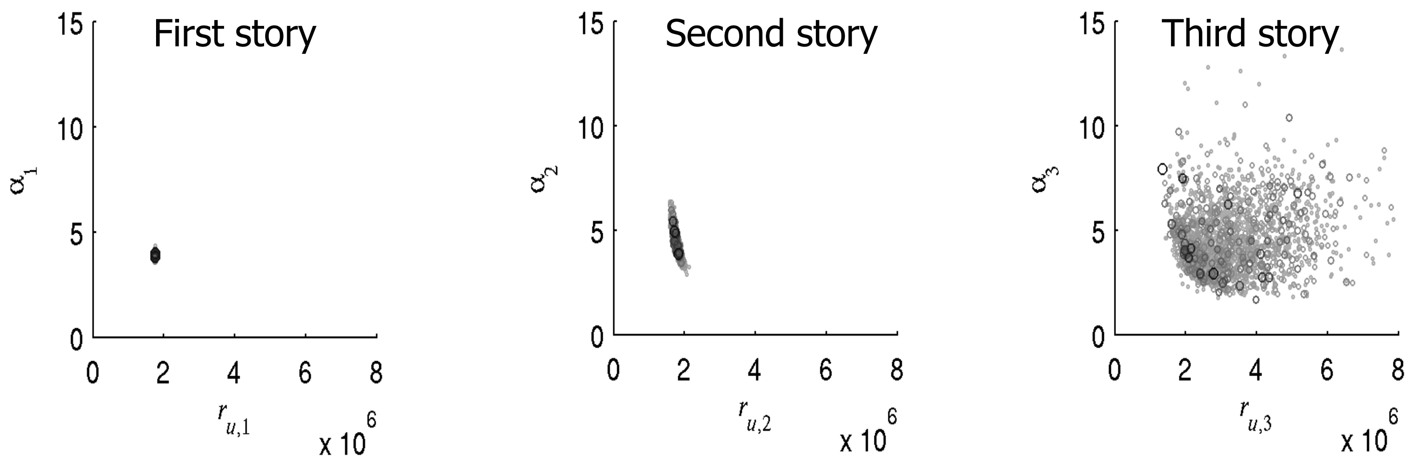
Target PDF: $\pi_m(\boldsymbol{\theta}) \propto p^{\beta_m}(\mathcal{D}\,|\,\boldsymbol{\theta},\mathcal{M})\,p(\boldsymbol{\theta}\,|\,\mathcal{M})$ where $0=\beta_1<\beta_2<..<\beta_M=1$



Each population is 1000 samples of stiffness and strength parameters for third story

Projection of 12-dimensional samples generated by Metropolis algorithm at different levels of one run of <u>TMCMC</u> <u>algorithm</u> [Ching & Chen: JEM 2007] when updating using model class $\mathcal{M}$ (Repeated samples are indicated by size of markers)

Posterior samples from TMCMC using inelastic seismic building data



Population is 1000 samples of yield transition and strength parameters

Projection of 12-dimensional posterior samples from final level of TMCMC when updating using the model class. Last 2 slides show that model class is <u>unidentifiable</u> based on data

<u>Robust predictions</u> based on <u>all</u> posterior samples should be used, not just for one parameter estimate; Laplace asymptotic approximation not applicable

Outline of Presentation

- **Introduction**
- Probability logic as the foundation for Bayesian probability
- Stochastic model classes & stochastic embedding
- Bayesian system identification with Ex. 1
- **Prior & posterior robust stochastic analyses with Ex. 2**
- Posterior model class assessment & selection with Ex.3
- Concluding remarks

Robust stochastic system analysis

- Use the whole model class $\mathcal{M}$ for robustness w.r.t. modeling uncertainties in the prediction of system performance:
  - **Combine all stochastic predictions** $p(Y_n|U_n,\theta,\mathcal{M})$ for each $\theta\in\Theta$ weighted by their probability conditional on the available information (Theorem of Total Probability)
  - **Prior robust analysis** uses weighting by prior PDF $p(\theta|\mathcal{M})$
  - **Posterior robust analysis** uses weighting by posterior PDF $p(\theta|\mathcal{D}_N,\mathcal{M})$
  - **Prior and posterior hyper-robust analyses** use multiple candidate model classes for a system, weighting each of their robust stochastic predictions by their prior and posterior probabilities, respectively (Bayesian model averaging)

Prior robust analysis for system design

- **Ingredients**:

  (1) Stochastic model class $\mathcal{M}$ defining set of I/O probability models $\{p(Y_n \mid U_n, \theta, \mathcal{M}) : \theta \in \Theta\}$ and prior $p(\theta|\mathcal{M})$ over set

  (2) Stochastic input model $\mathcal{U}$ defining PDF $p(U_n \mid \mathcal{U})$

- **Prior robust failure probability** (Total Probability Theorem):

  Failure: output time-history Y_n up to time t_n lies in F_n = region of unacceptable performance

$$\mathrm{P_F} = \mathrm{P}[Y_n \in F_n \mid \mathcal{U}, \mathcal{M}]$$

Modeling uncertainty   Input uncertainty

$$= \int \mathrm{I}(Y_n \in F_n)\, p(Y_n \mid U_n, \theta, \mathcal{M})\, p(\theta \mid \mathcal{M})\, p(U_n \mid \mathcal{U})\, dY_n\, d\theta\, dU_n$$

$$= \mathrm{E}[\mathrm{I}(Y_n \in F_n) \mid \mathcal{U}, \mathcal{M}] \approx \frac{1}{\mathrm{K}} \sum_{k=1}^{\mathrm{K}} \mathrm{I}(Y_n^{(k)} \in F_n)$$

- Evaluated by drawing K independent samples $(Y_n^{(k)}, U_n^{(k)}, \theta^{(k)})$ (Monte Carlo Simulation), or if $\mathrm{P_F} << 1$, use Subset Simulation or ISEE (Au & Beck:PEM 2001)

Posterior robust analysis for updated system performance predictions

- **Ingredients:**

  (1) Stochastic model class $\mathcal{M}$

  (2) Stochastic input model $\mathcal{U}$ defining PDF $p(U_n \mid \mathcal{U})$

  (3) System input and output data $\mathcal{D}_N$

- **Posterior robust failure probability** (Bayes Theorem and Total Probability Theorem):

Modeling uncertainty

Input uncertainty

$$\mathrm{P}[Y_n \in F_n \mid \mathcal{D}_N, \mathcal{U}, \mathcal{M}]$$

$$= \int \mathrm{I}(Y_n \in F_n)\, p(Y_n \mid U_n, \theta, \mathcal{M})\, p(\theta \mid \mathcal{D}_N, \mathcal{M})\, p(U_n \mid \mathcal{U})\, dY_n\, d\theta\, dU_n$$

$$= \mathrm{E}[\mathrm{I}(Y_n \in F_n) \big| \mathcal{D}_N, \mathcal{U}, \mathcal{M}] \approx \frac{1}{\mathrm{K}} \sum_{k=1}^{\mathrm{K}} \mathrm{I}(Y_n^{(k)} \in F_n)$$

- Evaluated by drawing independent samples $(Y_n^{(k)}, U_n^{(k)}, \theta^{(k)})$ where posterior samples are generated by using MCMC simulation, e.g. multi-level methods such as TMCMC based on Metropolis-Hastings algorithm

Posterior robust analysis for updated system performance predictions

- **Laplace's method for posterior robust analysis:**
  - Find <u>optimal</u> (MAP or MLE) <u>parameter values</u> $\hat{\theta}$, i.e. maximize posterior $p(\theta|\mathcal{D}_N,\mathcal{M})$ or likelihood $p(\mathcal{D}_N|\theta,\mathcal{M})$, then to $O(\frac{1}{N})$ (Laplace's asymptotic approximation):

$$\mathrm{P}[Y_n \in F_n \,|\, \mathcal{D}_N, \mathcal{U}, \mathcal{M}] \approx \mathrm{P}[Y_n \in F_n \,|\, \mathcal{D}_N, \mathcal{U}, \hat{\theta}, \mathcal{M}]$$
$$= \int \mathrm{I}(Y_n \in F_n)\, p(Y_n \,|\, U_n, \hat{\theta}, \mathcal{M})\, p(U_n \,|\, \mathcal{U})\, dY_n dU_n$$

