Inference, Calibration, and Prediction: Applying Bayesian Statistical Methods to the Muon Ionization Cooling Experiment (MICE)

Robert D. Ryne

Center for Beam Physics
Accelerator Technology & Applied Physics Division
Lawrence Berkeley National Laboratory

May 1, 2015

ATAP
ACCELERATOR TECHNOLOGY & APPLIED PHYSICS DIVISION

BERKELEY LAB

Acknowledgements

- Chris Rogers (Rutherford Appleton Laboratory)
- Elizabeth Kelly, Earl Lawrence, David Higdon (Los Alamos National Laboratory)

Outline

- Intro to MICE and concepts from beam physics
  - for applied mathematicians & statisticians
- Intro to statistical tools that combine expt'l measurement and simulation for inference
  - for beam physicists
- Application to MICE

Note

- The target audiences for this presentation include *both* accelerator physicists and applied mathematicians/statisticians
- The introductory sections are written with that in mind

Outline

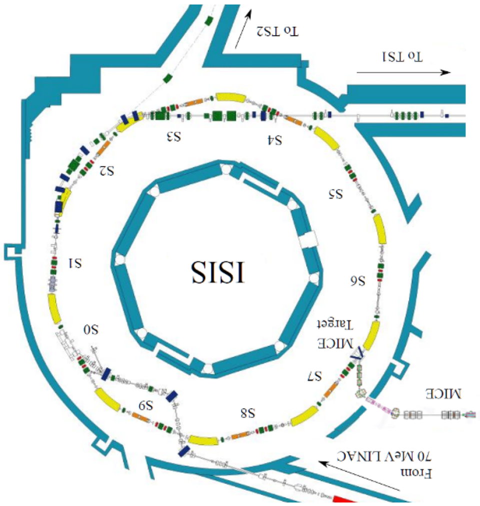
- Intro to MICE and concepts from beam physics
  - for applied mathematicians & statisticians
- Intro to statistical tools that combine expt'l measurement and simulation for inference
  - for beam physicists
- Application to MICE

What is MICE?

- An international experiment at the ISIS facility at STFC Rutherford Appleton Laboratory (Oxford, UK).

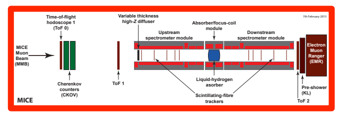
What is MICE?

- An international experiment at the ISIS facility at STFC Rutherford Appleton Laboratory (Oxford, UK).

