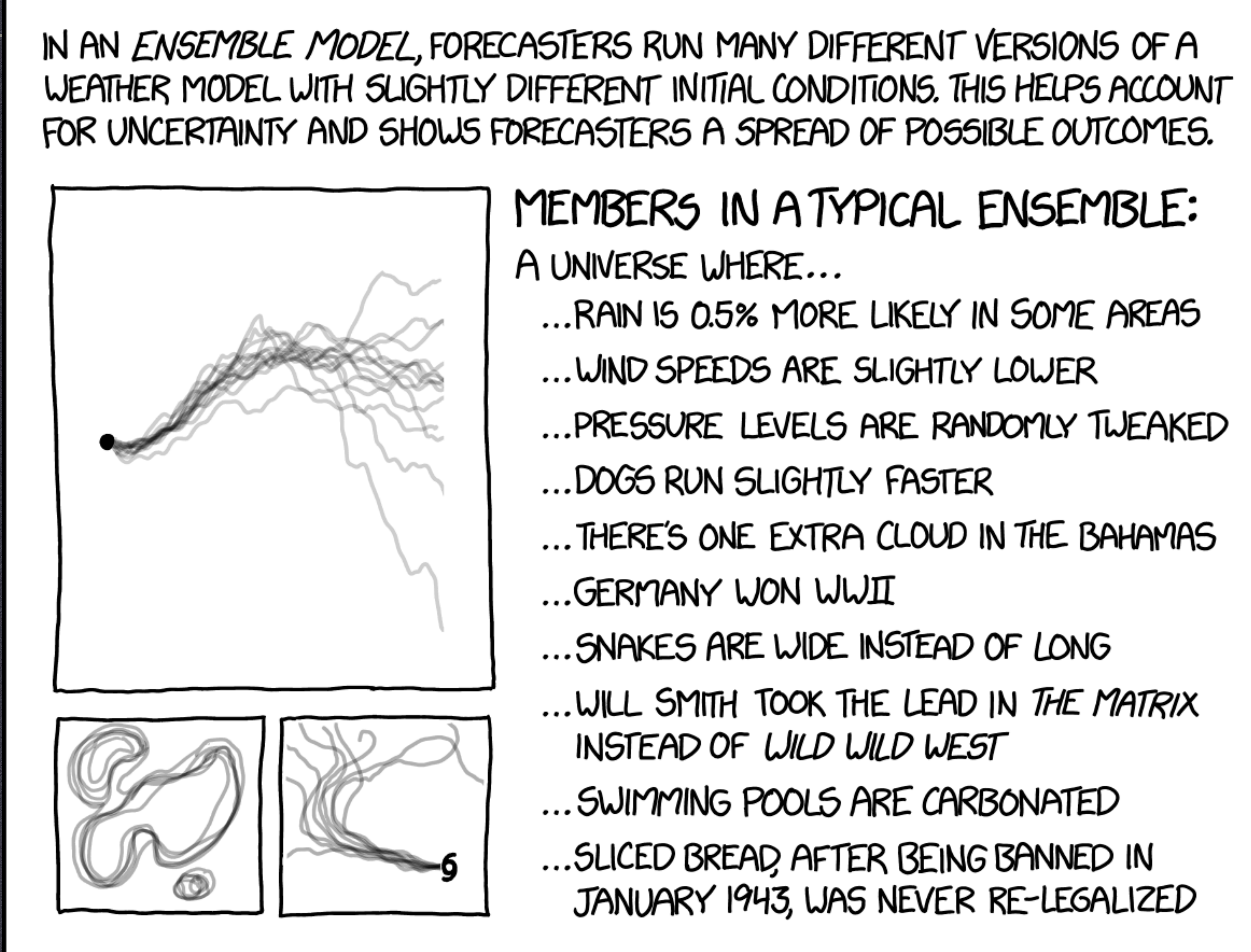


LESSON 6

PROPAGATING, ANALYZING, AND REDUCING UNCERTAINTY

Concepts

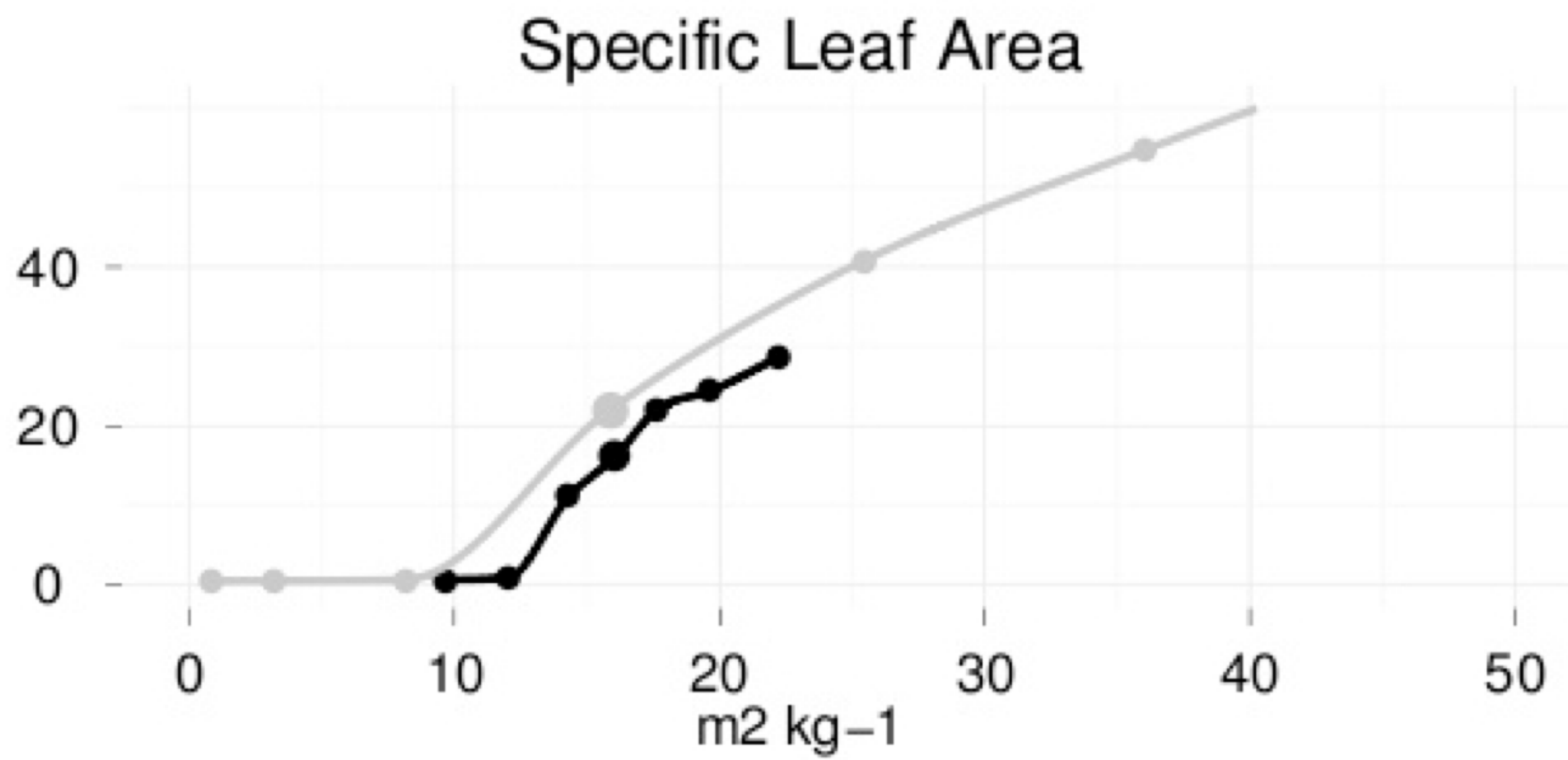
- Sensitivity Analysis
  How does a change in X translate into a change in Y?
- Uncertainty Propagation
  How do we forecast Y with uncertainty?
  How does the uncertainty in X affect the uncertainty in Y?
- Uncertainty Analysis
  which sources of uncertainty are most important?
- Optimal Design
  How do we best reduce the uncertainty in our forecast?

Sensitivity Methods

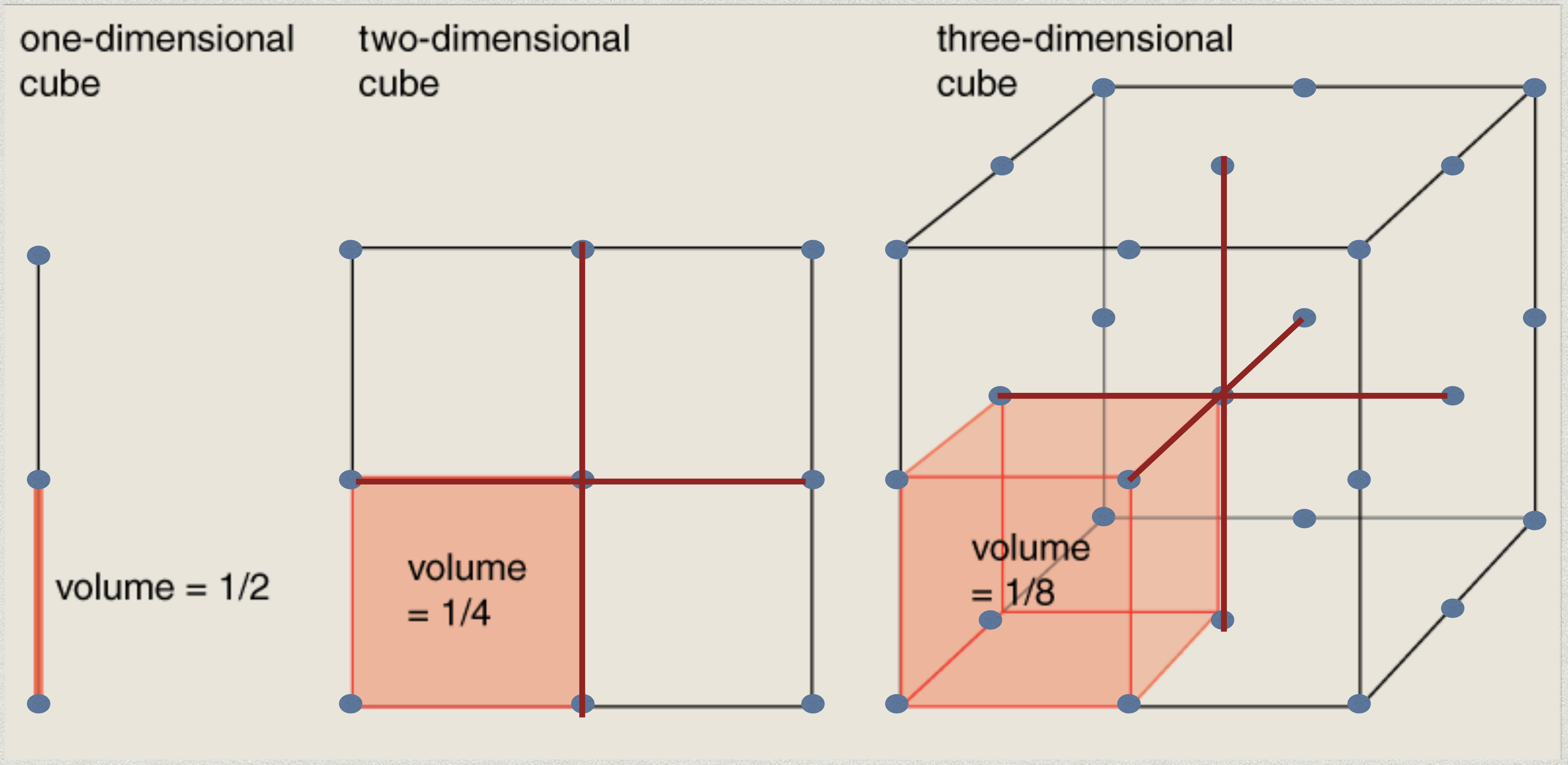
- Local
  - Analytical: $df/d\Theta$
  - One-at-a-time perturbations