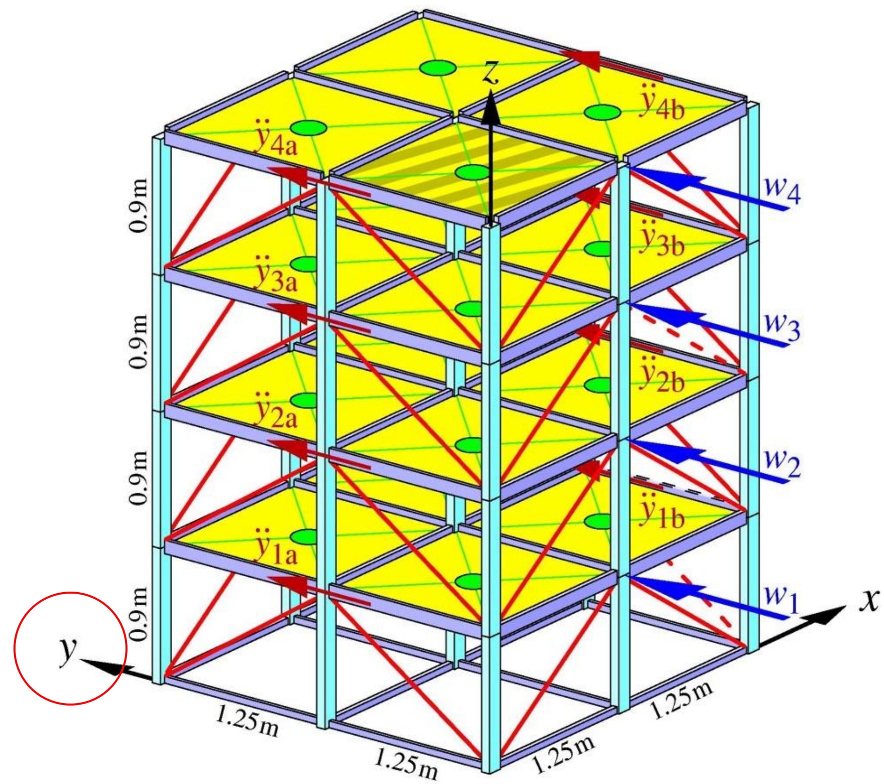
  - Acceptable approximation only if $\mathcal{M}$ is <u>globally identifiable</u> on $\mathcal{D}_N$ (MLE is unique) and large amount of data N (then MAP≈MLE)
  - <u>Parameter estimation</u> (i.e. using $\hat{\theta}$) is reasonable only in this case (otherwise spurious reductions in prediction uncertainty)
  - <u>Ref</u>: Beck & Katafygiotis: J. Eng. Mech. 1998

    Papadimitriou et al.: Prob. Eng. Mech. 2001

Example 2: Prior & posterior robust analysis using linear state-space models

- 4-story, 1/4-scale steel-frame ASCE SHM benchmark structure (JEM 2004)
- Data $\mathcal{D}$ =10 s (time-step=0.004s) of noisy horizontal acceleration (output) & wind force (input) at each floor (synthetic data from 120-DOF 3D benchmark finite-element model)


Photo courtesy Prof. Carlos Ventura, UBC



Stochastic embedding of state-space model

- **Stochastic linear state-space model of system:**

$$\mathbf{x}_n = \mathbf{A}(\boldsymbol{\theta}_s)\mathbf{x}_{n-1} + \mathbf{B}(\boldsymbol{\theta}_s)\mathbf{u}_{n-1} + \mathbf{w}_n, \quad n = 1, 2, \ldots$$
$$\mathbf{y}_n = \mathbf{C}(\boldsymbol{\theta}_s)\mathbf{x}_n + \mathbf{D}(\boldsymbol{\theta}_s)\mathbf{u}_n + \mathbf{v}_n, \quad n = 0, 1, 2, \ldots$$

  - System matrices $\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}$**:** use discrete-time version of equations of motion for shear-building model with 4 DOF: State $\mathbf{x}_n$= displacements & velocities, and output $\mathbf{y}_n$= accelerations, at each floor
  - **Stochastic embedding**: Apply Methods 1 and 2, so establish probability models for $\mathbf{x}_o$, $[\mathbf{w}_1 \ldots \mathbf{w}_n]$ & $[\mathbf{v}_1 \ldots \mathbf{v}_n]$ by using the Principle of Maximum Information Entropy: For all n,

    $\mathbf{w}_n \sim \mathrm{N}(\mathbf{0},\Sigma_w)$, $\mathbf{v}_n \sim \mathrm{N}(\mathbf{0},\Sigma_v)$, and $\mathbf{x}_o \sim \mathrm{N}(\mathbf{0},\Sigma_o)$ and all are independent with diagonal covariance matrices Σ_w, Σ_v and Σ_o

First stochastic model class for benchmark structure

- **First model class** $\mathcal{M}_1$ [Stochastic embedding: Method 1]:

  $\boldsymbol{\theta}$ = 13 parameters, i.e. $\boldsymbol{\theta}_s$ = 12 mass, stiffness & damping parameters, Σ_o=0 & Σ_w=0 (no state prediction errors), $\Sigma_v = \sigma_{acc}^2 I_4$ is 4x4 diagonal matrix with acceleration prediction-error variance the only parameter
- **Prior PDF** $p(\boldsymbol{\theta}|\mathcal{M}_1)$: $\boldsymbol{\theta}_s$ – independent lognormals with median = nominal value, and coefficients of variation= 0.1,0.3 & 0.5 for mass, stiffness and damping ratio parameters, respectively; $\sigma_{acc}^2 \sim U[0, \sigma_{max}^2]$
- **Likelihood function** for $\mathcal{M}_1$: With N=2500 and $\mathbf{q}_n(\boldsymbol{\theta}_s)$ = state-space model output for measured input & state prediction errors $\mathbf{v}_n=0$, $\mathbf{w}_n=0$:

$$p(\hat{\mathbf{Y}}_N \mid \hat{\mathbf{U}}_N, \boldsymbol{\theta}, \mathcal{M}_1) = \frac{1}{(2\pi\sigma_{acc}^2)^{\frac{N_O(N+1)}{2}}} \exp\left(-\frac{1}{2\sigma_{acc}^2}\sum_{n=0}^{N}[\hat{\mathbf{y}}_n - \boldsymbol{q}_n(\boldsymbol{\theta}_s)]^T[\hat{\mathbf{y}}_n - \boldsymbol{q}_n(\boldsymbol{\theta}_s)]\right)$$

Second stochastic model class for benchmark structure

- **Second model class** $\mathcal{M}_2$ [Stochastic embedding: Method 2]:

  $\boldsymbol{\theta}$ = 15 parameters – same as $\mathcal{M}_1$ parameters except that Σ_w is 8x8 diagonal matrix with 2 prediction-error variance parameters, σ_{dis}^2 and σ_{vel}^2, respectively, for displacement and velocity states

- **Prior PDF** $p(\boldsymbol{\theta}|\ \mathcal{M}_2)$: Physical parameters $\boldsymbol{\theta}_s$ are independent lognormals as for $\mathcal{M}_1$ and all 3 prediction-error variances are uniformly distributed and independent

Second stochastic model class for benchmark structure

- **Likelihood function** for $\mathcal{M}_2$:

$$p(\hat{\mathbf{Y}}_N \mid \hat{\mathbf{U}}_N, \boldsymbol{\theta}, \mathcal{M}_2) = \frac{1}{(2\pi)^{N_O(N+1)/2} \prod_{n=0}^{N} |\mathbf{S}_{n|n-1}|^{1/2}} \exp\left(-\frac{1}{2}\sum_{n=0}^{N} (\hat{\mathbf{y}}_n - \mathbf{y}_{n|n-1})^T \mathbf{S}_{n|n-1}^{-1} (\hat{\mathbf{y}}_n - \mathbf{y}_{n|n-1})\right)$$

$\mathbf{y}_{n|n-1}(\boldsymbol{\theta})$ = mean predicted acceleration (output) at each floor at time n conditioned on observed output at all previous times

$\mathbf{S}_{n|n-1}(\boldsymbol{\theta})$ = covariance matrix of predicted output $\mathbf{y}_n$ at time n conditioned on observed output at all previous times

- These can be evaluated efficiently using the Kalman filter equations from Bayesian sequential state updating

