SAND2014-16475PE

Uncertainty Quantification, Bayesian Inference, and Analysis of Models

Habib N. Najm

Sandia National Laboratories
Livermore, CA
hnnajm@sandia.gov

Twelfth International Workshop
on Measurement and Computation of Turbulent Flames (TNF12)
Pleasanton, CA
July 31-Aug 2, 2014

Acknowledgement

B.J. Debusschere, R.D. Berry, K. Sargsyan, C. Safta,
K. Chowdhary, M. Khalil — Sandia National Laboratories, CA

R.G. Ghanem — U. South. California, Los Angeles, CA
O.M. Knio — Duke Univ., Durham, NC
O.P. Le Maître — CNRS, Paris, France
Y.M. Marzouk — Mass. Inst. of Tech., Cambridge, MA

This work was supported by:

- DOE Office of Advanced Scientific Computing Research (ASCR), Scientific Discovery through Advanced Computing (SciDAC)
- DOE ASCR Applied Mathematics program.
- DOE Office of Basic Energy Sciences, Div. of Chem. Sci., Geosci., & Biosci.

Sandia National Laboratories is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94-AL85000.

Outline

1 Introduction

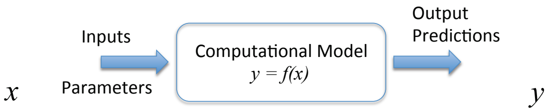
2 Bayesian Inference

3 Illustration in Chemical Ignition

4 Model Comparison, Validation, Averaging

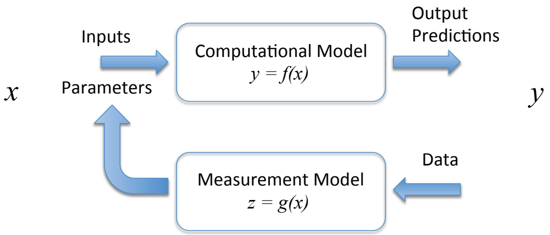
5 Closure

Uncertainty Quantification and Computational Science



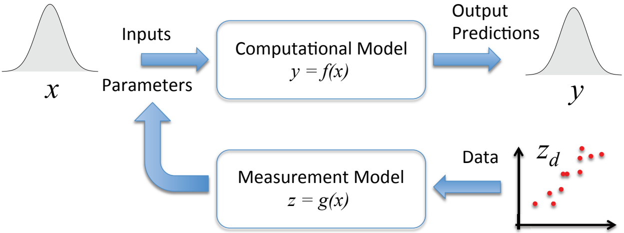
Forward problem

Uncertainty Quantification and Computational Science



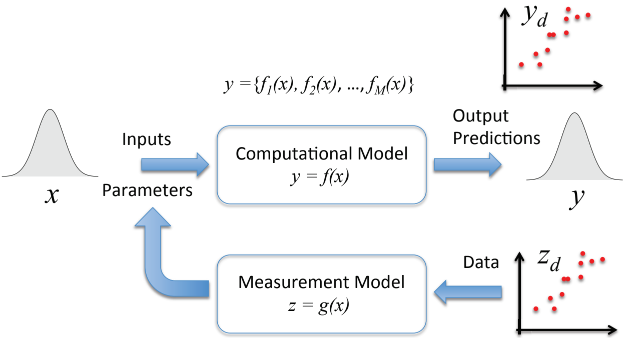
Inverse & Forward problems

Uncertainty Quantification and Computational Science



Inverse & Forward UQ

Uncertainty Quantification and Computational Science



Inverse & Forward UQ

Model validation & comparison, Hypothesis testing

Why UQ? Why in computational combustion?

Why UQ?

- Assessment of confidence in computational predictions
- Validation and comparison of scientific/engineering models
- Design optimization
- Use of computational predictions for decision-support
- Assimilation of observational data and model construction

Further ...

- Explore model response over range of parameter variation
- Enhanced understanding extracted from computations
- Particularly important given **cost** of computations

Specific uses of UQ in LES studies

Forward uncertainty propagation

- Given known uncertainties in model inputs, *e.g.*
  - subgrid model parameters, initial and boundary conditions

– estimate uncertainties in model output predictions
– evaluate global sensitivities to model inputs
– build surrogates for model output dependence on inputs

Inverse UQ – model calibration, parameter estimation

- Given data on model output observables, *e.g.*
  - experimental measurements, DNS computations

– estimate values of model inputs/parameters
– estimate plausibility of, compare, select among models
– validate models

Least-Squares Parameter Estimation

- Fit model $g()$; unknown parameters λ; measurement y
- Forward Problem:

$$g(\lambda) = y_m$$

- Estimate λ for best fit between $g(\lambda)$ and y :

$$\lambda_{\text{fit}} \quad = \quad g^{-1}(y)$$

- This is a classic inverse problem
  - Typically solved using least-squares regression
  - e.g. Newton's method

$$\lambda_{\text{rms}} \quad = \quad \underset{\lambda}{\text{argmin}}(||y - g(\lambda)||)$$

i.e. minimize the χ^2:

$$\chi^2 = \sum_{k=1}^{\mathcal{D}} \frac{((g(\lambda) - y)^2}{\sigma_k^2}$$

Issues with Least Squares (LS) Parameter Estimation

- Choice of optimal number of fit parameters (p)
  - χ^2 decreases with increased p
  - Danger of overfitting
- No general means for handling *nuisance* parameter
- LS best fit is the Maximum Likelihood Estimate (MLE) assuming gaussian noise in the data
- LS Estimation of Uncertainty in inferred parameter values relies on assumed linearity of the model in the parameters
- Support Planes method to estimate standard deviation
  - Variation of one parameter at a time
  - Solve the LS problem for remaining $p-1$ params
  - Re-evaluate χ^2
  - When χ^2 decreases by a predetermined factor, the parameter is $n\sigma$ away from best fit