Saltelli et al. 2008. Global Sensitivity Analysis

Sensitivity Analysis



Global Sensitivity



Curse of Dimensionality

Sensitivity Methods

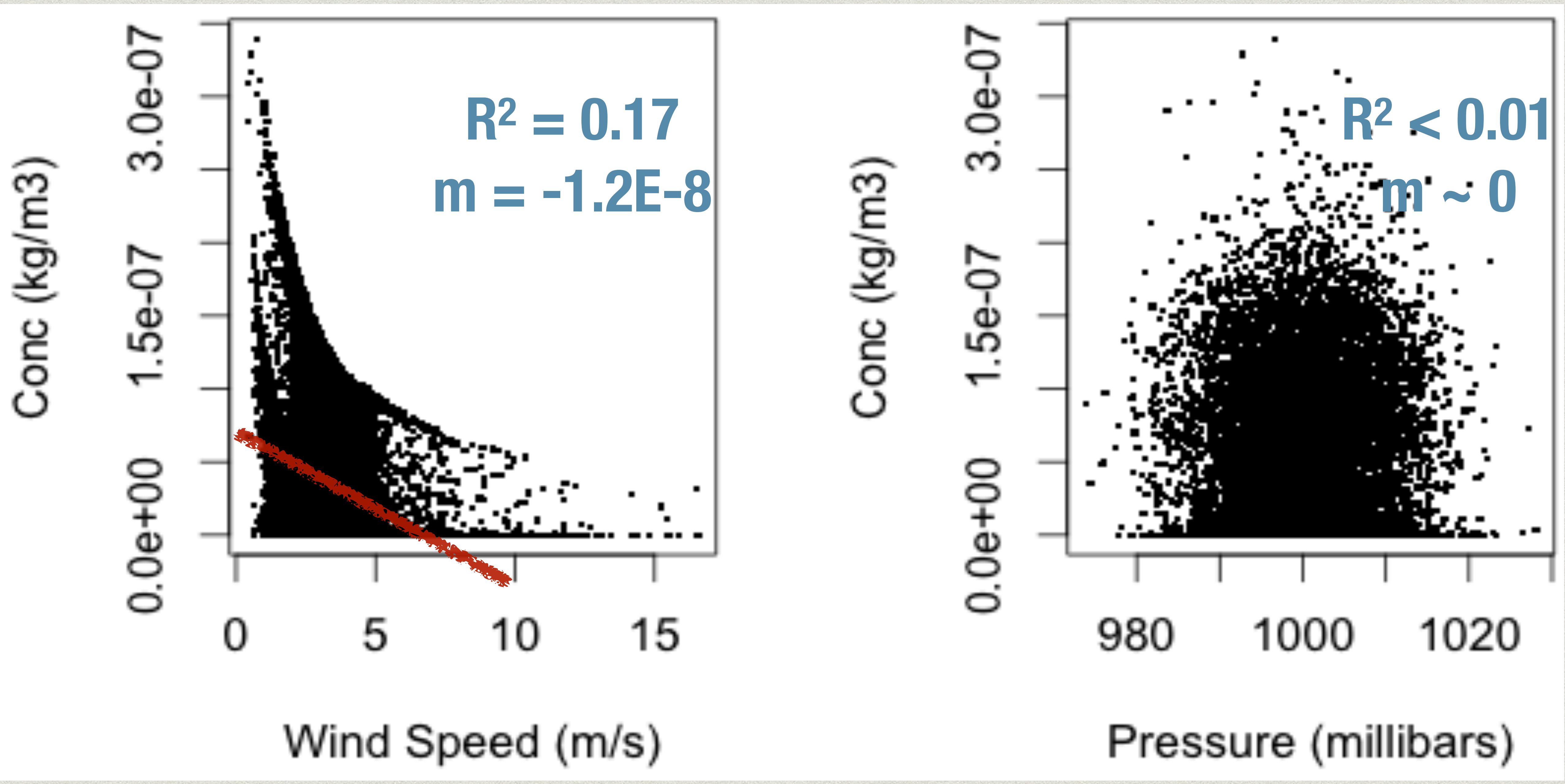
- Local
  - Analytical: df/dθ
  - One-at-a-time perturbations
- Global
  - Monte Carlo
  - Sobol
  - Emulators
  - Elementary Effects
  - Group Sampling