  (e.g. S.H. Cheung: Caltech PhD Thesis in CE 2009

  Beck: Structural Control & Health Monitoring 2010)

Posterior samples from TMCMC for benchmark structure

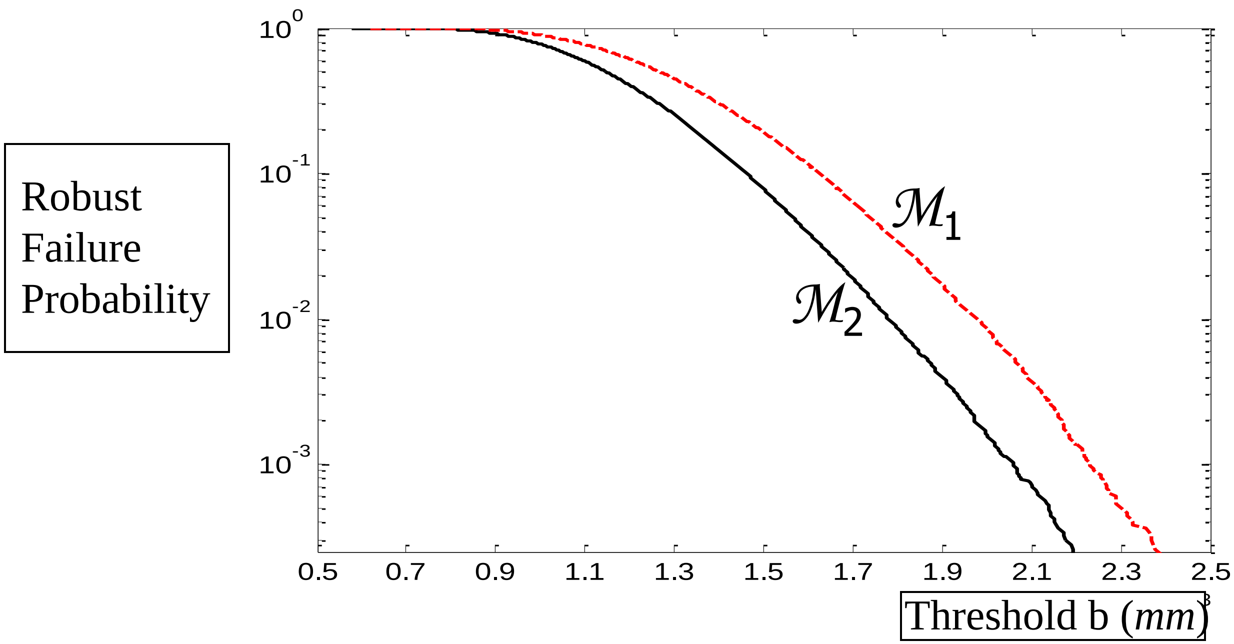
- **Posterior mean and coefficient of variation of model parameters**

		$\mathcal{M}_2$	$\mathcal{M}_1$
Normalized Masses	θ_1	0.97 (0.5%)	1.12 (0.9%)
	θ_2	0.98 (0.5%)	1.13 (1.0%)
	θ_3	0.99 (0.5%)	1.04 (0.9%)
	θ_4	1.07 (0.5%)	1.21 (1.0%)
Normalized Stiffnesses	θ_5	0.76 (0.7%)	0.81 (0.9%)
	θ_6	0.94 (0.6%)	1.10 (1.0%)
	θ_7	0.90 (0.7%)	1.03 (0.9%)
	θ_8	0.92 (0.5%)	0.95 (0.9%)
Normalized Damping Ratios	θ_9	1.11 (14.8%)	0.88 (2.7%)
	θ_{10}	1.42 (6.9%)	0.86 (1.6%)
	θ_{11}	1.89 (4.9%)	0.86 (1.4%)
	θ_{12}	1.23 (7.2%)	1.40 (2.1%)
	σ^2_{dis}	5.80×10^{-11} (3.7%)	Not applicable
	σ^2_{vel}	2.26×10^{-6} (10.1%)	
	σ^2_{acc}	0.103 (2.4%)	3.26 (1.4%)

Based on 1000 posterior samples from TMCMC

Robust system analysis for benchmark structure

- **Posterior robust failure probability**
  - Threshold exceedance probability for maximum inter-story drifts over all stories under future earthquake ground motions (Clough-Penzien stochastic ground motion model)

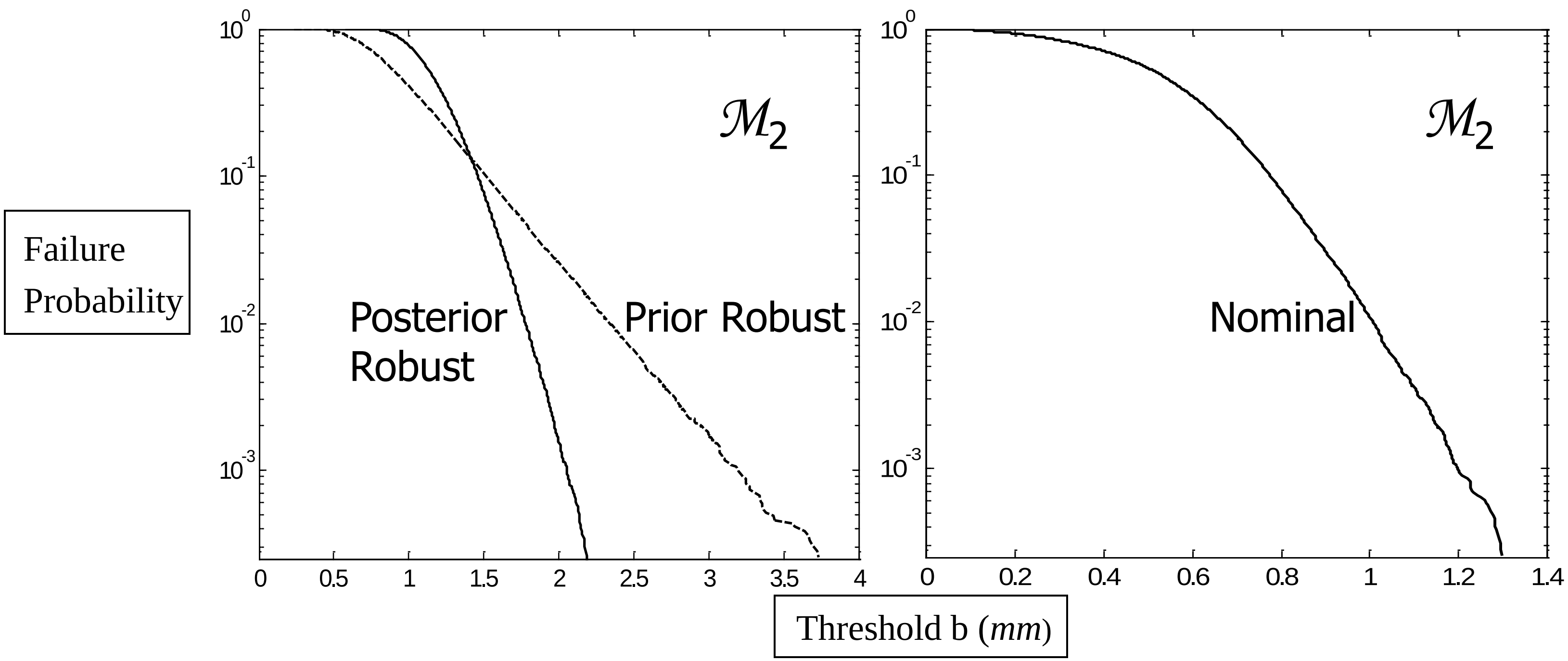


Stochastic simulation (ISEE+TMCMC) with no. of uncertain parameters 8x2501+4x2501+15 +2501=32528 for $\mathcal{M}_2$ and 12518 for $\mathcal{M}_1$

$$\text{Failure, F} = \max_{n\in\{0,1,\ldots,2500\}} \left\{ |\, x_1(t_n)\,|, |\, x_l(t_n) - x_{l-1}(t_n)\,|, l = 2,3,4 \right\} \geq \text{b}$$