Bayes formula for Parameter Inference

- Data Model (fit model + noise): $y = f(\lambda) + \epsilon$
- Bayes Formula:

$$p(\lambda, y) = p(\lambda|y)p(y) = p(y|\lambda)p(\lambda)$$

$$\underbrace{p(\lambda|y)}_{\text{Posterior}} = \frac{\overbrace{p(y|\lambda)}^{\text{Likelihood}}\ \overbrace{p(\lambda)}^{\text{Prior}}}{\underbrace{p(y)}_{\text{Evidence}}}$$

- Prior: knowledge of λ prior to data
- Likelihood: forward model and measurement noise
- Posterior: combines information from prior and data
- Evidence: normalizing constant for present context

Advantages of Bayesian Methods

- Formal means of logical inference and machine learning
- Means of incorporation of prior knowledge/measurements and heterogeneous data
- Full probabilistic description of parameters
- General means of handling nuisance parameters through marginalization
- Means of identification of *optimal* model complexity
  - Ockham's razor
  - Only as much complexity as is required by the physics, and no more
  - Avoid fitting to noise

The Prior

- Prior $p(\lambda)$ comes from
  - Physical constraints
  - Prior data
  - Prior knowledge
- The prior can be **uninformative**
- It can be chosen to impose **regularization**
- Unknown aspects of the prior can be added to the rest of the parameters as hyperparameters

Construction of the Likelihood $p(y|\lambda)$

- Where does probability enter the mapping $\lambda \to y$ in $p(y|\lambda)$?
- Through a presumed error model:
- Example:
    - Model:

$$y_m = g(\lambda)$$

    - Data: y
    - Error between data and model prediction: ϵ

$$y = g(\lambda) + \epsilon$$

- Model this error as a random variable
- Example
    - Error is due to instrument measurement noise
    - Instrument has Gaussian errors, with no bias

$$\epsilon \sim N(0, \sigma^2)$$

Construction of the Likelihood $p(y|\lambda)$ – cont'd

For any given λ, this implies

$$y|\lambda, \sigma \sim N(g(\lambda), \sigma^2)$$

or

$$p(y|\lambda, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(y - g(\lambda))^2}{2\sigma^2}\right)$$

Given N measurements $(y_1, \ldots, y_N)$, and presuming independent identically distributed (*iid*) noise

$$\begin{aligned} y_i &= g(\lambda) + \epsilon_i \\ \epsilon_i &\sim N(0, \sigma^2) \\ L(\lambda) = p(y_1, \ldots, y_N|\lambda, \sigma) &= \prod_{i=1}^{N} p(y_i|\lambda, \sigma) \end{aligned}$$

Construction of the Likelihood $p(y|\lambda)$ – cont'd

Recall that the weighted least-squares data mis-fit is given by

$$\chi^2 = \sum_{i=1}^{N} \left[\frac{y_i - g(\lambda)}{\sigma_i} \right]^2$$

and the best-fit estimate of λ is

$$\lambda_{\text{rms}} = \underset{\lambda}{\operatorname{argmin}}(\chi^2(\lambda))$$

Minimizing χ^2 is equivalent to maximizing the likelihood
Maximum Likelihood Estimate (MLE):

$$\lambda_{\text{MLE}} \equiv \lambda_{\text{rms}}$$

Exploration of the likelihood provides for a more general examination of quality of fit than χ^2

Experimental Data

- Empirical data error model structure can be informed based on knowledge of the experimental apparatus
- Both bias and noise models are typically available from instrument calibration
- Noise PDF structure
    - A counting instrument would exhibit Poisson noise
    - A measurement combining many noise sources would exhibit Gaussian noise
- Noise correlation structure
    - – Point measurement
    - – Field measurement

Line fitting example

Consider the fitting of a straight line

$$y_m = ax + b$$

to data $D = \{(x_i, y_i),\ i = 1, \ldots, N\}$.
Consider an (improper) uninformative prior

$$\pi(a, b) = \mathrm{Const}$$

providing no prior information on (a, b).
Assume *iid* additive unbiased Gaussian noise in y with a given constant noise variance σ^2, thus the data model is:

$$y = ax + b + \epsilon, \qquad \epsilon \sim N(0, \sigma^2)$$

with no noise in the independent variable x.

Line fitting example

Presuming σ known, we have the likelihood,

$$L(a,b) = p(D|a,b) = \prod_{i=1}^{N} p(y_i|a,b)$$

where

$$p(y_i|a,b) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(y_i - ax_i - b)^2}{2\sigma^2}\right)$$

and, per Bayes formula, the posterior density $p(a,b|D)$ is

$$p(a,b|D) \quad = \quad \frac{p(D|a,b)\pi(a,b)}{p(D)} \propto p(D|a,b)\pi(a,b)$$