Extensive but Costly

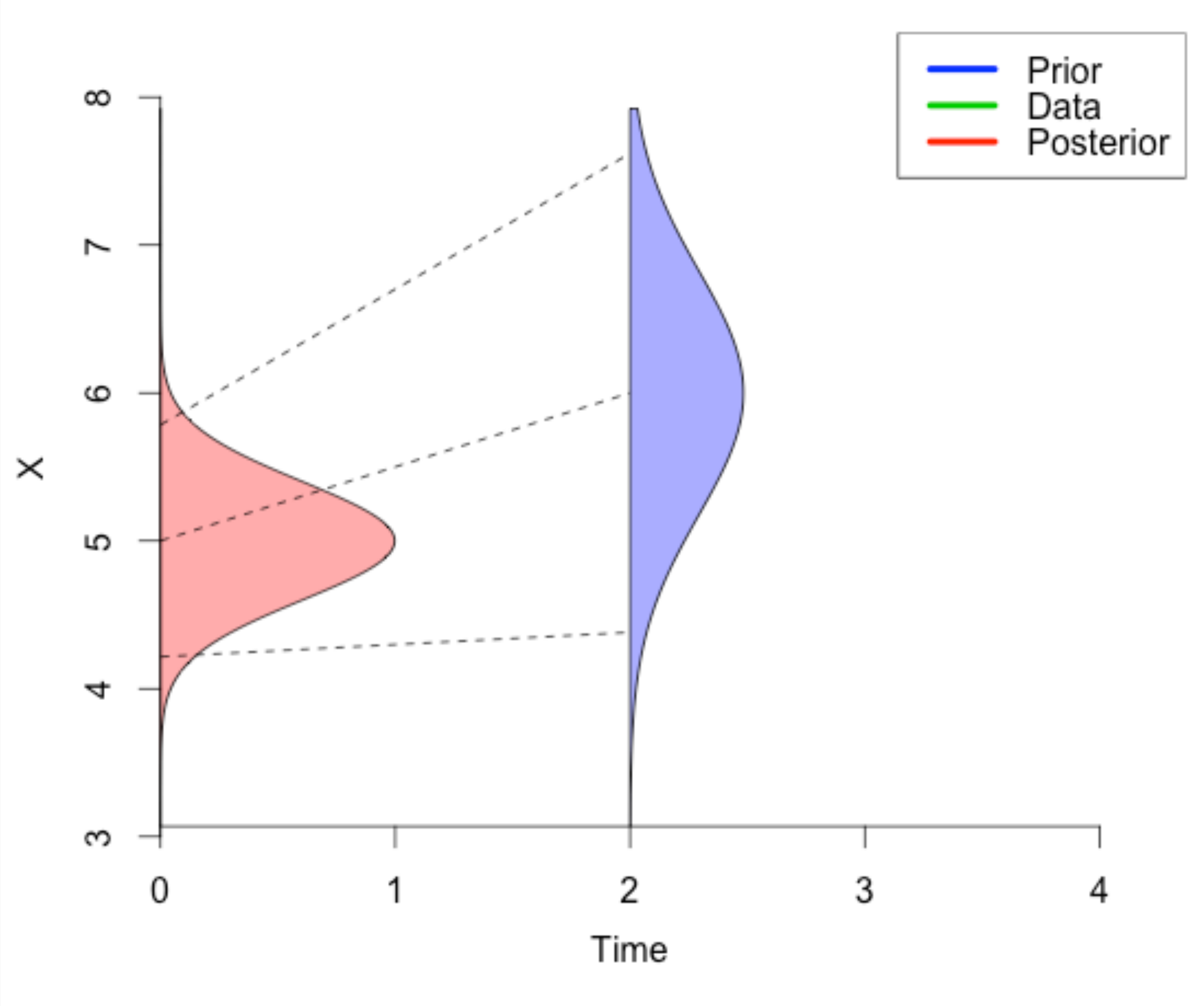
Sparse but Cheap

Saltelli et al. 2008. Global Sensitivity Analysis

Monte Carlo Sensitivity

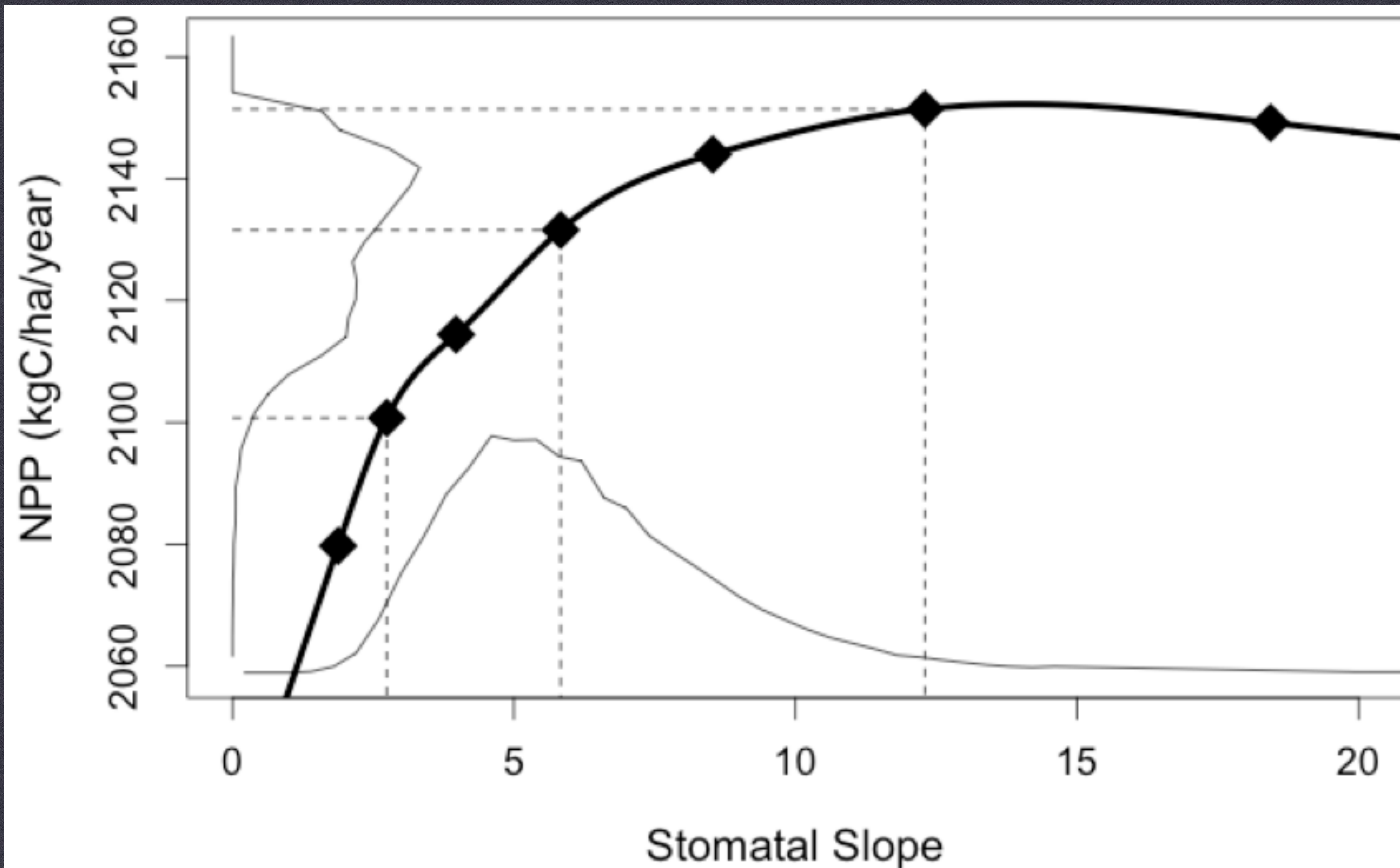


Free if you do MC uncertainty propagation or MCMC



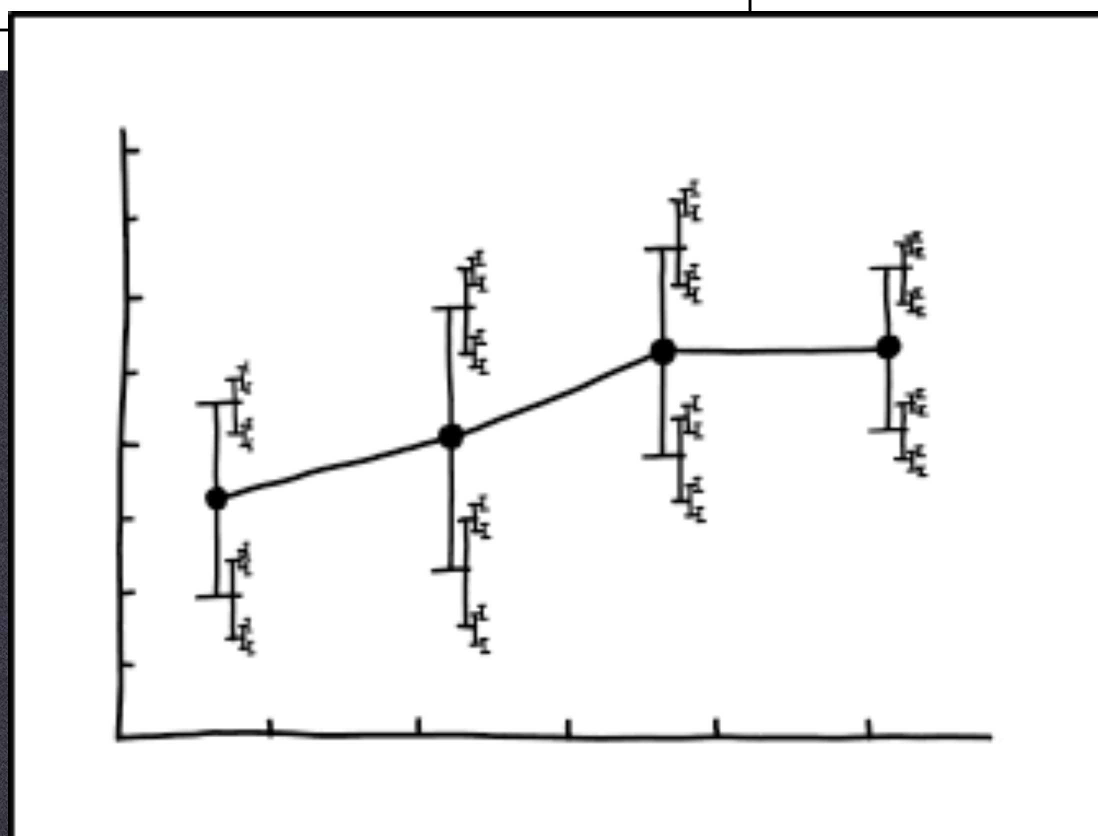
Prior
Data
Posterior
X
8
7
6
5
4
3
0
1
2
3
4
Time

UNCERTAINTY PROPAGATION



UNCERTAINTY PROPAGATION

Approach		Output: Distribution	Output: Moments
	Analytic	Variable Transform	Analytical Moments Taylor Series
	Numeric	Monte Carlo	Ensemble



I DON'T KNOW HOW TO PROPAGATE ERROR CORRECTLY, SO I JUST PUT ERROR BARS ON ALL MY ERROR BARS.

VARIABLE TRANSFORM

$$P_Y[y] = P_\theta[f^{-1}(y)] \cdot \left| \frac{d f^{-1}(y)}{dy} \right|$$

Analytical Moments

$$Var(aX) = a^2 Var(X)$$

$$Var(X+b) = Var(X)$$

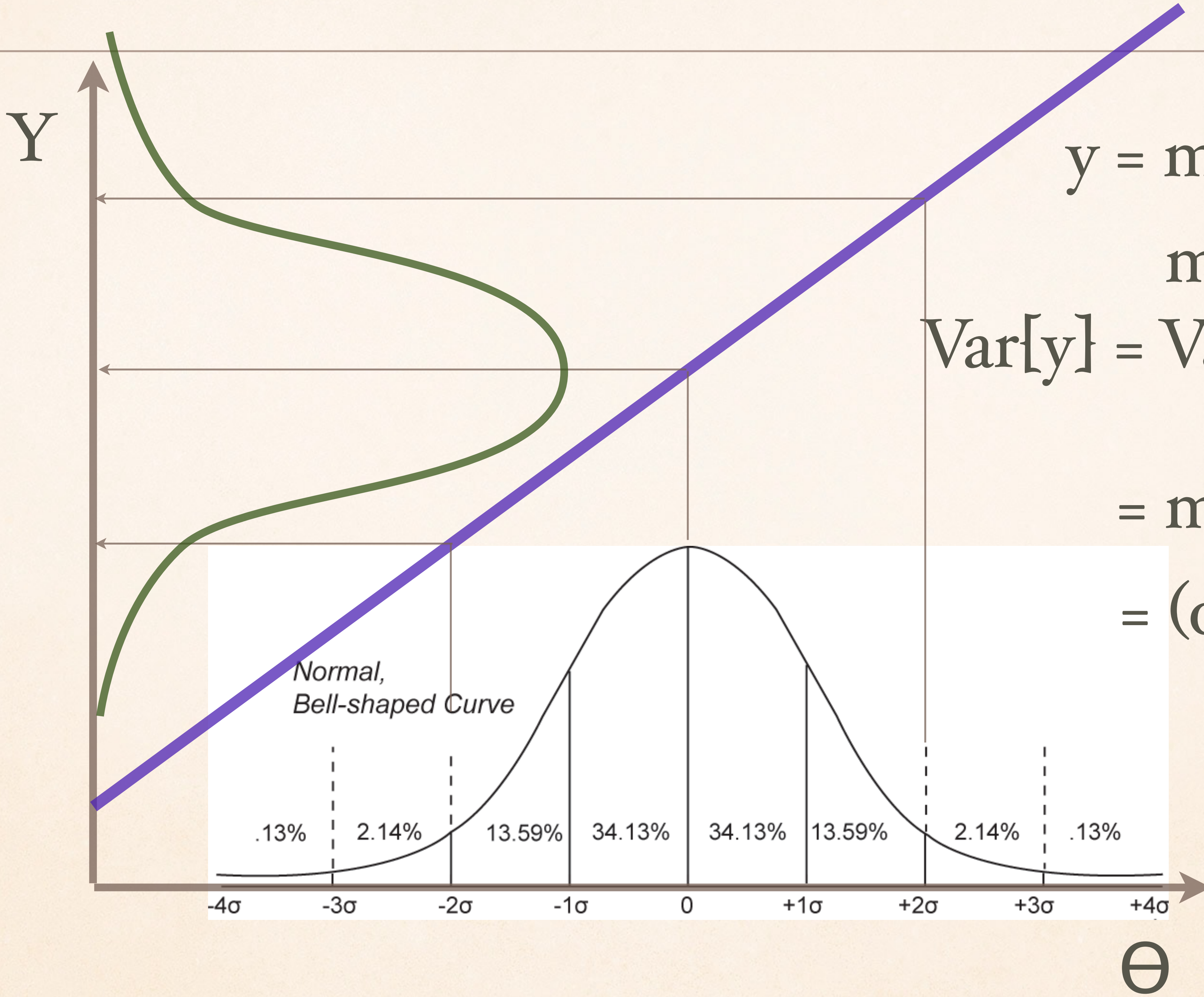
$$Var(X+Y) = Var(X) + Var(Y) + 2\mathrm{Cov}(X,Y)$$

$$Var(aX+bY) = a^2 Var(X) + b^2 Var(Y) + 2\mathrm{abCov}(X,Y)$$

$$Var\left(\sum X\right) = \sum Var(X_i) + 2\sum_{i<j} Cov(X_i, X_j)$$

$$Var(X) = Var(E[X|Y]) + E[Var(X|Y)]$$

REL'N TO SENSITIVITY



$y = m\Theta + b$

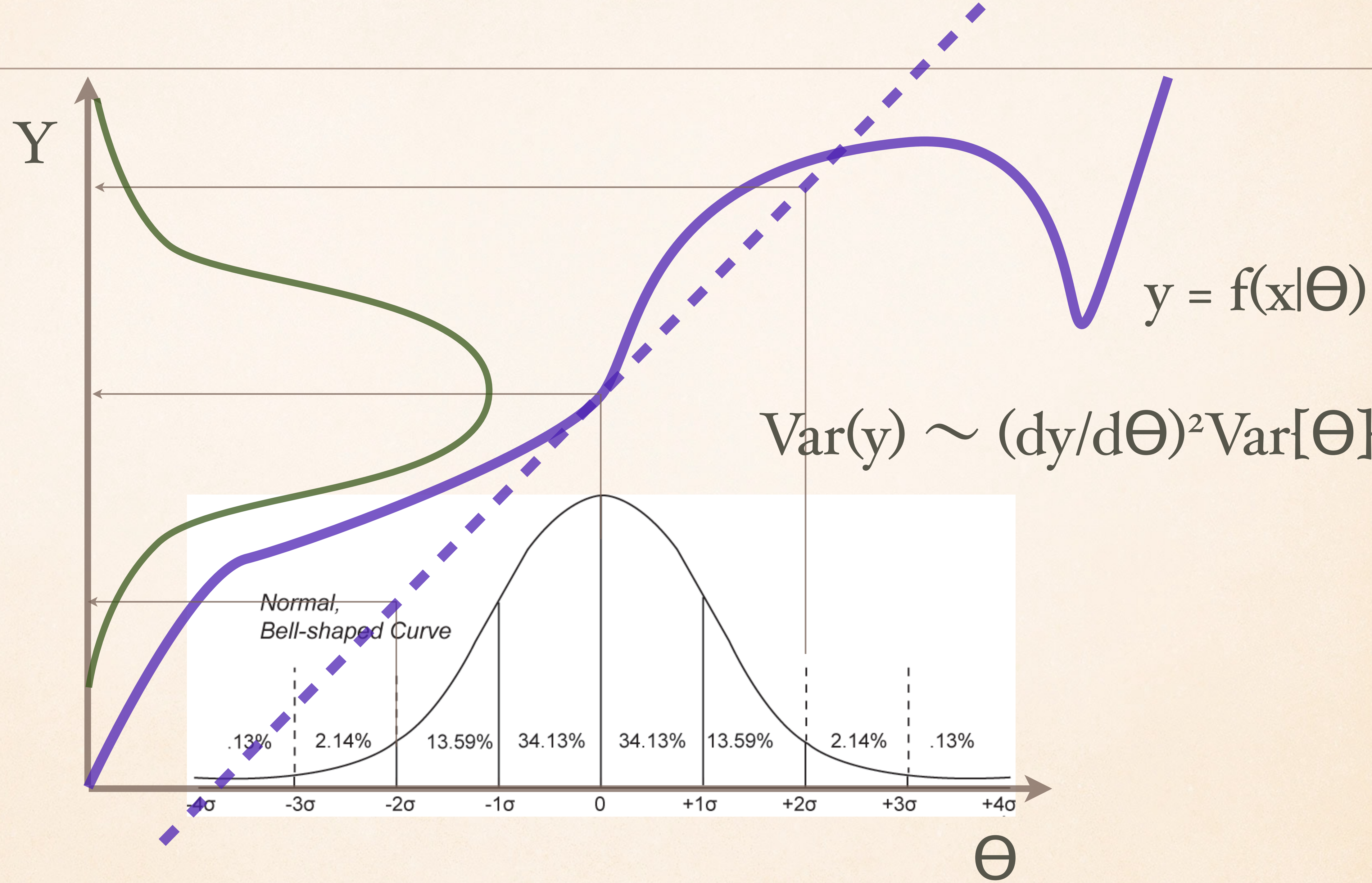
$m = dy/d\Theta$

$\mathrm{Var}\{y\} = \mathrm{Var}[m\Theta + b]$

$= m^2 \mathrm{Var}[\Theta]$

$= (dy/d\Theta)^2 \mathrm{Var}[\Theta]$

TAYLOR SERIES



LINEAR APPROX

$$Var[f(x|\theta)] \approx Var\left[f(x|\bar{\theta}) + \frac{\frac{df}{d\theta}(x|\bar{\theta})}{1!}(\theta - \bar{\theta}) + \ldots \right]$$

$$var[f(x)] \approx \left(\frac{\partial f}{\partial \theta_i} \right)^2 var[\theta]$$

