Time Series

Why Time is Important

- Explicit in many environmental models
  - Can generate complex/chaotic feedbacks
  - External/environmental factors change over time
- Measurements often made repeatedly over time
  - Data usually correlated in time
  - Response to treatments
- Importance of separating process and measurement error
  - Measurement error does not propagate

Characteristics of Time Series Data

- Single/small number of long time series
  - Often concerned with identifying trends, periodicity, autocorrelation, cross-correlation, etc.
- Longitudinal / repeated measures
  - Many short time series
  - Intervention analysis
  - Mark-recapture (boolean data)

Dynamic process models

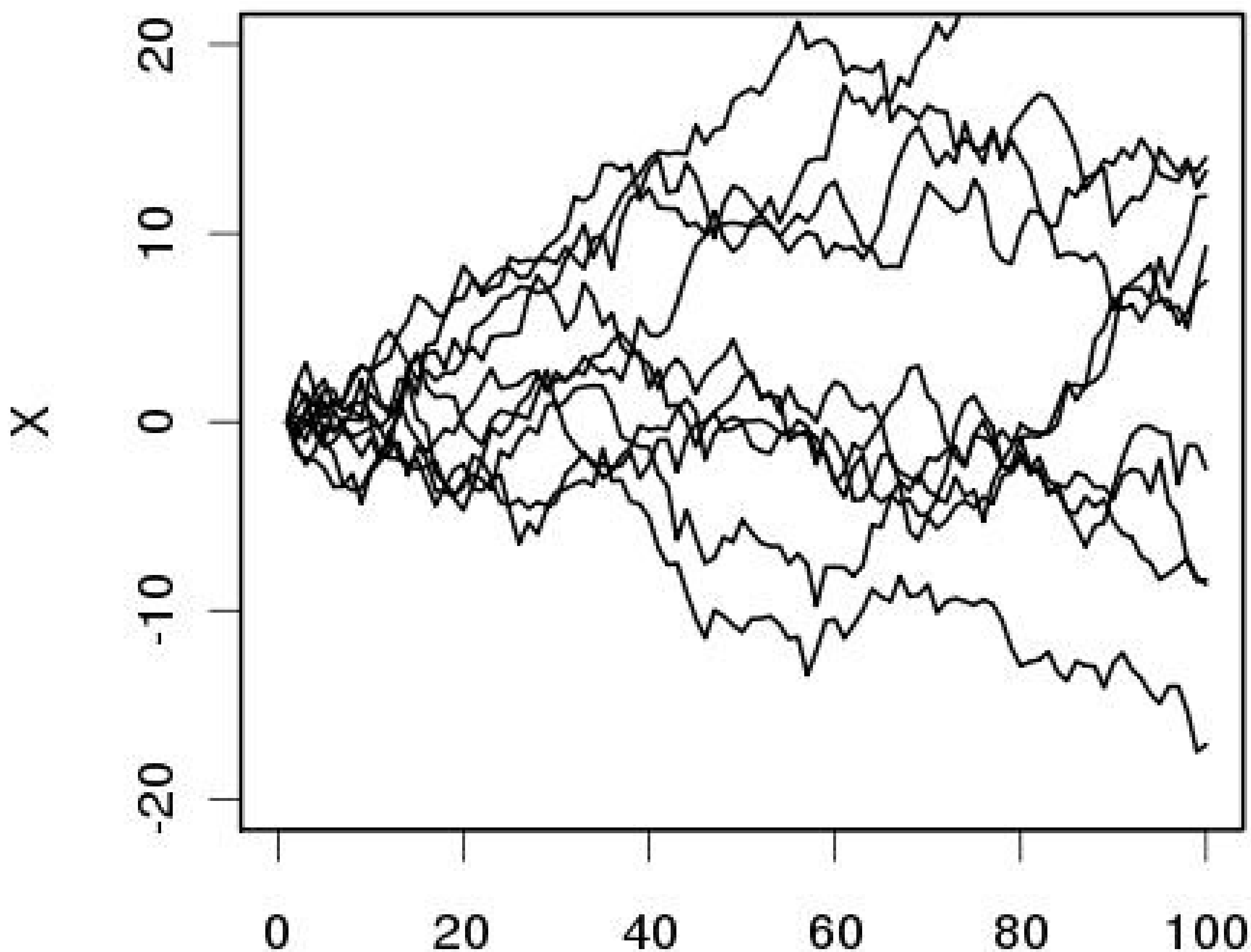
$$x_t = f(x_{t-1}, x_{t-2}, \ldots, \theta)$$

- Recursive: state at current time point a function of the previous state
- Any model that depends only on the most recent state (x_{t-1}) is called a Markov model
- Higher order models (additional lags) introduce memory to the system