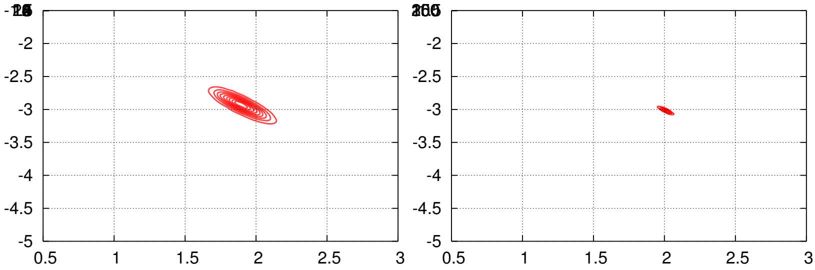
Line fitting example – cont'd

The posterior on (a, b) is the two-dimensional Multivariate Normal (MVN) distribution

$$
\begin{aligned}
p(a, b|D) &\propto (2\pi\sigma^2)^{-N/2} \prod_{i=1}^{N} \exp\left(-\frac{(y_i - ax_i - b)^2}{2\sigma^2}\right) \\
&\propto (2\pi\sigma^2)^{-N/2} \exp\left(-\sum_{i=1}^{N} \frac{(y_i - ax_i - b)^2}{2\sigma^2}\right)
\end{aligned}
$$

Linear model, Gaussian noise, σ-given, and a Gaussian or constant-uninformative prior.

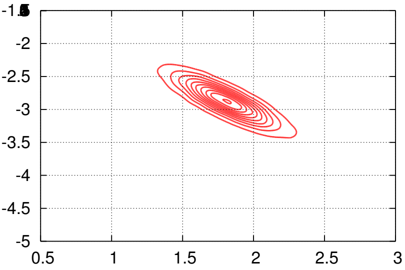
Line fitting example – Effect of data size on $p(a, b|D)$ Low data noise: $\sigma = 0.25$



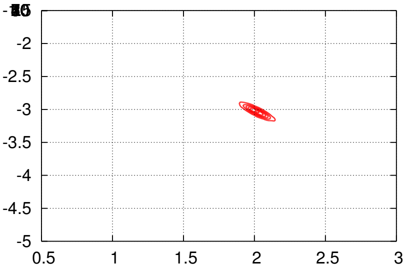
$N = 20$ $N = 200$

- More data $\Rightarrow$ more accurate parameter estimates

Line fitting example – Effect of data size on $p(a, b|D)$
Medium data noise: $\sigma = 0.5$



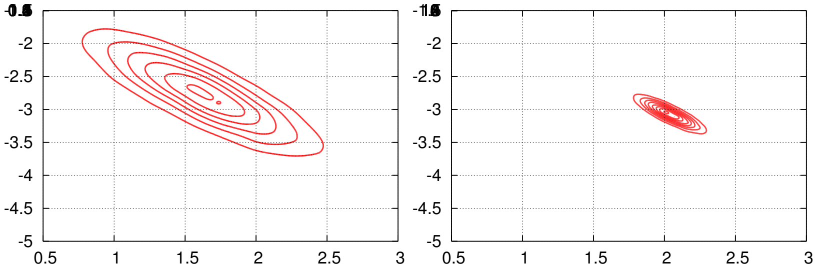
$N = 20$



$N = 200$

- More data $\Rightarrow$ more accurate parameter estimates
- Higher noise amplitude $\Rightarrow$ higher uncertainty

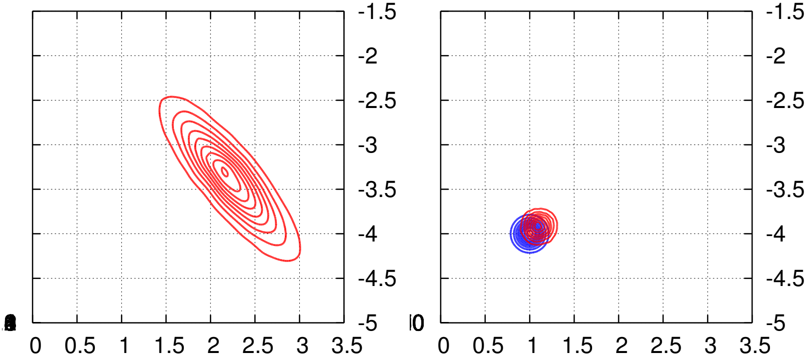
Line fitting example – Effect of data size on $p(a, b|D)$ High data noise: $\sigma = 1.0$



$N = 20$ $N = 200$

- More data $\Rightarrow$ more accurate parameter estimates
- Higher noise amplitude $\Rightarrow$ higher uncertainty

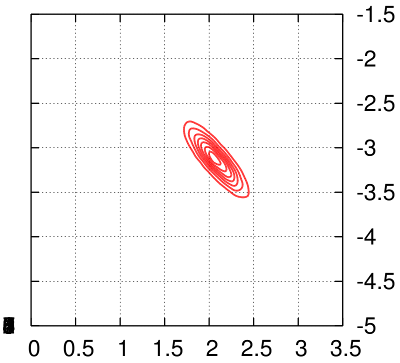
Line fitting example – prior vs. data-size
20 data points



