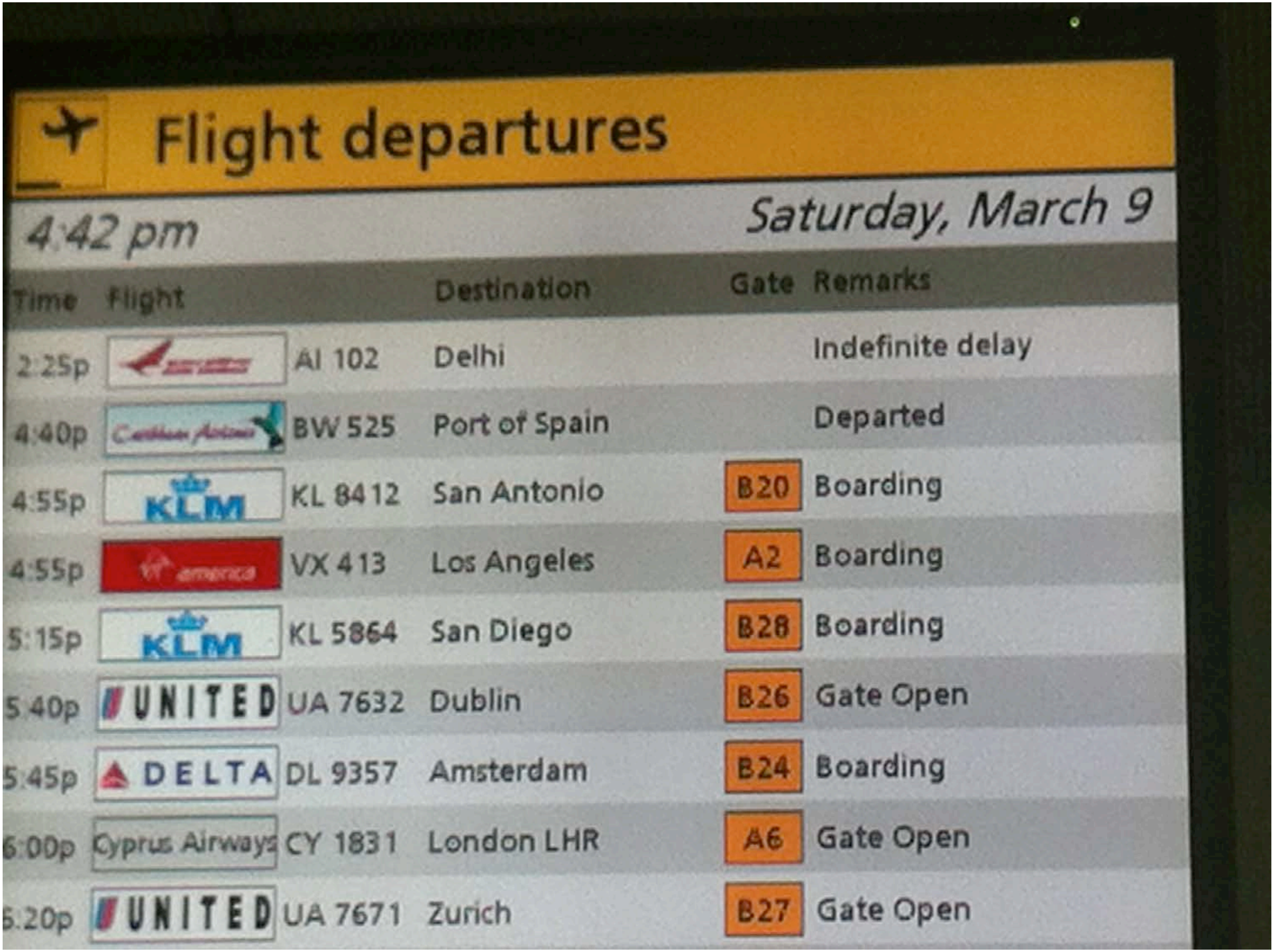
Quantifying Uncertainty

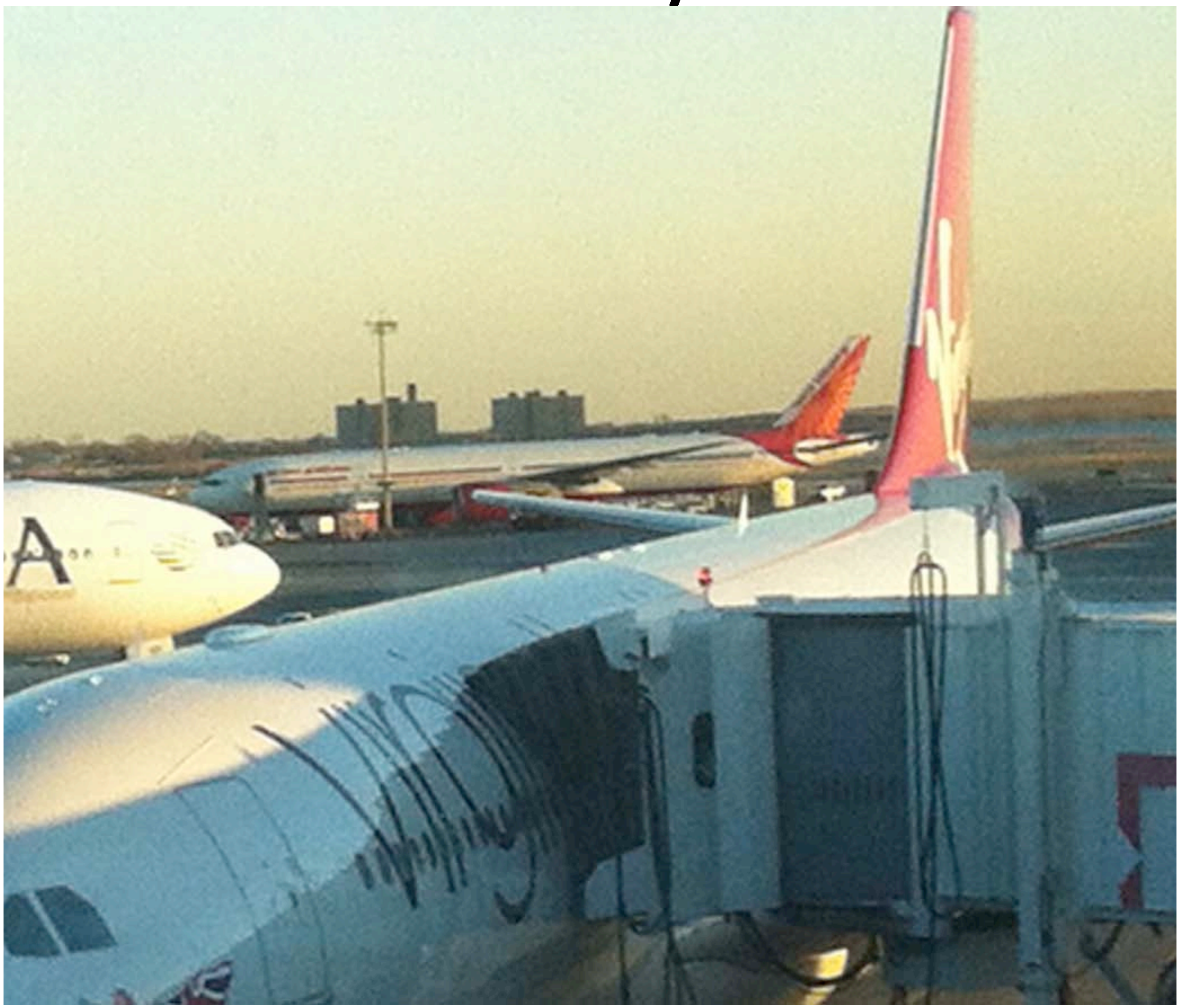
Sai Ravela
Massachusetts Institute of Technology

the many sources of uncertainty!