Robust system analysis for benchmark structure

- **Left: Posterior** (solid curve) and **prior** (dashed) **robust failure probabilities** for $\mathcal{M}_2$ for threshold exceedance of maximum inter-story drifts over all stories
- **Right: Nominal failure probability** for $\mathcal{M}_2$ (ignores parametric and non-parametric modeling uncertainty)



Outline of Presentation

- **Introduction**
- Probability logic as the foundation for Bayesian probability
- Stochastic model classes & stochastic embedding
- Bayesian system identification with Ex. 1
- Prior & posterior robust stochastic analyses with Ex. 2
- **Posterior model class assessment & selection with Ex.3**
- Concluding remarks

Posterior model class assessment/selection

- **Ingredients:**

  (1) $\mathbf{M}$ defining a set of candidate stochastic model classes for a system $\{\mathcal{M}_1,\ldots,\mathcal{M}_\mathrm{J}\}$ and a prior $\mathrm{P}[\mathcal{M}_j|\mathbf{M}]$ over this set

  (2) Input and output data $\mathcal{D}$ from system

- **Posterior probability** $\mathrm{P}[\mathcal{M}_j|\mathcal{D},\mathbf{M}]$, $j = 1,\ldots,\mathrm{J}$, from Bayes' Theorem at <u>model-class level</u>:

Evidence for model class $\mathcal{M}_j$

$$\mathrm{P}[\mathcal{M}_j \mid \mathcal{D},\mathbf{M}] = \frac{p(\mathcal{D} \mid \mathcal{M}_j)\mathrm{P}[\mathcal{M}_j \mid \mathbf{M}]}{p(\mathcal{D} \mid \mathbf{M})} \propto p(\mathcal{D} \mid \mathcal{M}_j) \ \text{ if } \ \mathrm{P}[\mathcal{M}_j \mid \mathbf{M}] = \tfrac{1}{\mathrm{J}}$$

Calculation of evidence for model class $\mathcal{M}_j$ based on $\mathcal{D}$

- **Evidence** (or marginal likelihood) for $\mathcal{M}_j$ based on data $\mathcal{D}$:

Likelihood Prior

$$p(\mathcal{D}\,|\,\mathcal{M}_j) = \int p(\mathcal{D}\,|\,\theta_j, \mathcal{M}_j)\, p(\theta_j\,|\,\mathcal{M}_j)\, d\theta_j$$

  - Calculate using Laplace's asymptotic method if $\mathcal{M}_j$ is globally identifiable based on data $\mathcal{D}$ [Beck & Yuen: J. Eng. Mech. 2004]
or by TMCMC [Ching & Chen: J. Eng. Mech. 2007]
or using posterior (not prior) samples for $\mathcal{M}_j$ by the stationarity method [Cheung & Beck: CACAIE 2010]

Information-theoretic interpretation of evidence for model class $\mathcal{M}_j$

- Bayes' Theorem at model-class level automatically gives a quantitative Ockham's Razor that avoids over-fitting of data [Gull 1988; Mackay 1992] ["Entities should not be multiplied unnecessarily" - William of Ockham, 1285-1349]

- Recently shown [Beck & Yuen 2004; Muto & Beck 2008]:

Log evidence = Mean data fit of $\mathcal{M}_j$ [posterior mean of log likelihood]

– Expected information gain about model parameters θ from data $\mathcal{D}$ [relative entropy/Kullback-Leibler info]

= Measure of consistency of model class with the data

– Penalty for more complex model classes that extract more information from the data

- Well-known BIC [Schwartz 1978] neglects significant terms of O(1) w.r.t. N in Laplace asymptotic approximation of the log evidence

Posterior robust analysis with multiple candidate model classes

- **Ingredients:**
  - (1) Set of stochastic model classes $\{\mathcal{M}_1, \ldots, \mathcal{M}_J\}$ with prior $P[\mathcal{M}_j|\mathbf{M}]$
  - (2) Stochastic input model $\mathcal{U}$ defining PDF $p(U_n | \mathcal{U})$
  - (3) System input and output data $\mathcal{D}$
- **Posterior hyper-robust failure probability:**

$$P[Y_n \in F_n | \mathcal{D}, \mathcal{U}, \mathbf{M}] = \sum_{j=1}^{J} P[Y_n \in F_n | \mathcal{D}, \mathcal{U}, \mathcal{M}_j]\, P[\mathcal{M}_j | \mathcal{D}, \mathbf{M}]$$

Posterior hyper-robust failure probability for model set of $\mathbf{M}$

Posterior robust failure probability for model class $\mathcal{M}_j$

Posterior probability for model class $\mathcal{M}_j$

Stochastic model class assessment and selection for benchmark structure

- **Posterior probability of model classes:**

	$\mathcal{M}_2$	$\mathcal{M}_1$
$E[\ln(p(\mathcal{D}\|\theta,\mathcal{M}_i))]$ (mean data-fit)	$-1.5762\text{x}10^4$	$-2.0251\text{x}10^4$
Expected info gain	76.12	63.52
$\ln p(\mathcal{D}\|\mathcal{M}_i)$ (log evidence)	$-1.5838\text{x}10^4$	$-2.0315\text{x}10^4$
$P[\mathcal{M}_i\|\mathcal{D},\mathbf{M}]$	1.0000	0.0000

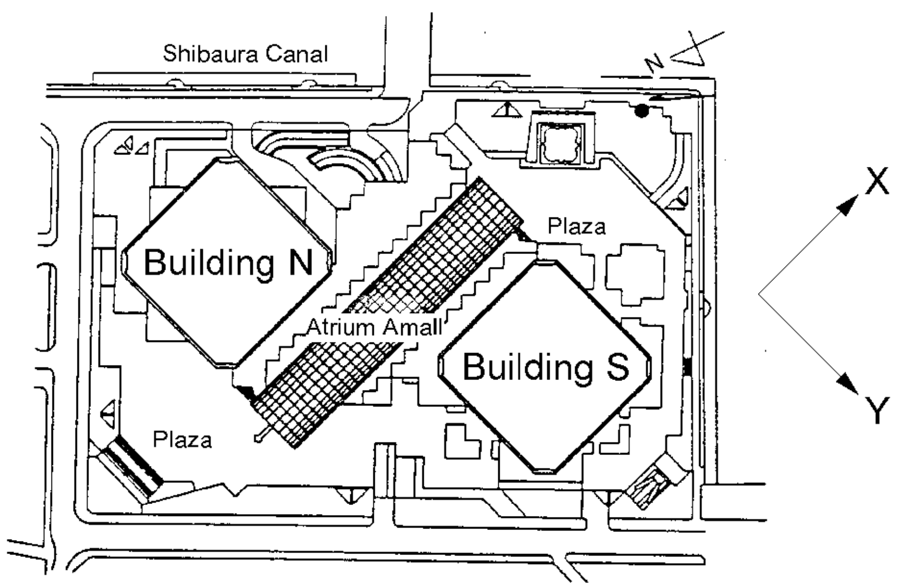
- **Posterior hyper-robust failure probability:**

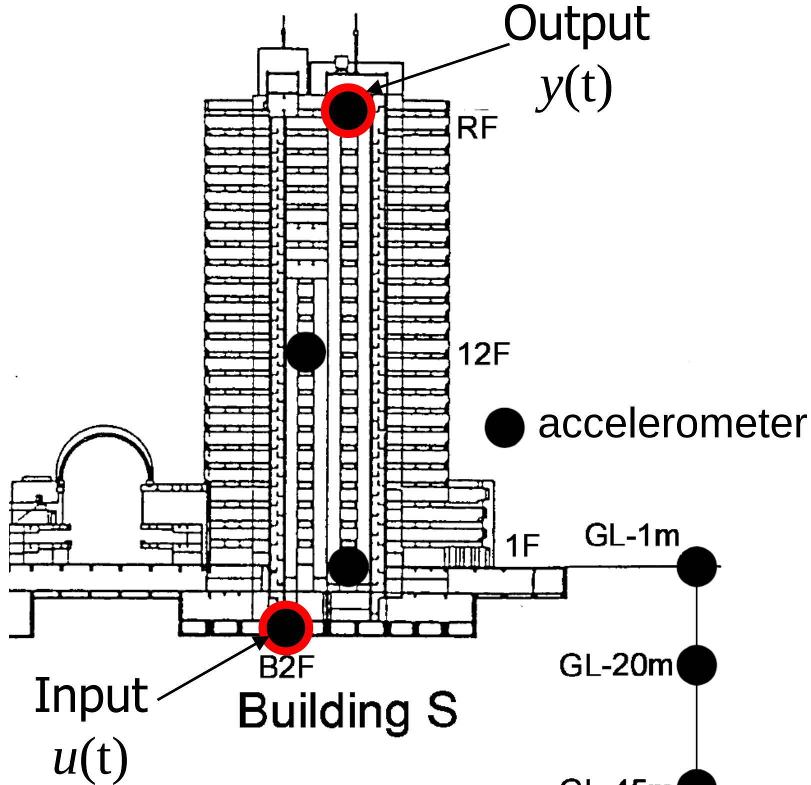
$$\mathrm{P}[F \mid \mathcal{D}, \mathcal{U}, \mathbf{M}] = \sum_{j=1}^{2} \mathrm{P}[F \mid \mathcal{D}, \mathcal{U}, \mathcal{M}_j]\, \mathrm{P}[\mathcal{M}_j \mid \mathcal{D}, \mathbf{M}]$$