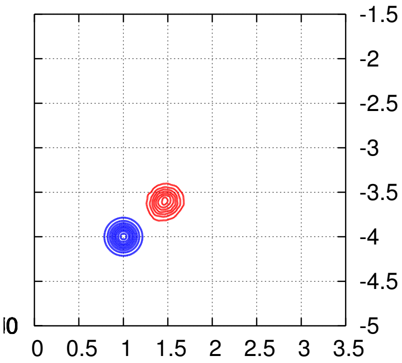
Constant uninformative prior

Gaussian prior

Line fitting example – prior vs. data-size
80 data points

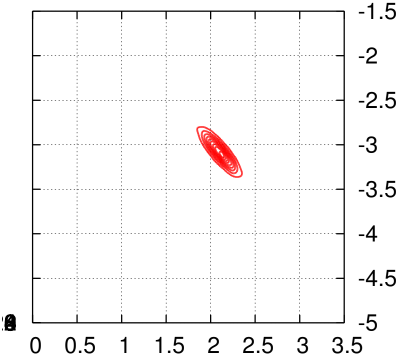


Constant uninformative prior

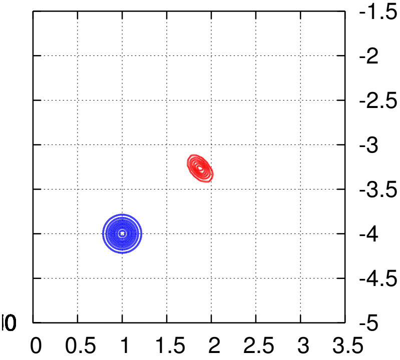


Gaussian prior

Line fitting example – prior vs. data-size
200 data points

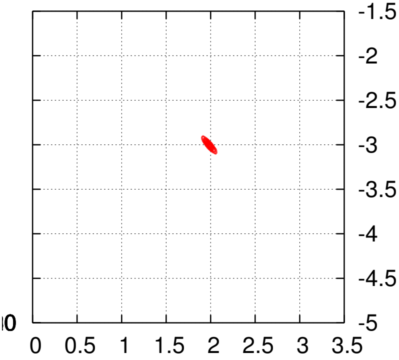


Constant uninformative prior

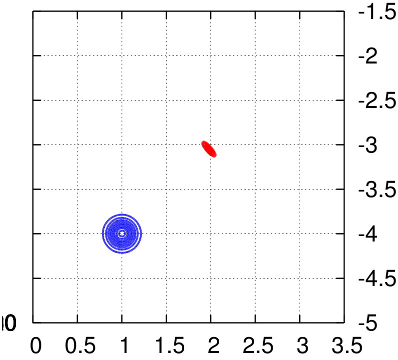


Gaussian prior

Line fitting example – prior vs. data-size 2000 data points



Constant uninformative prior



Gaussian prior

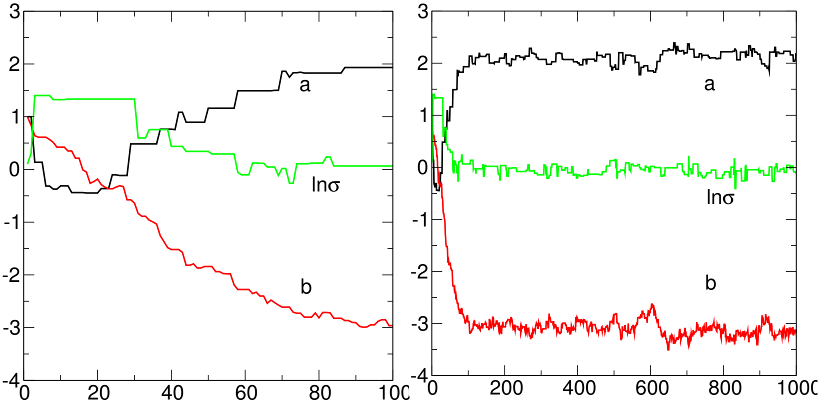
Exploring the Posterior

- Given any sample λ, the un-normalized posterior probability can be easily computed

$$p(\lambda|y) \propto p(y|\lambda)p(\lambda)$$

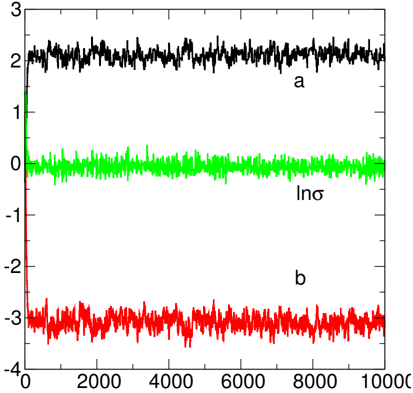
- Explore posterior w/ Markov Chain Monte Carlo (MCMC)
  - Metropolis-Hastings algorithm:
    - Random walk with proposal PDF & rejection rules
  - Computationally intensive, $\mathcal{O}(10^5)$ samples
  - Each sample: evaluation of the forward model
    - Surrogate models
- Evaluate moments/marginals from the MCMC statistics

Line fitting example – MCMC – $(a, b, \ln \sigma)$ samples



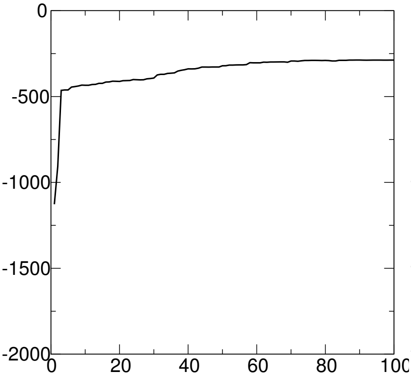
- Initial transient "Burn-in" period, ≈ 100 steps
- Problem and initial condition dependent

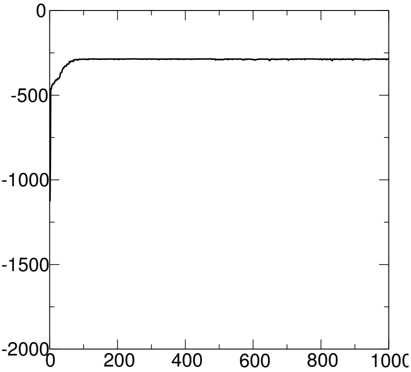
Line fitting example – MCMC – $(a, b, \ln\sigma)$ samples



- Visual inspection reveals “good mixing”
- No significant long-term correlation or periodicity

