Characterizing Uncertainty

Lesson 6

SYNOPSIS

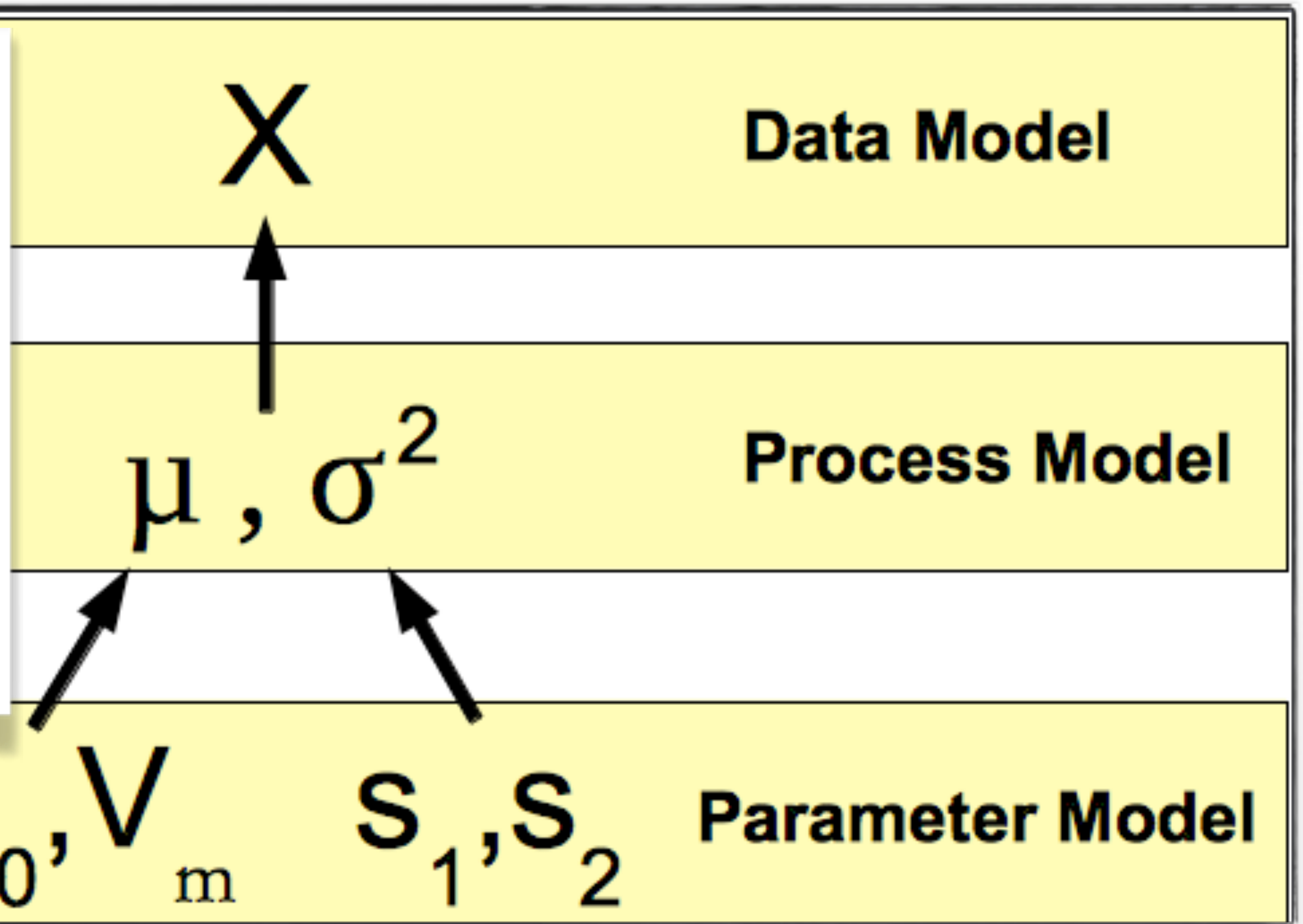
This section dives into the Bayesian methods for characterizing and partitioning sources of error that take us far beyond the classic assumption of a constant Normal variance.

- Non-Gaussian
- **Errors in Variables**
- **Missing Data**
- **Hierarchical models**
- State-Space
- Autocorrelation