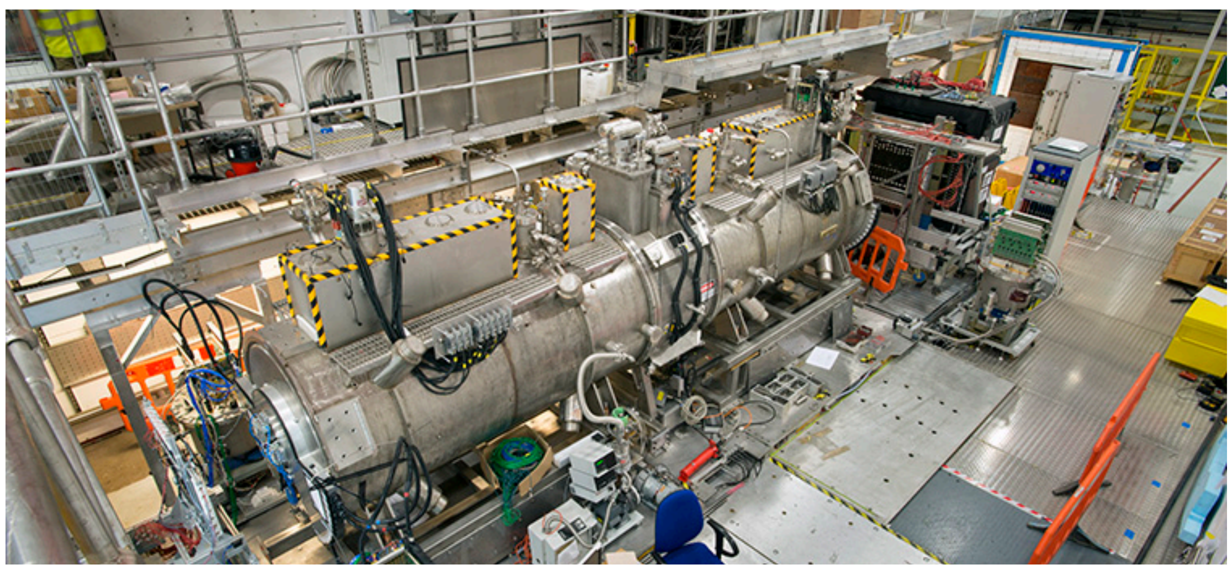


muons
from ISIS

MICE Beamline



cooled muons



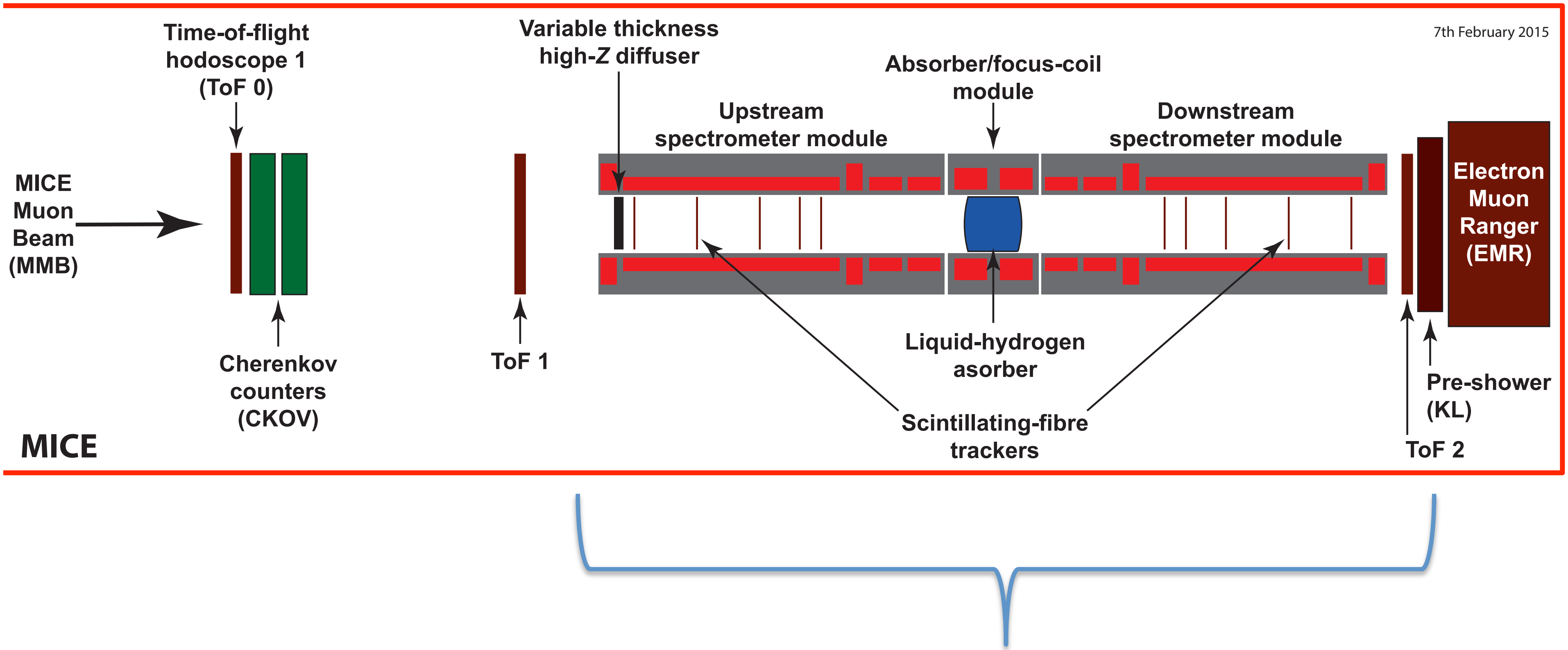
MICE is an Extraordinary Experiment

- Initial and final state of *every particle* will be measured individually
  - not just the usual beam centroid, rms size,...
- Unprecedented opportunity to combine measurements and simulation for inference and insight

MICE is happening *now*

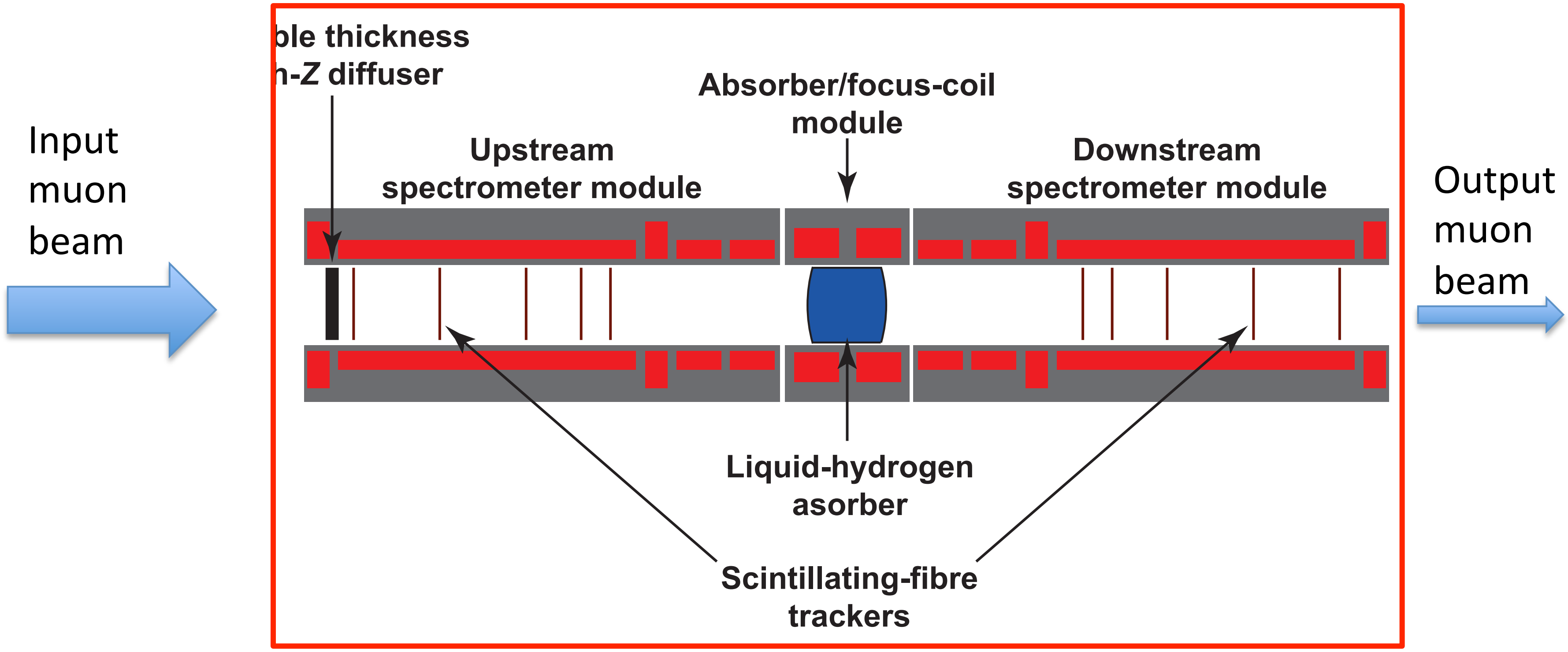
- Originally envisioned to have several steps
  - Planning and early steps underway since early 2000's
- MICE Step IV is upon us
  - Commissioning and data collection 2015 thru May 2016
  - Study the impact of material properties, optics, and input-beam properties on cooling
- Final Step: Data collection May 2017 to March 2018
  - Cooling demonstration
- Step IV provides first opportunity to collect data to calibrate a computer model of the beamline