Line fitting example – MCMC – posterior density





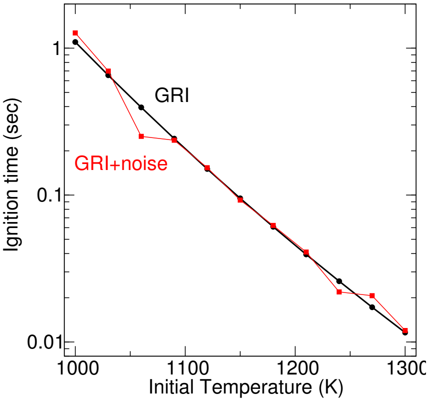
- Chain finds high posterior density (HPD) region
- stays there generating many random samples

Chemical Rate Parameter Estimation example

Synthetic ignition data generated using a detailed model+noise

- Ignition using GRImech3.0 methane-air chemistry
- Ignition time versus Initial Temperature
- Multiplicative noise error model
- 11 data points:

$$\tau_i^d = \tau^{\text{GRI}}(T_i^o)\,(1+\sigma\epsilon_i)$$
$$\epsilon \sim N(0,1)$$



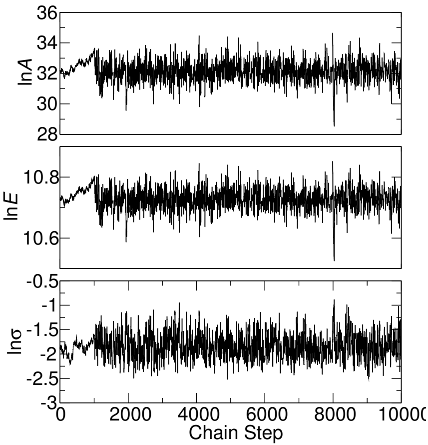
Fitting with a simple chemical model

- Fit a global single-step irreversible chemical model

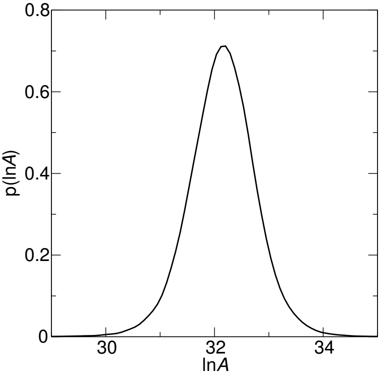
$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$$

$$\begin{aligned} \mathfrak{R} &= [CH_4][O_2]k_f \\ k_f &= A\exp(-E/R^oT) \end{aligned}$$

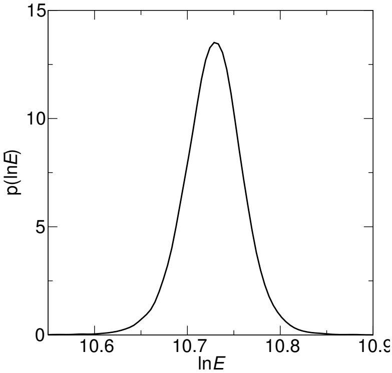
- Infer 3-D parameter vector $(\ln A, \ln E, \ln\sigma)$
- Good mixing with adaptive MCMC when start at MLE



Marginal Posteriors on $\ln A$ and $\ln E$

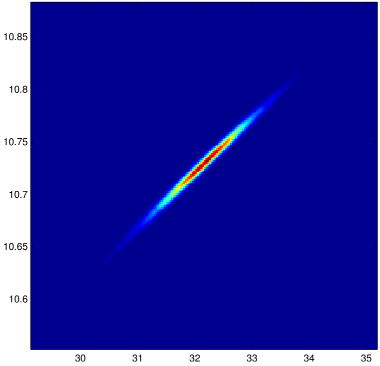


$\ln A = 32.15 \pm 3 \times 0.61$

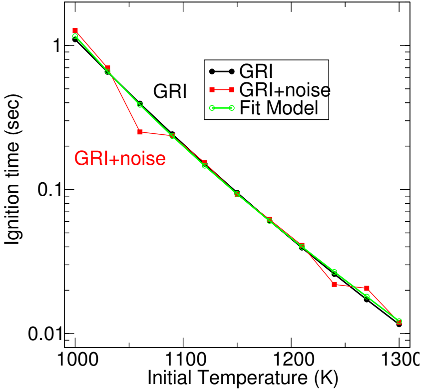


$\ln E = 10.73 \pm 3 \times 0.032$

Bayesian Inference Posterior and Nominal Prediction



Marginal joint posterior on $(\ln A, \ln E)$ exhibits strong correlation



Nominal fit model is consistent with the true model

Model UQ

- No model of a physical system is strictly true
- The probability of a model being strictly true is zero
- Given limited information, some models may be relied upon for describing the system

Let $\mathcal{M} = \{M_1, M_2, \ldots\}$ be the set of all models

- $p(M_k|I)$ is the probability that M_k is the model behind the available information
  - Model Plausibility
- Parameter estimation from data is conditioned on the model

$$p(\theta|D, M_k) = \frac{p(D|\theta, M_k)\pi(\theta|M_k)}{p(D|M_k)}$$

Bayesian Model Comparison

Evidence (marginal likelihood) for M_k:

$$p(D|M_k) = \int p(D|\theta, M_k)\pi(\theta|M_k)\mathrm{d}\theta$$

Bayes Factor B_{ij}:

$$B_{ij} = \frac{p(D|M_i)}{p(D|M_j)}$$

Plausibility of M_k:

$$p(M_k|D,\mathcal{M}) = \frac{p(D|M_k)\ \pi(M_k|\mathcal{M})}{\sum_s p(D|M_s)\pi(M_s|\mathcal{M})} \qquad k = 1,\ldots$$