Random Walk

$$X_t = X_{t-1} + \epsilon_t$$

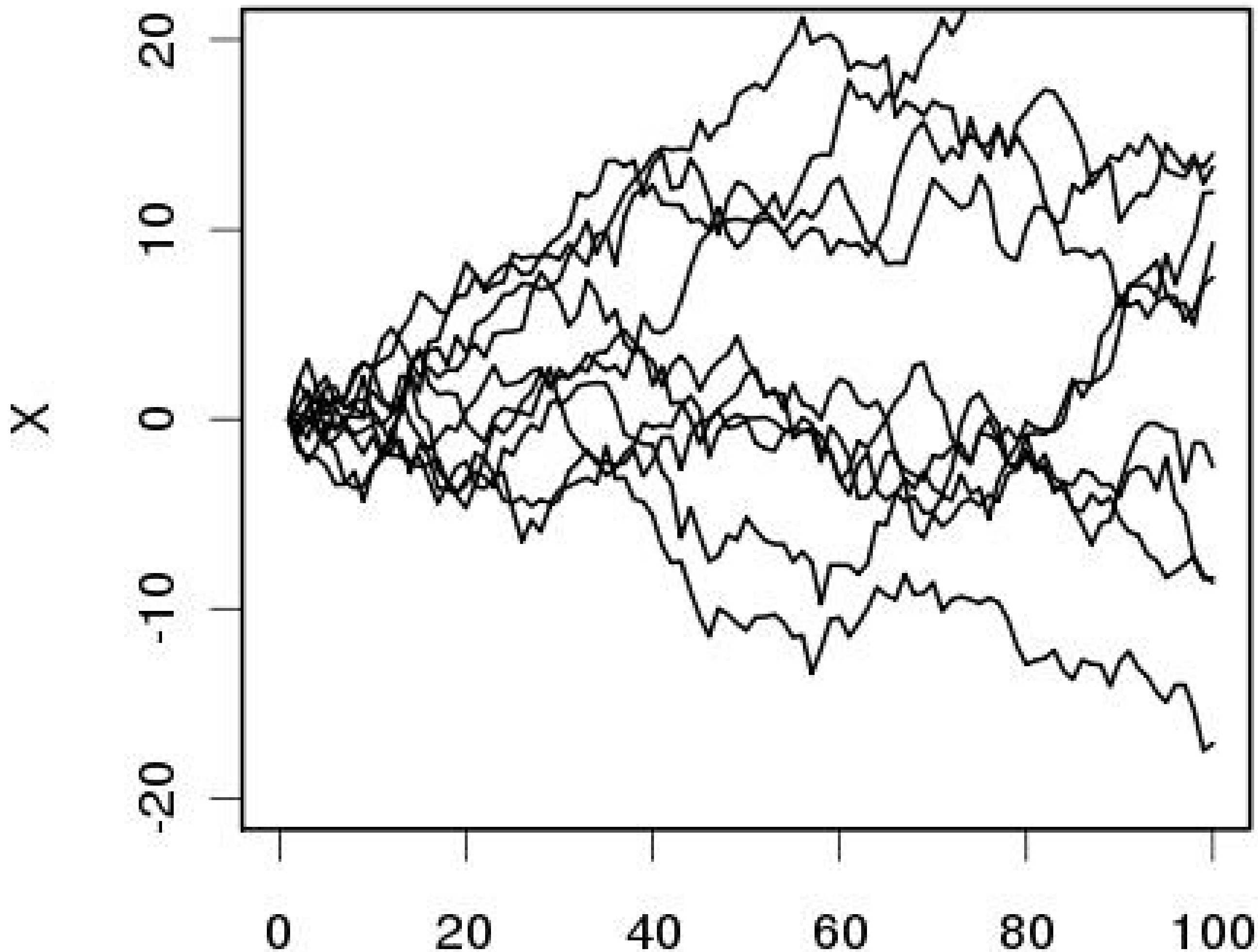
- Mean = X_0
- Var = $t\,\sigma^2$



Random Walk

$$X_t = X_{t-1} + \epsilon_t$$

- Mean = 0
- Var = t σ^2



<u>Approaches</u>

→ Random effects?

→ Autocorrelation?

→ State space?