MICE Step IV



We want to calibrate a computer model of this

Parameters of the simulation model



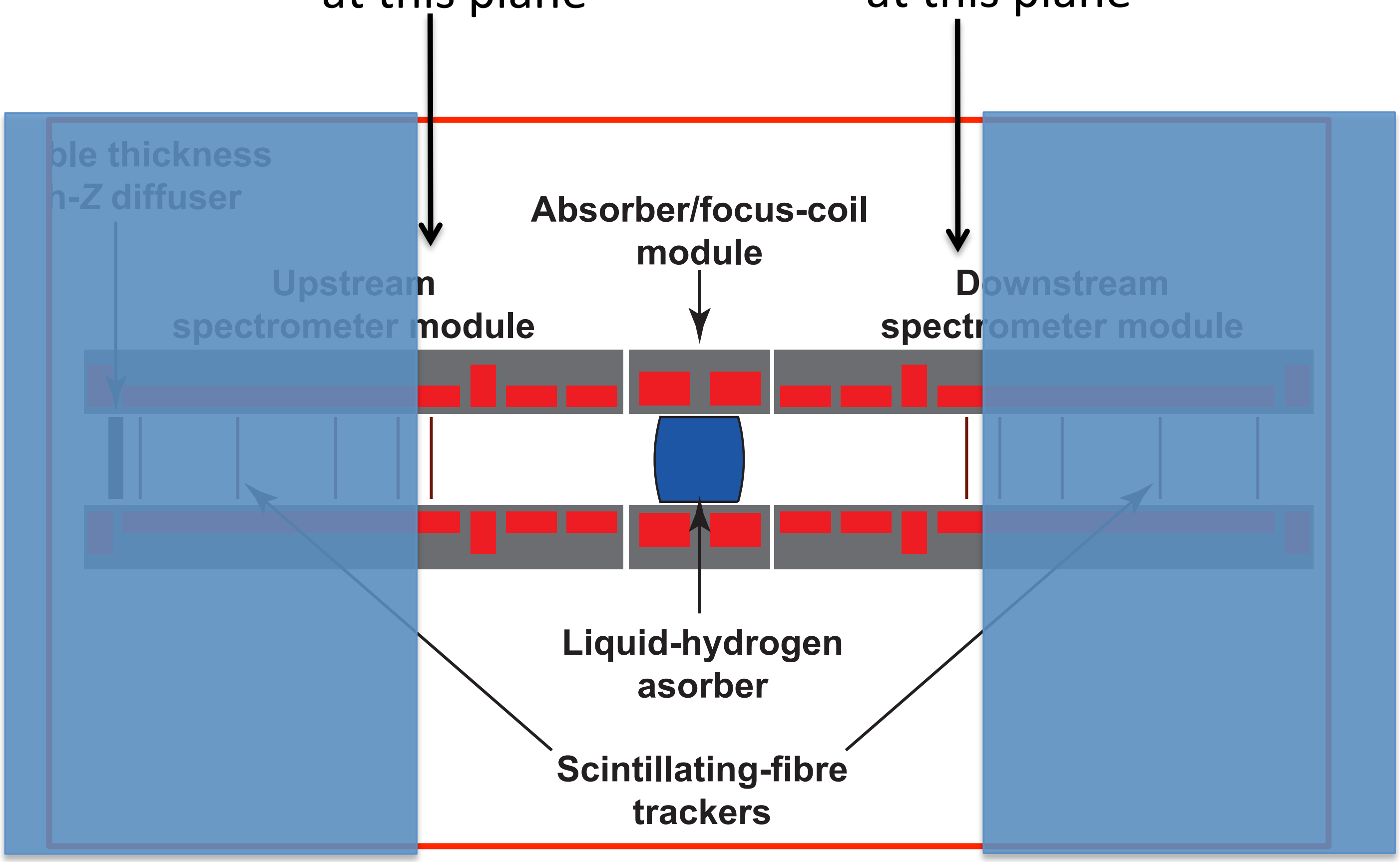
- 12 solenoids each described by 5 parameters (J, r_1, r_2, z_c, l)
- Don't include the absorber for now

Current J can be changed in the control room.
The other 4 (geometrical) parameters are fixed.

- Don't include magnet imperfections, misalignment, etc.
  - Can add by using MAUS as the simulator
  - In the meantime, resulting calibration parameters would still be aware of these, since calibration is based on measurements

muon input values
"X" are measured
at this plane

muon output values
"Y" are measured
at this plane



- The currents in the magnets affect the magnetic field that guides the muons.
- The currents are likely to be fixed for a day or several days, then changed.
- Each day, order ~1M muons from ISIS will pass through MICE and be measured.

Experimental Observables and derived quantities

Observables:

- Muon input and output 6-vectors at the I/O planes
  - 6 vector = (x, p_x, y, p_y, t, E)

(x,y)=transverse position
(p_x, p_y)=transverse momentum
t=arrival time at z-plane
E=muon energy

Derived quantities:

- Beam 2nd moment matrix
- Beam rms **emittance**

The primary goal of MICE is to demonstrate our understanding of, and validity of, a technique for *cooling*, i.e., *emittance reduction*

The muon input beam can be changed by adjusting beamline elements upstream of MICE

3π 140 MeV/c	6π 140 MeV/c	10π 140 MeV/c
3π 200 MeV/c	6π 200 MeV/c	10π 200 MeV/c
3π 240 MeV/c	6π 240 MeV/c	10π 240 MeV/c

- The nominal setting (center of table) has emittance ε=6π mm-rad and momeneum p=200 MeV/c
- Other settings will be explored to see their impact on the muon cooling process

What is beam rms emittance?

- Roughly, rms emittance, ε, is a measure of macroscopic* beam area (or volume) in phase space
  - 2D phase space = (x,p_x)
  - 4D phase space = (x,p_x,y,p_y)
  - 6D phase space=(x,p_x,y,p_y,t,E)
- Let M = beam rms 2nd moment matrix

  $m_{11}=\langle x\ x\rangle$, $m_{12}=\langle x\ p_x\rangle$, $m_{13}=\langle x,y\rangle$, $m_{14}=\langle x,p_y\rangle$, etc.
  - If M_{4x4} is 2x2 block diagonal, then $\varepsilon_x^2=\langle x^2\rangle\langle p_x^2\rangle - \langle x\ p_x\rangle^2$
- In MICE, cooling can be quantified by sqrt(det[M_{4x4}])

*not to be confused w/ microscopic volume, which is constant in a Hamiltonian system (Liouville's Theorem)

What can affect rms emittance?

- 2 things:

(1)Non-Hamiltonian effects

This is what MICE is about: inserting absorber material in a carefully designed beamline to reduce the emittance through ionization cooling

(2)Nonlinear effects

This can cause emittance growth (heating) that will counteract the reduction (cooling) that we want to occur and that we want to measure

Nonlinear effects can obscure the beam-material interaction (i.e., the cooling) that we want to measure

Nonlinearities

- Motion of a particle propagating near the axis is linear
- Motion becomes increasingly nonlinear as it moves away from the axis and as its transverse velocities increase
  – large (x,y) causes nonlinear dynamics
  – large (p_x,p_y) causes nonlinear dynamics

- Question: if we will measure every muon, why not do the analysis with just the particles near the axis? If we throw out large amplitude particles, nonlinearities won't matter.
  – Answer: we need large (p_x,p_y) at the absorber for the cooling to work.

Nonlinearities, cont.