LINEAR APPROX

$$Var[f(x|\theta)] \approx Var\left[f(x|\bar{\theta}) + \frac{\frac{df}{d\theta}(x|\bar{\theta})}{1!}(\theta - \bar{\theta}) + \ldots \right]$$

$$var[f(x)] \approx \sum \left(\frac{\partial f}{\partial \theta_i} \right)^2 var[\theta_i] +$$

$$\sum_{i \neq j} \left(\frac{\partial f}{\partial \theta_i} \right) \left(\frac{\partial f}{\partial \theta_j} \right) cov[\theta_i, \theta_j]$$

$$Y_{t+1} = f(Y_t, X_t|\theta) + \varepsilon$$

$$Var[Y_{t+1}] \approx \underbrace{\left(\frac{df}{dY}\right)^2}_{stability} \underbrace{Var[Y_t]}_{\substack{IC\\uncert}} + \underbrace{\left(\frac{df}{dX}\right)^2}_{\substack{driver\\sens}} \underbrace{Var[X]}_{\substack{driver\\uncert}} + \underbrace{\left(\frac{df}{d\theta}\right)^2}_{\substack{param\\sens}} \underbrace{Var[\theta]}_{\substack{param\\uncert}} + \underbrace{Var[\varepsilon]}_{\substack{process\\error}}$$

COV & SCALING

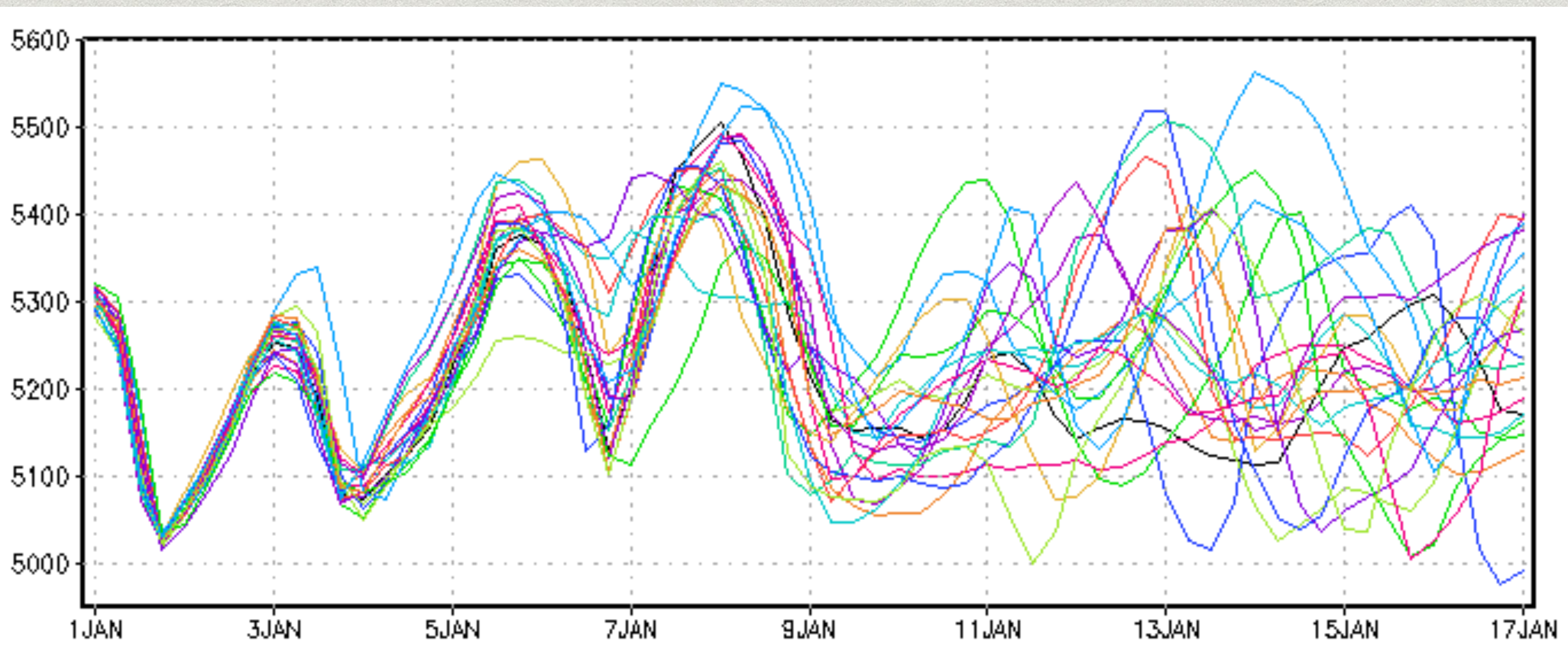
- **Scaling very dependent on spatial and temporal auto- & cross-correlation**

$$\sum\sum\frac{\partial f}{\partial X_i}\frac{\partial f}{\partial X_j}COV[X_i, X_j]$$

UNCERTAINTY PROPAGATION

		Output	
		Distribution	Moments
Approach	Analytic	Variable Transform	Analytical Moments Taylor Series
	Numeric	Monte Carlo	Ensemble

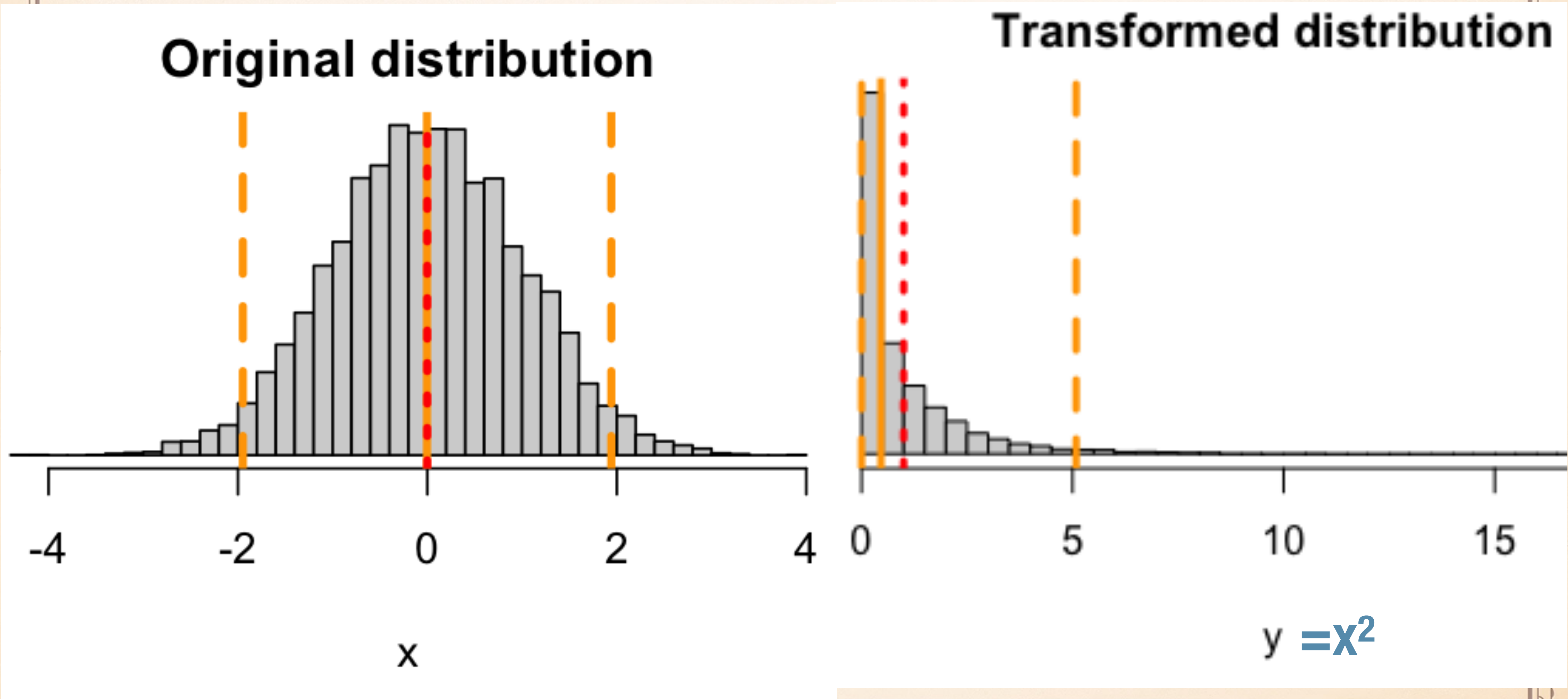
Numerical Approximation



- Monte Carlo Simulation --> Distribution
- Ensemble Analysis --> Moments

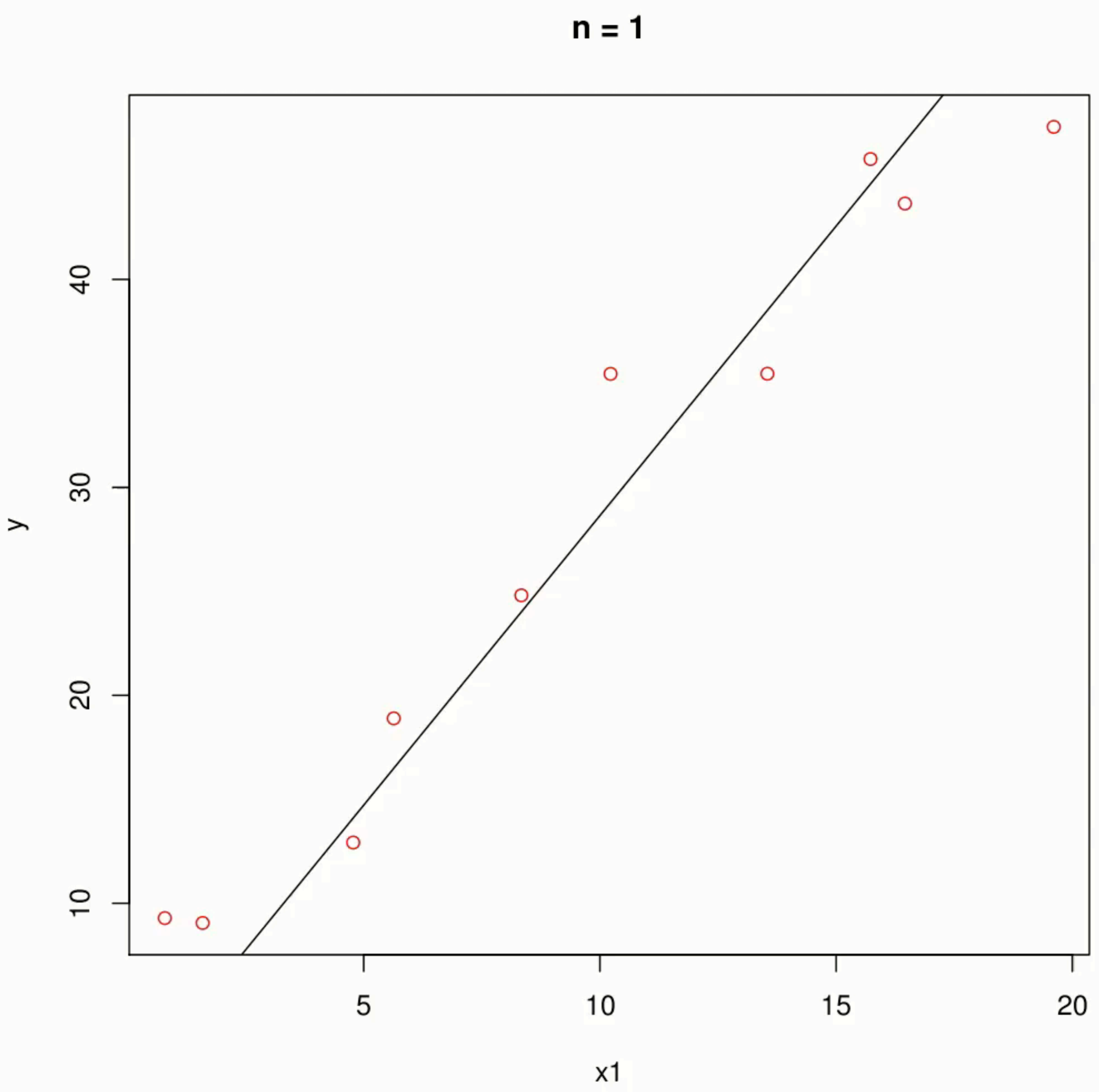
JENSEN'S INEQUALITY

$$f(\bar{x}) \neq \overline{f(x)}$$



MONTE CARLO UNCERTAINTY

- for (i in 1:n)
  - draw random values from distributions
  - run model
  - save results
- summarize distributions