$$X \sim N(\mu, \sigma^2)$$

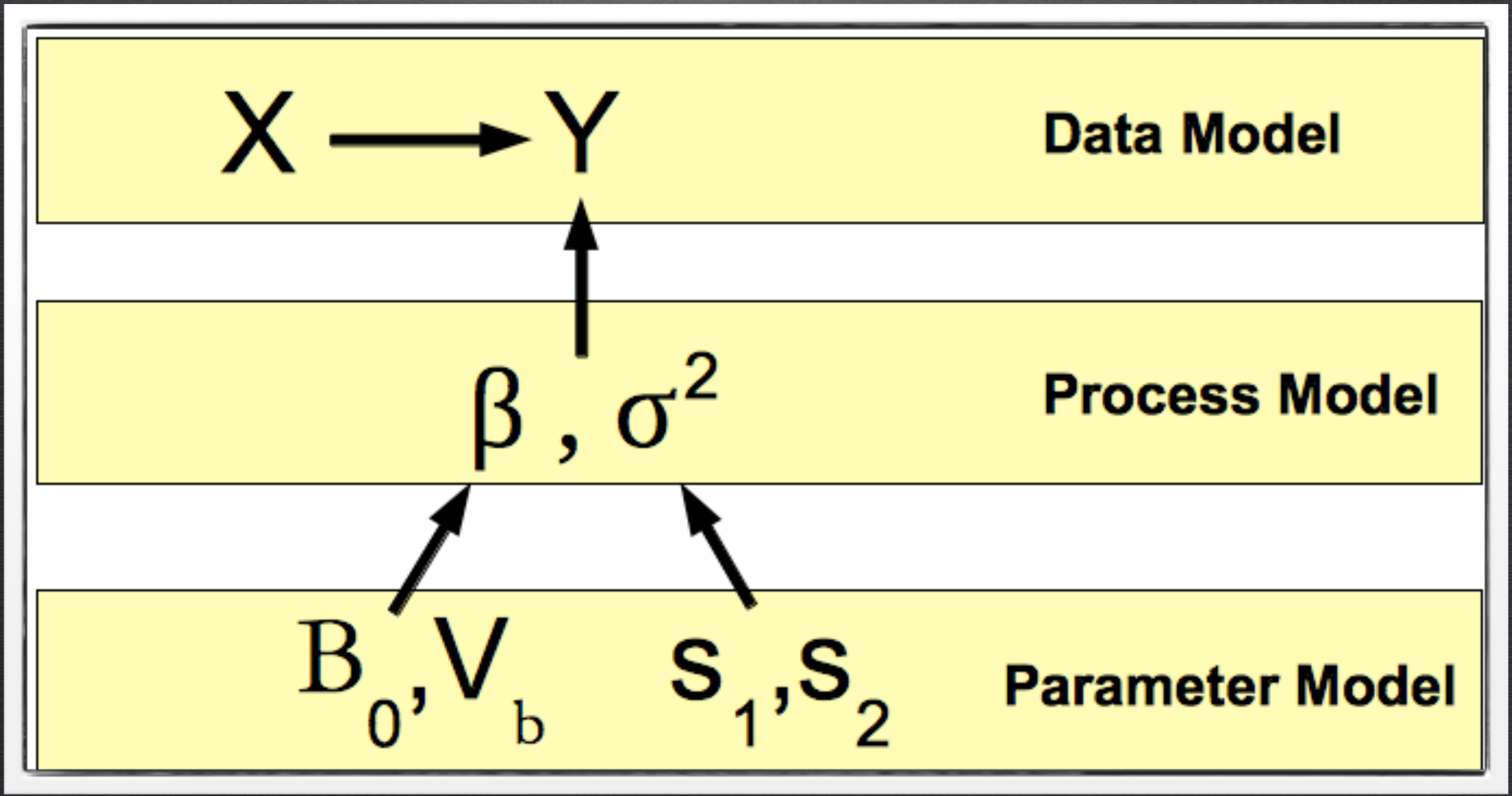
$$\mu \sim N(\mu_0, V_\mu)$$

$$\sigma^2 \sim IG(s_1, s_2)$$



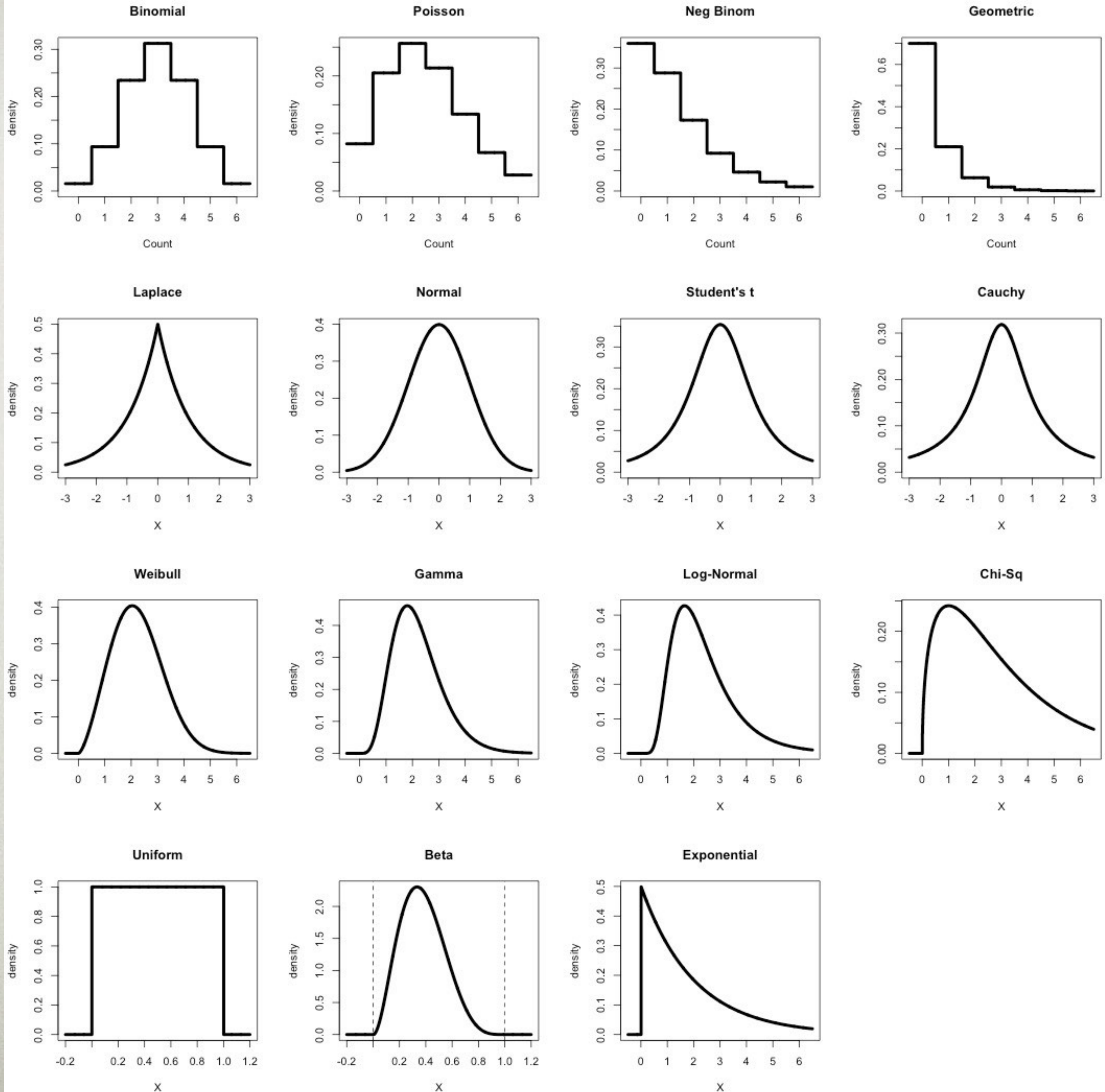
Graph Notation

$$\vec{y} \sim N(\boldsymbol{X}\vec{\beta}, \sigma^2)$$

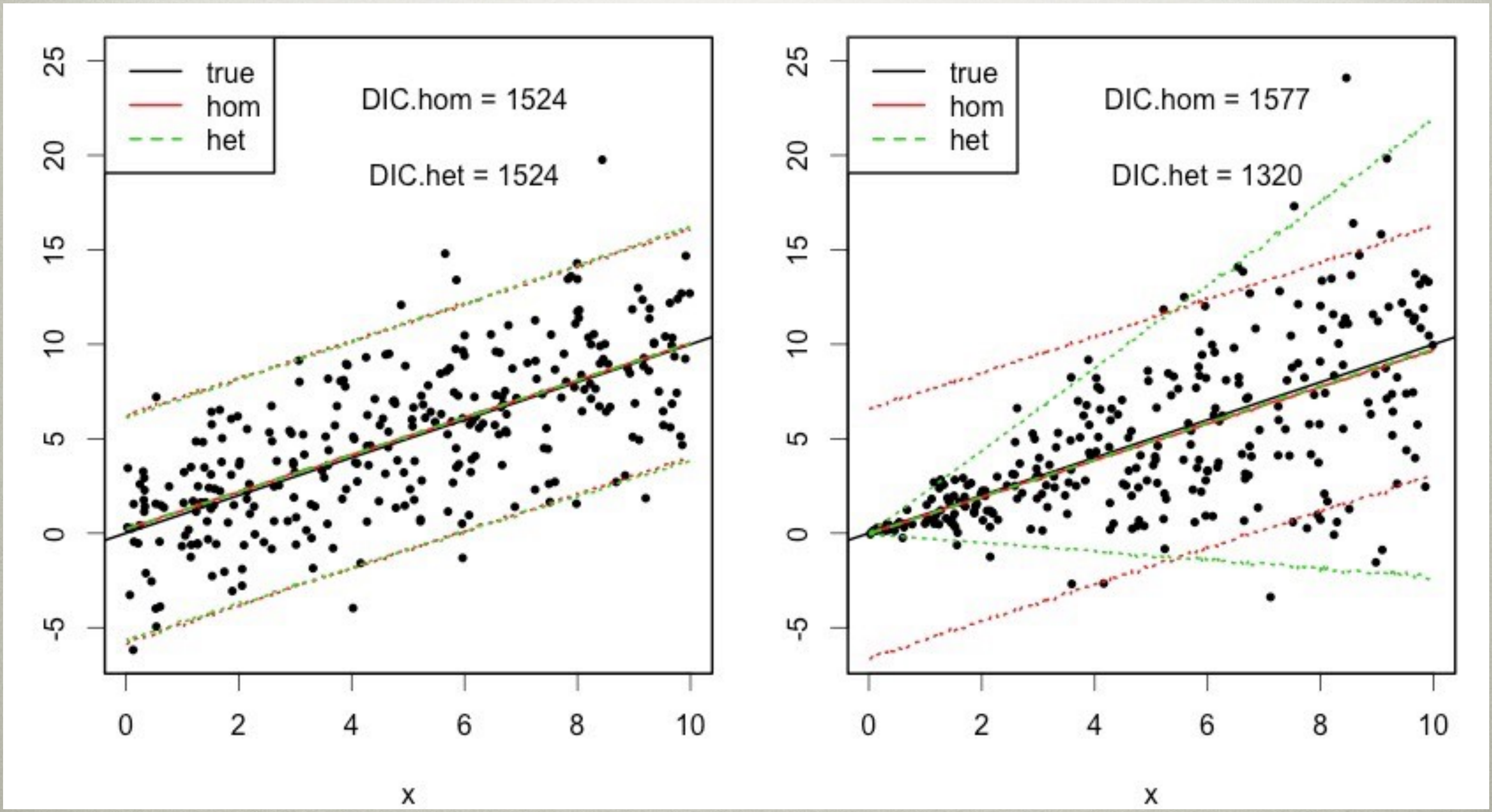