- There is a region of parameter space for the input particle distribution where
  - particles won't experience nonlinear effects
  - (p_x,p_y) is large enough at the absorber to measurably cool the beam
- This is where we want to operate, but it requires careful control of the input beam that will be difficult to achieve

- We need to be prepared that ionization cooling and nonlinear heating might both be present

A calibrated model of the beamline is needed

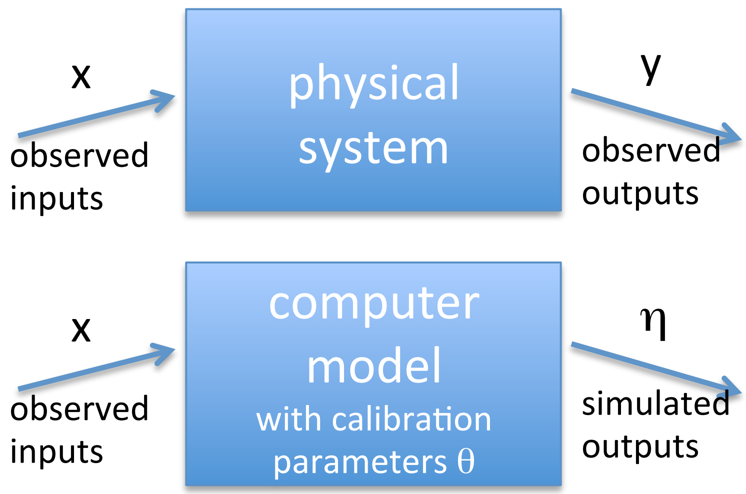
Importance of a well-understood, calibrated computer model for MICE

- Measurement alone can show cooling
  - But it doesn't demonstrate understanding
  - A model is needed to quantitatively <u>predict</u> cooling
- A model will also allow us to
  - disentangle nonlinear heating from ionization cooling
  - explore and diagnose unexpected observations
  - test ideas more quickly than changing the expt'l configuration
  - guide toward advantageous regions of the parameter space for the expt to be a success

Outline

- Intro to MICE and concepts from beam physics
  - – for applied mathematicians & statisticians
- Intro to Bayesian statistical tools that combine expt'l measurement and simulation for inference
  - – for beam physicists
- Application to MICE

Bayesian Inference for Code Calibration*



$$y(x_i)=\eta(x_i,\theta)+\varepsilon(x_i)+\delta(x_i) \quad i=1,\ldots,n$$

i denotes a measurement

- At various inputs x, we have measurements y, w/ meas error ε and model error δ
  - For now we ignore model error $\delta(x_i)$ and assume the x_i are known perfectly
- We have a sampling model or likelihood, L, for the observations y
  - For example, assume ε is multivariate $N(0,\Sigma)$.

*D. Higdon et al., "Combining Field Data and Computer Simulations for Calibration and Prediction," SIAM J. Sci. Comput. Vol. 26, No. 2, pp. 448-466 (2004)

Bayesian Inference, cont.

- Suppose we have some prior knowledge about what we think the model parameters, θ, must be for the simulator to agree with measurements. Let $\pi(\theta)$ denote the prior distribution.
- The Bayesian formulation states:

$$\pi(\theta|y) \propto L(y|\eta(x,\theta)) \times \pi(\theta)$$

The posterior is proportional to the likelihood times the prior

- Instead of thinking of θ simply as an ordinary scalar or vector quantity, we think of it as a random variable with a distribution associated with it
  - initially, this is the prior $\pi(\theta)$
- We want to find the posterior distribution of θ given the data y, $\pi(\theta|y)$
  - Then we can determine the moments of $\pi(\theta|y)$, Bayesian credible intervals, etc.
- Also can determine posterior for the variance parameters of the sampling model, $\pi(\Sigma|y)$
- All this involves combining observations with computer simulations

Markov Chain Monte Carlo

- In general problems will be nonlinear & multi-dimensional
  - Closed form solutions for the posterior not available
- How to obtain the posterior's properties?
  - Generate a sequence of samples drawn from $\pi(\theta|y)$
  - Can use to obtain moments, histograms, Bayesian credible intervals, etc.
- One way to accomplish this is with a Markov Chain Monte Carlo (MCMC) method
  - Produces a sequence of values equivalent to a random draw from the posterior conditional (on y) distribution
  - In the following we use the Metropolis algorithm

Prediction

- Having obtained the posterior $\pi(\theta|Y)$ to calibrate the computer model in the previous examples, how to predict what will happen with new X values (i.e. new muons in MICE)?
  - Simply using the posterior means of θ does not include info about our uncertainty in θ
- The correct way is to use the posterior predictive distribution:

$$p(Y_{new} \mid Y) = \int_\theta \pi(Y_{new} \mid \theta)\pi(\theta \mid Y)\, d\theta$$

where
$Y=Y(X)$,
$Y_{new}=Y_{new}(X_{new})$

- Given N new input observations, predict the output by doing the following in simulation:

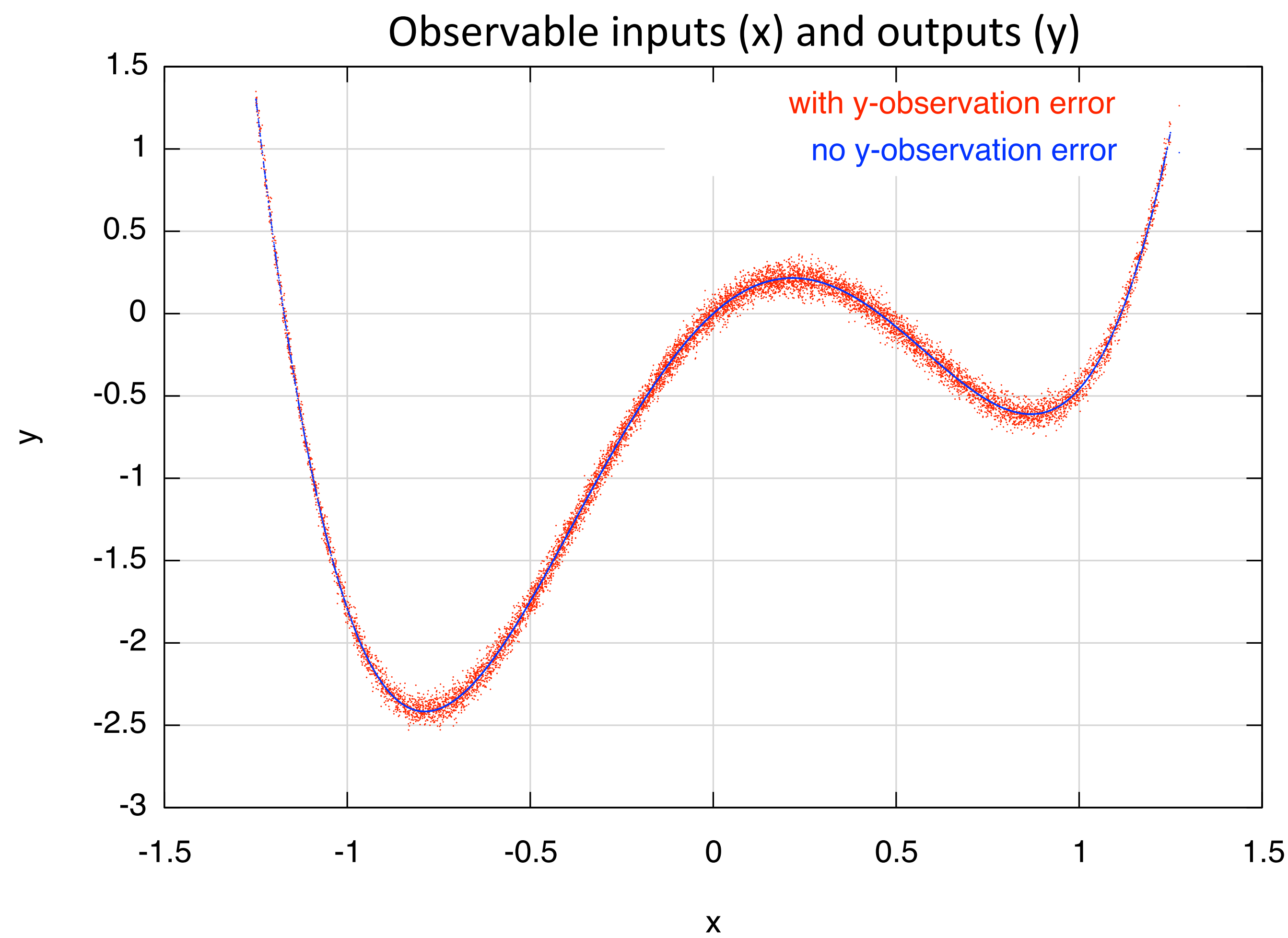
DO I=1,...,N

- draw θ' from $\pi(\theta|Y)$ [e.g. draw from the MCMC chain]
- forward simulate $X_{new,i}$ using θ' to produce $Y_{new,i}$

ENDDO

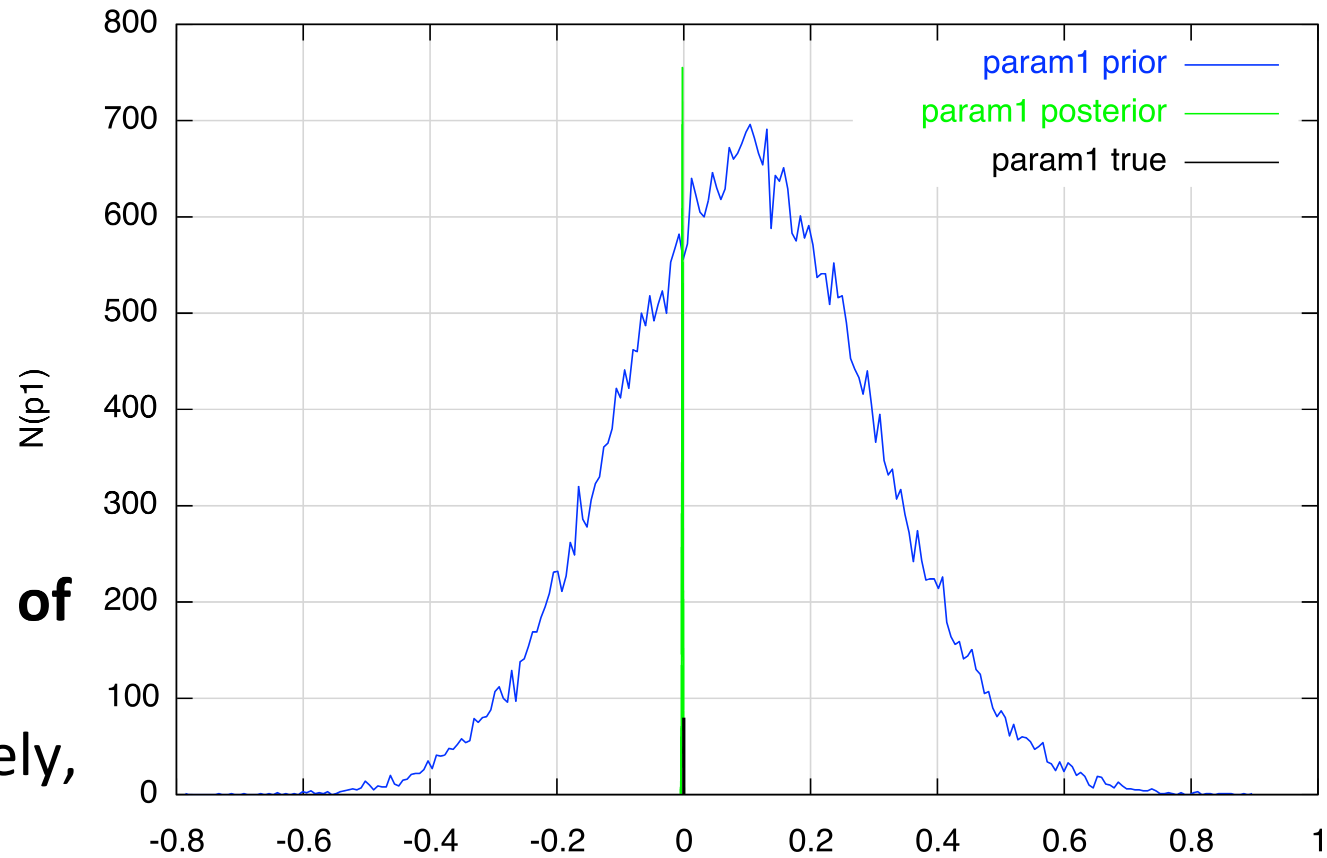
Example: Application of MCMC to a 1D regression problem

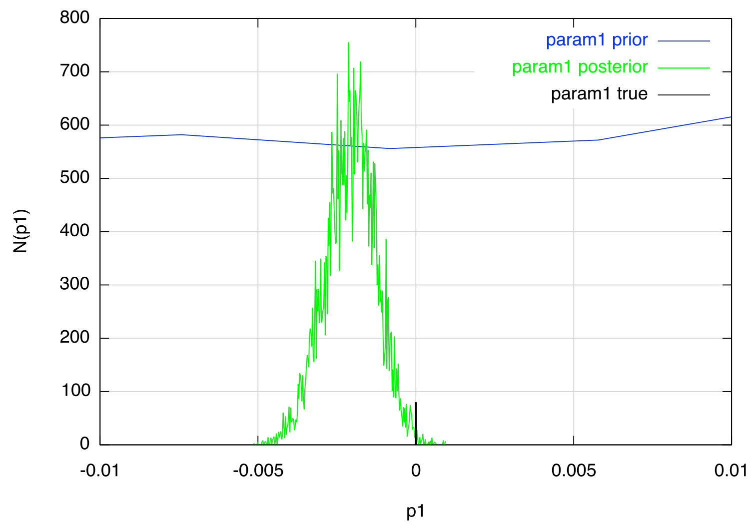
- Let y=b0 + b1 x + b2 x^2 + b3 x^3 + b4 x^4 +ε
  - Treat b0 thru b4 as calibration parameters
- b_{true}=(0, 2., -4.5, -1.33333, 3.375)
- 10000 observations
- Observation error on y: σ = 0.05
- Perform MCMC for 100K steps