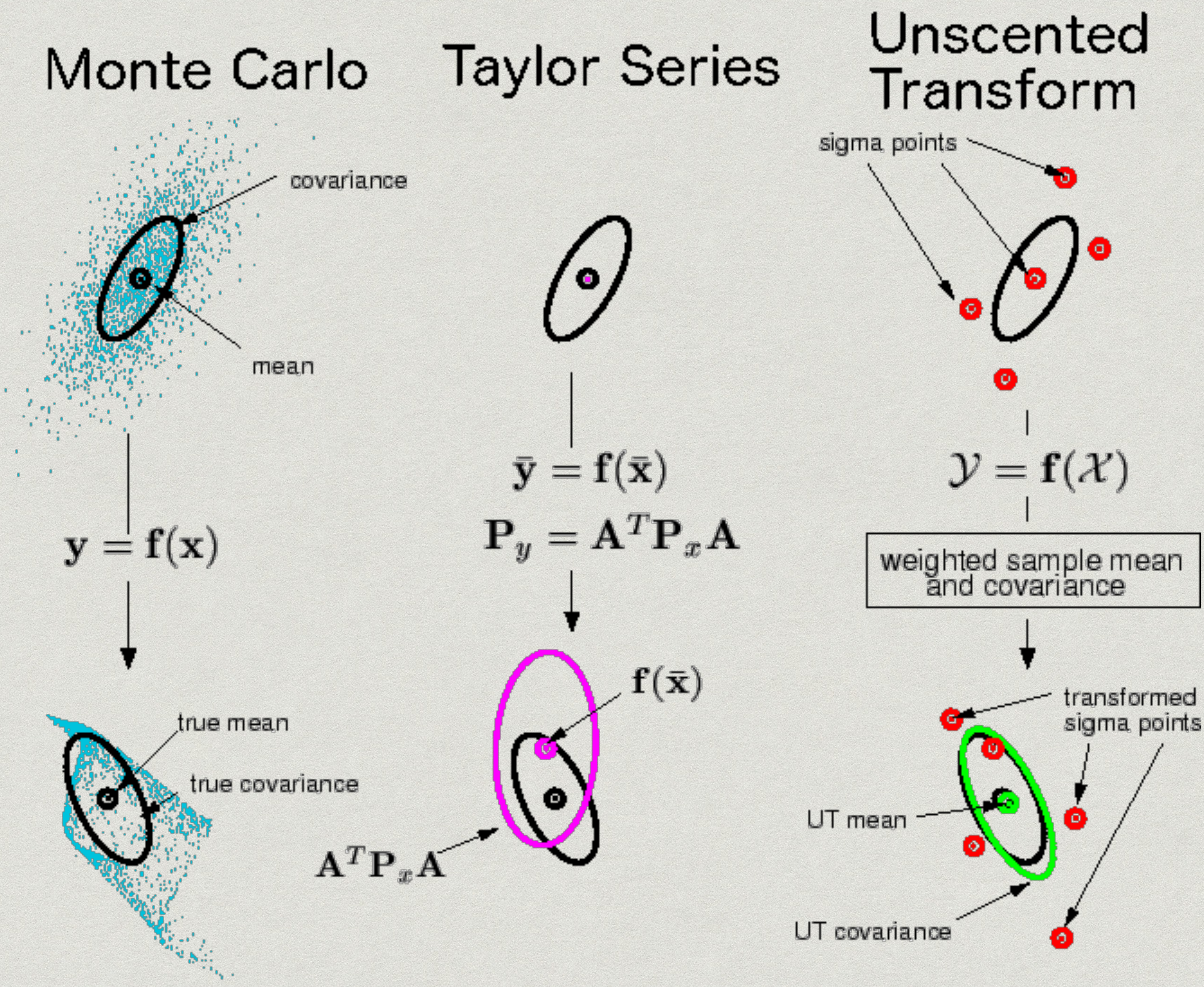
n = 1
y
40
30
20
10
5
10
15
20
x1

ENSEMBLE UNCERTAINTY

- for (i in 1:n)
  - draw random values from distributions
  - run model
  - save results
- **Fit PDF to results**
- **Use PDF for intervals, etc.**

Requires smaller N to estimate moments than to approximate full PDF

Already have this from MCMC!


Monte Carlo
Taylor Series
Unscented Transform
covariance
mean
$\mathbf{y} = \mathbf{f}(\mathbf{x})$
true mean
true covariance
$\bar{\mathbf{y}} = \mathbf{f}(\bar{\mathbf{x}})$
$\mathbf{P}_y = \mathbf{A}^T \mathbf{P}_x \mathbf{A}$
$\mathbf{f}(\bar{\mathbf{x}})$
$\mathbf{A}^T \mathbf{P}_x \mathbf{A}$
sigma points
$\mathcal{Y} = \mathbf{f}(\mathcal{X})$
weighted sample mean and covariance
transformed sigma points
UT mean
UT covariance

UNCERTAINTY PROPAGATION
AND ITERATIVE DATA ASSIMILATION

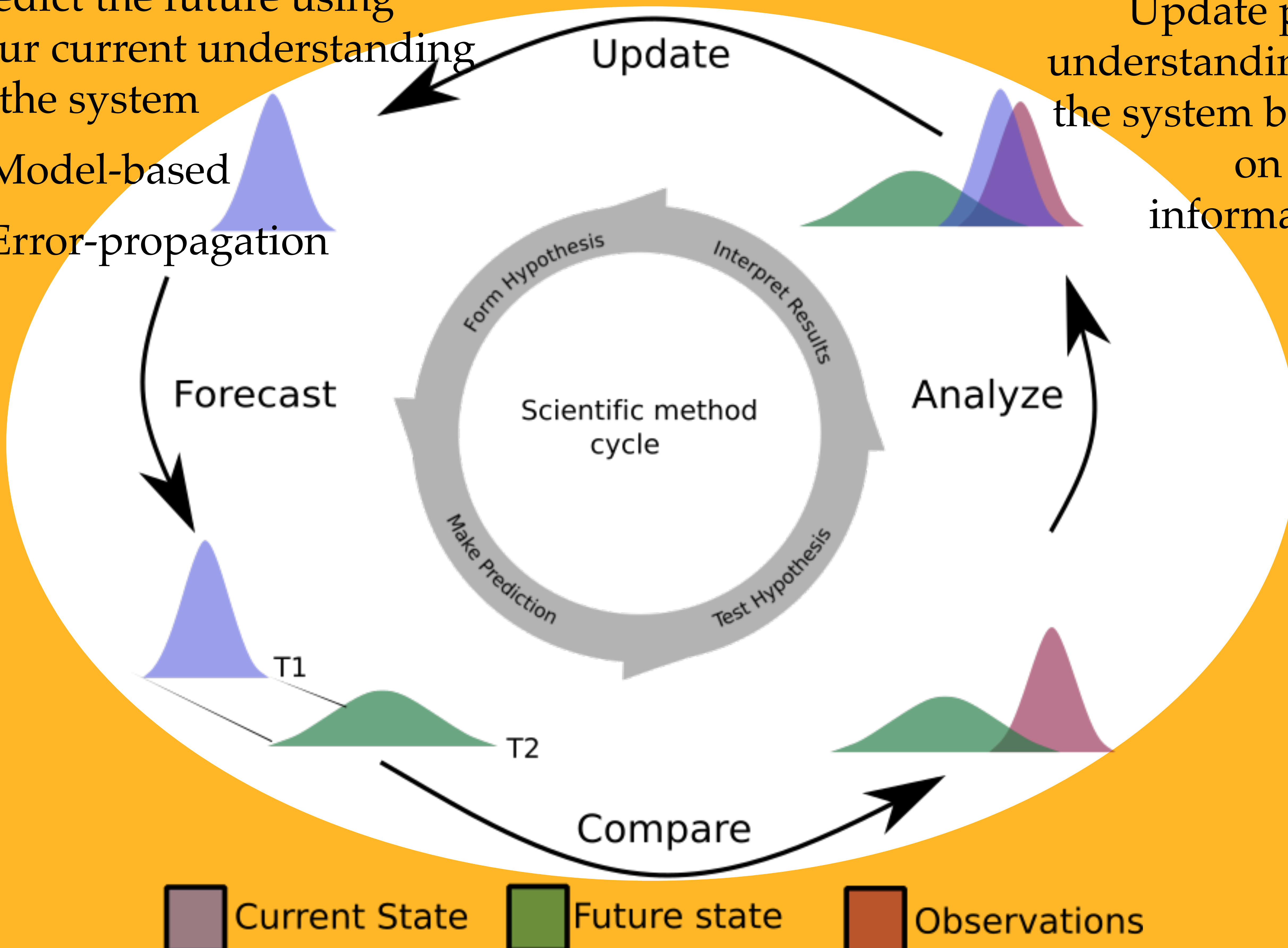
		Output	
		Distribution	Moments
Approach	Analytic	Variable Transform	Analytical Moments KF Taylor Series EKF
	Numeric	Monte Carlo PF	Ensemble EnKF