Posterior odds:

$$\frac{p(M_i|D,\mathcal{M})}{p(M_j|D,\mathcal{M})} = B_{ij}\ \frac{\pi(M_i|\mathcal{M})}{\pi(M_j|\mathcal{M})}$$

Marginal Likelihood example

- Consider Fitting with data from a truth model

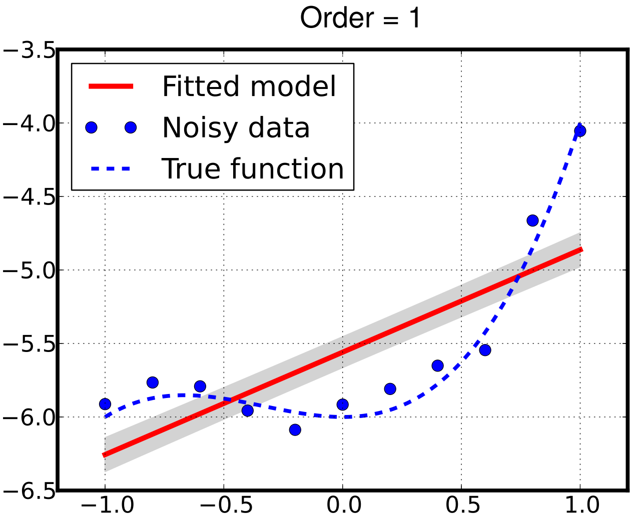
$$y_t = x^3 + x^2 - 6$$

- Gaussian *iid* additive noise model with fixed variance s
- Bayesian regression with a Gaussian Likelihood, *iid* and given s
- Consider a set of Legendre Polynomial expansion models, order 1-10

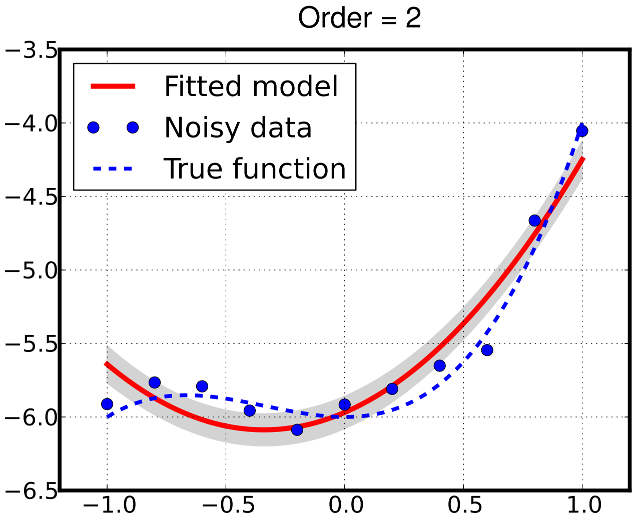
$$y_m = \sum_{k=0}^{P} c_k \psi_k(x)$$

- Uniform priors $[-D, D]$ on all coefficients

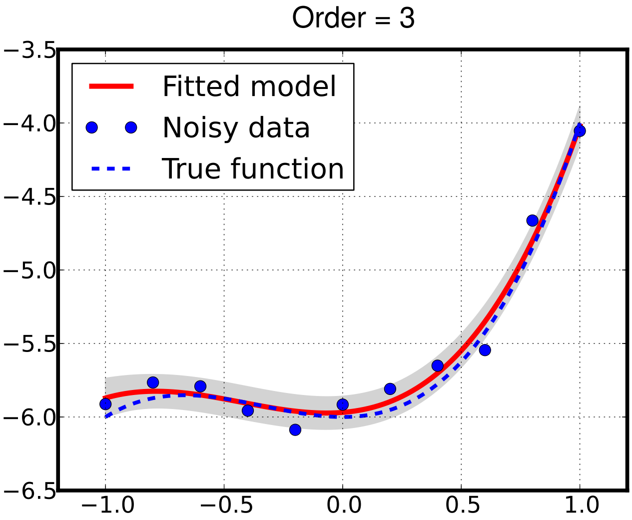
Too much model complexity leads to overfitting



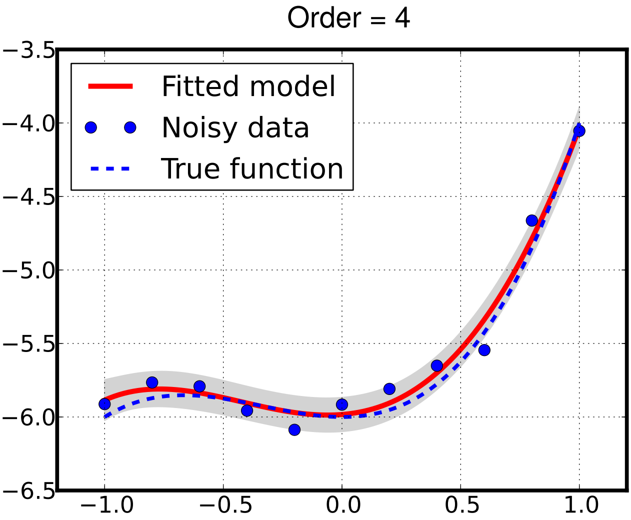
Too much model complexity leads to overfitting