Linear Regression

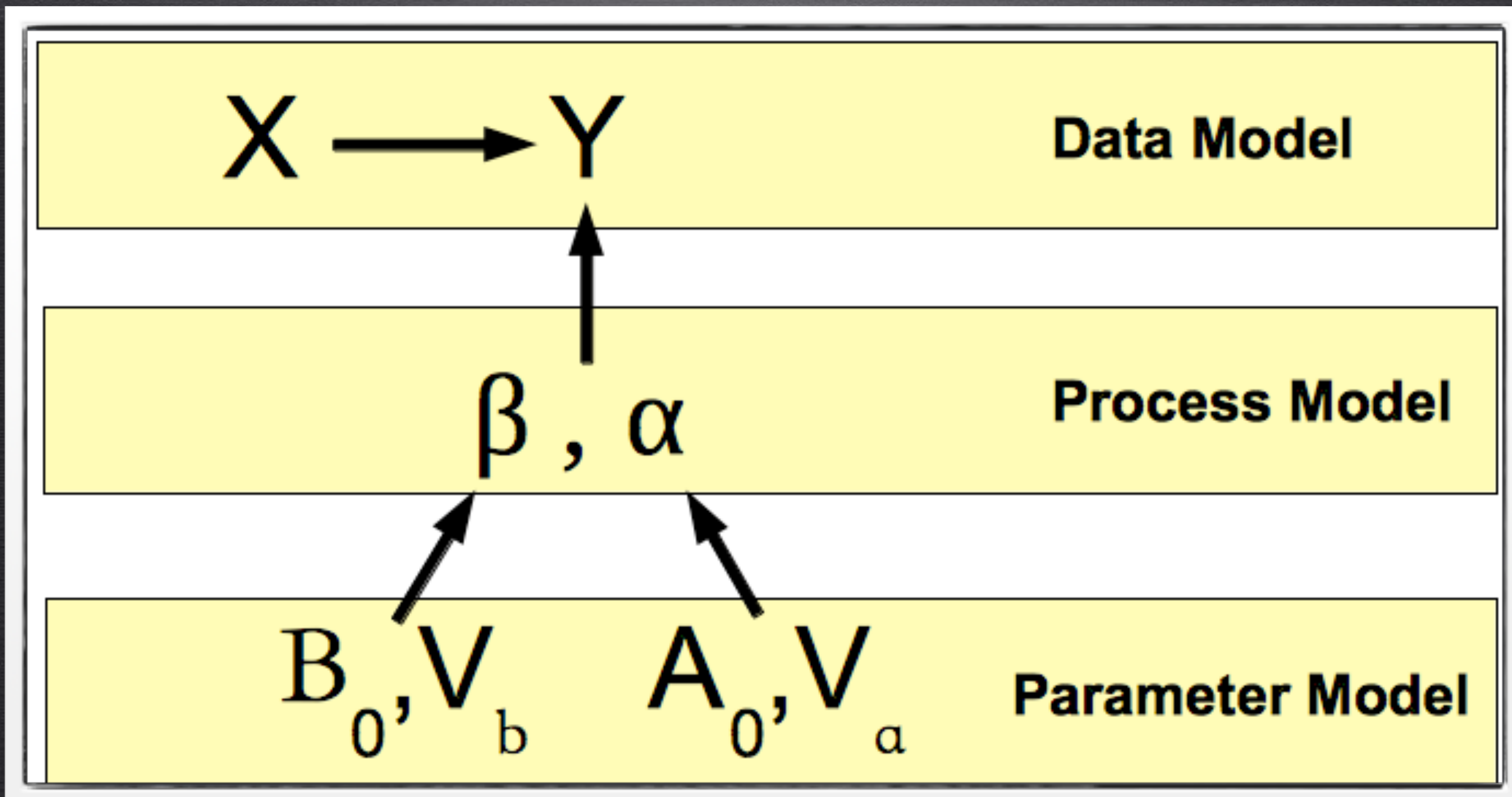
Choosing a Distribution



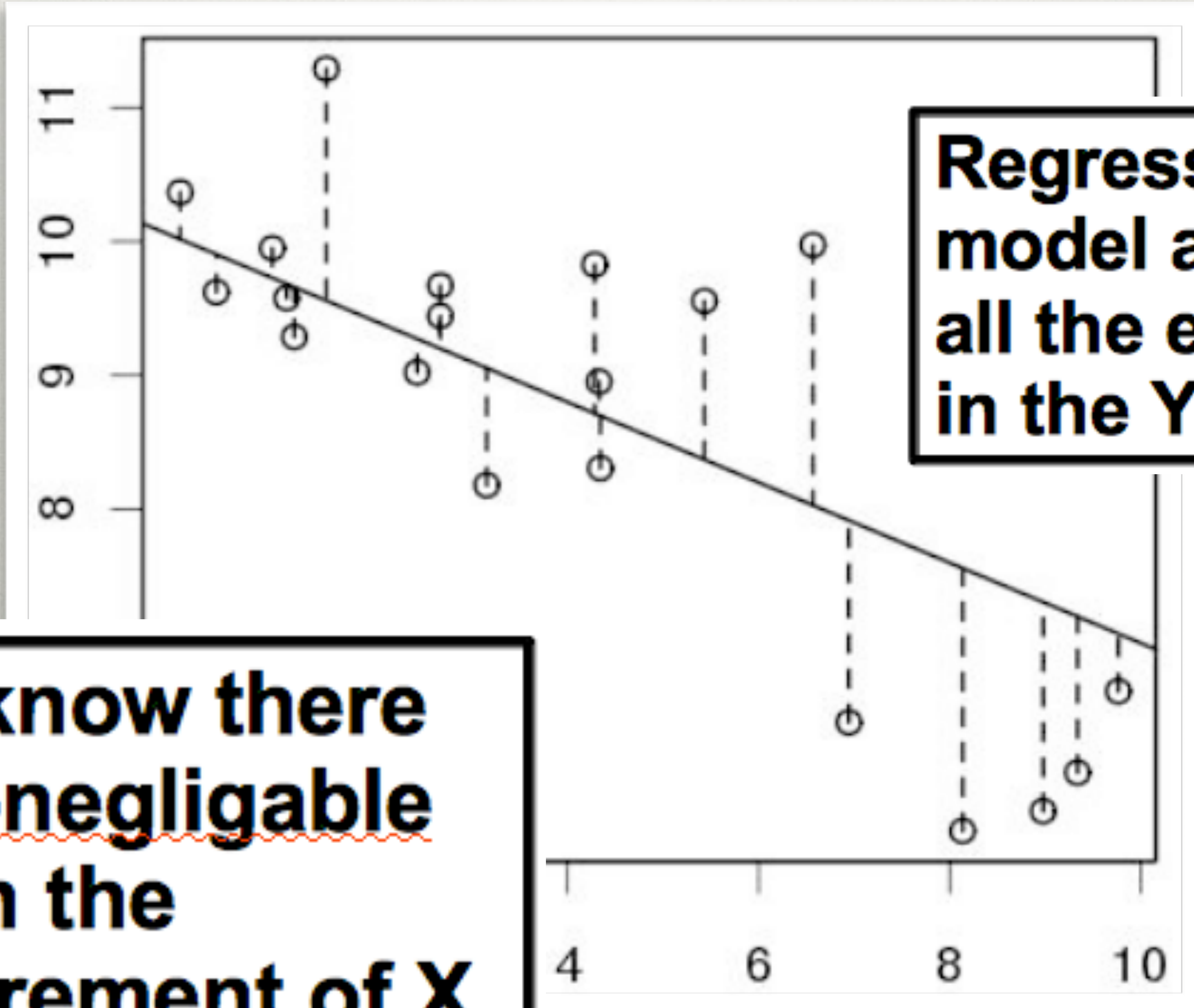
Heteroskedasticity



$$y \sim N(\beta_1 + \beta_2 x, (\alpha_1 + \alpha_2 x)^2)$$



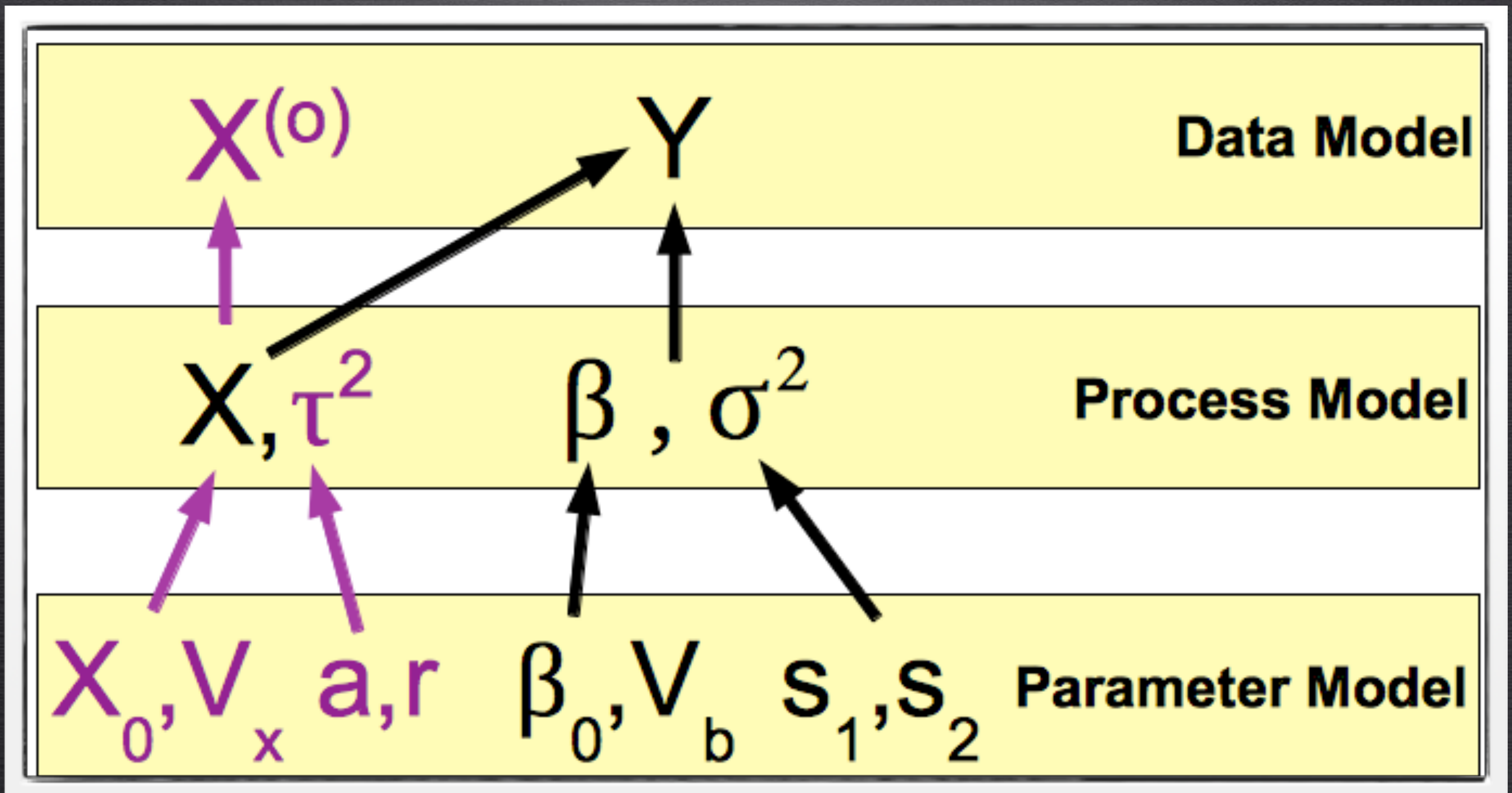
Errors in Variables