Two days ago…


Flight departures
4:42 pm
Saturday, March 9
Time
Flight
Destination
Gate
Remarks
2:25p
AI 102
Delhi
Indefinite delay
4:40p
BW 525
Port of Spain
Departed
4:55p
KLM
KL 8412
San Antonio
B20
Boarding
4:55p
VX 413
Los Angeles
A2
Boarding
5:15p
KLM
KL 5864
San Diego
B28
Boarding
5:40p
UNITED
UA 7632
Dublin
B26
Gate Open
5:45p
DELTA
DL 9357
Amsterdam
B24
Boarding
6:00p
Cyprus Airways
CY 1831
London LHR
A6
Gate Open
6:20p
UNITED
UA 7671
Zurich
B27
Gate Open

Finally

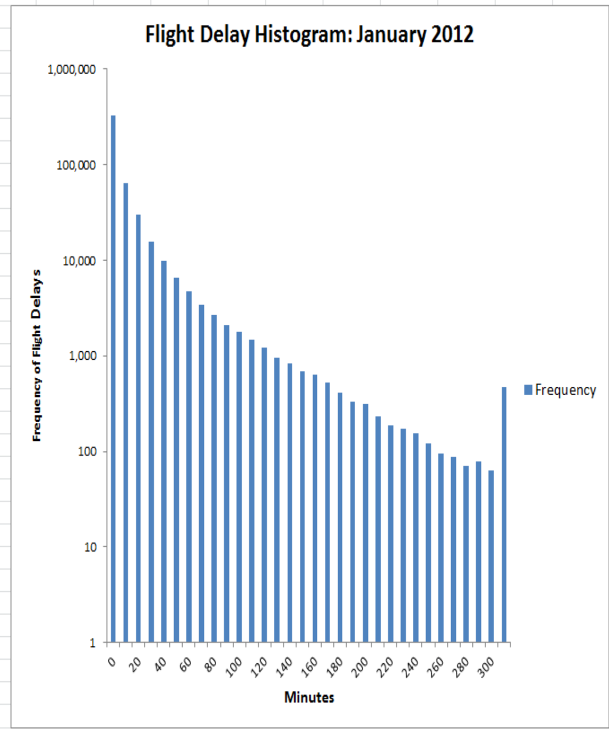


Quantifying Indefinite Delay

- P(X=*delay*| M="Indefinite Delay")

- P(Z=*cancel* |
  M="Indefinite Delay")

Bin	Frequency
0	328989
10	63659
20	29692
30	15653
40	9716
50	6501
60	4710
70	3427
80	2679
90	2080
100	1791
110	1484
120	1216
130	947
140	832
150	686
160	640
170	530
180	416
190	333
200	314
210	229
220	186
230	171
240	153
250	121
260	96
270	88
280	70
290	79
300	63
More	474



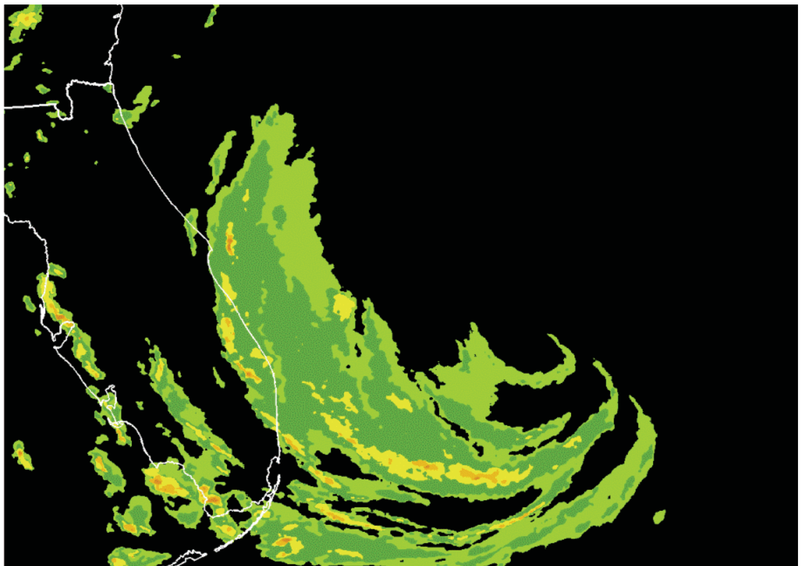
www.simple-talk.com

Motivation

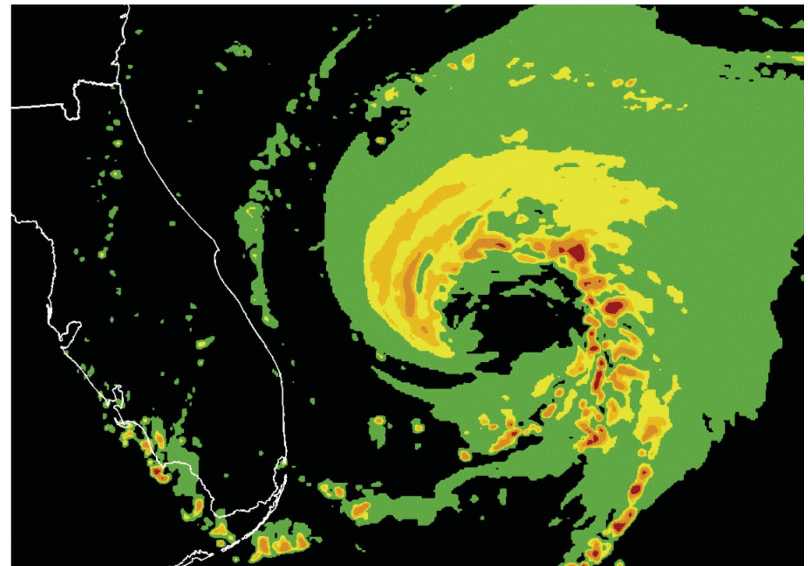
- Prediction, Estimation, Inference, Decision and Control must contend with Uncertainty.
- Uncertainty quantification is a rigorous subject in its own right with much recent progress.
- Widely relevant, including Climate, Weather, Environmental Sampling, Hazard Mitigation, Geophysics, Oceans, Geochemistry, Planetary Science, Economics, Engineering, among others.
- “Quantifying Uncertainty” is overdue as a systematic course.