Can drop $\mathcal{M}_1$ since its contribution is negligible relative to that of $\mathcal{M}_2$

Example 3: Bayesian modal identification of 24-story building in Tokyo, Japan

- Steel framed and 97.7m high
- 24 and 2 stories above and below the ground

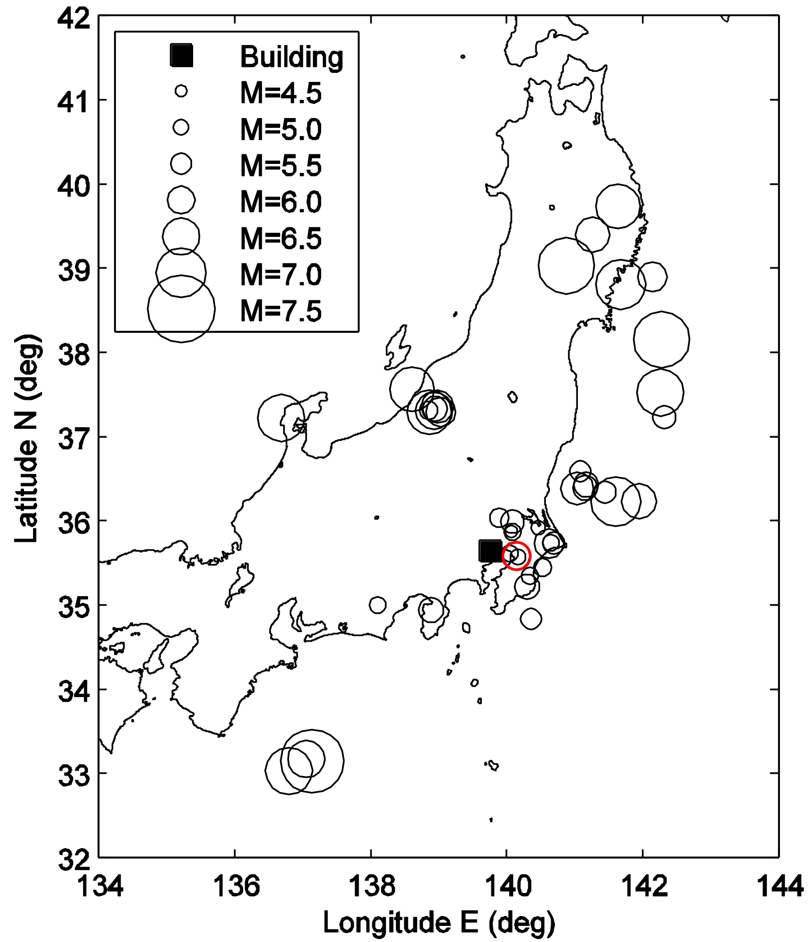




Saito & Beck: Eq. Eng. Struc. Dyn. 2010

43 earthquake records over 9 years

No.	Date (y/m/d)	Time (GMT+9h)	Location N (deg)	Location E (deg)	Epicentral distance (km)	Depth (km)	M	JMA seismic intensity
1	1999/07/15	07:56	35.917	140.467	71	56	4.9	1.0
2	1999/09/13	07:56	35.567	140.167	38	77	5.0	1.9
3	2001/04/03	23:57	34.995	138.108	166	33	5.1	1.0
4	2001/12/02	22:01	39.395	141.267	436	122	6.4	0.8
5	2002/02/12	22:44	36.585	141.085	158	48	5.5	1.2
6	2002/07/24	05:05	37.228	142.318	289	30	5.7	0.9
7	2002/11/03	12:37	38.893	142.142	417	46	6.1	0.9
8	2003/05/12	00:57	35.865	140.088	38	47	5.2	2.6
9	2003/05/12	00:59	35.872	140.072	38	50	4.6	1.7
10	2003/05/17	23:33	35.735	140.653	82	47	5.1	0.9
11	2003/05/26	18:24	38.805	141.682	389	71	7.0	2.3
12	2003/09/20	12:54	35.215	140.303	69	70	5.8	2.3
13	2003/10/15	16:30	35.610	140.052	27	74	5.1	2.7
14	2003/11/12	17:26	33.170	137.057	370	398	6.5	1.6
15	2003/11/15	03:43	36.428	141.168	154	48	5.8	1.2
16	2004/04/04	08:02	36.387	141.157	150	49	5.8	1.3
17	2004/07/17	15:10	34.833	140.358	106	69	5.5	1.5
18	2004/09/05	19:07	33.028	136.800	398	38	6.9	1.0
19	2004/09/05	23:57	33.143	137.142	367	44	7.4	1.5
20	2004/10/06	23:40	35.988	140.088	48	66	5.7	2.8
21	2004/10/23	17:56	37.292	138.867	199	13	6.8	2.0
22	2004/10/23	17:59	37.312	138.855	201	16	5.3	0.8
23	2004/10/23	18:03	37.353	138.983	201	9	6.3	1.4
24	2004/10/23	18:34	37.305	138.930	198	14	6.5	2.1
25	2004/10/25	06:04	37.330	138.947	200	15	5.8	0.9
26	2004/10/27	10:40	37.292	139.033	193	12	6.1	1.6
27	2005/02/16	04:46	36.035	139.895	45	45	5.4	2.5
28	2005/04/11	07:22	35.727	140.620	79	52	6.1	2.1
29	2005/06/20	01:15	35.733	140.693	85	51	5.6	1.3
30	2005/07/23	16:34	35.582	140.138	35	73	6.0	3.5
31	2005/08/08	00:06	36.338	141.445	170	46	5.6	0.7
32	2005/08/16	11:46	38.150	142.278	357	42	7.2	2.9
33	2005/10/19	20:44	36.382	141.042	141	48	6.3	1.8
34	2007/01/16	03:17	34.937	138.892	111	175	5.8	1.4
35	2007/03/25	09:41	37.220	136.685	326	11	6.9	0.9
36	2007/07/16	10:13	37.557	138.608	235	17	6.8	2.0
37	2007/08/16	04:15	35.443	140.530	74	31	5.3	1.5
38	2007/08/18	16:55	35.342	140.345	63	20	5.2	1.0
39	2008/05/08	01:02	36.230	141.948	208	60	6.4	1.2
40	2008/05/08	01:45	36.227	141.607	179	51	7.0	2.4
41	2008/06/14	08:43	39.028	140.880	388	8	7.2	1.7
42	2008/07/19	11:39	37.520	142.263	305	32	6.9	1.6
43	2008/07/24	00:26	39.732	141.635	482	108	6.8	1.6



ARX model class $\mathcal{M}_d$ of order d

- **Auto-Regressive eXogenous model:**

Prediction error

$$\mathrm{y}_n = -\sum_{j=1}^{d} a_j \mathrm{y}_{n-j} + \sum_{j=0}^{d} b_j \mathrm{u}_{n-j} + \mathrm{e}_n$$

Output

Input

$$y_n = y(n\Delta t) \qquad u_n = u(n\Delta t)$$

$$e_n = e(n\Delta t) \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2)$$

  - Gaussian likelihood based on I/O data $\mathcal{D}_N = \{\hat{Y}_N, \hat{U}_N\}$

- **Model parameters:** $\theta = \left[a_1, \ldots, a_d, b_0, \ldots, b_d, \sigma^2\right]^{\mathrm{T}} \in R^{2d+2}$