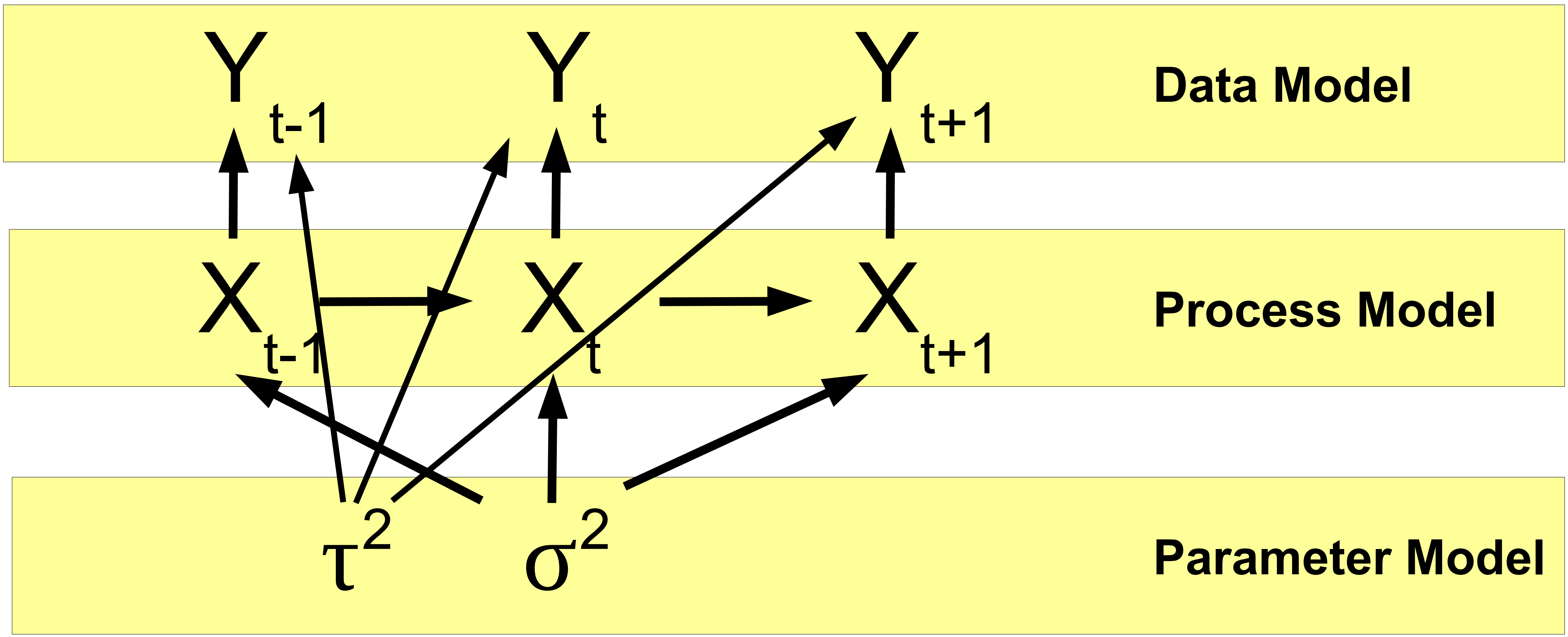
Bayesian State Space Model

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

$$Y_t = g(X_t) + \omega_t$$ **Data Model**

- X = <u>latent</u> time series
- Y = observed data
- ε = process error
- ω = observation error

Random Walk State Space Model



Y's are conditionally independent given the X's

Random Walk State Space Model

$$X_t \sim N(X_{t-1}, \sigma^2)$$ **Process Model**

$$Y_t \sim N(X_t, \tau^2)$$ **Data Model**

$$\sigma^2 \sim IG(s1, s2)$$ **Process Error prior**

$$\tau^2 \sim IG(t1, t2)$$ **Observation Error prior**

$$X_0 \sim N(X_{ic}, V_X)$$ **Initial Condition prior**

- What are the parameters?
- What is the joint (full) posterior?
- What are the conditional distributions for each parameter?

Random Walk State Space Model

$X_t \sim N(X_{t-1}, \sigma^2)$	**Process Model**
$Y_t \sim N(X_t, \tau^2)$	**Data Model**
$\sigma^2 \sim IG(s1, s2)$	**Process Error prior**
$\tau^2 \sim IG(t1, t2)$	**Observation Error prior**
$X_0 \sim N(X_{ic}, V_X)$	**Initial Condition prior**

- What are the parameters?
  - X's, σ^2, τ^2

Random Walk State Space Model

- What is the joint (full) posterior?

$$X_t \sim N(X_{t-1}, \sigma^2)$$

$$Y_t \sim N(X_t, \tau^2)$$

$$\sigma^2 \sim IG(s1, s2)$$

$$\tau^2 \sim IG(t1, t2)$$

$$X_0 \sim N(X_{ic}, V_X)$$

Random Walk State Space Model

- What is the joint (full) posterior?

$$p(\vec{X}, \sigma^2, \tau^2 | \vec{Y}, \ldots) \sim$$

$$\prod_{t=1}^{n} N(Y_t | X_t, \tau^2) \times$$

$$\prod_{t=1}^{n} N(X_t | X_{t-1}, \sigma^2) \times$$

$$IG(\tau^2 | t1, t2) \times IG(\sigma^2 | s1, s2) \times N(X_0 | X_{ic}, V_X)$$

$$X_t \sim N(X_{t-1}, \sigma^2)$$

$$Y_t \sim N(X_t, \tau^2)$$

$$\sigma^2 \sim IG(s1, s2)$$

$$\tau^2 \sim IG(t1, t2)$$

$$X_0 \sim N(X_{ic}, V_X)$$

Random Walk State Space Model

- What are the conditional distributions?

$$X_t \sim N(X_{t-1}, \sigma^2)$$
$$Y_t \sim N(X_t, \tau^2)$$
$$\sigma^2 \sim IG(s1, s2)$$
$$\tau^2 \sim IG(t1, t2)$$
$$X_0 \sim N(X_{ic}, V_X)$$

Random Walk State Space Model

- What are the conditional distributions?

$$\sigma^2 \sim IG(s1, s2) \times \prod_{t=1}^{n} N(X_t | X_{t-1}, \sigma^2)$$

$$\tau^2 \sim IG(\tau^2 | t1, t2) \times \prod_{t=1}^{n} N(Y_t | X_t, \tau^2)$$

$$X_t \sim N(X_{t-1}, \sigma^2)$$

$$Y_t \sim N(X_t, \tau^2)$$

$$\sigma^2 \sim IG(s1, s2)$$

$$\tau^2 \sim IG(t1, t2)$$

$$X_0 \sim N(X_{ic}, V_X)$$

Random Walk State Space Model

- What are the conditional distributions?

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

- Three special cases
  - First
  - Last
  - Missing Y

$$X_t \sim N(X_{t-1}, \sigma^2)$$
$$Y_t \sim N(X_t, \tau^2)$$
$$\sigma^2 \sim IG(s1, s2)$$
$$\tau^2 \sim IG(t1, t2)$$
$$X_0 \sim N(X_{ic}, V_X)$$

Random Walk State Space Model

- First

$$X_0 \sim N(X_0 | X_{ic}, V_X) \times N(X_1 | X_{0,} \sigma^2)$$

- Last

$$X_n \sim N(X_n | X_{n-1}, \sigma^2) \times N(Y_n | X_n, \tau^2)$$

- Missing Y

$$X_t \sim N(X_t | X_{t-1}, \sigma^2) \times N(X_{t+1} | X_t, \sigma^2)$$

$$X_t \sim N(X_{t-1}, \sigma^2)$$
$$Y_t \sim N(X_t, \tau^2)$$
$$\sigma^2 \sim IG(s1, s2)$$
$$\tau^2 \sim IG(t1, t2)$$
$$X_0 \sim N(X_{ic}, V_X)$$

How do we sample?