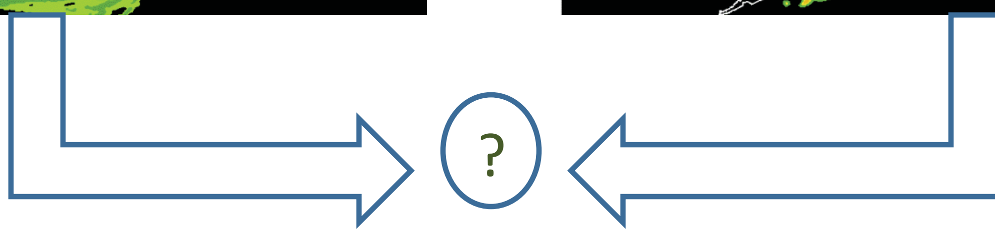
Data, Models and Inference

Observations

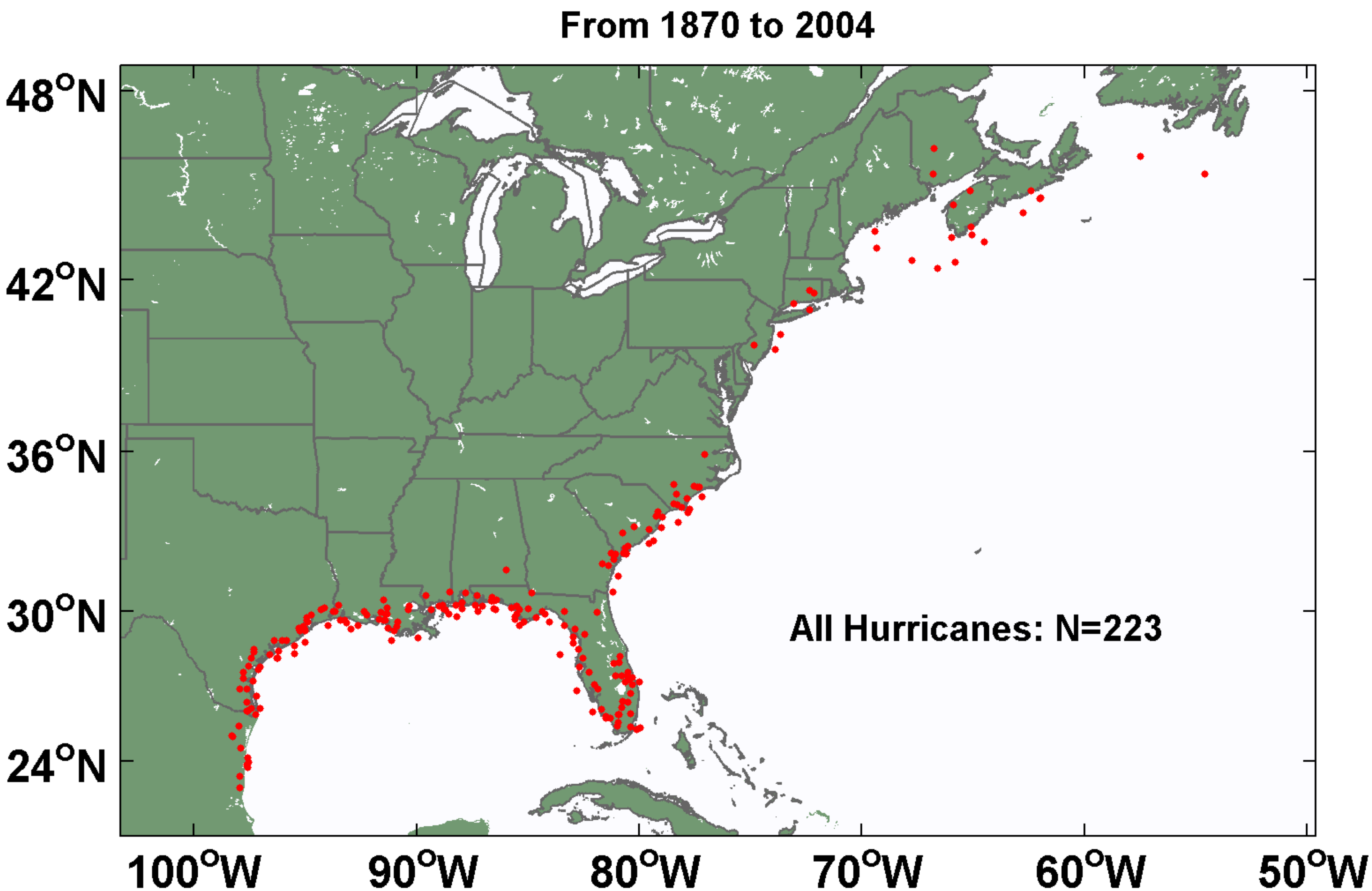


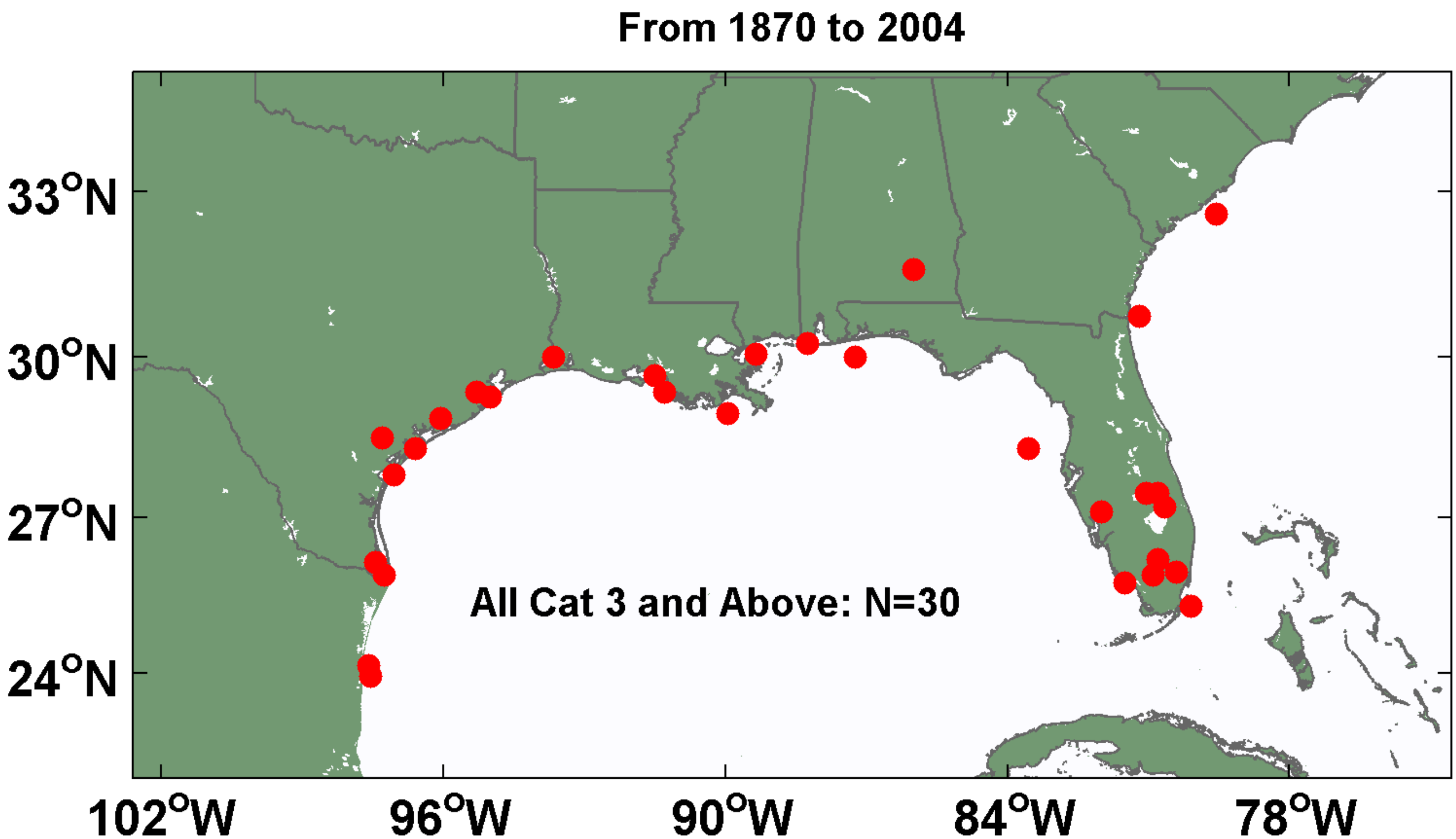
Numerical Model Forecast





How can we combine Data and Models to produce an estimate of nature better than either source alone?


From 1870 to 2004
48°N
42°N
36°N
30°N
24°N
100°W
90°W
80°W
70°W
60°W
50°W
All Hurricanes: N=223


From 1870 to 2004
33°N
30°N
27°N
24°N
102°W
96°W
90°W
84°W
78°W
All Cat 3 and Above: N=30

Objectives

- To understand the methods by which we can **represent**, **propagate** and **estimate** uncertainty.

- To explore "interdisciplinary" application.

- To build a community around Uncertainty Quantification.

Course Series

- Useful preparation for this course:
  - Some Probability & Statistics
  - Some Linear Algebra
  - Some Dynamics/Physics.

Structure

- Meets **Every day** @ 3:30pm.
- Course web: **LOCALLY ARRANGED**

Expectations

- Attend class.
- Read the assigned papers/readings
- Do the take home questions.

Some Projects

- Wind Ensemble Active Sensing
- Modeling response of GEOS-Chem simulations to model parameter uncertainties
- Towards the understating of tropospheric mercury (Hg) decline
- Emissions estimation
- Classifying Cloud Particles
- Quantifying Uncertainty in simple paleoclimate models
- Robust non-parametric fits to heteroskedastic data
- MCMC in search for exoplanets
- MCMC Source Localization
- Transit Estimation
- Estimation of Chemical Species