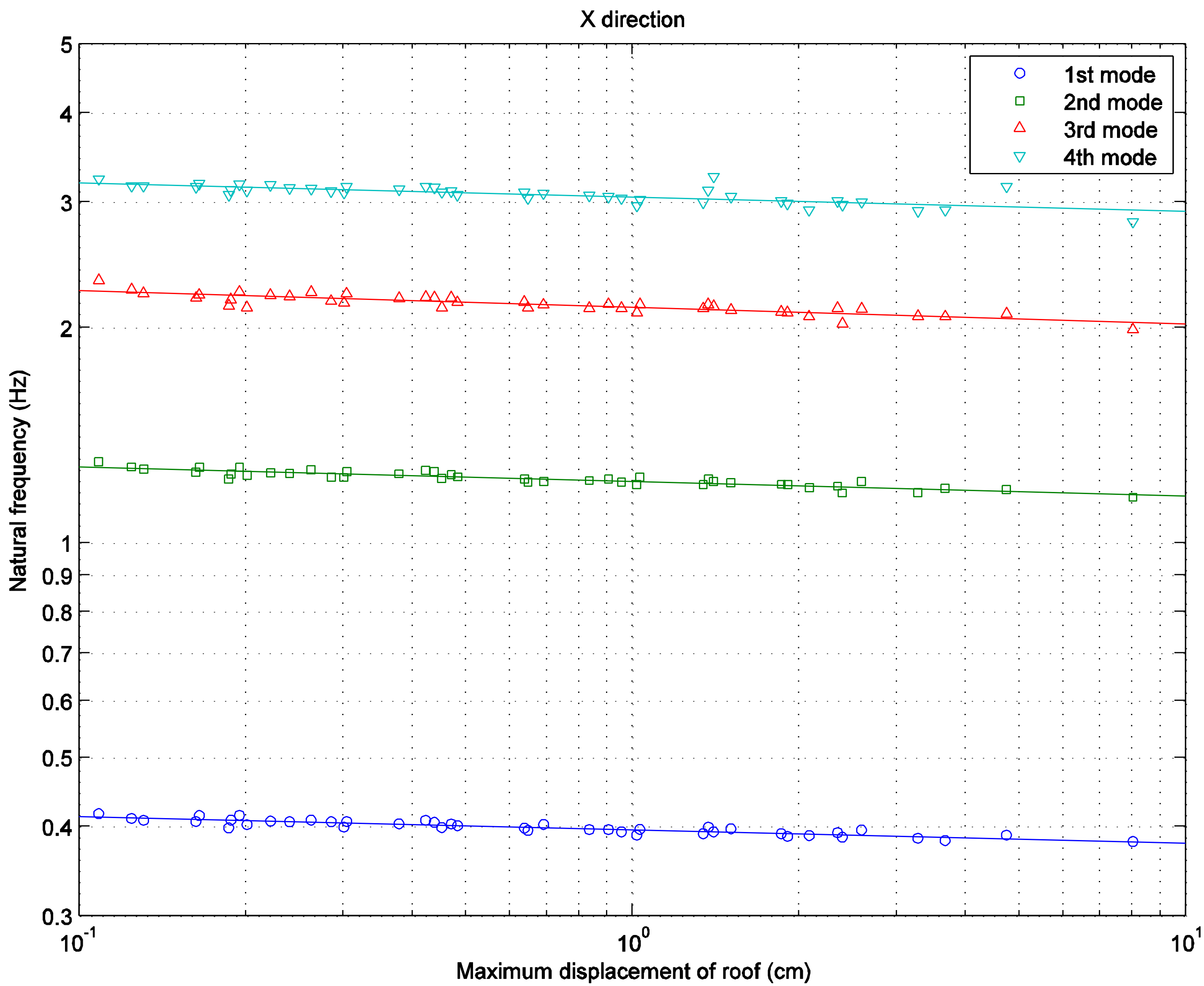
  - Gaussian priors on coefficients, lognormal prior on σ^2

Probability of each model class: Record #30

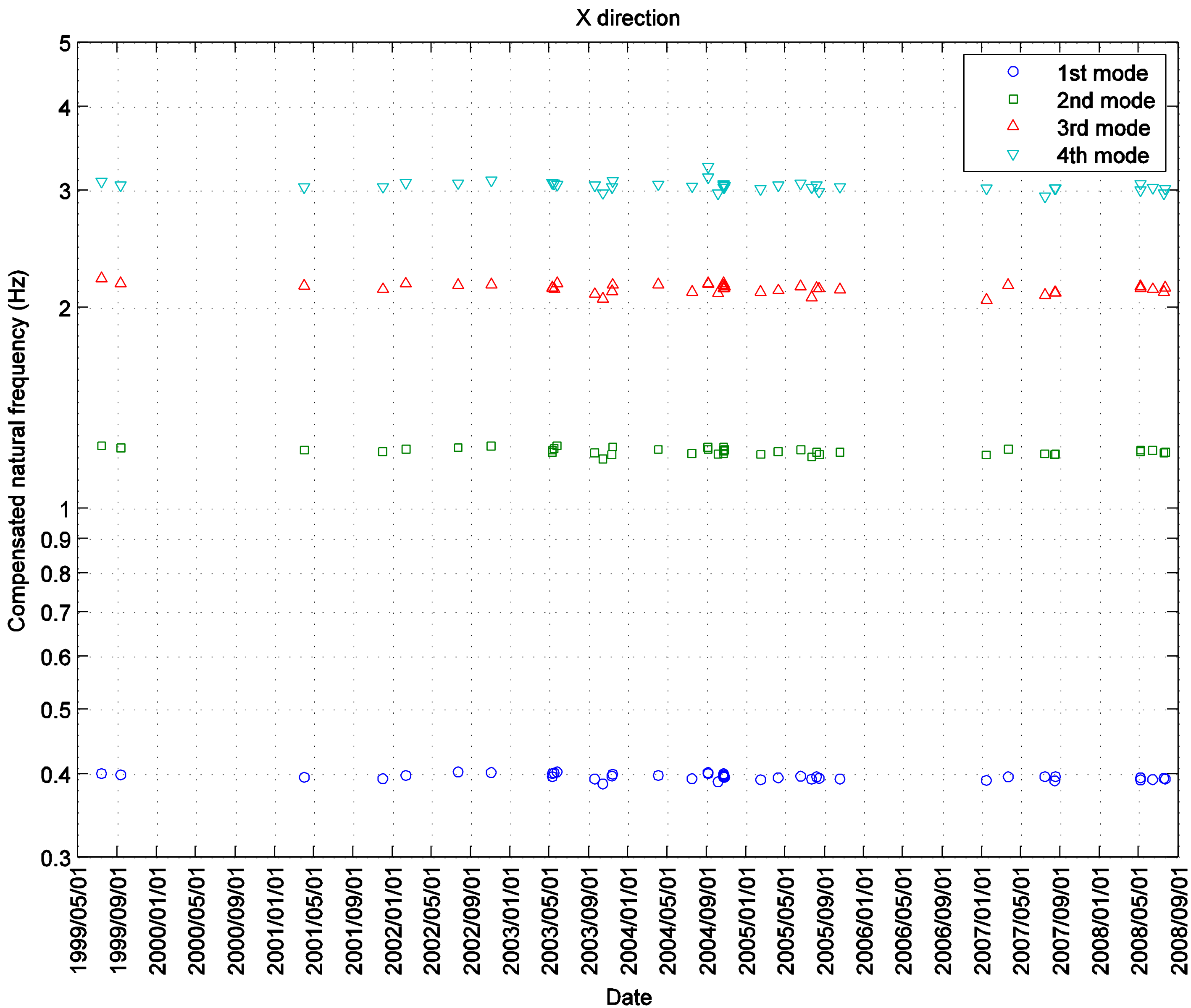
Model order d	…	26	28	30	32	…
Posterior probability of model class $\mathcal{M}_d$	…	0.0	0.93	0.07	0.0	…

- Equal prior probabilities are chosen for each $\mathcal{M}_d$
- Most probable a posteriori ARX order is 28

Natural frequencies vs response amplitude



Natural frequencies after amplitude compensation



Concluding Remarks #1

- **Rigorous framework for Bayesian System ID:** based on probability logic for quantifying plausible reasoning
  - Treats uncertainty due to missing information (epistemic); the assumption of inherent randomness is not needed (aleatory)
  - Uses <u>only</u> the probability axioms and the probability models defined by a chosen <u>stochastic</u> <u>model</u> <u>class</u> for the system
- **Prior/posterior robust analysis:** combine the stochastic predictions of all of the models corresponding to prior/ posterior samples generated by MC/MCMC algorithms (rather than selecting a single nominal or "optimal" MAP or MLE model to represent the system)

Concluding Remarks #2

- **Posterior model class assessment:** use the posterior probability of each model class to assess multiple candidate model classes based on system data
  - This assessment provides a quantitative Ockham's razor to avoid over-fitting since the posterior probability of each model class trades-off its data-fit against its "complexity" (amount of information it extracts from the data)

Thank you!