- All amenable to Gibbs sampling (NxN, NxIG)
- X's need to be updated sequentially because each X is conditioned on the one BEFORE it and the one AFTER it

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

Bayesian State Space Model

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

$$Y_t = g(X_t) + \omega_t$$ **Data Model**

- Y = observed data
- X = latent time series
- ε = process error
- ω = observation error

Population growth

$$\frac{dN}{dt} = rN$$

Exponential Growth

$$N_t = N_0 e^{rt}$$

Let X = log(N)

$$X_t = X_{t-1} + r + \epsilon_t$$

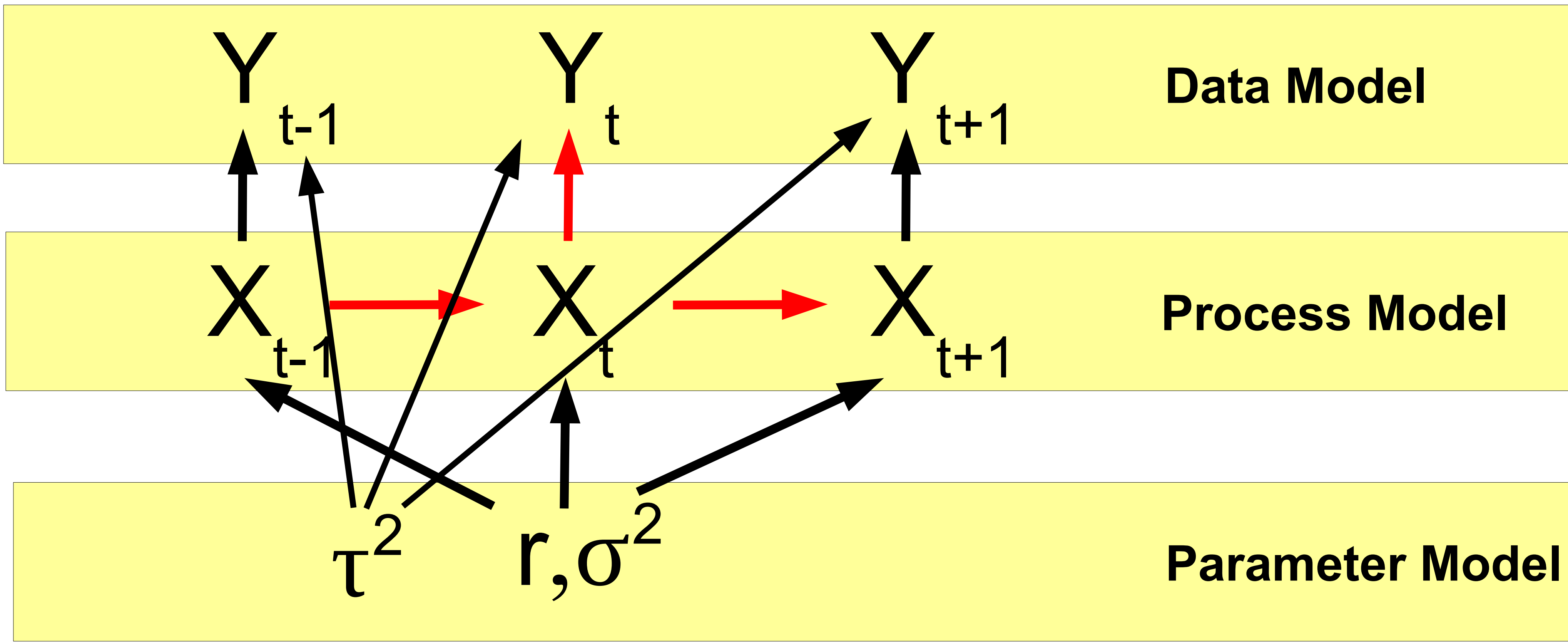
Discrete time recursion

$$Y_t = X_t + \omega_t$$

Observation error model

- r = intrinsic growth rate
- Y = log(observation)
- Can interpret ε as a random effect on r