FORECAST-ANALYSIS CYCLE

Predict the future using your current understanding of the system

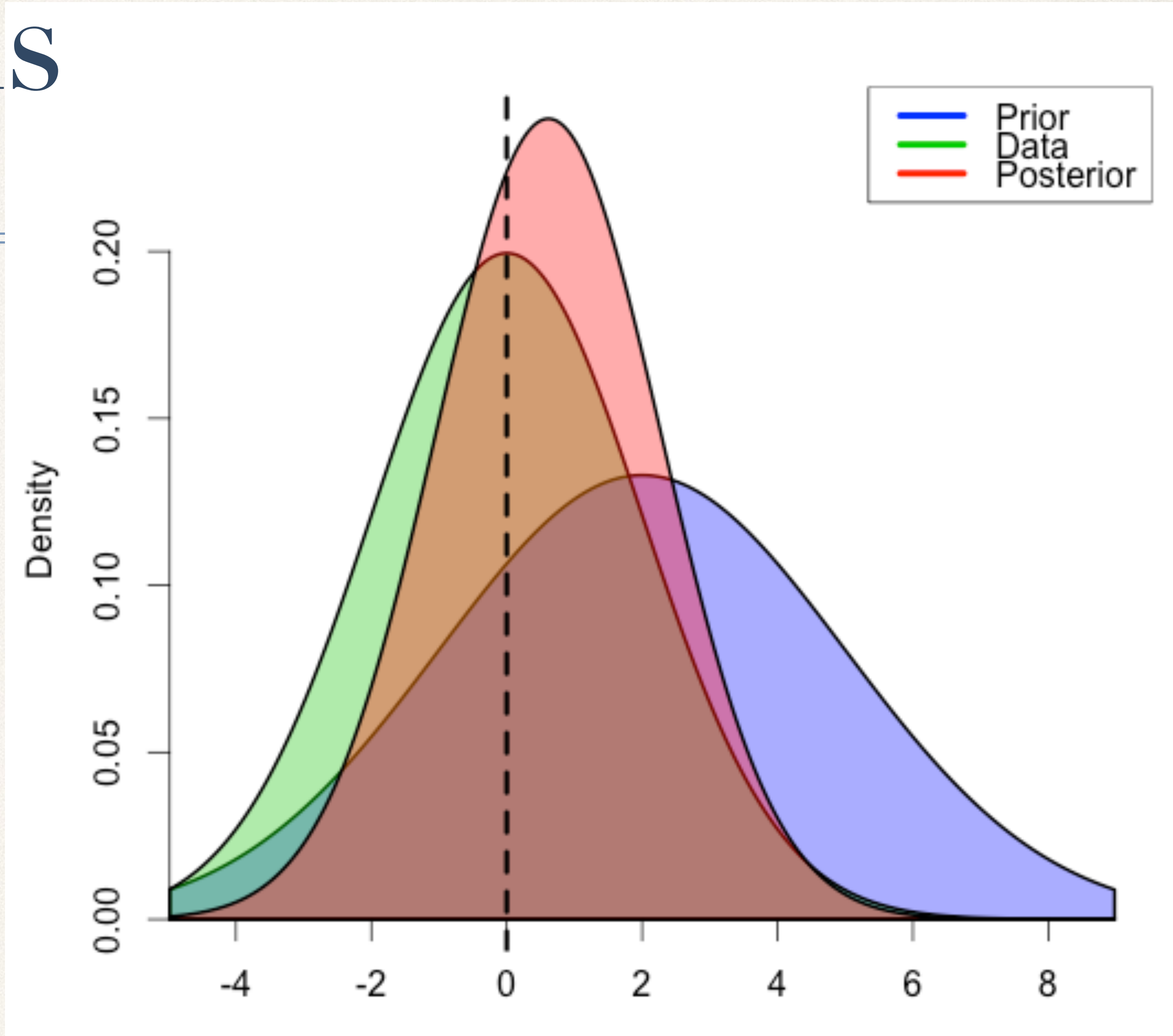
- ➡Model-based
- ➡Error-propagation

Update prior understanding of the system based on new information



Simplest Analysis

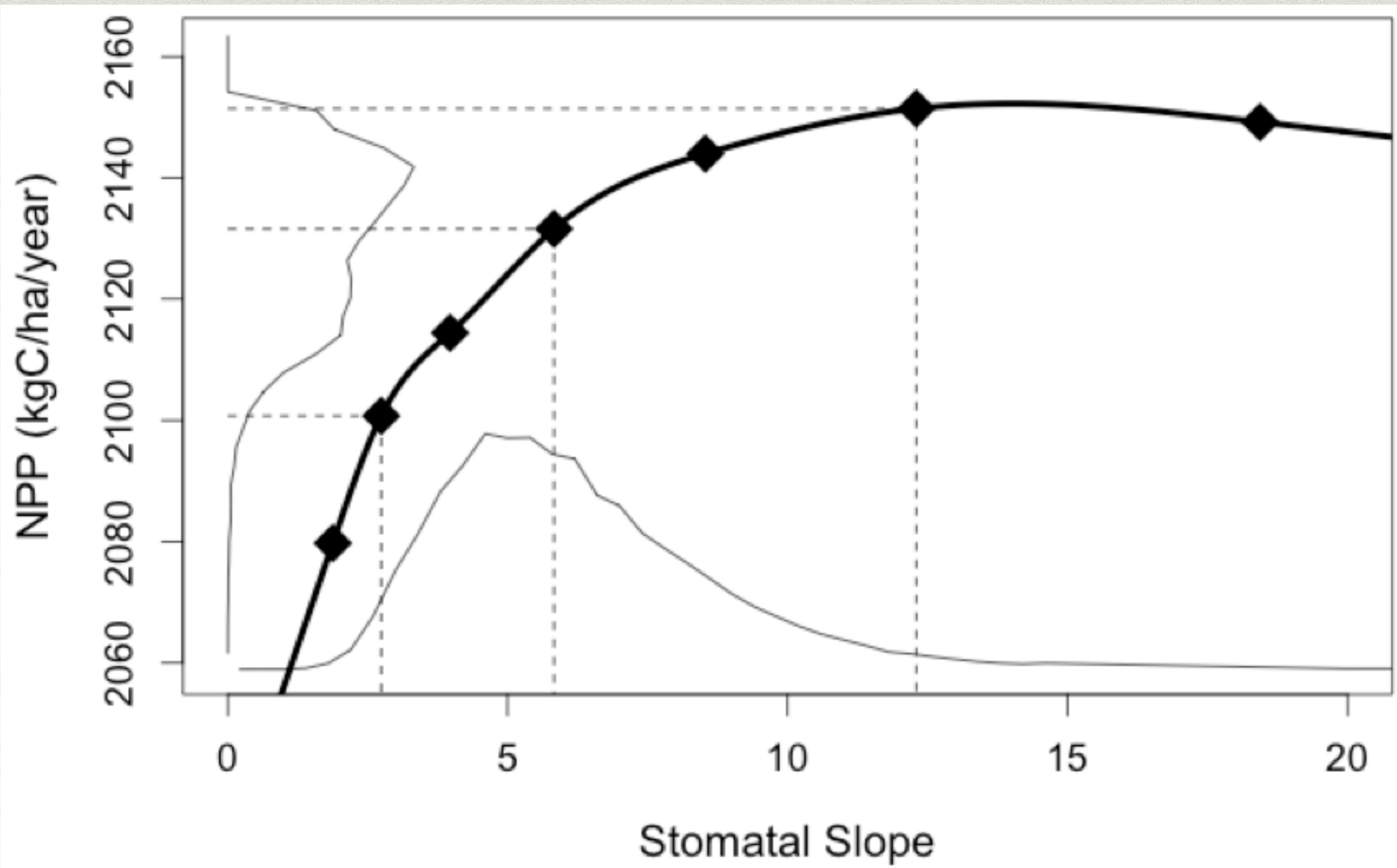
- Forecast:
  Assume $P(X_{t+1}) \sim N(\mu_f, p_f)$
- Observation error:
  Assume $P(Y_{t+1} \mid X_{t+1}) \sim N(X_{t+1}, r)$
  - Likelihood = Data model
- Assume Y, μ_f, p_f and r are known
- $P(X_{t+1} \mid Y_{t+1}) \sim N(\mu_a, p_a)$

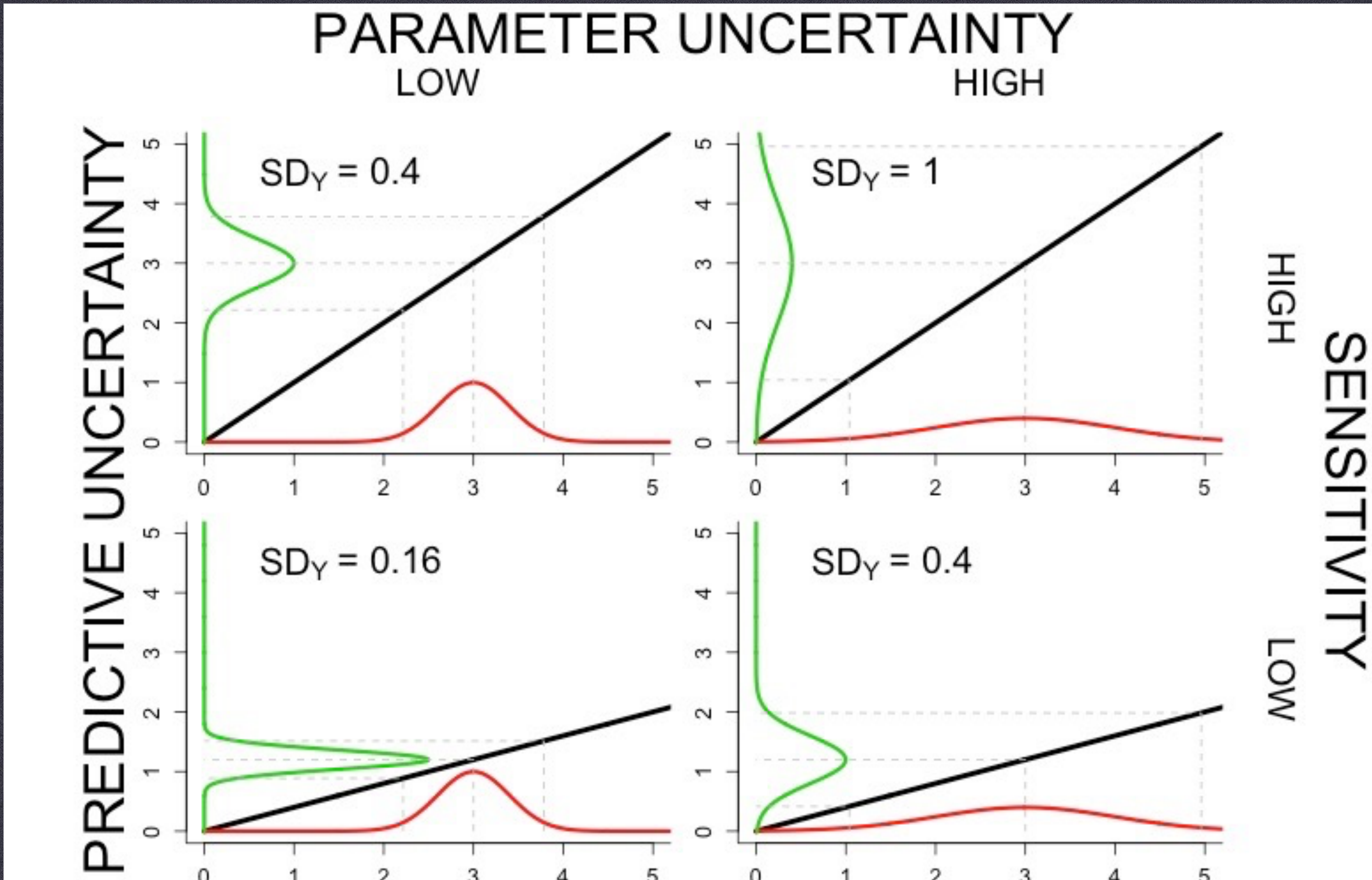


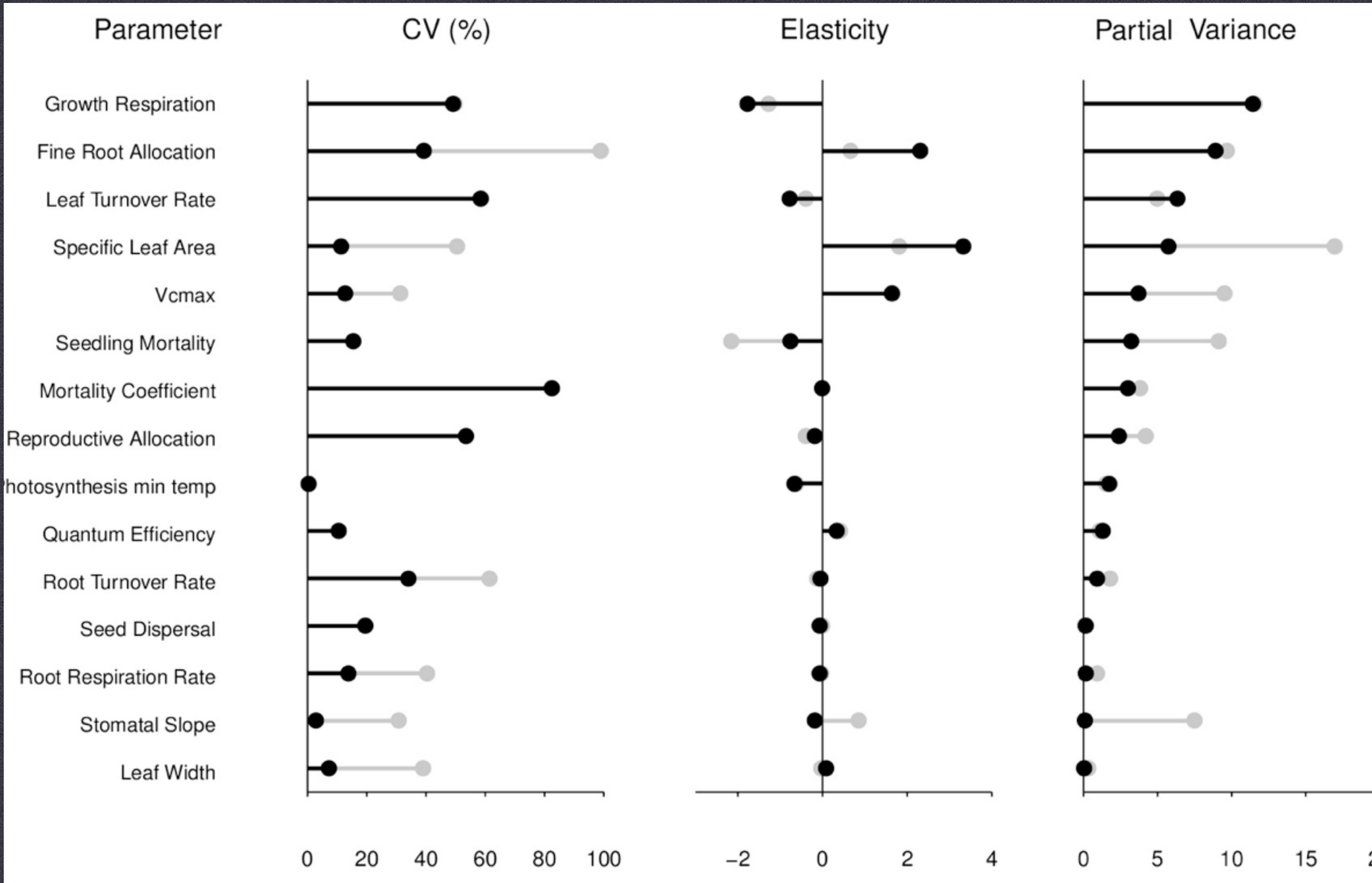
$$\rho = 1/r \qquad \phi = 1/p_f$$

$$X|Y \sim N\left(\frac{\rho}{n\rho+\phi}n\bar{Y} + \frac{\phi}{n\rho+\phi}\mu_f, n\rho+\phi\right)$$