Contact me if you are interested in doing a project

A bit of History

- Science of Uncertainty > 300 years old
  - Jacob Bernoulli: Father of Uncertainty Quantification.
  - Abraham DeMoivre
  - Francis Galton

From H. Wainer, Picturing the Uncertain World

Uncertainty Gymnastics

- Classic example:
  - A die is rolled 6 times and you get, say: 1,5,4,4,3,2
  - I ask "what number comes next?" to my
    - 5 year old son:
      - Response: 45-one-hundred-million
    - 93 year old grand mother:
      - Response: How do you know it's a fair die?
    - My mathematician friend:
      - Response: Uniform in 1-6 and iid assuming fair.
    - An unnamed statistical learning colleague
      - P(X={1,2,3,5}) = 1/6, P(X=4) = 1/3 and draws from P(X).
    - Second bright bulb decides there's something about the sequence that's important
      - P(X[n] | X[n-1])

The Types of Uncertainty

- Epistemic Uncertainty:
  - Unknown value of parameter must be *imputed*.
  - The model structure must be estimated.

- Aleatory Variability:
  - Parameter may have several outcomes.

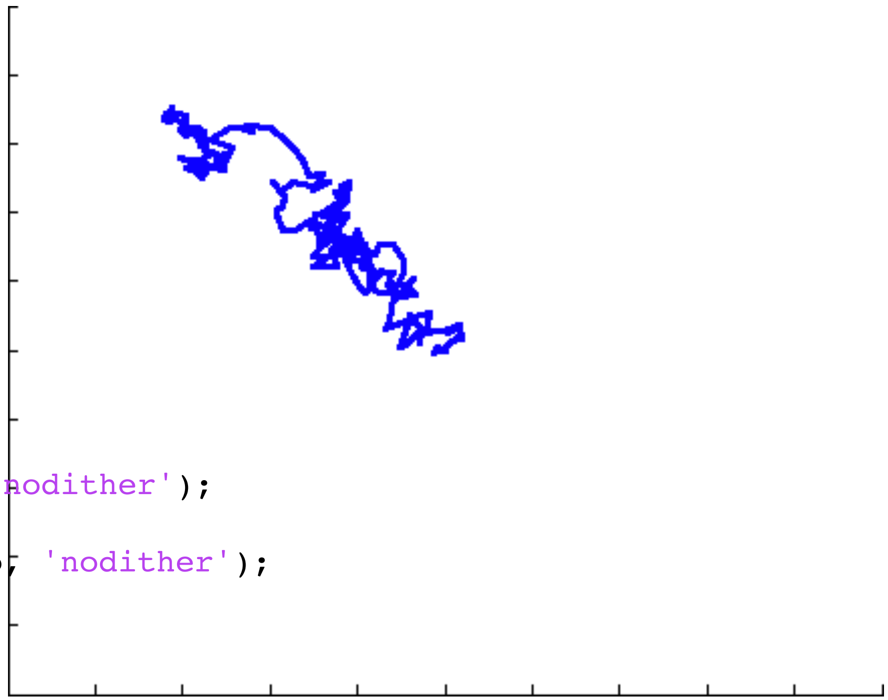
Quantifying Uncertainty

- We quantify uncertainty to convert:
  - Epistemic Uncertainty ➔ Aleatory Variability

- An unknown parameter (that we must impute) is represented by a probability distribution that captures uncertainty of its knowledge.
- This is indistinguishable from an aleatory variability in an inherently probabilistic outcome.