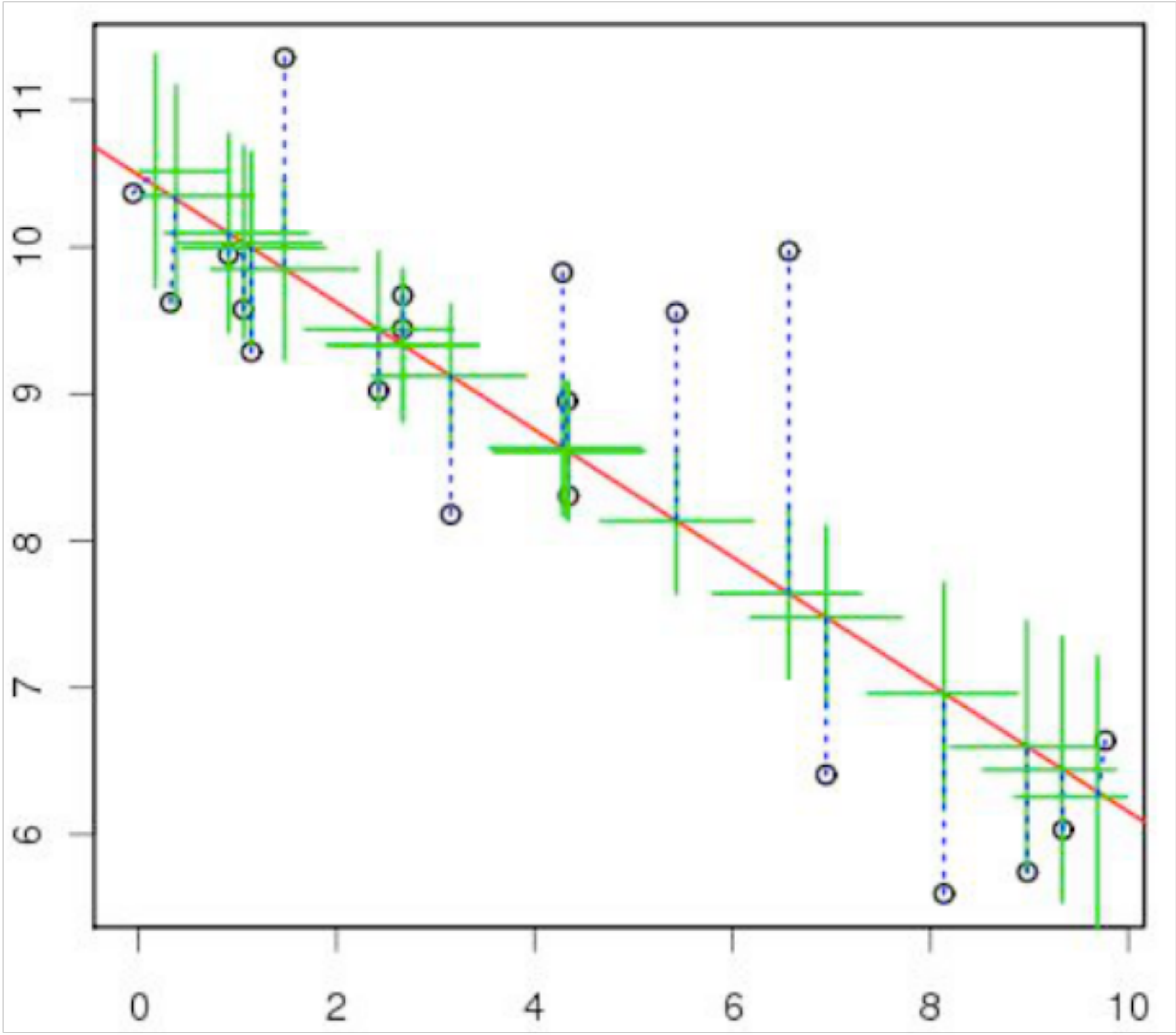


Regression model assumes all the error is in the Y

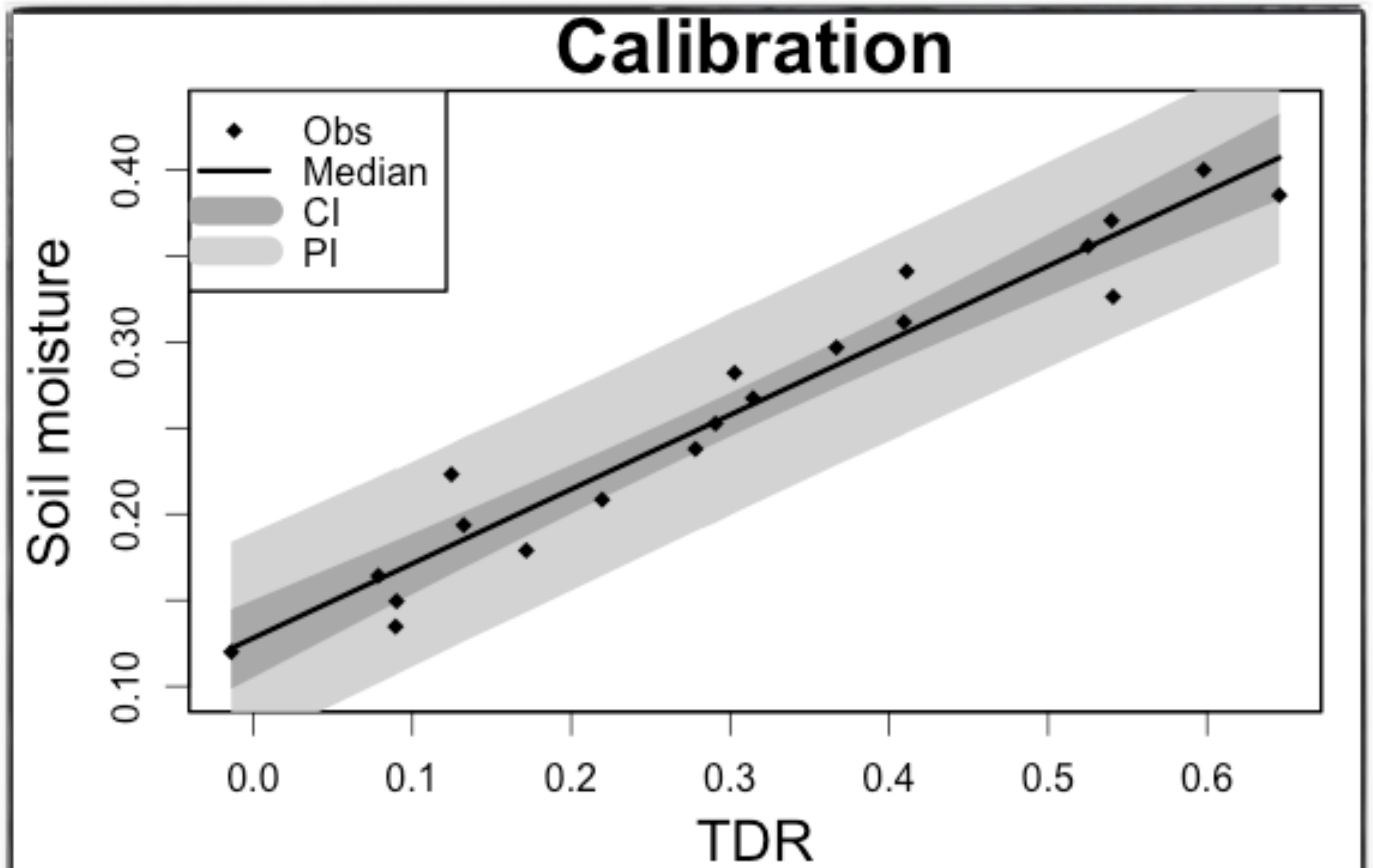
Often know there is non-negligable error in the measurement of X

$$\vec{y} \sim N(\boldsymbol{X}\vec{\beta}, \sigma^2)$$
$$x^{(o)} \sim N(x, \tau^2)$$