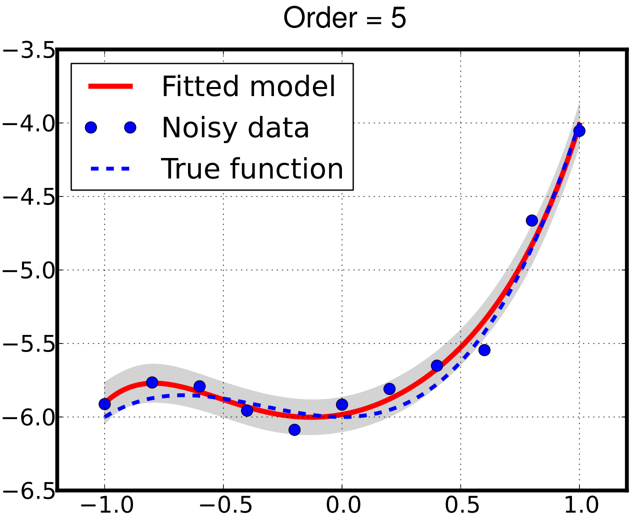
Too much model complexity leads to overfitting



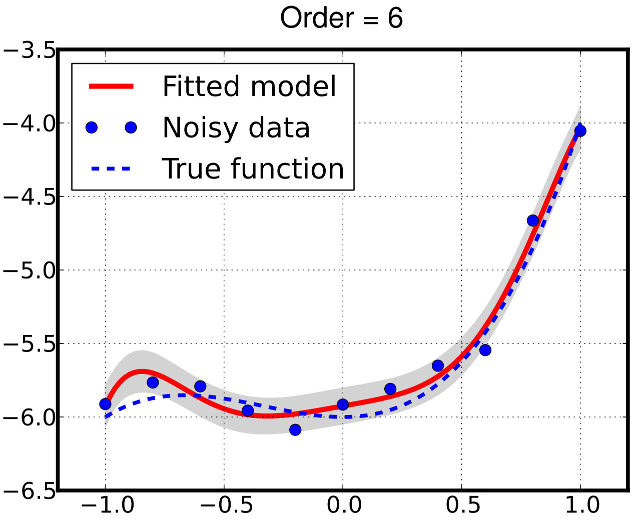
Too much model complexity leads to overfitting



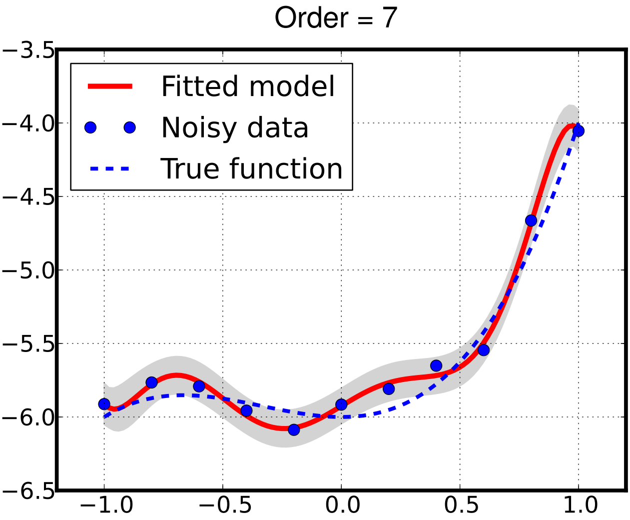
Too much model complexity leads to overfitting



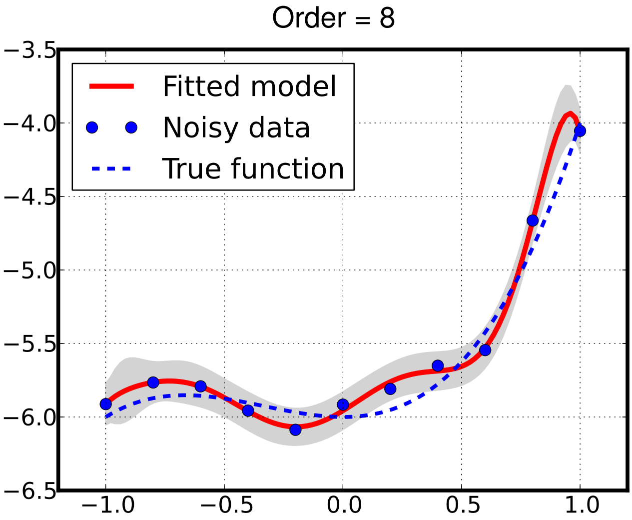
Too much model complexity leads to overfitting



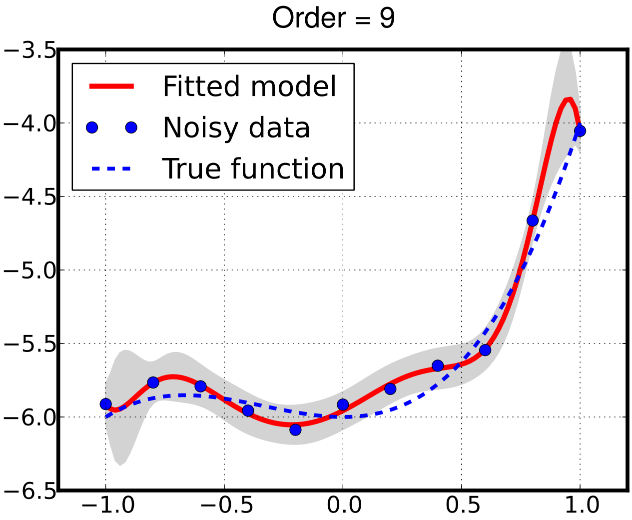
Too much model complexity leads to overfitting