A Process Perspective

- Random Walk – AR(1)

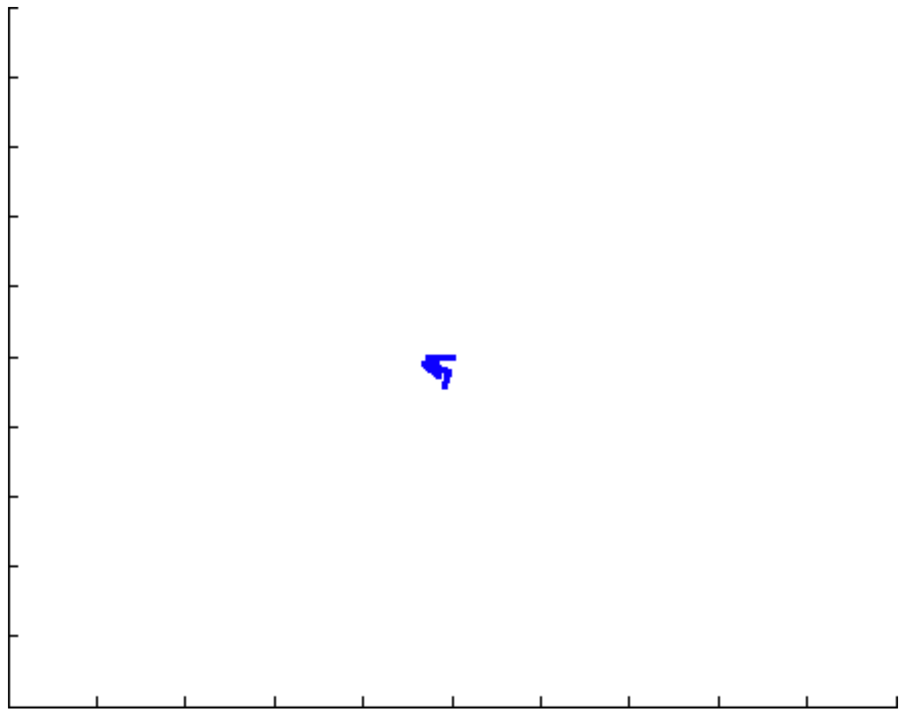


```
clear;
close all;
numFrames = 250;
p = zeros(numFrames,2);
animated(1,1,1,numFrames) =0;
for i = 2:numFrames,
    p(i,:) = p(i-1,:)+randn(1,2)*.25;
    line(p(i-1:i,1),p(i-1:i,2),'LineWidth',2);
    axis([-10 10 -10 10]);
    frame = getframe;
    if i == 2
      [animated, cmap] = rgb2ind(frame.cdata, 256,'nodither');
   else
      animated(:,:,1,i) = rgb2ind(frame.cdata, cmap, 'nodither');
   end
end
filename = 'randwalk.gif';
imwrite(animated, cmap, filename, 'DelayTime', 0.05, ...
   'LoopCount', inf);
web(filename)
```

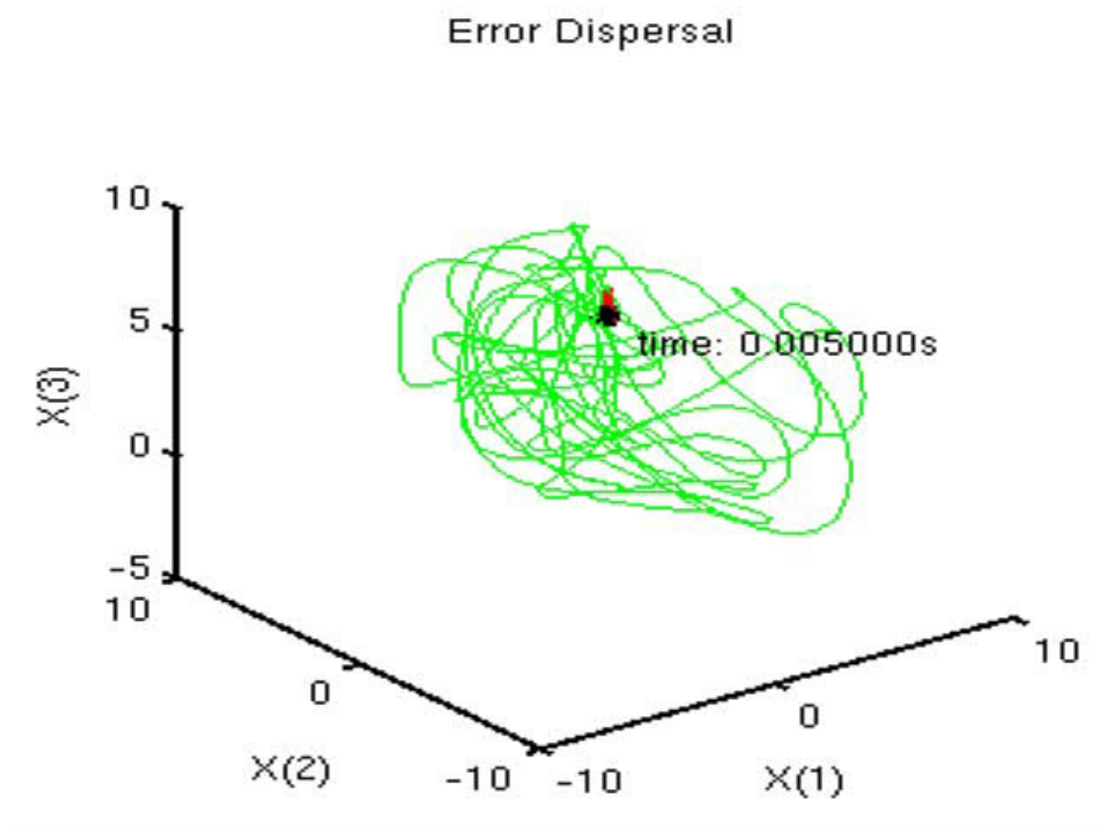
Random Process

- Random Walk with "more memory"

```
clear;
close all;
numFrames = 250;
p = zeros(numFrames,2);
animated(1,1,1,numFrames) =0;
for i = 4:numFrames,
    p(i,:) = .5*p(i-1,:)+.3*p(i-2,:)+.
2*p(i-3,:)+randn(1,2)*.25;
    line(p(i-1:i,1),p(i-1:i,2),'LineWidth',2);
    axis([-10 10 -10 10]);
    frame = getframe;
    if i == 4
      [animated, cmap] = rgb2ind(frame.cdata, 256,
'nodither');
   else
      animated(:,:,1,i) = rgb2ind(frame.cdata, cmap,
'nodither');
   end
end
filename = 'ar3walk.gif';
imwrite(animated, cmap, filename, 'DelayTime', 0.05, ...
   'LoopCount', inf);
web(filename)
```