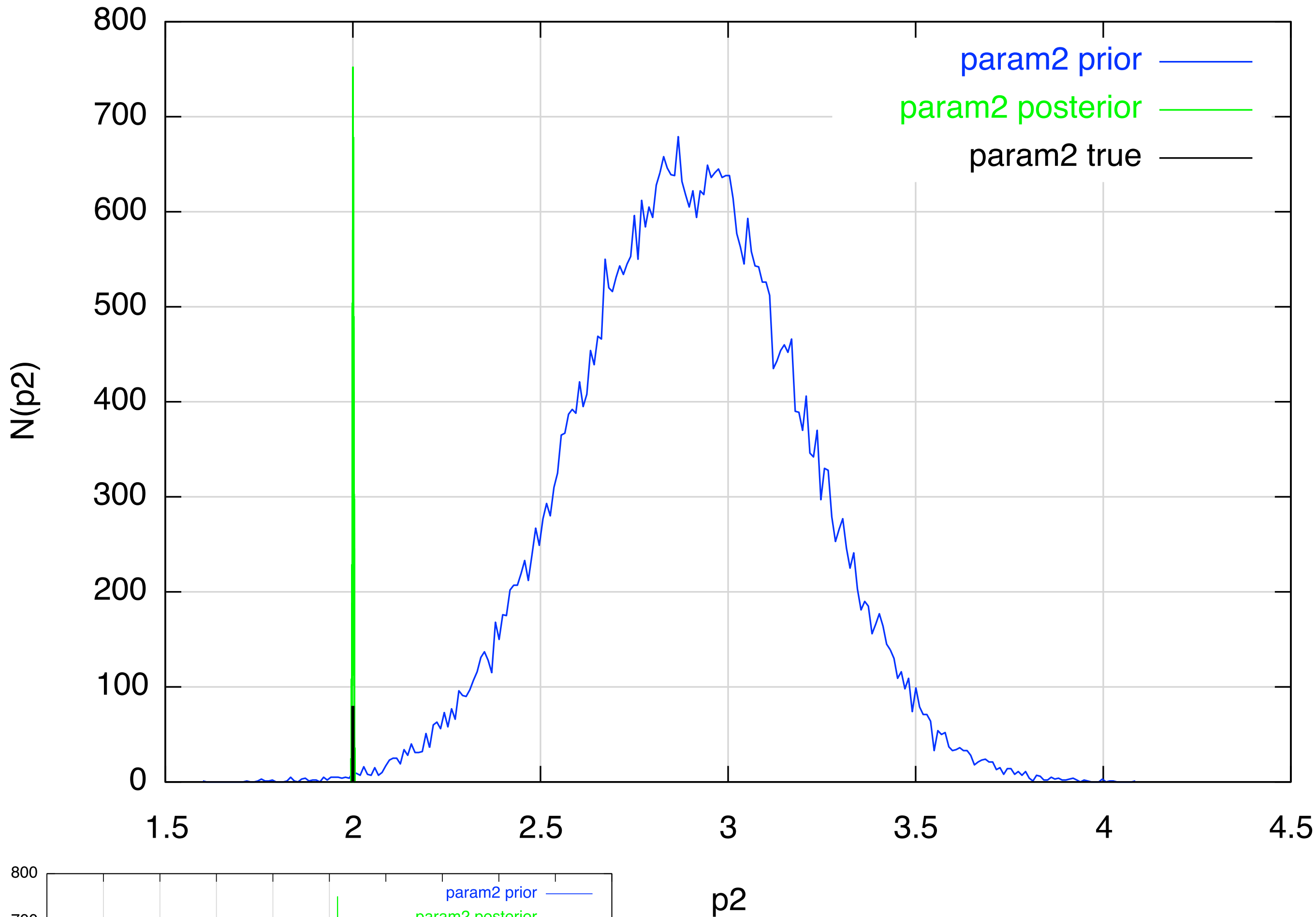
Posterior density (θ|y) of parameter 1

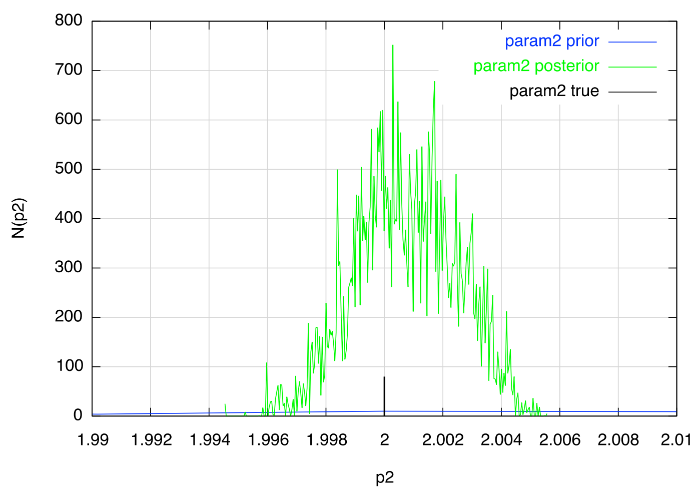
(more precisely, the posterior density of parameter 1 with other parameters integrated out)





Posterior density of parameter 2





Results Summary

param	exact	μ_{prior}	μ_{post}	σ_{prior}	σ_{post}
1	0	0.1	-2.e-3	0.2	8.2e-4
2	2	2.9	2.0008	0.3	1.8e-3
3	-4.5	-5.5	-4.497	0.9	3.4e-3
4	1.3333333	-1.0	-1.3353	0.4	1.9e-3
5	3.375	3.0	3.37512	0.8	2.5e-3

σ_{obs}=0.05

Comment:
With a lot of data, the priors don't have to be that good, instead we can use a diffuse prior

These results are encouraging, but...

- Not all problems are so easy
  - dependence on # of observations, length of Markov chain, MCMC proposal function, correlations among inferred quantities θ (calibration parameters, regressions parameters,...) , complexity of simulation model η,...
- Use requires care, experience,...
  - Wise to enlist the help of statistical scientists!