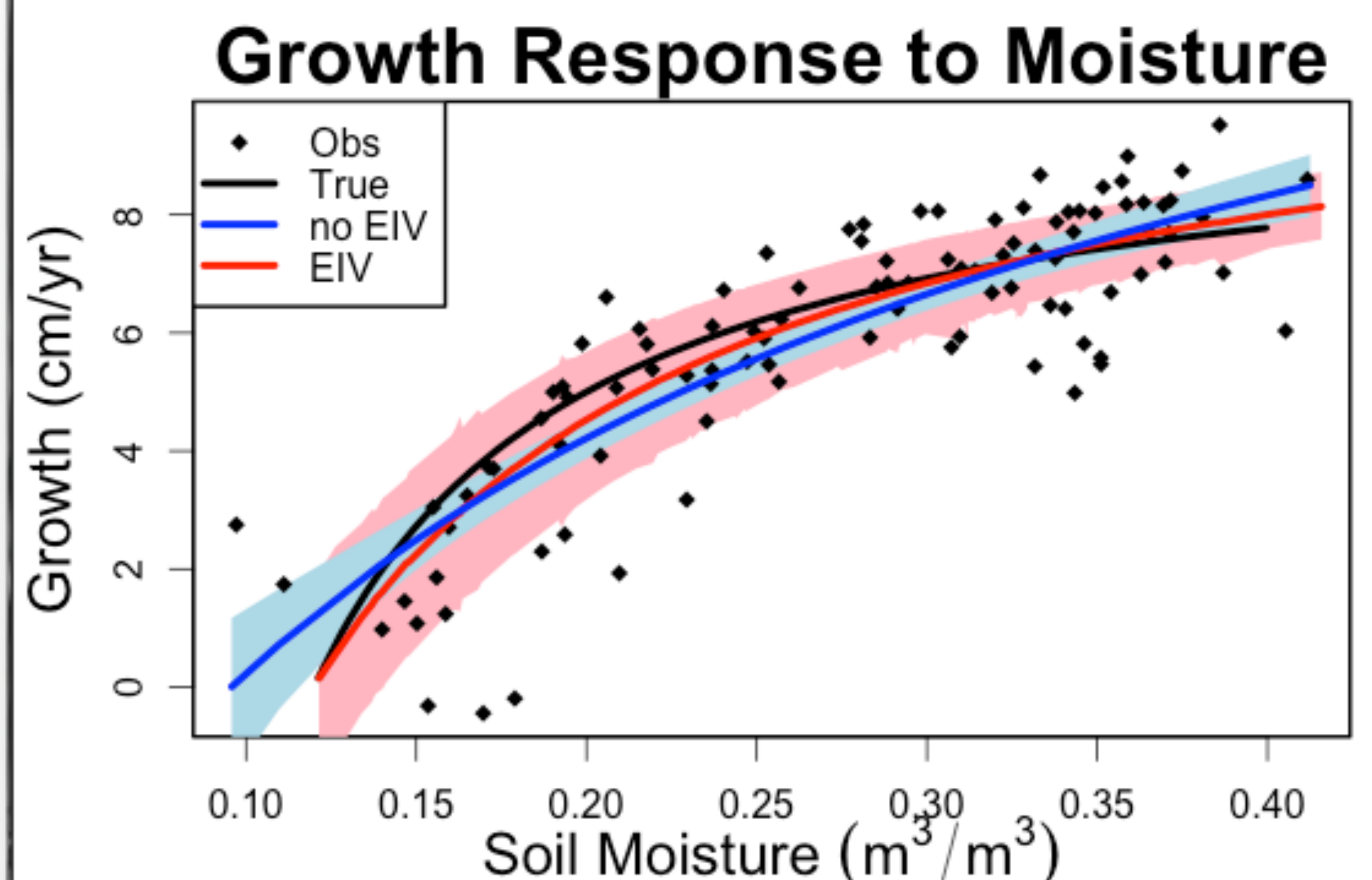




11
10
9
8
7
6
0
2
4
6
8
10

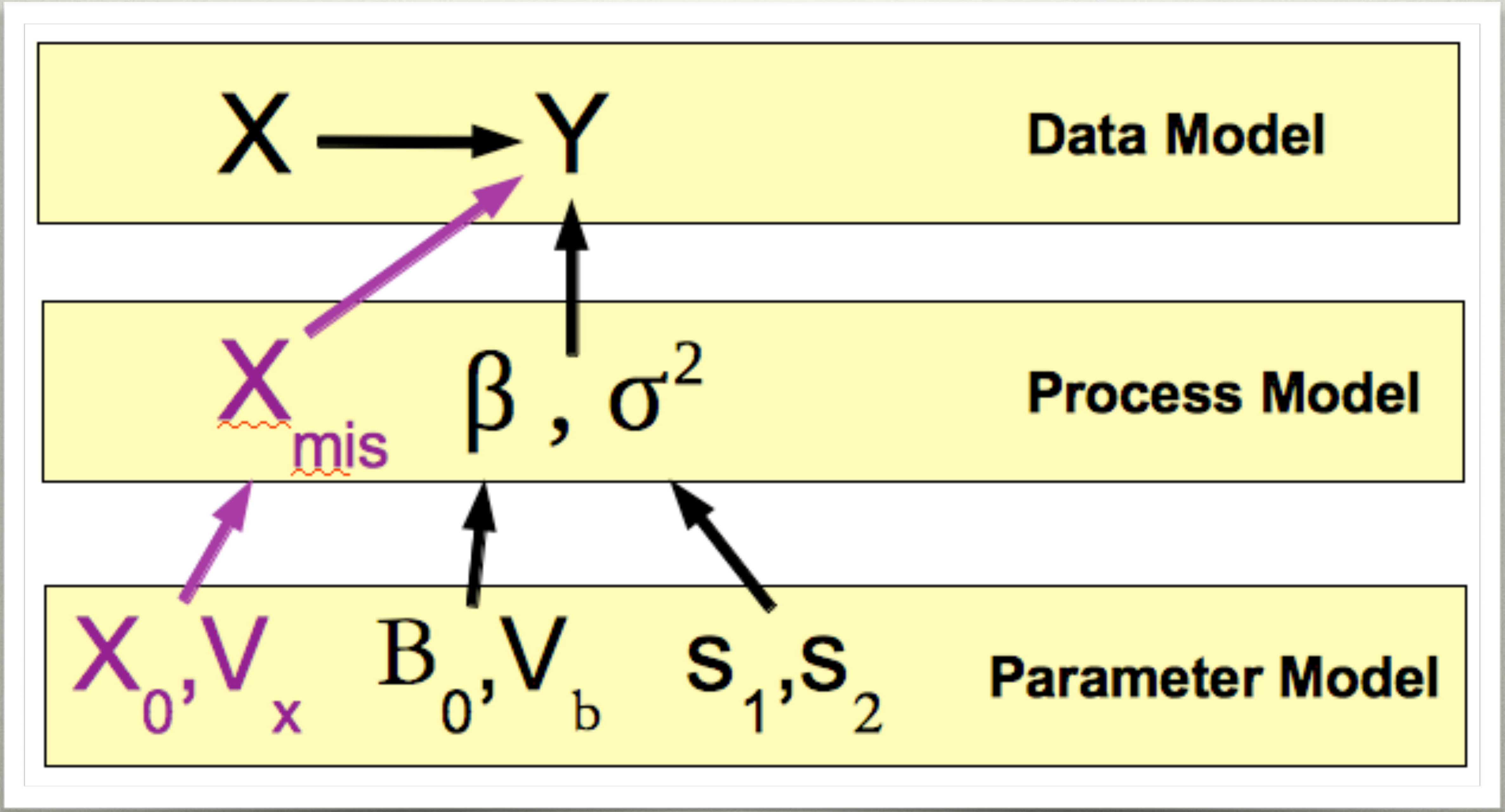


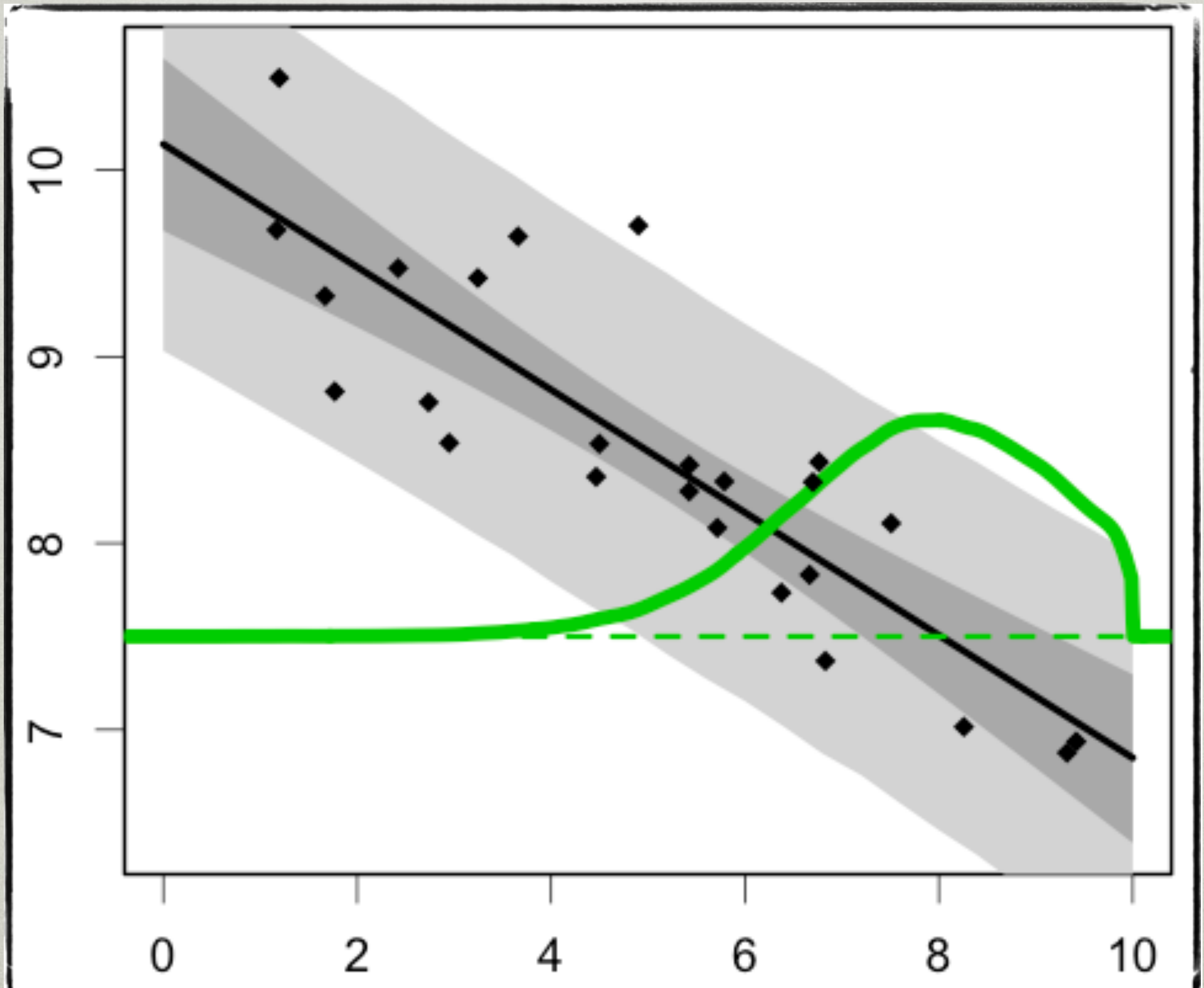
Calibration
Obs
Median
CI
PI
Soil moisture
0.10
0.20
0.30
0.40
0.0
0.1
0.2
0.3
0.4
0.5
0.6
TDR



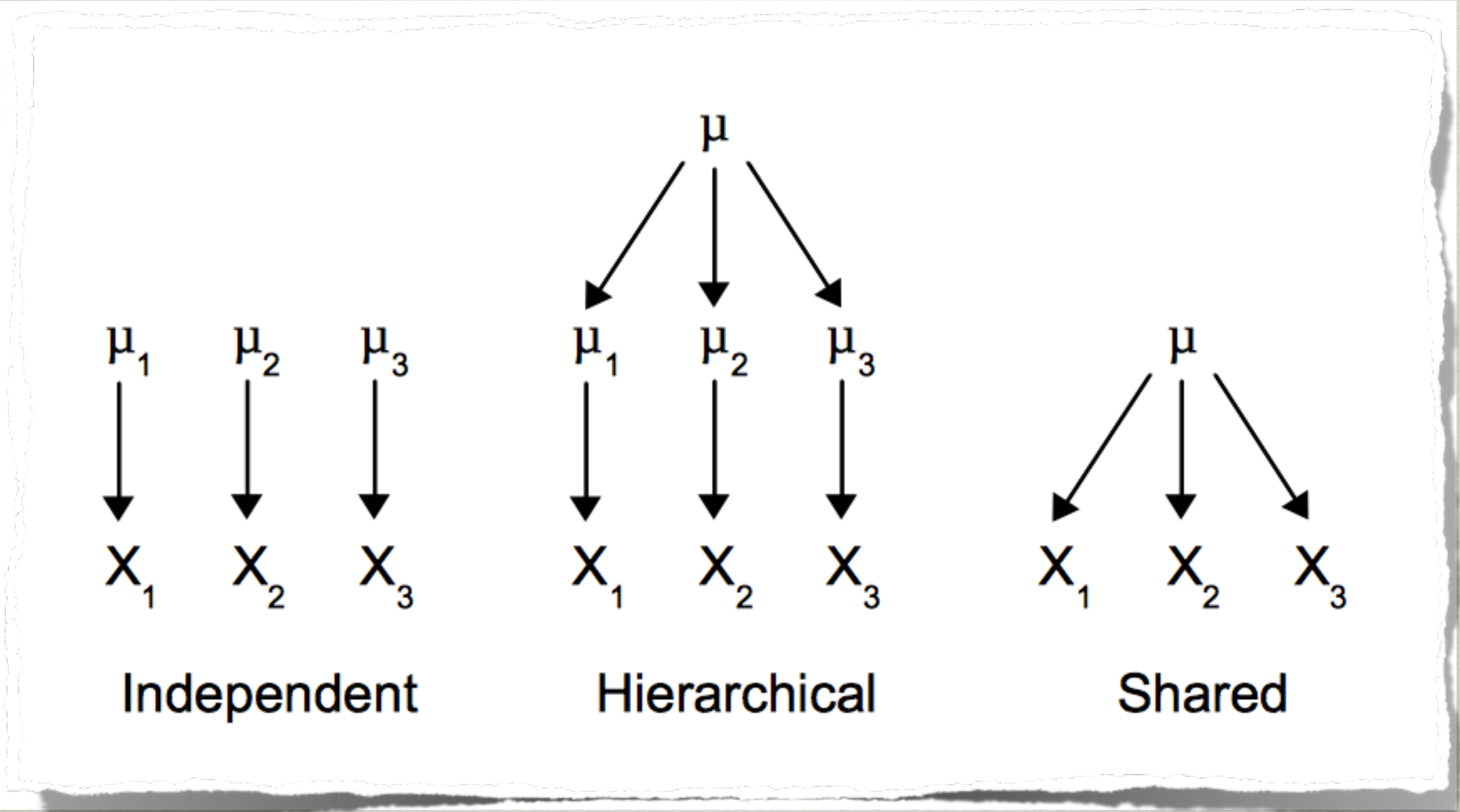
Growth Response to Moisture
Obs
True
no EIV
EIV
Growth (cm/yr)
0
2
4
6
8
0.10
0.15
0.20
0.25
0.30
0.35
0.40
Soil Moisture (m3/m3)