Example: Random vs. Chaotic Process

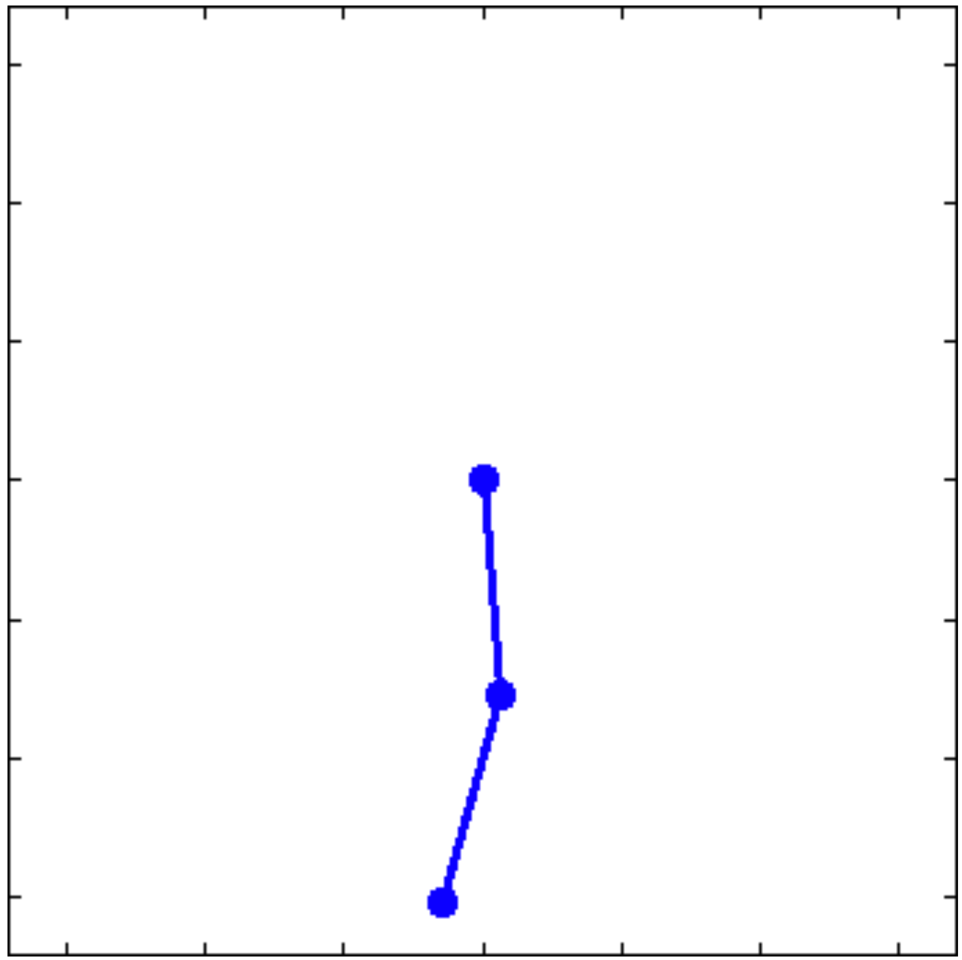


What's the difference?

- The outcome at every step in the random process is random – there is an aleatory variability.

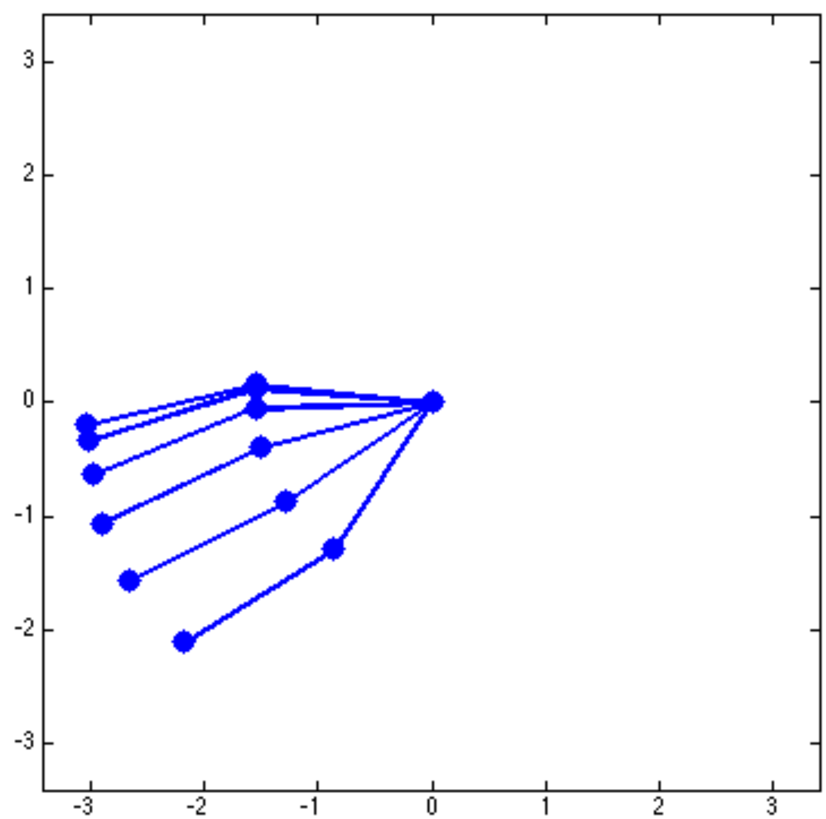
- The outcome at every step in the chaotic process is deterministic – but we don't precisely know the step; an epistemic uncertainty. We represent, propagate and estimate this uncertainty in Quantifying Uncertainty.

Another View

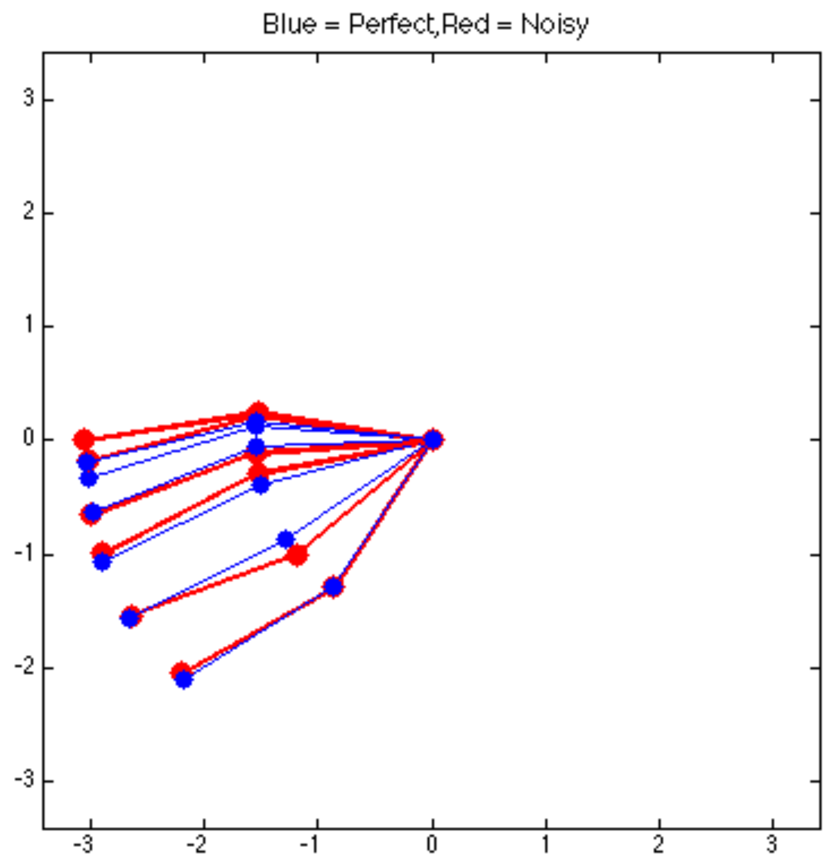


Perfect Reconstruction

Can we answer: what was the initial condition? what is the state at a future time?