Application to MICE

- Overview of the MICE model used here

i. Inference of 10 solenoid parameters (current J) and 4 measurement uncertainties

ii. Inference of the linear transfer map

The Computer Model

- MICE simulations normally make use of the MICE User Analysis and Software (MAUS) framework
- This study uses a simplified model, but the techniques presented could be applied with *any* model
- Like MAUS, the computer model used here propagates muon input 6-vectors and produces output 6-vectors
  - Numerical integration of Lorentz force equations for $(x,\gamma\beta_x,y,\gamma\beta_y,t,\gamma)$ (z)
    - the longitudinal coordinate, z, is the independent variable
    - (x,y) are transverse coordinates
    - $(\gamma\beta_x,\gamma\beta_y)$ are proportional to the transverse momentum
    - (t, γ) : t is arrival time (irrelevant for this study); γ is proportional to the particle energy
  - Beamline consisting of 12 solenoids, each modeled as a current block, with overlapping fields

Lorentz Force Equations with z as the Independent Variable

- Consider quantities $(x,\gamma\beta_x,y,\gamma\beta_y,t,\gamma)$ (z)
- Equations of motion are:

$$\frac{dx}{dz}=\frac{\gamma\beta_x}{\gamma\beta_z}$$

$$\frac{d(\gamma\beta_x)}{dz}=\gamma\frac{q/mc^2}{\gamma\beta_z}\left[E_x+\frac{c}{\gamma}\left(\gamma\beta\times B\right)_x\right]$$

$$\frac{dy}{dz}=\frac{\gamma\beta_y}{\gamma\beta_z}$$

$$\frac{d(\gamma\beta_y)}{dz}=\gamma\frac{q/mc^2}{\gamma\beta_z}\left[E_y+\frac{c}{\gamma}\left(\gamma\beta\times B\right)_y\right]$$

$$\frac{dt}{dz}=\frac{1}{\beta_z c}$$

$$\frac{d\gamma}{dz}=\frac{q/mc^2}{\gamma\beta_z}\left[\gamma\beta\bullet E\right]$$

where $\gamma\beta_z=\sqrt{\gamma^2-\left(\gamma\beta_x\right)^2-\left(\gamma\beta_y\right)^2-1}$

Solenoid Channel Magnetic Field

- Superposition of 12 solenoids modeled via

$B(x=0,y=0,z)=\mu_0\ J\ (r_2-r_1)[f(z-z_1)-f(z-z_2)]$

$$f(z)=\left[\frac{1}{2}+\frac{z}{2(r_2-r_1)}\log\frac{r_2+\sqrt{r_2^2+z^2}}{r_1+\sqrt{r_1^2+z^2}}\right]$$

$B_x=-(B'/2)x+(B'''/16)r^2x-(B^{(5)}/384)r^4x+(B^{(7)}/18432)r^6x$
$B_y=-(B'/2)y+(B'''/16)r^2y-(B^{(5)}/384)r^4y+(B^{(7)}/18432)r^6y$
$B_z=B-(B''/4)r^2+(B^{(4)}/64)r^4-(B^{(6)}/2304)r^6-...$

Solenoid Parameters

$z_{center}(m)$	$z_{length}(m)$	$r_1(m)$	$r_2(m)$
-3200.d-3	110.6d-3	258.0d-3	325.8d-3
-2450.d-3	1314.3d-3	258.0d-3	280.1d-3
-1700.d-3	110.6d-3	258.0d-3	318.9d-3
-1300.d-3	199.5d-3	258.0d-3	288.9d-3
-861.d-3	201.3d-3	258.0d-3	304.2d-3
-202.75d-3	213.3d-3	267.0d-3	361.8d-3
202.75d-3	213.3d-3	267.0d-3	361.8d-3
861.d-3	201.3d-3	258.0d-3	304.2d-3
1300.d-3	199.5d-3	258.0d-3	288.9d-3
1700.d-3	110.6d-3	258.0d-3	318.9d-3
2450.d-3	1314.3d-3	258.0d-3	280.1d-3
3200.d-3	110.6d-3	258.0d-3	325.8d-3

$200 MeV/c$
134.d6
147.d6
131.d6
135.d6
113.d6
104.d6
-104.d6
-112.d6
-140.d6
-131.d6
-131.d6
-134.d6

Nominal current densities for <p>=200 MeV/c

Generation of synthetic "experimental" observations

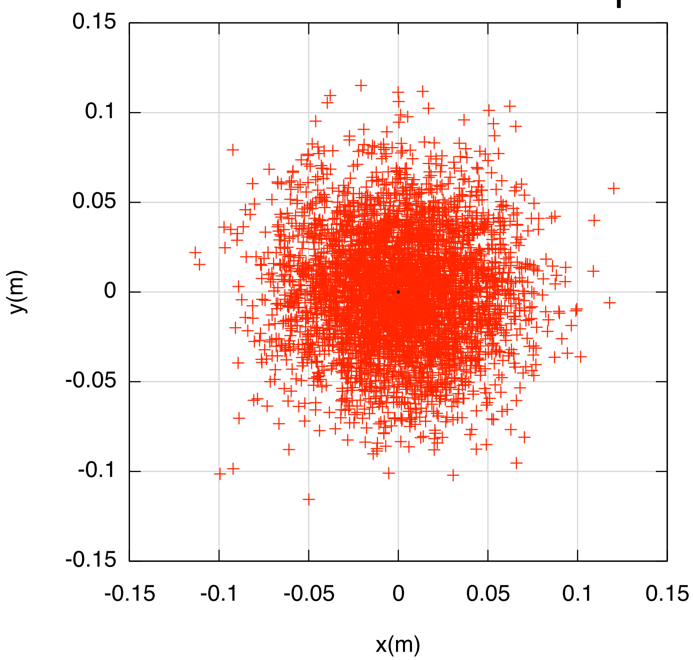
- Set magnet currents to values near (but ≠) nominal values
- Generate 6-vector inputs, "X" based on a matched beam in constant solenoid channel*
  - To simplify examples $\Delta E=0$, but this not necessary for the analysis
- Run the simulator to get outputs "Y"
- Add observation error to "Y"
  - errors ~ $N(0,\Sigma)$ where $\Sigma_{11}=\Sigma_{33}=0.4\text{e-}3$ m for simulated output coordinates (x,y) and $\Sigma_{22}=\Sigma_{44}=0.02$ for simulated output momenta $(\gamma\beta_x,\gamma\beta_y)$.

*Greg Penn MICE note

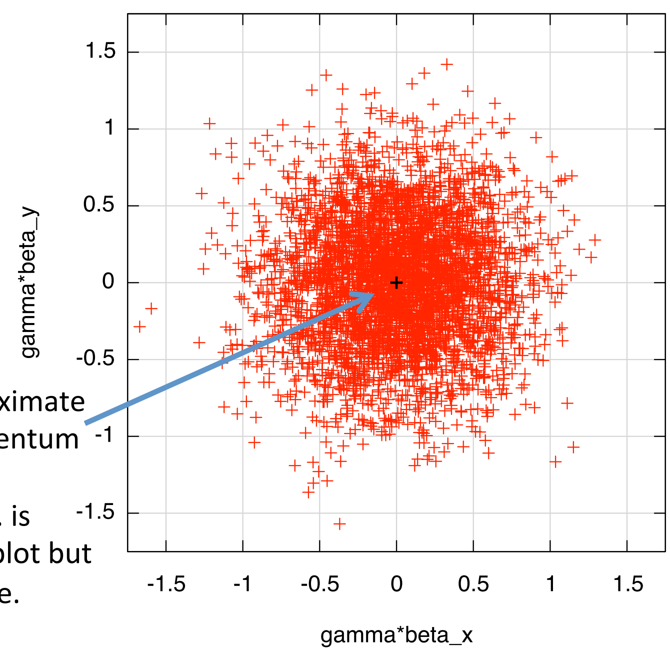
Typical simulation outputs

- Observed inputs, X, consists of ~10K - 1M 6-vectors of coordinates and momenta
  - in what follows we ignore last 2 components, so the problem involves 4-vectors
- Outputs, Y, are the result of simulating transport thru MICE beamline

simulated coordinate output



simulated momentum output



Cross shows approximate +/- 2σ s.d. in momentum measurement.
The coordinate s.d. is shown on the left plot but it is too small to see.