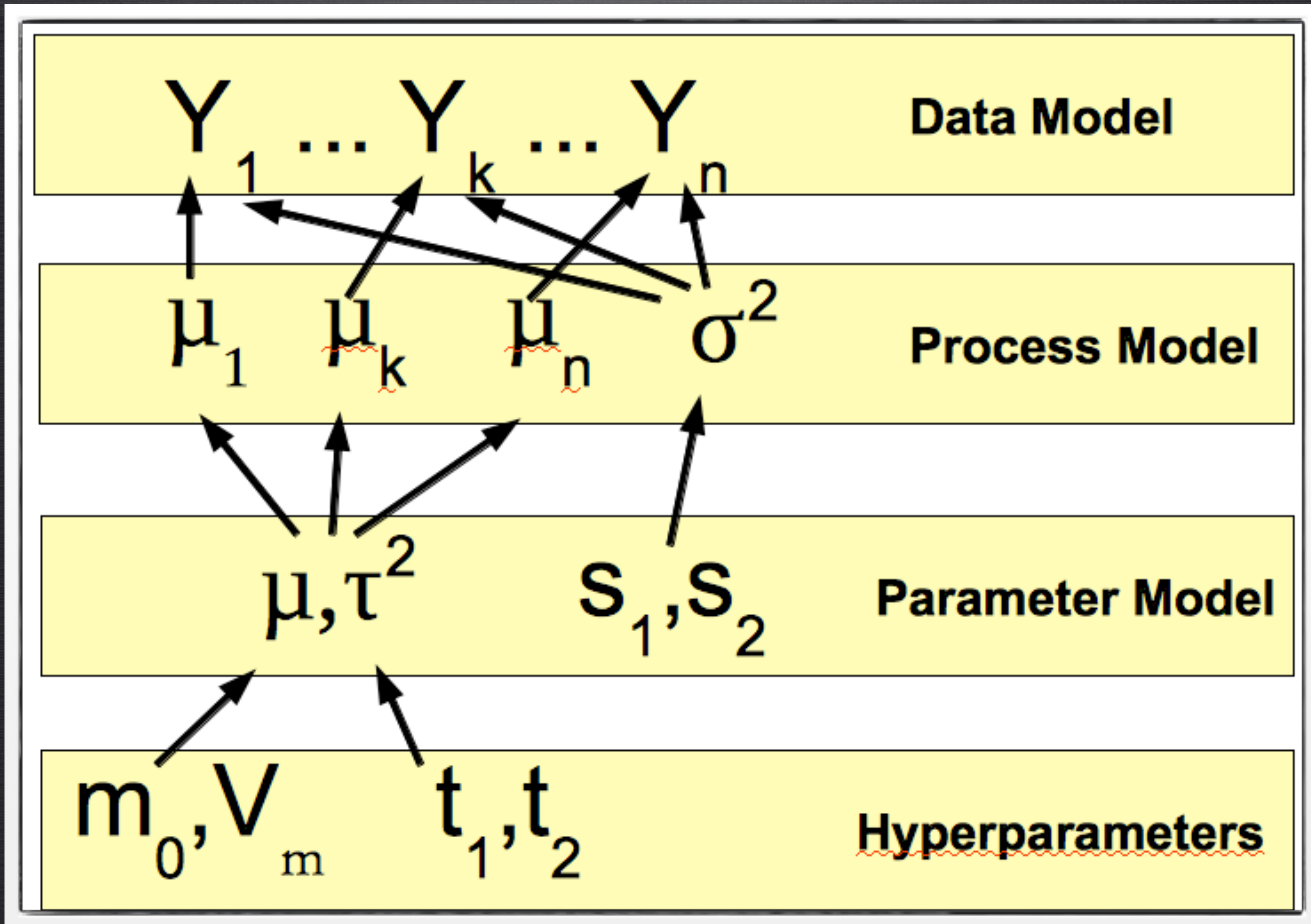
MISSING DATA

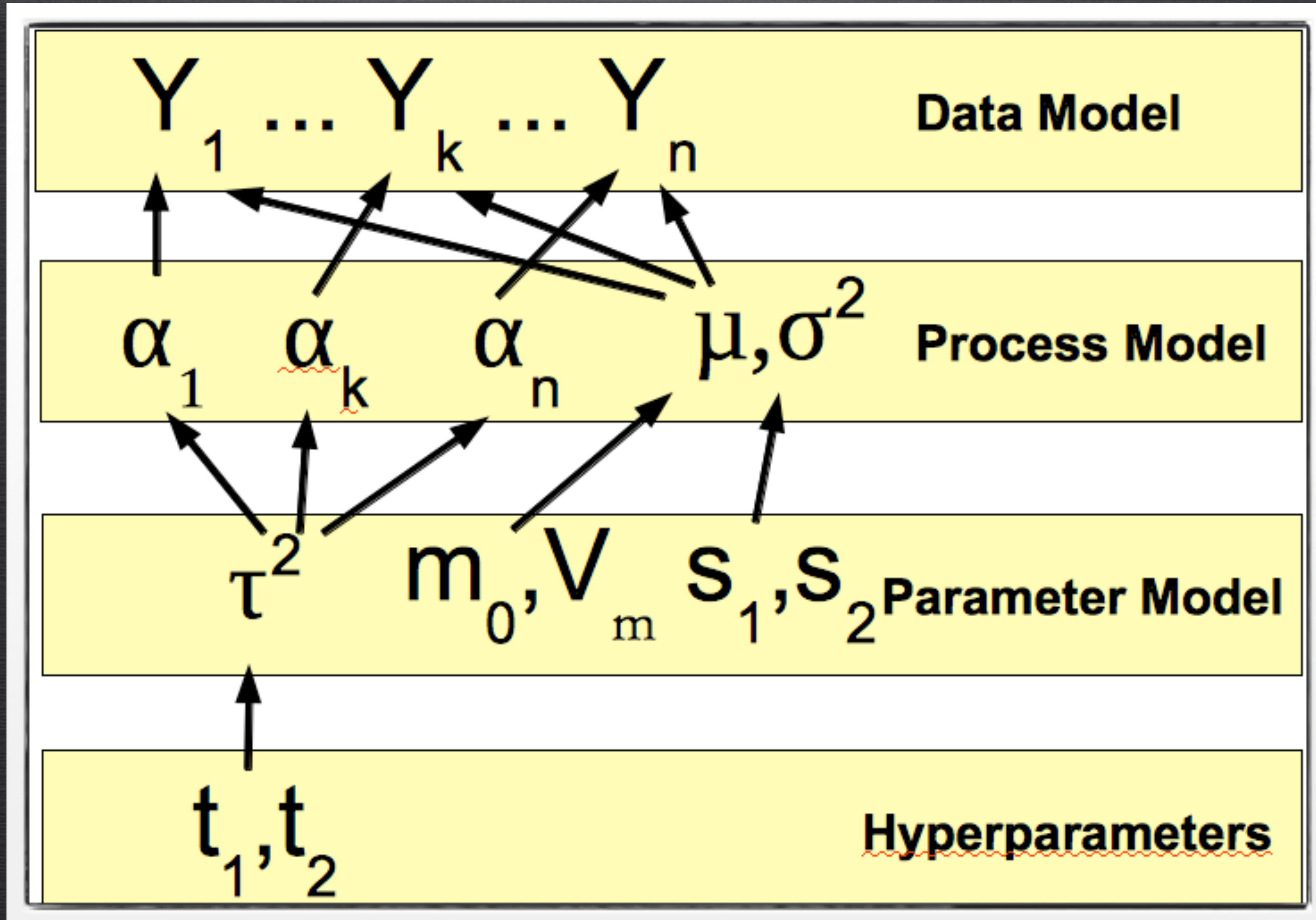



10
9
8
7
0
2
4
6
8
10

Hierarchical Models



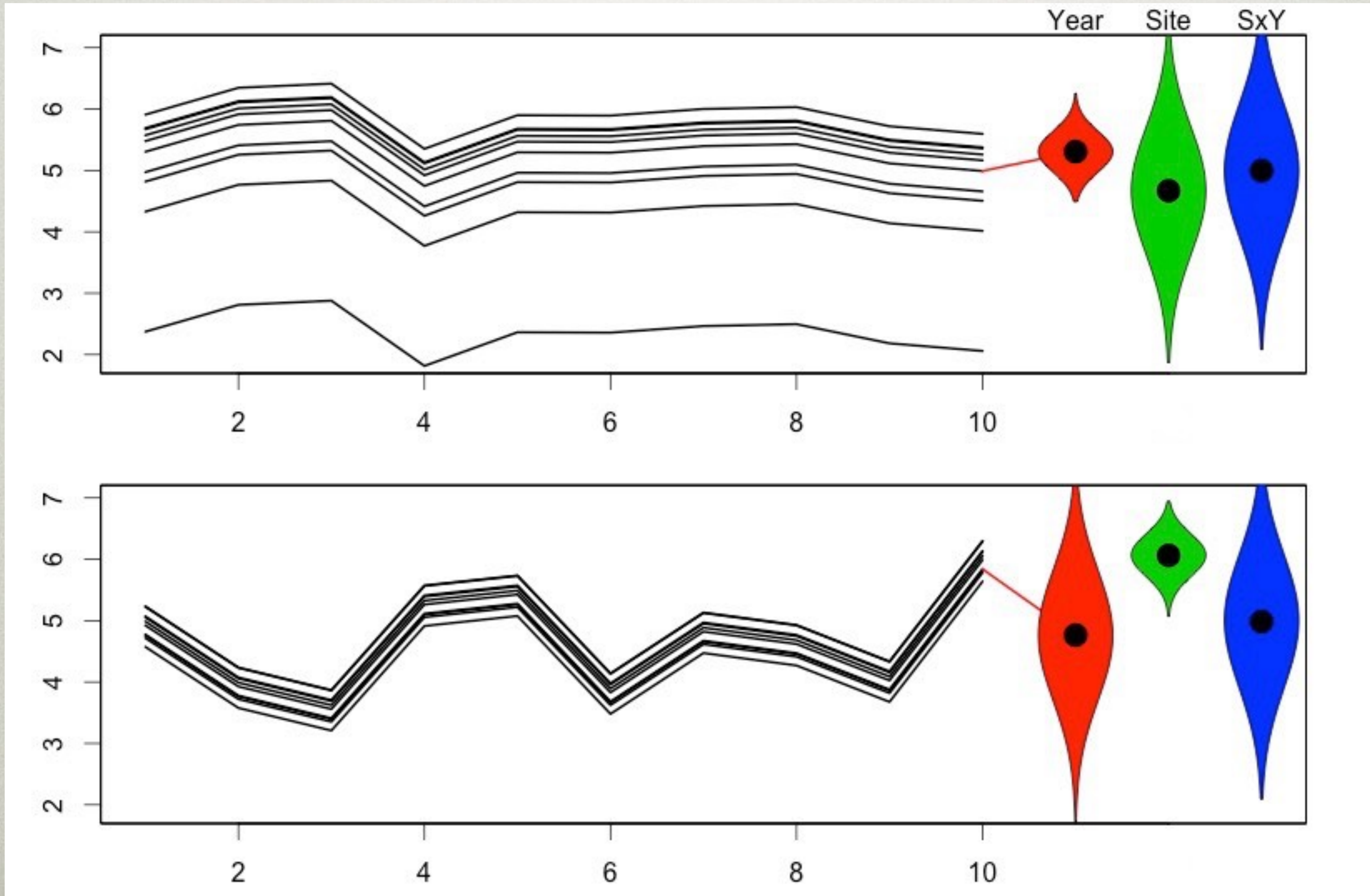

$Y_1 \ldots Y_k \ldots Y_n$
Data Model
μ_1 μ_k μ_n σ^2
Process Model
μ, τ^2
s_1, s_2
Parameter Model
m_0, V_m
t_1, t_2
Hyperparameters