Uncertainty Analysis



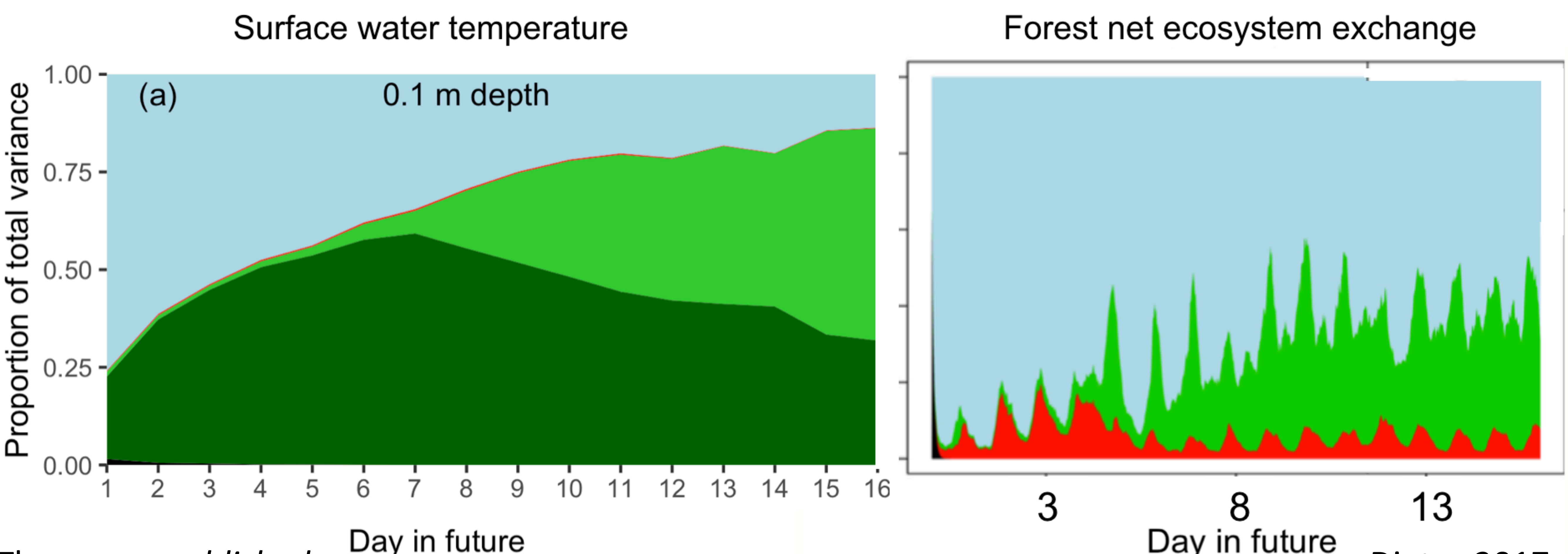

PARAMETER UNCERTAINTY
LOW
HIGH
PREDICTIVE UNCERTAINTY
SENSITIVITY
HIGH
LOW
SDY = 0.4
SDY = 1
SDY = 0.16
SDY = 0.4
0
1
2
3
4
5



VARIANCE DECOMPOSITION

SWITCHGRASS YIELD, CENTRAL ILLINOIS

How do the drivers of forecast uncertainty vary across ecological system?



Thomas *unpublished*

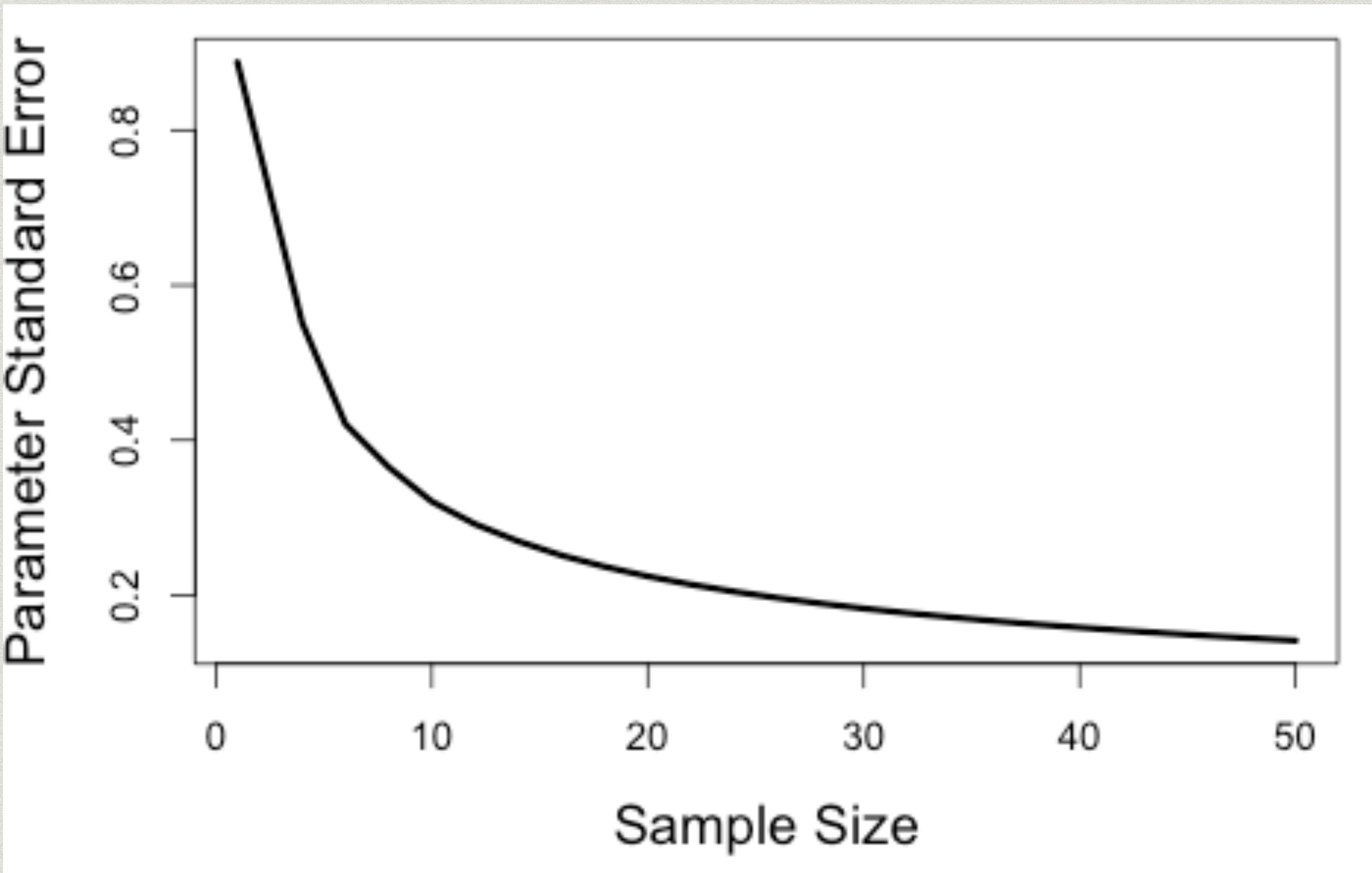
Dietze 2017

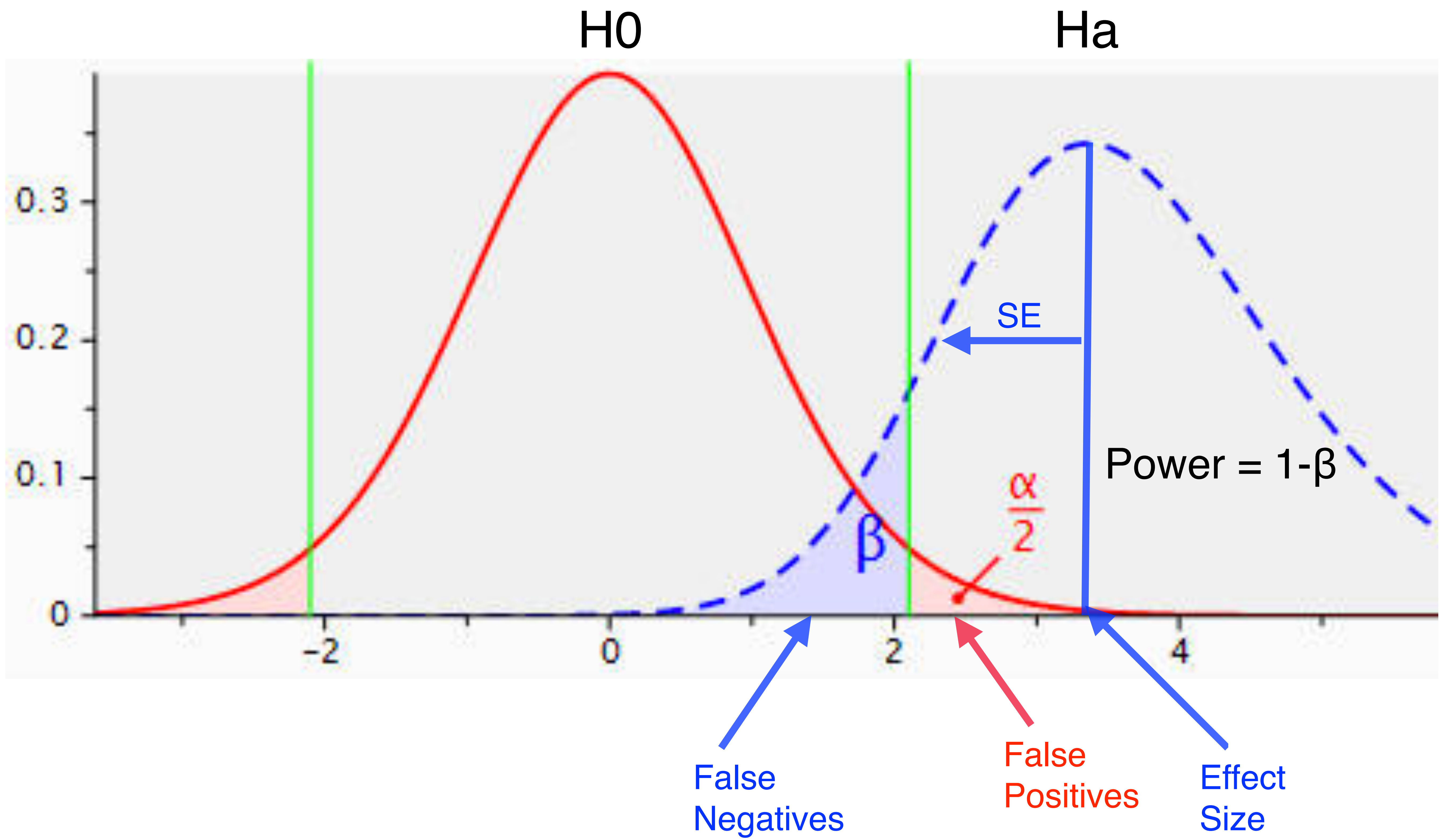
UNCERTAINTY: driver: meteorology | driver: meteorology downscaling | initial conditions (IC) | parameter | process

Tools for model-data feedbacks

* **Power analysis**
  * Sample size needed to detect an effect size
  * Minimum effect size detectable given a size
* **Observational design**
  * What do I need to measure?
  * Where should I collect new data?
  * How do I gain new info most efficiently?