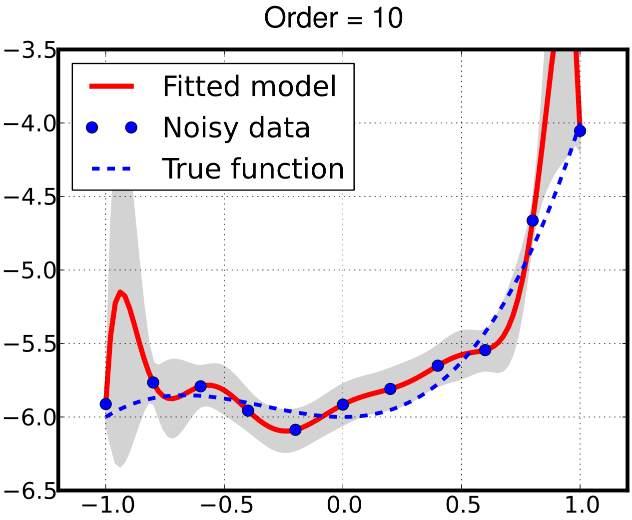
Too much model complexity leads to overfitting



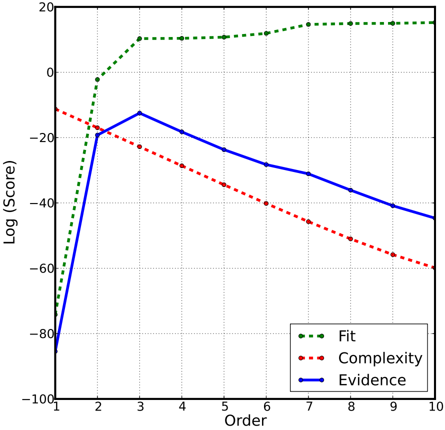
Too much model complexity leads to overfitting



Too much model complexity leads to overfitting

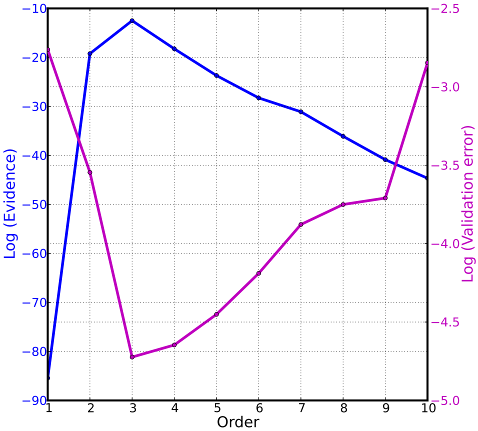


Evidence – Marginal Likelihood



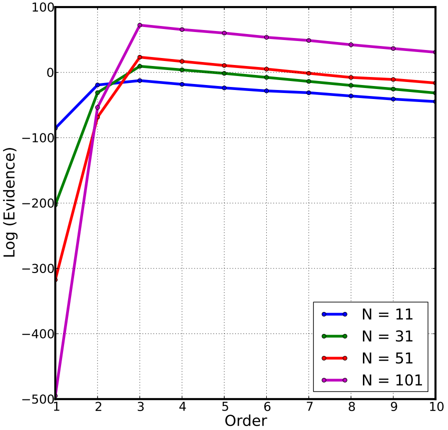
- Log evidence: sum of two scores, balances complexity & fit
- Peaks at 3rd order

Evidence and Validation Error



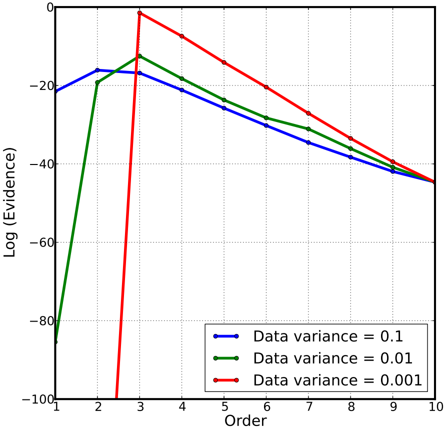
- Validation error – ℓ_2 error for a random set of 1000 points
- Validation error is minimal at the 3rd-order evidence peak

Evidence



- Discrimination among models is more clear-cut with higher amount of data

Evidence



- Discrimination among models is more clear-cut with less data noise

Bayesian Model Comparison in Physical Models

Bayesian methods for model comparison and selection have been used in

- Combustion
  - Turbulent combustion modeling (Cheung, Rel. Eng. 2011)
  - Graphite nitridation chemistry (Miki, AIAA Conf. 2012)
  - Syngas chemistry (Braman, CTM 2013)
- Cosmology
- Social science
- Biology – phylogenetics
- Climate modeling
- ...

Validation

- Validity is a statement of model utility for predicting a given observable under given conditions
- Inspection of model utility requires accounting for uncertainty
- Statistical tool-chest for model validation
  - Cross-validation
  - Bayes Factor
  - Model Plausibility
  - Posterior Odds
  - Posterior predictive:

$$p(\tilde{D}|D, M_k) = \int p(\tilde{D}|\theta, M_k)p(\theta|D, M_k)d\theta$$

Model Averaging

- When multiple models are acceptable, and no model is a clear winner, model averaging can be used to provide a prediction of interest
- If prediction errors among models are uncorrelated, then averaging is expected to reduce prediction errors
  - Not likely if models are dependent, or if they have comparable large bias errors in a given observable of interest
- Bayesian Model Averaging

$$p(\phi|D,\mathcal{M}) = \sum_k p(\phi|D,M_k)p(M_k|D,\mathcal{M})$$

Closure

- Probabilistic UQ framework
  - Forward and Inverse UQ
- Bayesian inference
  - Model calibration: parameter estimation
  - Parametric and model uncertainty
- Model comparison, validation, averaging
- Prediction with uncertainty