Inference of 10 parameters and 4 measurement uncertainties

param	exact	μ_{prior}	μ_{post}	σ_{prior}	σ_{post}
θ_1	151.634	147.	151.623	40.	.0185
θ_2	123.807	131.	123.752	40.	.0615
θ_3	142.602	135.	142.762	40.	.0722
θ_4	118.863	113.	118.930	40.	.0496
θ_5	103.874	104.	103.743	40.	.0652
θ_6	-101.920	-104.	-101.668	40.	.0918
θ_7	-108.330	-112.	-108.203	40.	.0753
θ_8	-132.950	-140.	-132.786	40.	.0976
θ_9	-127.378	-131.	-127.736	40.	.1266
θ_{10}	-133.948	-147.	-134.162	40.	.0669
τ_1	6.250e6	5.e6	6.256e6	1.0e6	.0903e6
τ_2	2500.	5000.	2434.	2236.	33.8
τ_3	6.250e6	5.e6	6.351e6	1.0e6	.0867
τ_4	2500.	5000.	2508.	2236.	36.7

# of observations=10000
length of MCMC chain = 5000 after 5000 burn-in

Comments

- The preceding examples are illustrative
- In real applications, the MCMC chain lengths should be longer, more observation data points could be used, the MCMC proposal draws should be tuned to have more reliable acceptance rates,...
- Still, the results successfully demonstrate the use of Bayesian inference techniques that combine measurements and simulation

Inference of the Linear Transfer Map

- The symplectic map describing the beamline can be expressed (in the absence of absorber) as a factored product of Lie transformations, *M*=exp(:f2:) exp(:f3:) exp(:f4:) ...
- The term exp(:f2:) corresponds to a linear map M
- We would like to measure (infer) the transfer map
- For illustration, consider just the linear map M
  – first step toward solving the nonlinear problem

Procedure for Linear Map Test Problem

- Use the Lorentz force code to generate a matched initial beam
  - convert to canonical variables and write it to a file
- Read the file into MaryLie/IMPACT and transport thru MICE with linear map; write the output
  - initial (X) and final (Y) canonical "observations"
- Prepare a metropolis MCMC code where the simulator is simply a 4x4 matrix-vector multiply
- Run the MCMC code
  - read X and Y; add observational error to Y
  - provide subroutines for likelyhood, priors, proposals

Digression: How to represent M?

- We could represent M as an arbitrary 4x4 matrix with 16 parameters
  - But the MCMC code would not result in a symplectic matrix, so we would have to symplectify it
    - no big deal, but an added complication and source of error
- Instead we can write* $M=\exp(JS_A)\exp(JS_C)$ where S_A , S_C are symmetric matrices that commute/anti-commute with J, resp.
  - S_A is 6D, S_C is 4D, so there are 10 parameters
  - 4x4 basis elements are known ($F_1,F_2,F_3,G_1,G_2,G_3,B_0,B_1,B_2,B_3$)
  - $M=\exp(\theta_1 F_1+\theta_2 F_2+\theta_3 F_3+\theta_4 G_1+\theta_5 G_2+\theta_6 G_3)\ \exp(\theta_7 B_0+\theta_8 B_1+\theta_9 B_2+\theta_{10} B_3)$
    - Priors for θ are not obvious; but the procedure works fine with broad prior

A. Dragt, "Lie Methods for Nonlinear Dynamics with Applications to Accelerator Physics," http://physics.umd.edu/dsat/

Bayesian inference of the 4x4 Linear Map: Results

MaryLie/Impact ("true") 4x4 matrix:

$$\begin{bmatrix} 4.81344E-01 & -1.55775E-01 & -9.60235E-03 & 3.10757E-03 \\ 4.83031E+00 & 5.13476E-01 & -9.63599E-02 & -1.02433E-02 \\ 9.60235E-03 & -3.10757E-03 & 4.81344E-01 & -1.55775E-01 \\ 9.63599E-02 & 1.02433E-02 & 4.83031E+00 & 5.13476E-01 \end{bmatrix}$$

MCMC results (10K observations; MCMC length=75K+75K burn-in:

$$\begin{bmatrix} 4.81349E-01 & -1.55757E-01 & -9.60723E-03 & 3.11229E-03 \\ 4.82935E+00 & 5.13961E-01 & -9.65443E-02 & -1.03595E-02 \\ 9.65088E-03 & -3.11962E-03 & 4.81362E-01 & -1.55776E-01 \\ 9.66604E-02 & 1.02018E-03 & 4.82893E+00 & 5.13895E-01 \end{bmatrix}$$

Both of these matrices are symplectic

	μ_{post}	σ_{post}
θ_2	1.556	2.04e-4
θ_5	-0.299	2.86e-5
θ_7	-1.374	1.00e-4
θ_9	-2.000e-2	3.83e-5

	μ_{post}	σ_{post}
θ_1	-9.e-6	2.8e-5
θ_3	-1.5e-5	2.9e-5
θ_4	4.6e-5	1.5e-4
θ_6	-8.8e-6	2.0e-4
θ_8	-1.3e-5	9.8e-5
θ_{10}	-2.6e-6	7.1e-5

Turns out (presumably due to symmetry associated w/ MICE channel) that only 4 regression coefficients matter. Note the small σ compared to the mean of these 4.

Comments on Linear Map example

- Could have done this simple example with more traditional techniques
- But eventually want to do the full (linear +nonlinear) map
  - Bayesian approach very flexible
  - Extension to nonlinear map involves
    - Replace matrix transport with transport via 4th order Lie map
    - Extend inferred quantities beyond the coefficients of matrix basis elements to also include coefficients of Lie polynomials f3 and f4
- Bayesian approach also allows inference of observational uncertainty

Conclusions

- Bayesian techniques are extremely powerful and flexible
- Applied to MICE, they can be used for computer model calibration, to infer the transfer map, to predict the impact of changes, to test ideas, and to provide insight
- The examples here demonstrate how measurements and simulation can be combined to
  - infer model parameters, (e.g. magnet current settings) so that the computer model agrees with expt, *including distributions that describe the uncertainty of inferred parameters*
  - infer the measurement uncertainty
  - infer the transfer map
- The techniques should be broadly applicable to other accelerator experiments as well