Noise & Uncertainty



Systems Perspective

- There is a true system state.
- There is a model of the system.
- If the model is "perfect" and we knew the initial condition of the true system exactly, we can predict forever.
- If the model is perfect but we did know the initial condition,
  - predictions have epistemic uncertainty that must be quantified.

Further

- If the model is imperfect (it is always)
  - Joints have friction whose coefficient we don't know.
  - The rods can flex etc.
- Then
  - Model Calibration: The parameters (e.g. friction) of the model must also be estimated.
  - Model Selection: The model equations themselves might have to be adjusted. Difficult problem for uncertainty quantification.
    - At some point: we don't know what we don't know.

From Systems Perspective

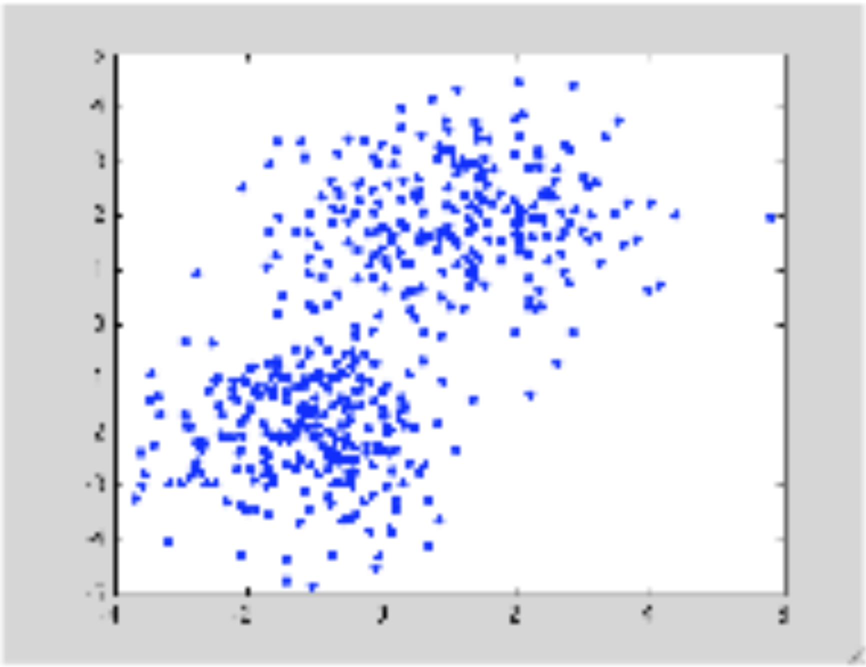
- Joint state and parameter estimation.
  - Complicated in real-world application (e.g. Climate and Weather) because the models are often
    - NONLINEAR
    - HIGH-DIMENSIONAL
  - Uncertainty in state (future, initial, current) and confidence in parameter estimates are to be quantified
    - Can be a complicated (non-Gaussian) function.

Data Analysis Perspective

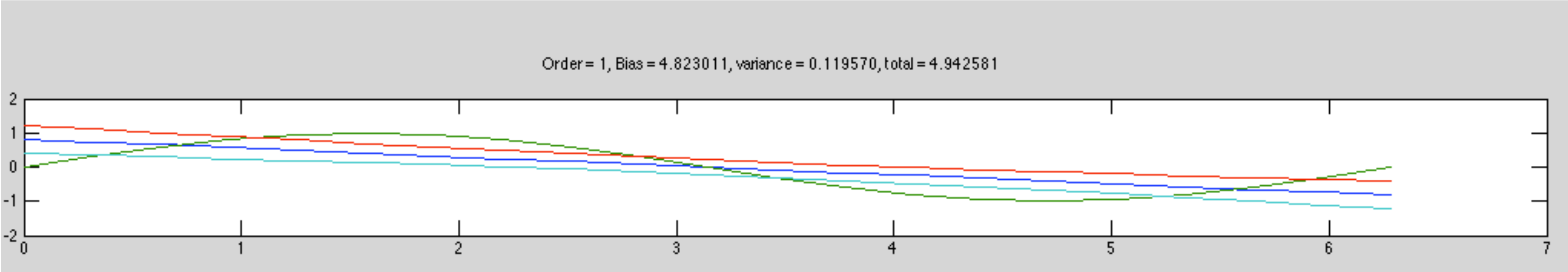
- Quantifying Uncertainty important in data analysis
- For example, consider a few prototypical problems:

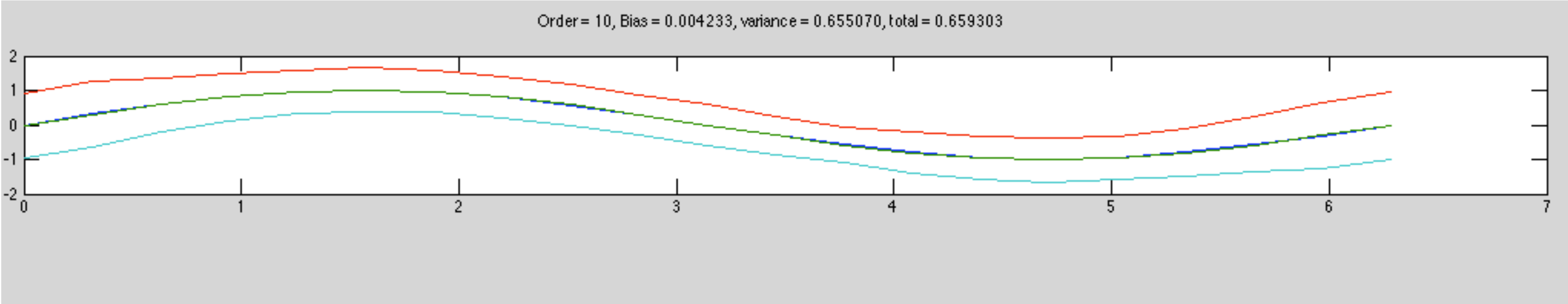
Problem	Estimate	Uncertainty	Model Selection
Density Estimation	Parameters of Probability Density (Mass) Function	A pdf of the parameters represents uncertainty	Choice of pdf, number of parameters
Regression	Parameters of regression function	Confidence intervals, pdf of parameters	Degrees of freedom
Missing Data	Impute missing values	A distribution	Rank reduction