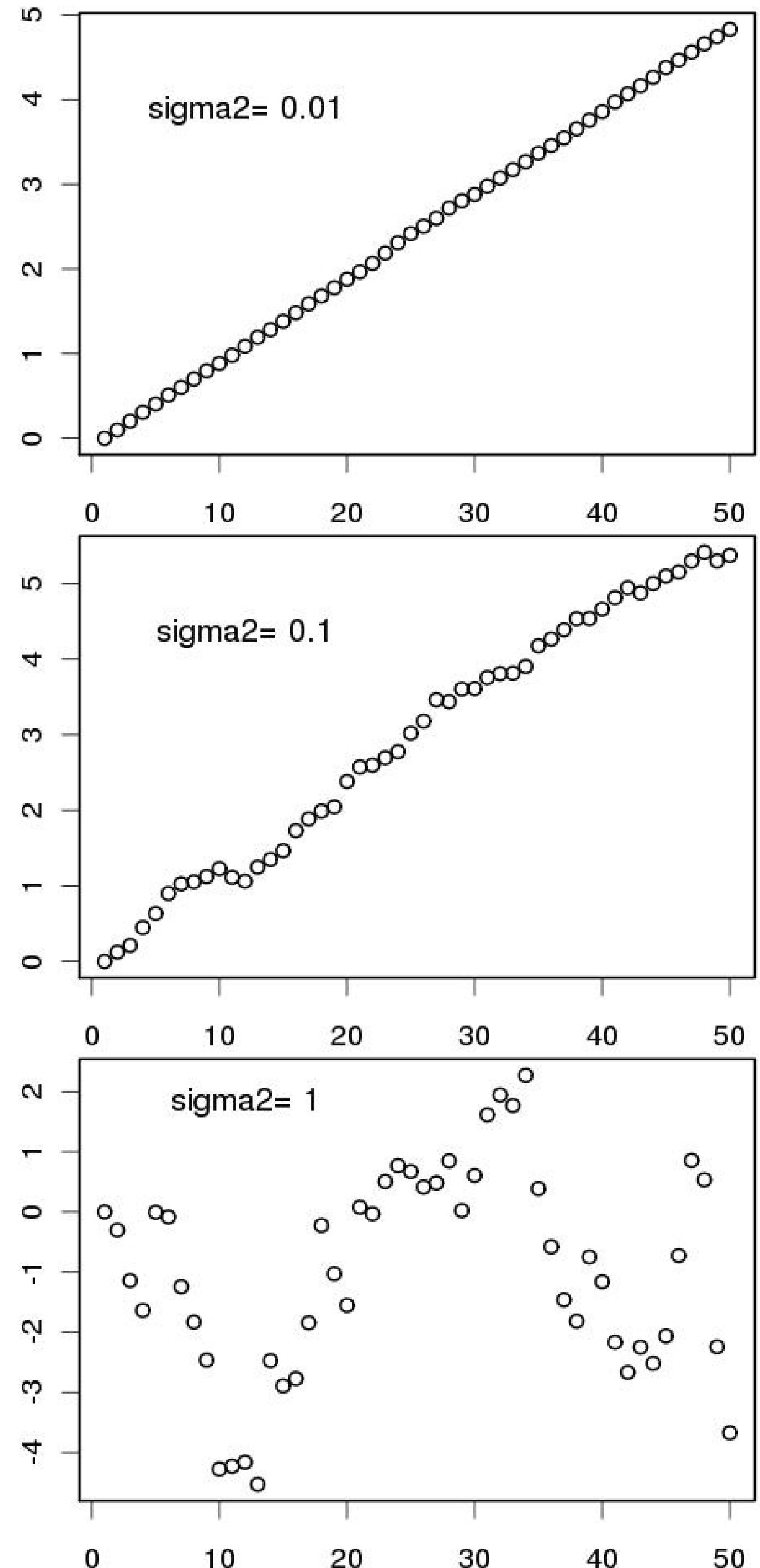
Exponential Growth State Space

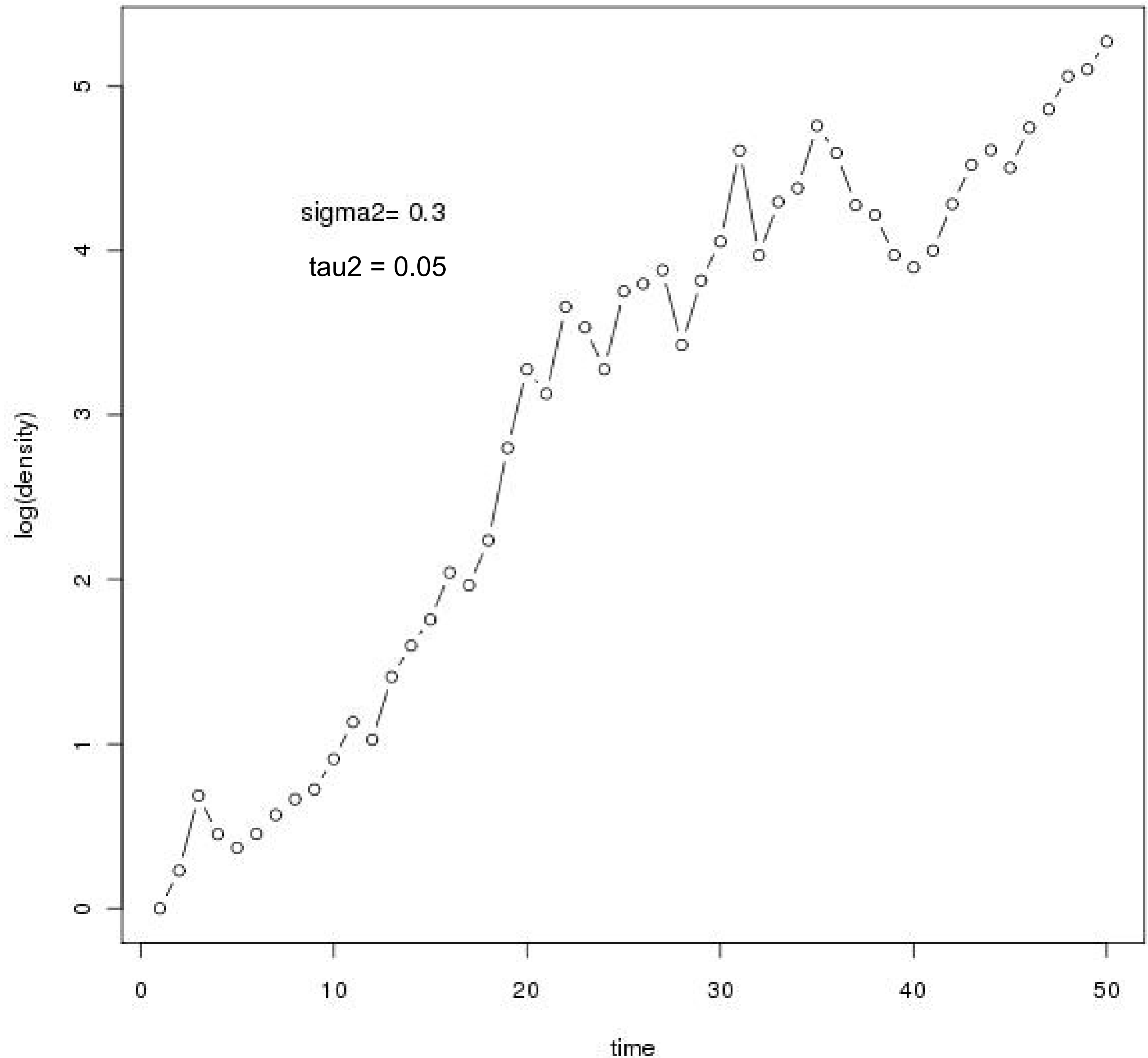


Y's are conditionally independent given the X's