$$Y_k \sim N(\mu_g + \alpha_k, \sigma^2)$$

$$\alpha_k \sim N(0, \tau^2)$$

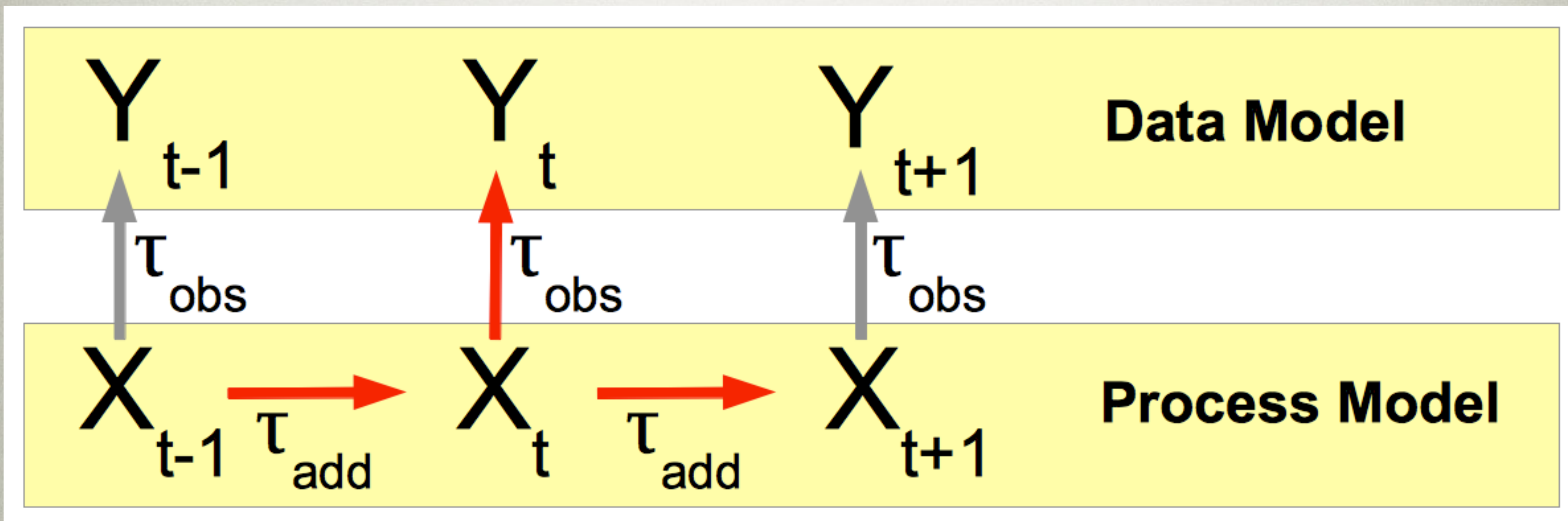
Impacts on Inference



Explaining unexplained variance

- Random effects attempt to account for the unexplained variance associated with some group (plot, year, etc.) due to all the things that were not measured
- May point to scales that need additional explanation
- Adding covariates may explain some portion of this variance, but there's always something you didn't measure
- Sometimes additional fixed effects not justified (model selection)

State Space



$$Y_t = g(X_t | \phi)$$ **Data Model**

$$X_t = f(X_{t-1} | \theta)$$ **Process Model**

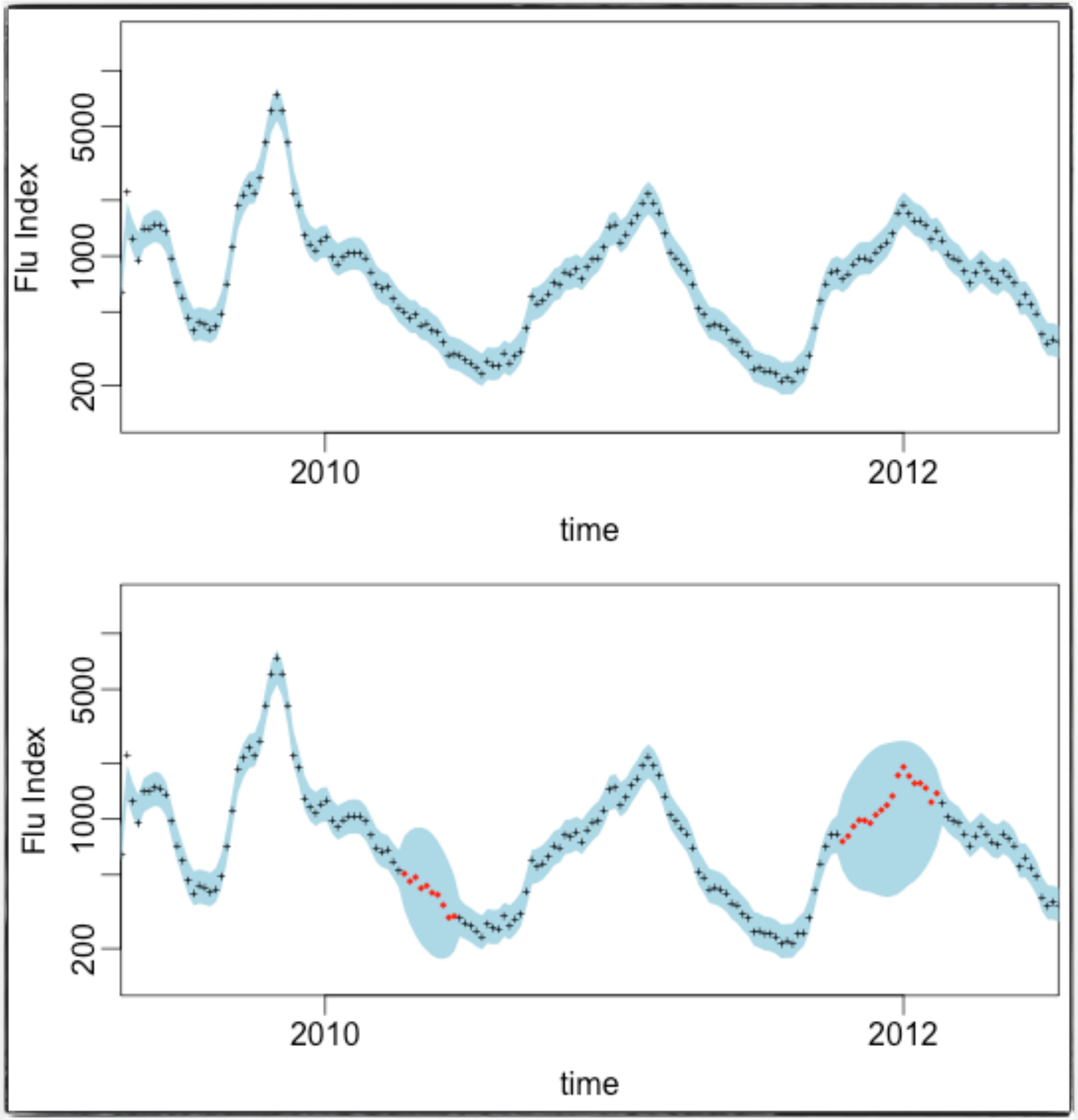
Random Walk

- What is the conditional

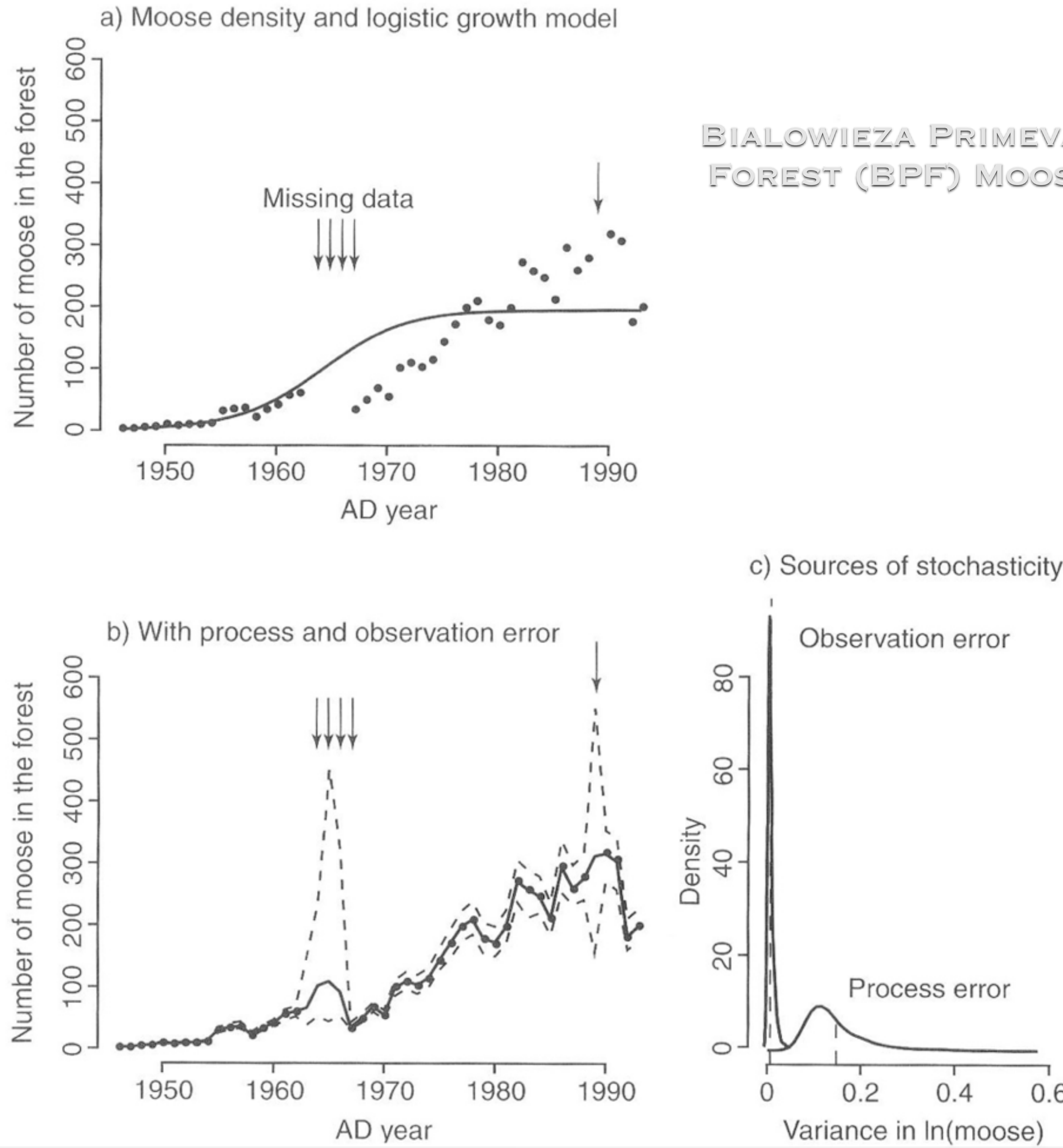
$$X_t \sim N(X_t | X_{t-1}, \tau^2_{add}) \times$$
$$N(X_{t+1} | X_t, \tau^2_{add}) \times$$
$$N(Y_t | X_t, \tau^2_{obs})$$