SE ∝ 1/sqrt(n)


Parameter Standard Error
0.2
0.4
0.6
0.8
0
10
20
30
40
50
Sample Size



Power = f(effect size, SE)

Pseudo-data simulation

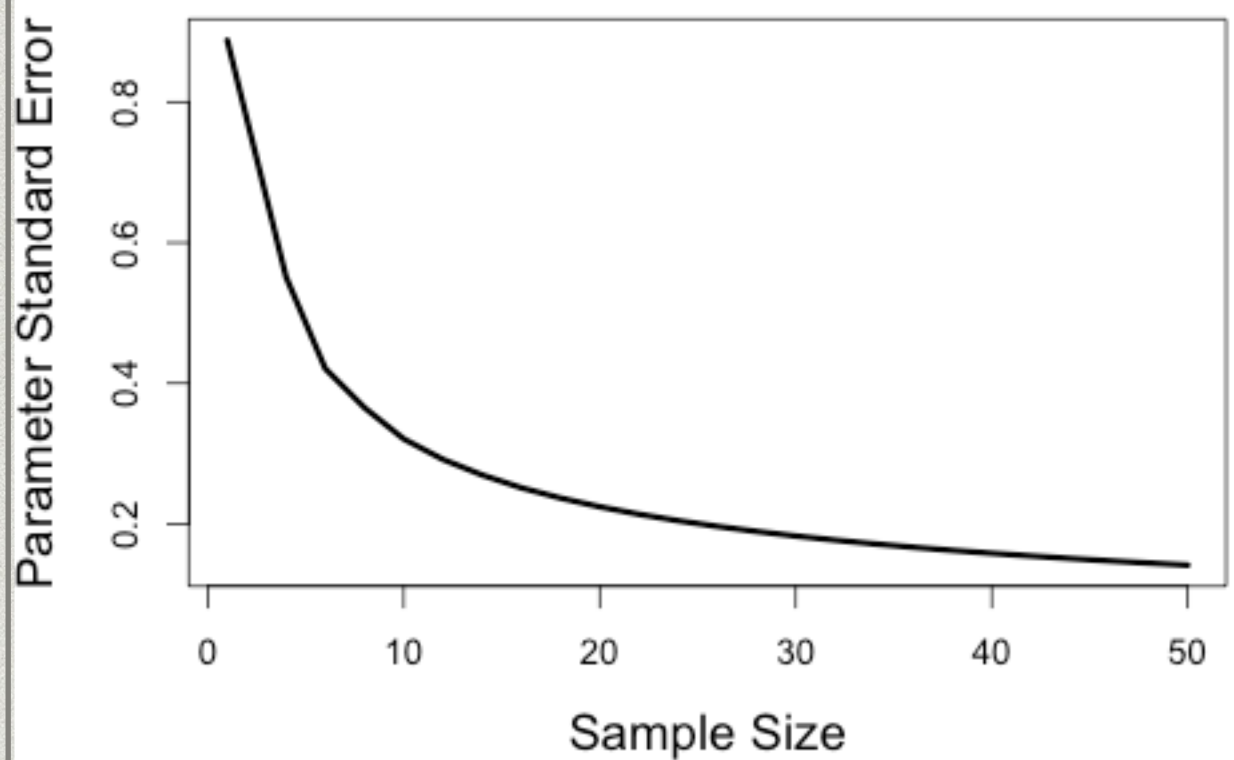
for(k in 1:M)

Draw random data of size N

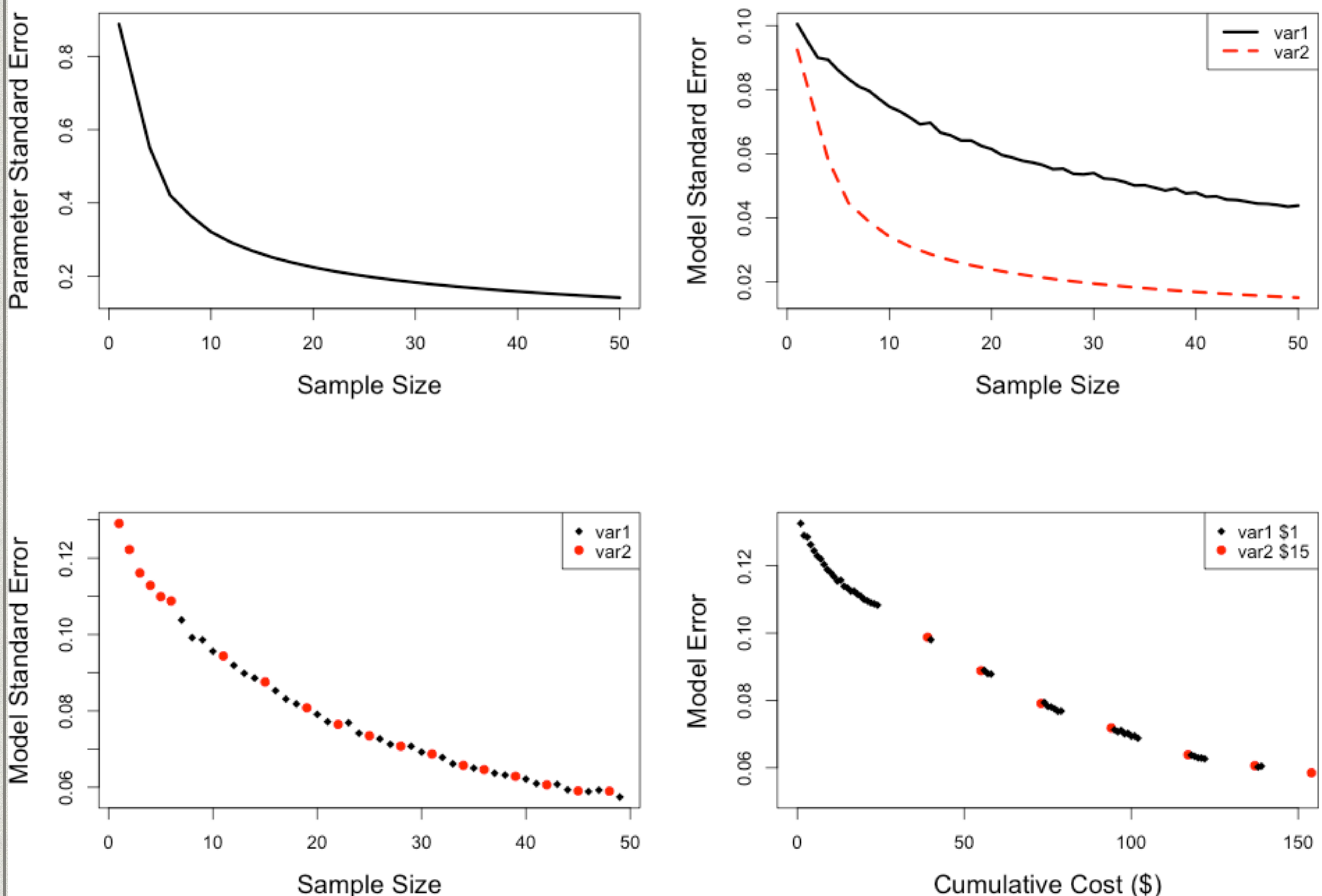
Fit model

Save Parameters

- Nonparameteric bootstrap: resample data
- Parameteric bootstrap: assume param, sim data
- Embed in overall loop over N or different effect sizes
- Summarize distribution



Parameter Standard Error
0.2
0.4
0.6
0.8
0
10
20
30
40
50
Sample Size

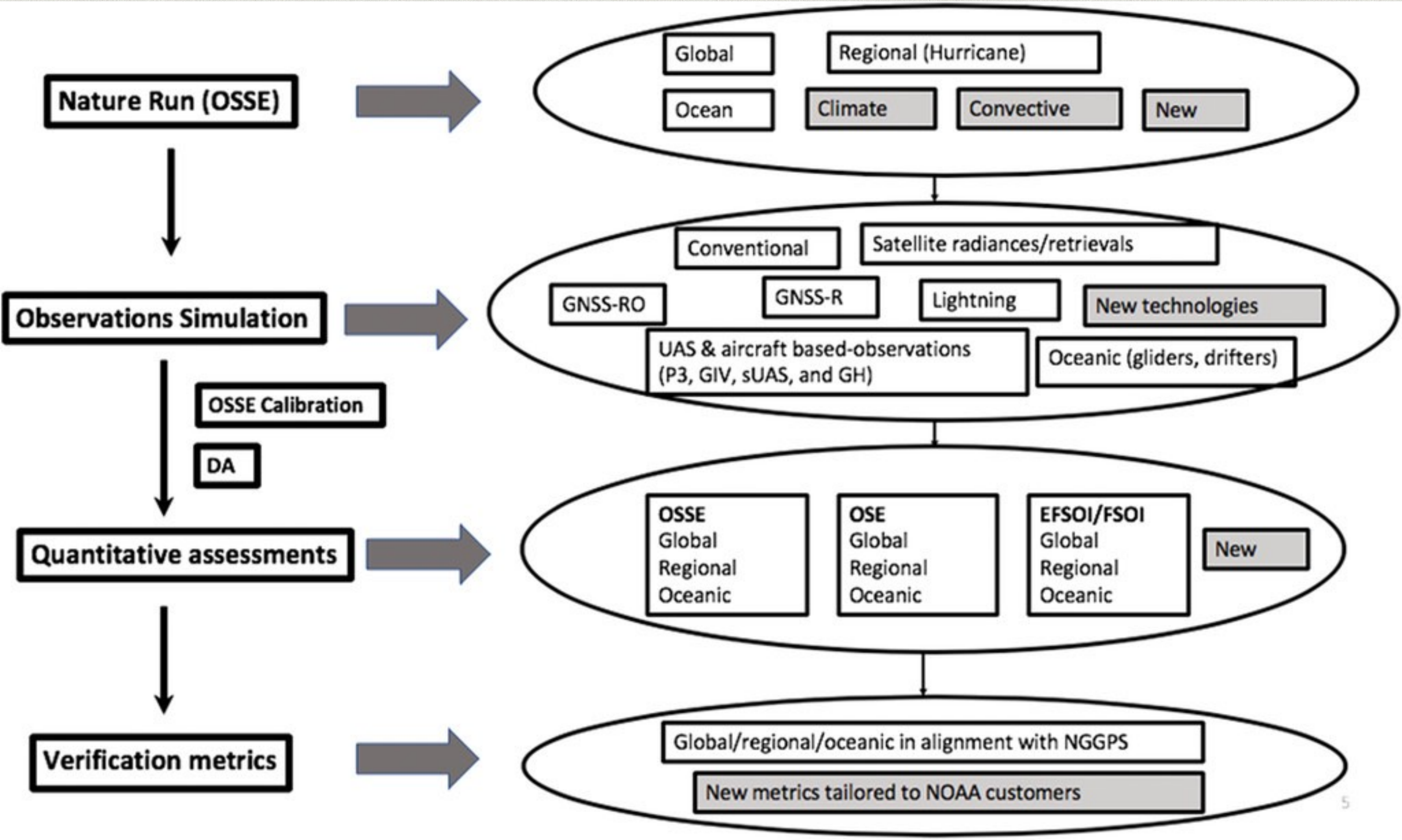


Parameter Standard Error
Sample Size
Model Standard Error
var1
var2
Sample Size
Model Standard Error
var1
var2
Sample Size
Model Error
var1 $1
var2 $15
Cumulative Cost ($)

Observing System Simulation Experiments

- Simulate “true” system
- Simulate pseudo-observations
- Assimilate pseudo-observations
- Assess impact on estimates

- **Augment an existing network**
  - **Additional locations**
  - **New Sensors**
- **Common in Weather, Remote Sensing, Oceanography**



Zeng et al 2020 "Use of Observing System Simulation Experiments in the United States" BAMS https://doi.org/10.1175/BAMS-D-19-0155.1