Example



Model Selection and the Bias-Variance Dilemma





A Bayesian Perspective

- The distribution of the estimated variable (state, parameter) of interest forms the basis for quantifying uncertainty (reported as error bars, confidence intervals, moments etc.)
- We will study estimation in a Bayesian context:
  - Generically,
    - P(X|Y) P(Y) = P(Y|X) P(X)

Bayesian Formulations

$$P(X_t|Y_1 \ldots Y_t) \propto P(Y_t|X_t) \sum_{X_{t-1}} P(X_t|X_{t-1}) P(X_{t-1}|Y_1 \ldots Y_{t-1})$$

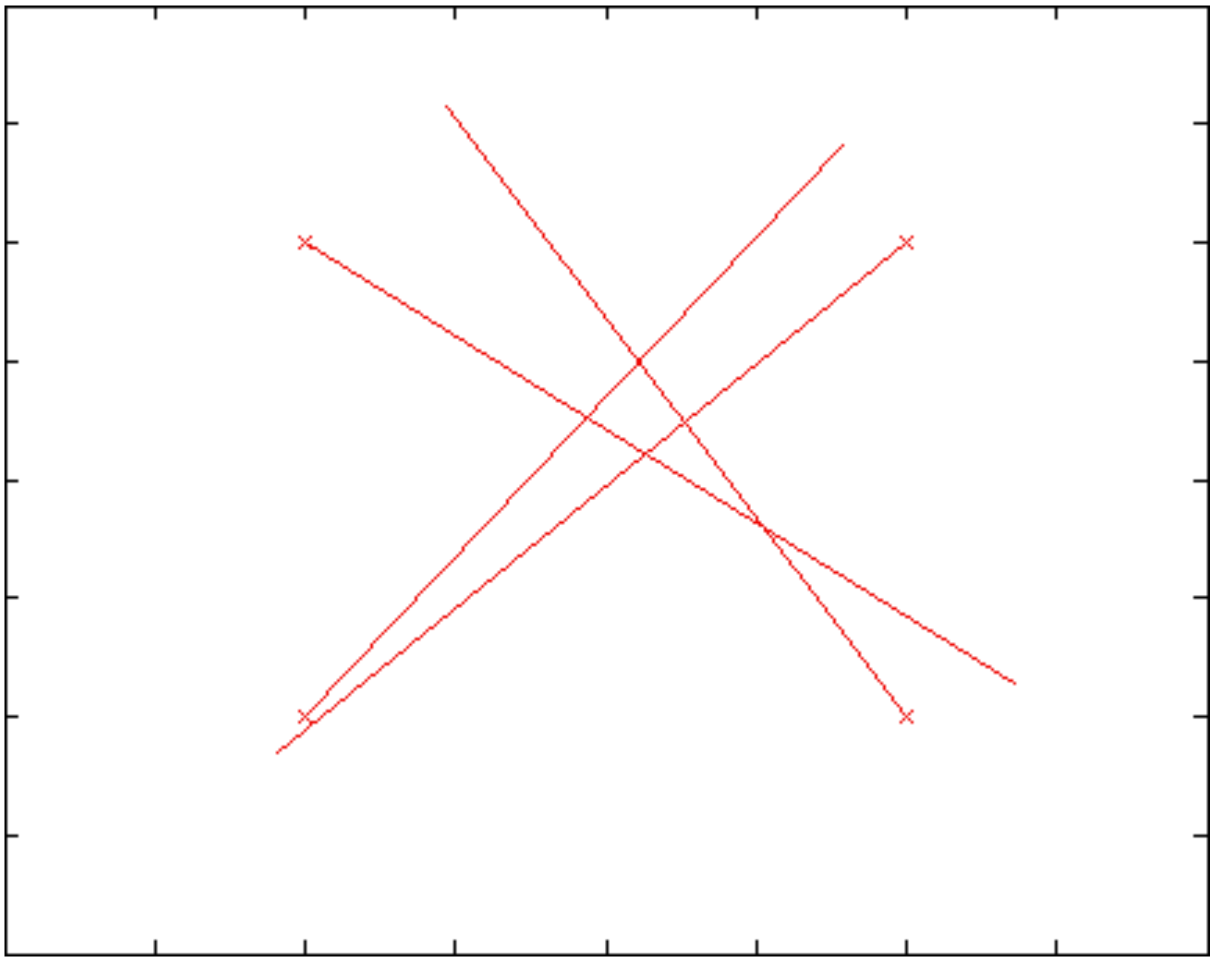
$$P(\alpha|Y) \propto P(Y|\alpha)P(\alpha)$$

$$P(\alpha|Y) \propto P(Y|\alpha) \sum_{\beta} P(\alpha|\beta)P(\beta)$$

Estimation Procedures

- Optimization
- Expectation Maximization
- Sampling
  - Markov Chain Monte Carlo

MCMC in Inverse Problems



Quantifying Uncertainty

- The objective of this course is to introduce methods that can be used to quantify uncertainty in a variety of estimation problems, but particularly those connected with physical sciences.

- Uncertainty is almost always represented as a probability density function; through samples, parameters or kernels. Our objective is to represent, propagate and estimate this density.

Structure

- Representation of Uncertainty
  - Samples, parametric forms, mixtures, kernels
- Propagation of Uncertainty
  - Chain Rule, Fokker Plank
    - Sampling approaches
  - POD/ROM, Multiscale, Polynomial Chaos
    - Model Reduction approaches

Structure Continued

- Uncertainty Estimation
  - Bayesian Estimation (MAP/MLE)
    - Variational Inference
      - Parametric Density Estimation
      - Two-point Boundary Value Problems
      - Ensemble Kalman Filter and Smoother
    - Expectation Maximization
      - Mixture Density Estimation
      - Imputation of Missing Data
    - Sampling
      - Particle Filtering
      - Markov-Chain Monte-Carlo

Content we hope to cover

① Introduction
② In a Linear Gaussian world
③ Two-Point Boundary Value Problems
④ Ensemble Kalman Filter and Smoother
⑤ Markov Chain Monte Carlo
⑥ Applications of MCMC
⑦ Particle Filter
⑧ Dimensionality Reduction
⑨ Density Estimation
⑩ Model Selection Criteria

Not Covered in this class

① Model Reduction
② Polynomial Chaos
③ Nonlinear Dimensionality Reduction
④ Regression Machines
⑤ Clustering and Classification
⑥ Markov Processes
⑦ Graphical Models
⑧ Entropy and Information Forms
⑨ Compressive Sensing

12.S990 Quantifying Uncertainty
Fall 2012