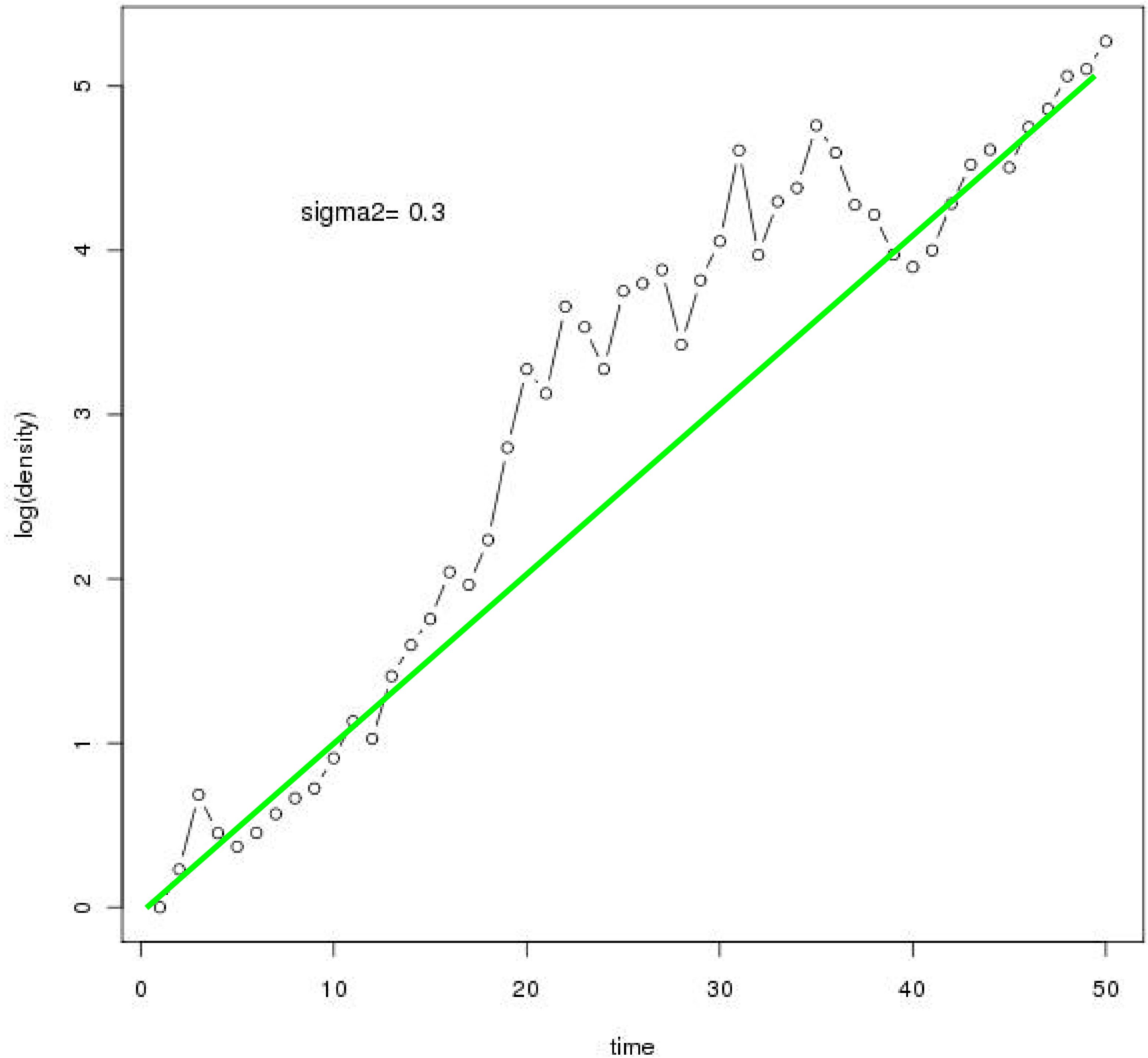
Example

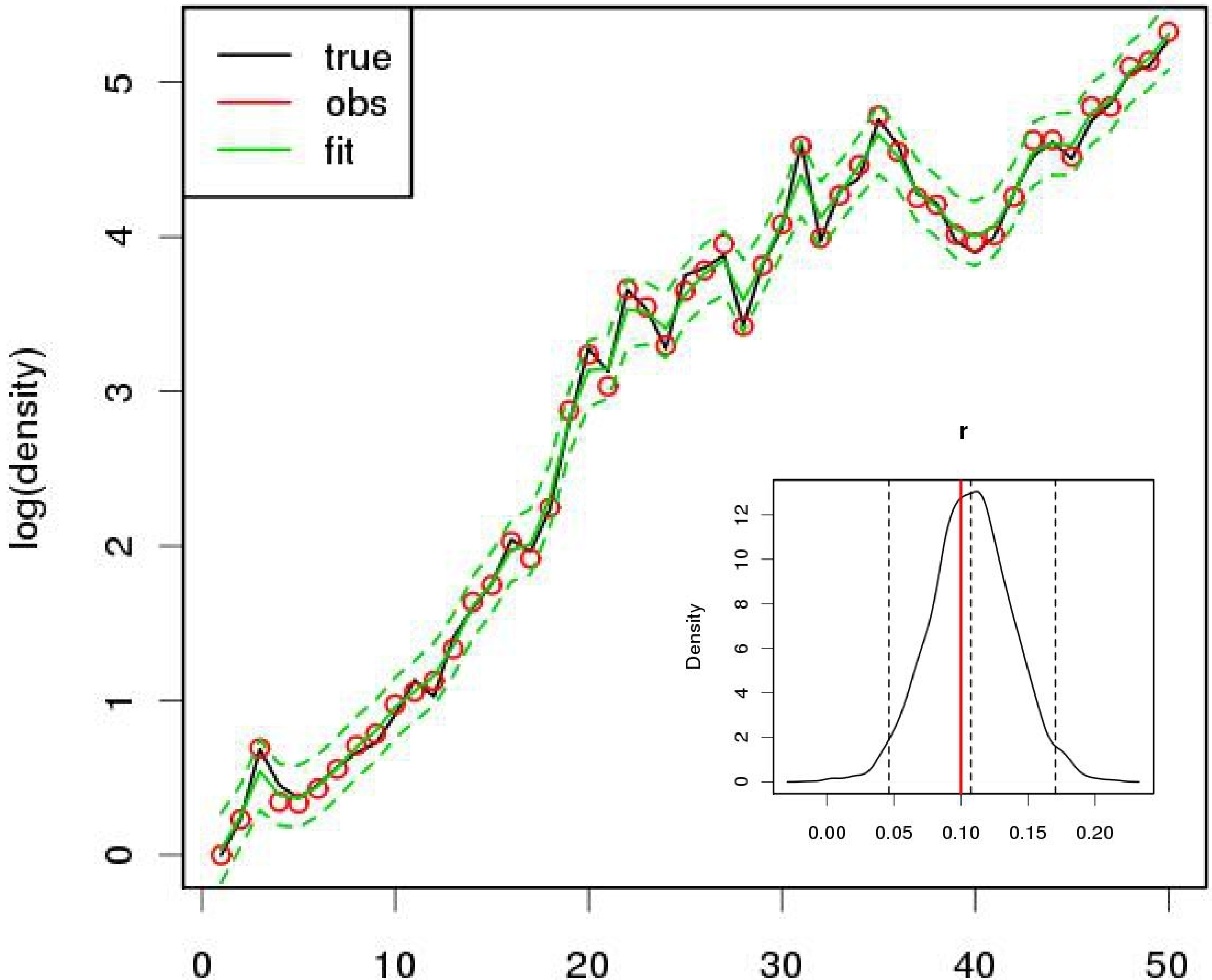
- r = 0.1
- Process s.d. varied from 0.01 to 1.0
- Run 1 = dominated by process model
- Run 2 = process model and process error similar
- Run 3 = dominated by process error
  - Env or wrong model