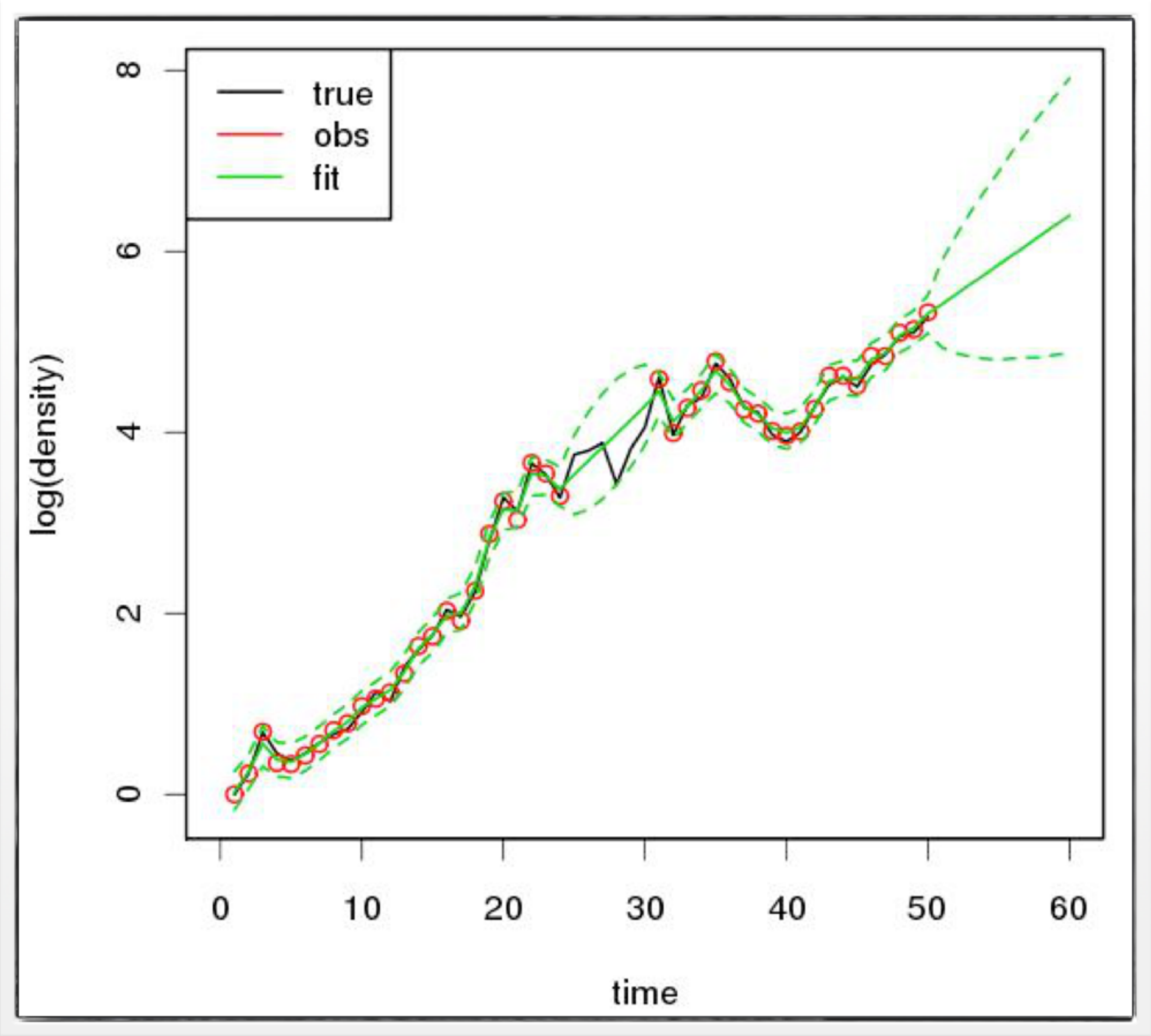
$$X_t \sim N(X_{t-1}, \tau^2_{add})$$
$$Y_t \sim N(X_t, \tau^2_{obs})$$
$$\tau^2_{obs} \sim IG(a_{obs}, r_{obs})$$
$$\tau^2_{add} \sim IG(a_{add}, r_{add})$$
$$X_0 \sim N(X_{ic}, \tau_{IC})$$

- Special Cases
  - First
  - Last
  - Missing Y

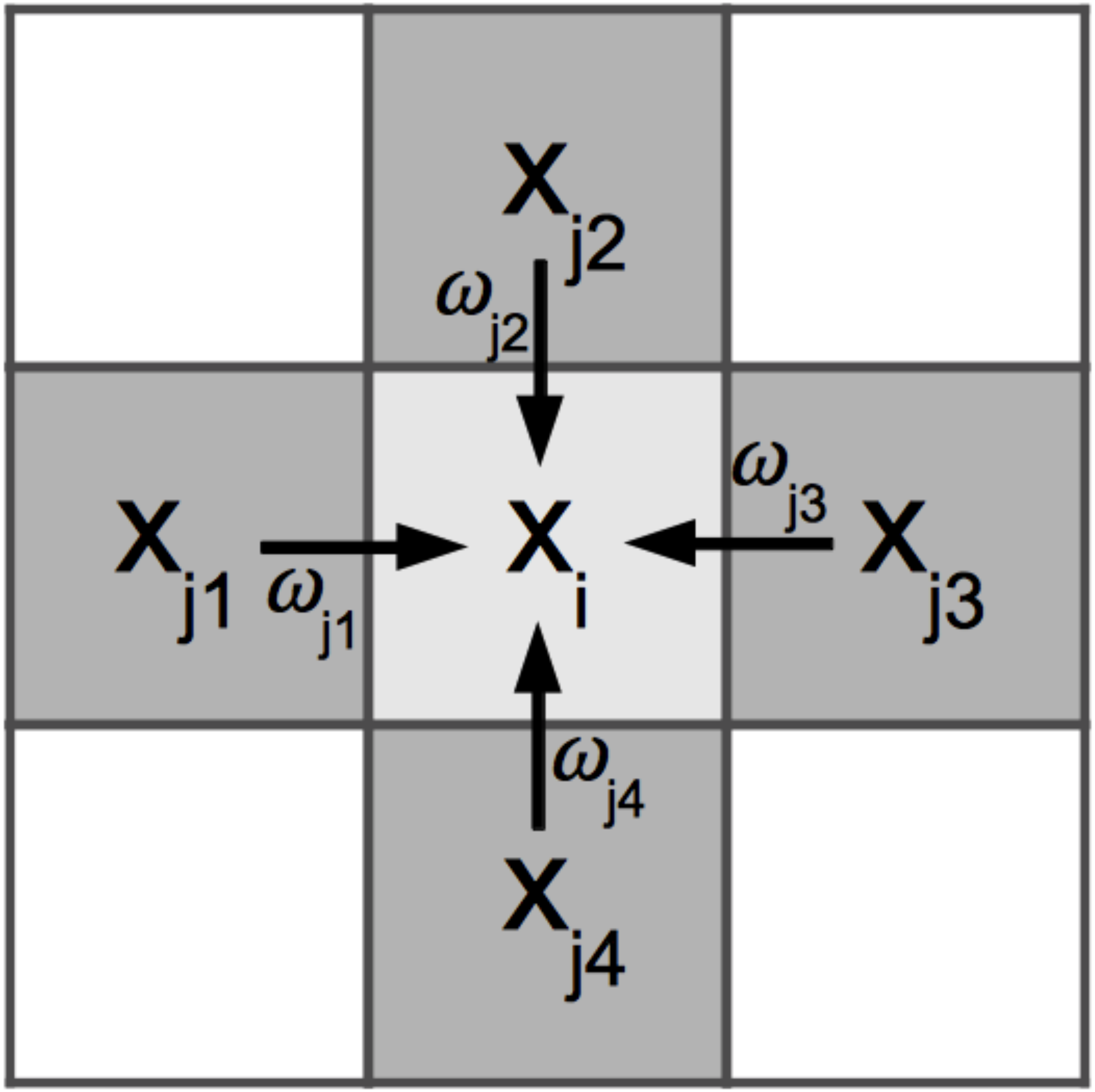

Flu Index
5000
1000
200
2010
2012
time
Flu Index
5000
1000
200
2010
2012
time

Bialowieza Primeval Forest (BPF) Moose




true
obs
fit
log(density)
time
0
2
4
6
8
0
10
20
30
40
50
60

Spatial Process



$$x_i = \frac{1}{\sum \omega_j} \sum \omega_j x_j$$

Generality of the State Space Framework

- Neither X nor Y need be Normal
- X and Y don't need to be the same type of data
- X and Y don't need to have the same time scale
- Handles missing data (gaps) and irregularly spaced data
- Handles multiple data sources (Y's), which don't need to be the same type or synchronous
- Handles time-integrated observations