"Bayesian System Identification based on Probability Logic", J.L. Beck, International J. Structural Control and Health Monitoring 2010

"Prior and Posterior Robust Stochastic Predictions for Dynamical Systems using Probability Logic", J.L. Beck and A.A. Taflanidis, International J. Uncertainty Quantification, 2013

Developing stochastic I/O models by stochastic embedding: Method 1

- e.g. choosing prediction-error covariance matrix $\sigma^2 I_{N_o}$ gives:

$$p(Y_N \mid U_N, \boldsymbol{\theta}, \mathcal{M}) = \prod_{n=1}^{N} p(y_n \mid U_n, \boldsymbol{\theta}, \mathcal{M})$$

$$= (2\pi\sigma^2)^{-\frac{N_O N}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{n=1}^{N}\left\| y_n - q_n(U_n, \boldsymbol{\theta})\right\|^2\right)$$

  - defines a <u>set of stochastic I/O models</u> where $\theta \in \Theta \subset R^{N_p}$, $\sigma > 0$

  Specification of a prior PDF for (θ, σ) completes the <u>stochastic model class</u> $\mathcal{M}$; e.g. a Gaussian prior $N(\bar{\theta}, \Sigma_\theta)$

- If system data $\mathcal{D}_N = \{\hat{Y}_N, \hat{U}_N\}$ is available, then <u>Bayes' Theorem</u> for the model class $\mathcal{M}$ gives for the log posterior PDF:

$$\ln p(\theta \mid \mathcal{D}_N, \mathcal{M}) = \ln p(\hat{Y}_N \mid \hat{U}_N, \theta, \mathcal{M}) + \ln p(\theta \mid \mathcal{M}) + \text{const.}$$

$$= -\,[\text{ data-fit term } + \text{ regularization term }]$$

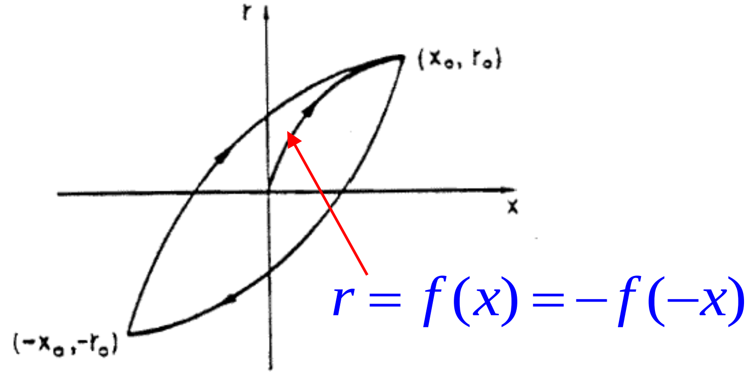
  e.g. choosing the MAP (<u>maximum a posteriori</u>) values for (θ, σ) is equivalent to <u>regularized least-squares estimation</u> of θ

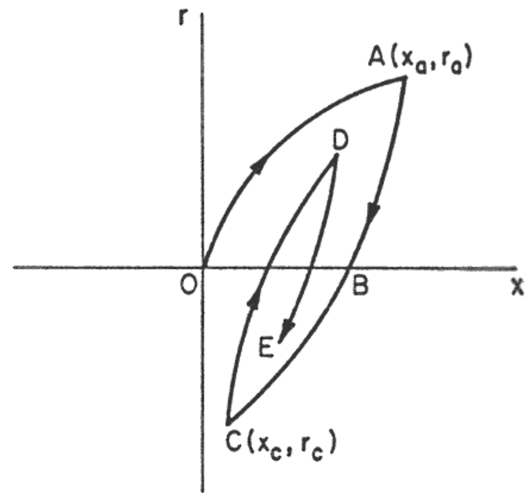
Masing hysteretic models

- Masing's postulate (1926) for steady-state hysteretic behavior of metals: Given an initial loading curve $r = f(x)$ then for unloading:

$$\frac{r - r_0}{2} = f\left(\frac{x - x_0}{2}\right)$$

- For transient loading, Jayakumar (1987) showed that two hysteretic rules (extending Masing's rule) give the same hysteretic behavior as Iwan's DEM [J. App. Mech. 1966,67]
- Models are defined by initial loading curve $r = f(x)$ only (or yield strength distribution function for DEM)
- No non-physical behavior since DEM corresponds to a parallel system of elasto-plastic elements





Masing hysteretic shear building model

- Model multi-story structure as a shear building with story shear forces r_i related to inter-story drifts (x_i-x_{i-1}) by Masing model with initial loading curve (similar to Bouc-Wen):

$$\dot{r}_i = K_i[1-(r_i / r_{u,i})^{\alpha_i}](\dot{x}_i - \dot{x}_{i-1})$$

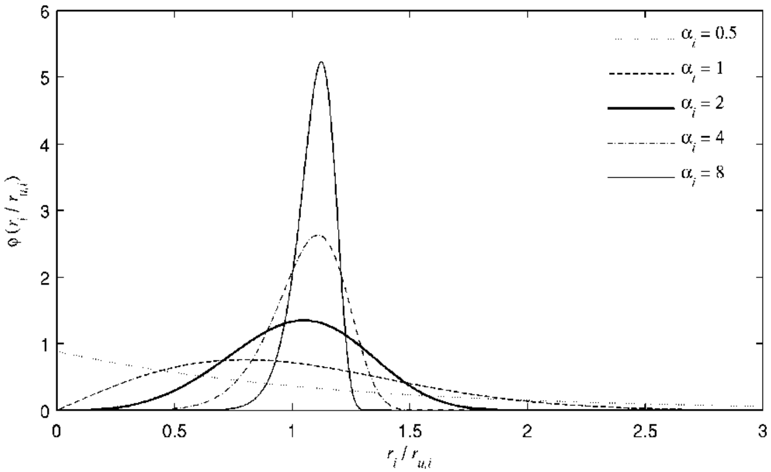
K_i = small-amplitude interstory stiffness

$r_{u,i}$ = ultimate strength of story

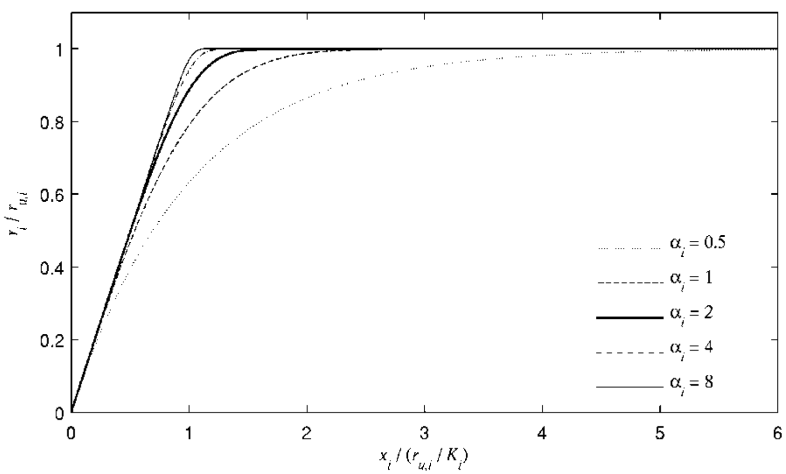
α_i = yield transition parameter

Corresponding to $r_i = f_i(x_i - x_{i-1})$:

Yield strength distribution



(Shear force)/Strength



Normalized deflection

Initial force-deflection curve

Prior system analysis for system design

- If failure probabilities $\mathrm{P_F} = \mathrm{P}[Y_n \in F_n \mid \mathcal{U}, \mathcal{M}] << 1$, we can improve efficiency by using advanced simulation methods:
  - Review article for reliability calculations in higher dimensions: Schueller et al., Prob. Eng. Mechanics (2004)
  - Special issue of Structural Safety (2007) on benchmark reliability problems in higher dimensions: Schueller & Pradlwarter (Eds)
  - e.g. **Subset Simulation** for general model dynamics

    (Au & Beck: Prob. Eng. Mech. 2001, J. Eng. Mech. 2003) - uses successive batches of MCMC samples to adaptively create a nested sequence of subsets converging onto the failure region in input space
  - e.g. **ISEE** for linear model dynamics with Gaussian excitation

    (Au & Beck: Prob. Eng. Mech. 2001) – uses importance sampling with a weighted sum of Gaussians, one for each elementary "failure" event at all discrete times

Accuracy of MCS, Subset Simulation and ISEE for failure probability $\mathrm{P_F}$

- Coefficient of variation of estimators for K samples

$$\text{c.o.v.}, \delta = \frac{\text{standard deviation}}{\text{mean}} = \frac{\Delta}{\sqrt{K}}$$

MCS: $\Delta \approx \dfrac{1}{\sqrt{\mathrm{P_F}}}$

Subset Simulation: $\Delta = \alpha \,|\log_{10} \mathrm{P_F}\,|,\ \alpha = O(1)$

ISEE: $\Delta \approx 1$ [**decreases** slightly as $\mathrm{P_F}$ decreases!]

e.g. to get a c.o.v. of 35% for $\mathrm{P_F} = 10^{-3}$ (resp. 10^{-4}):
MCS requires about 10,000 (resp. 100,000) samples;
Subset Simulation requires 1500 (resp. 2000) samples;
ISEE requires 10 samples for any $\mathrm{P_F}$!

Posterior probabilities of $\mathcal{M}_d$ for various d

- Probability of $\mathcal{M}_d$ based on data $\mathcal{D}$ (Bayes' Theorem):

$$P(\mathcal{M}_d \mid \mathcal{D}, \bigcup \mathcal{M}_d) = \frac{\overbrace{\boxed{p(\mathcal{D} \mid \mathcal{M}_d)}}^{\text{Evidence}} P(\mathcal{M}_d \mid \bigcup \mathcal{M}_d)}{p(\mathcal{D} \mid \bigcup \mathcal{M}_d)}$$

- Evidence:

$$\mathrm{EV}(\mathcal{M}_d \mid \mathcal{D}) = p(\mathcal{D} \mid \mathcal{M}_d) = \int p(\mathcal{D} \mid \theta, \mathcal{M}_d)\, p(\theta \mid \mathcal{M}_d)\, d\theta$$

- Laplace's asymptotic approximation about MLE $\hat{\theta}_{\mathrm{L}}$

$$\mathrm{EV}(\mathcal{M}_d \mid \mathcal{D}) \approx \frac{\overbrace{\boxed{p(\mathcal{D} \mid \hat{\boldsymbol{\theta}}_{\mathrm{L}}, \mathcal{M}_d)}}^{\text{Likelihood}}\, \overbrace{\boxed{p(\hat{\boldsymbol{\theta}}_{\mathrm{L}} \mid \mathcal{M}_d)}}^{\text{Prior}}\, (2\pi)^{Np/2}}{\det\left[\boxed{H(\hat{\boldsymbol{\theta}}_{\mathrm{L}})}\right]^{1/2}}$$

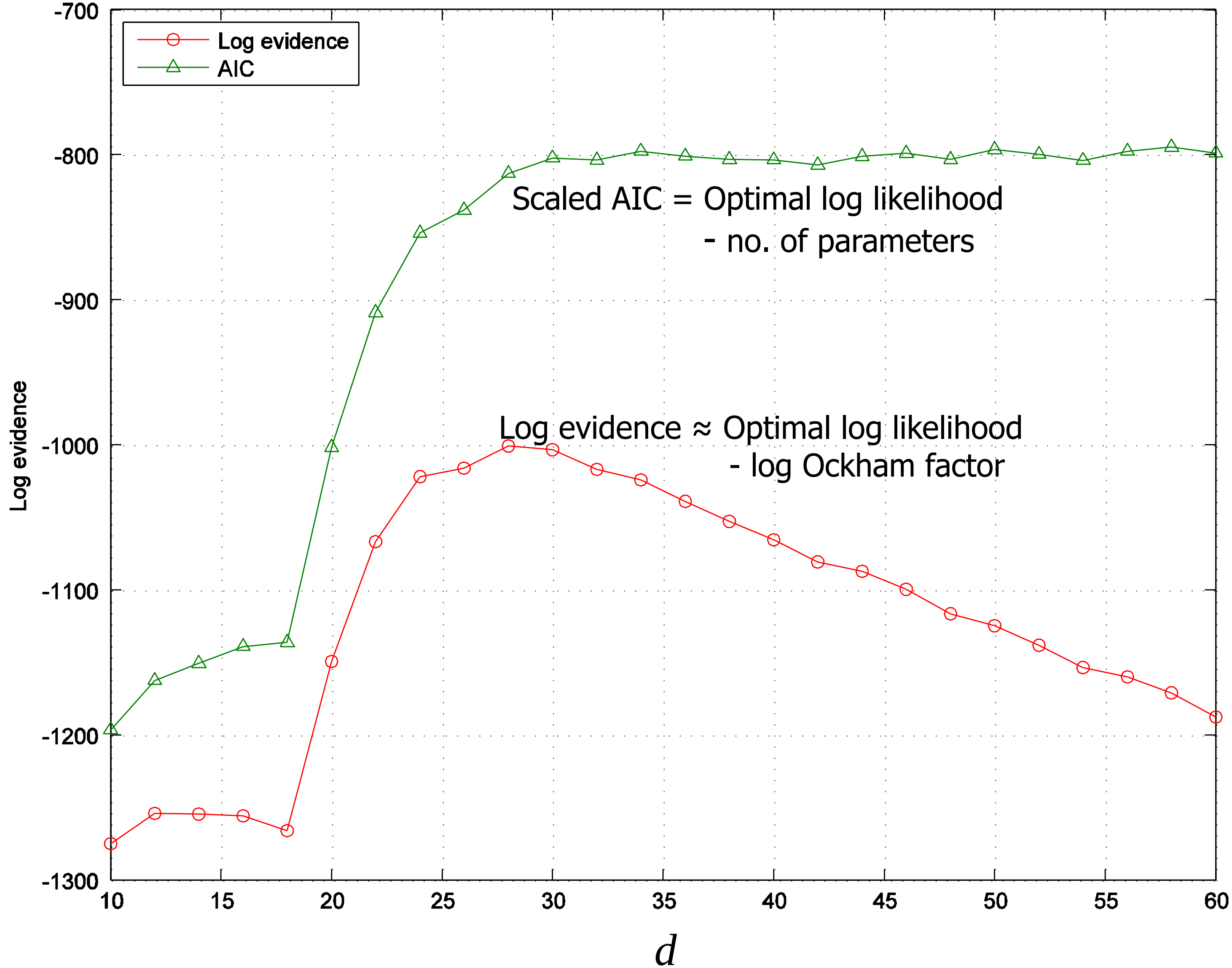
Hessian matrix

= Optimal likelihood x Ockham factor

Hessian matrix: $H(\boldsymbol{\theta}) \triangleq -\nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}} \ln p(\mathcal{D} \mid \boldsymbol{\theta}, \mathcal{M}_d)$

Log evidence and AIC: Record #30 (X direction)

$\mathcal{M}_d$: d=10,12,…,58,60



Modal parameters: MAP estimates, precision and correlation from Record #30

- MAP estimates and coefficient of variation (CV)

	f_1(Hz)	h_1(%)	γ_1	f_2	h_2	γ_2	f_3	h_3	γ_3	f_4	h_4	γ_4
MAP	0.386	2.369	1.466	1.172	3.300	−0.621	2.025	3.650	0.363	2.971	3.785	−0.191
CV (%)	0.528	22.950	5.444	0.134	3.951	1.548	0.286	7.983	3.799	0.676	18.227	17.340

- Absolute values of correlation coefficients

$|\rho_{jk}|$