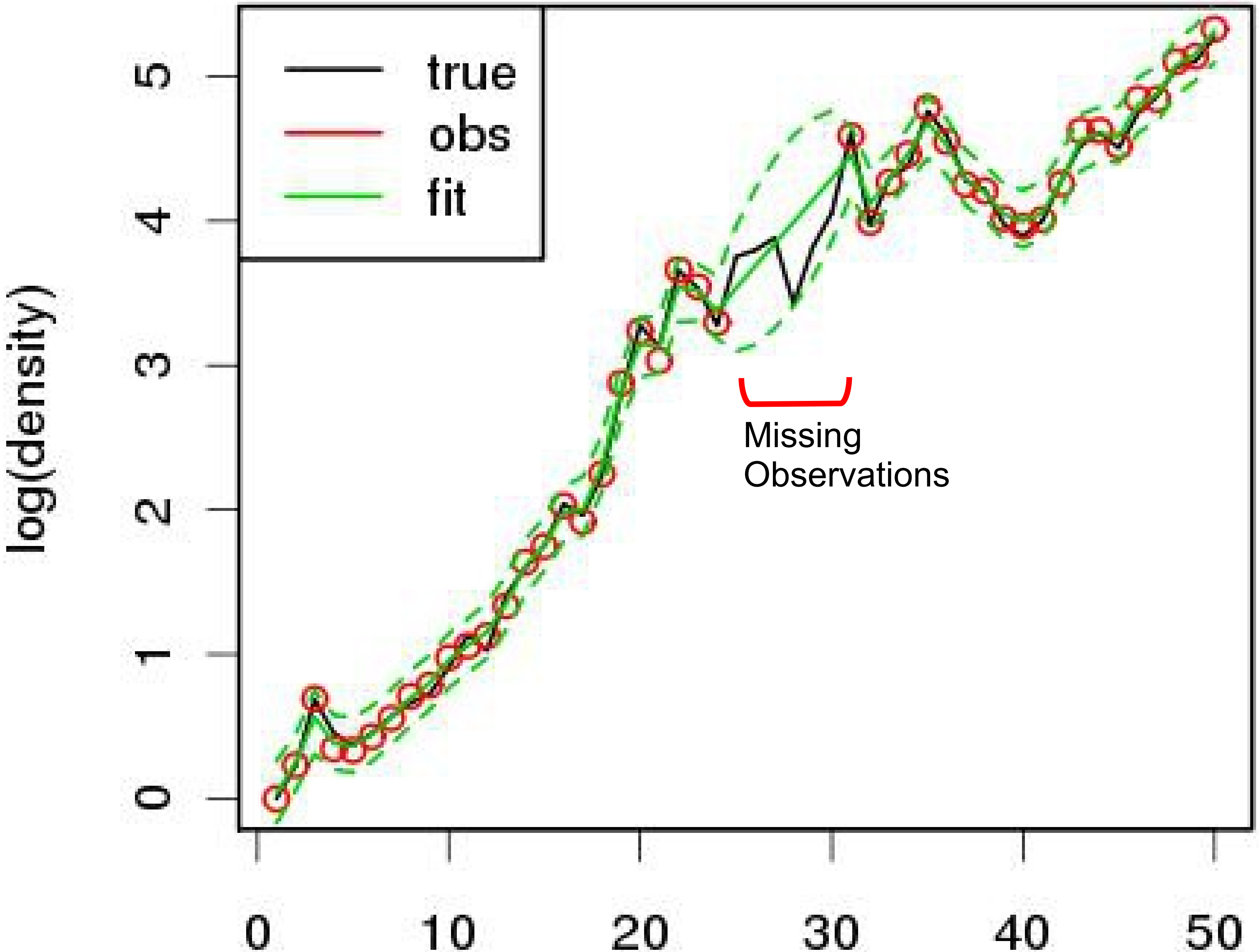


sigma2= 0.3
tau2 = 0.05
log(density)
time
0
1
2
3
4
5
0
10
20
30
40
50

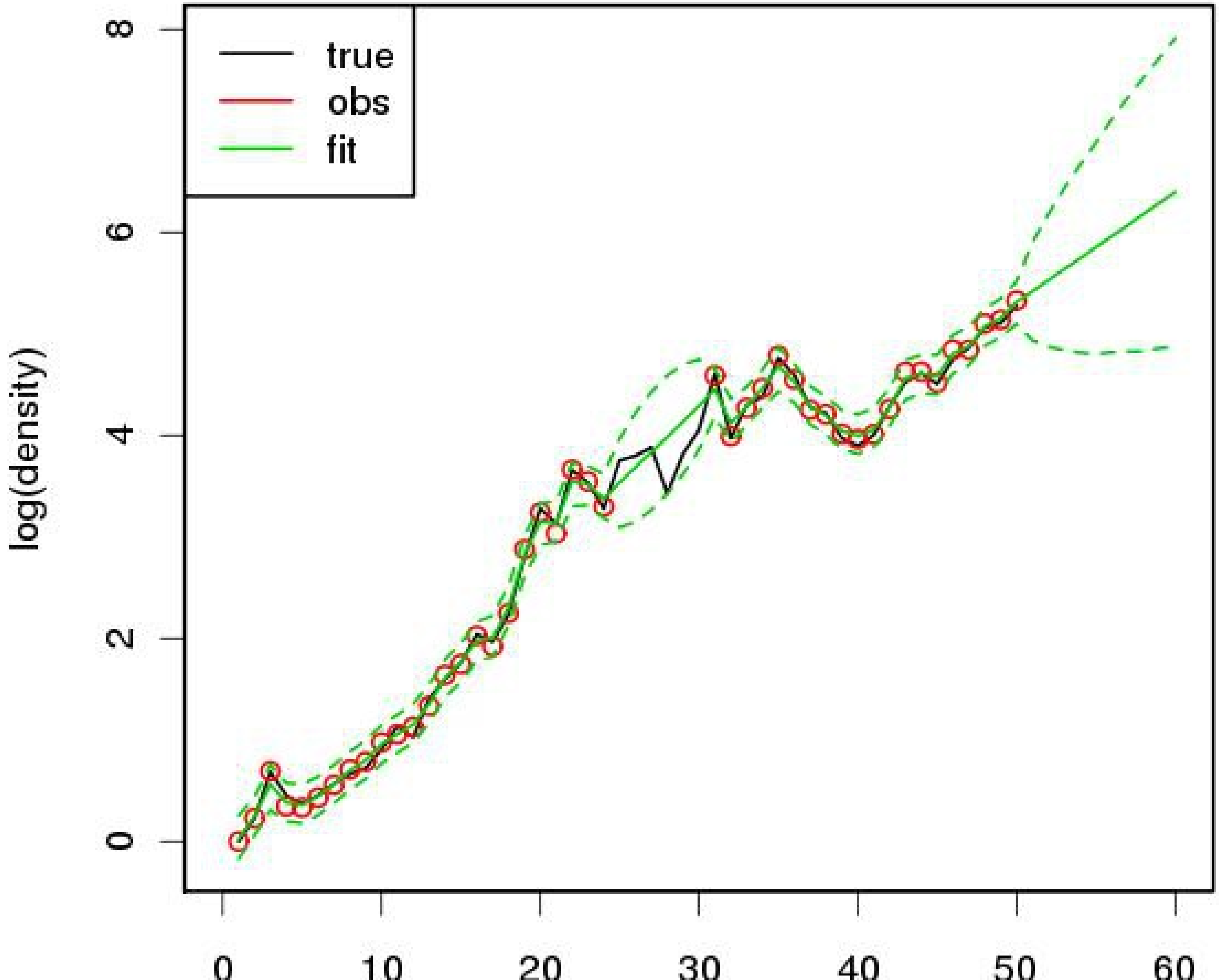

sigma2= 0.3
log(density)
time
0
1
2
3
4
5
0
10
20
30
40
50


true
obs
fit
log(density)
0
1
2
3
4
5
0
10
20
30
40
50
r
Density
0
2
4
6
8
10
12
0.00
0.05
0.10
0.15
0.20

Missing Data



Prediction



Generality of the State Space Model

- 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
- Easily handles missing data (gaps) and irregularly spaced data
- Easily handles multiple data sources (Y's), which don't need to be the same type or synchronous
- Easily handles time-integrated observations