Stochastic Models & Parameter Estimation

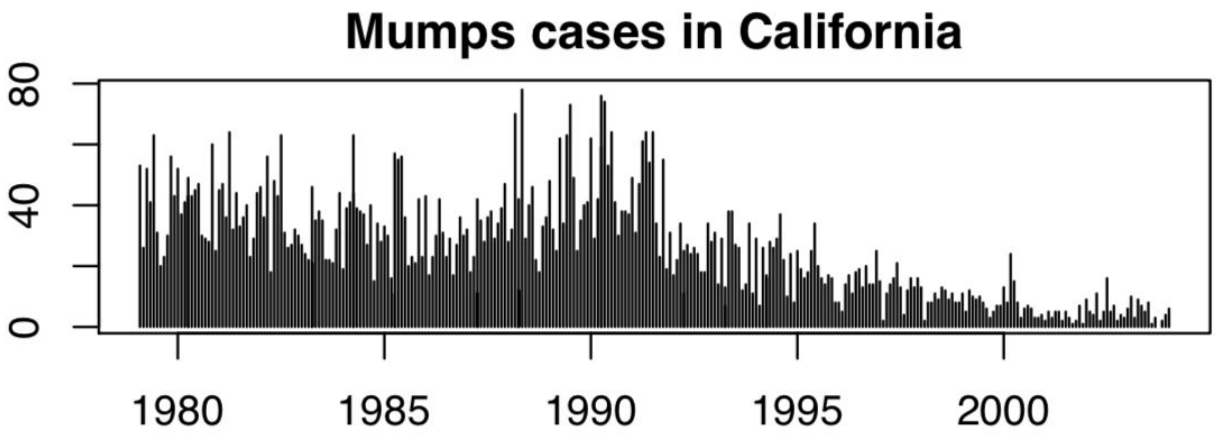
Micaela E. Martinez

Emory University

Epidemiological Data are Noisy

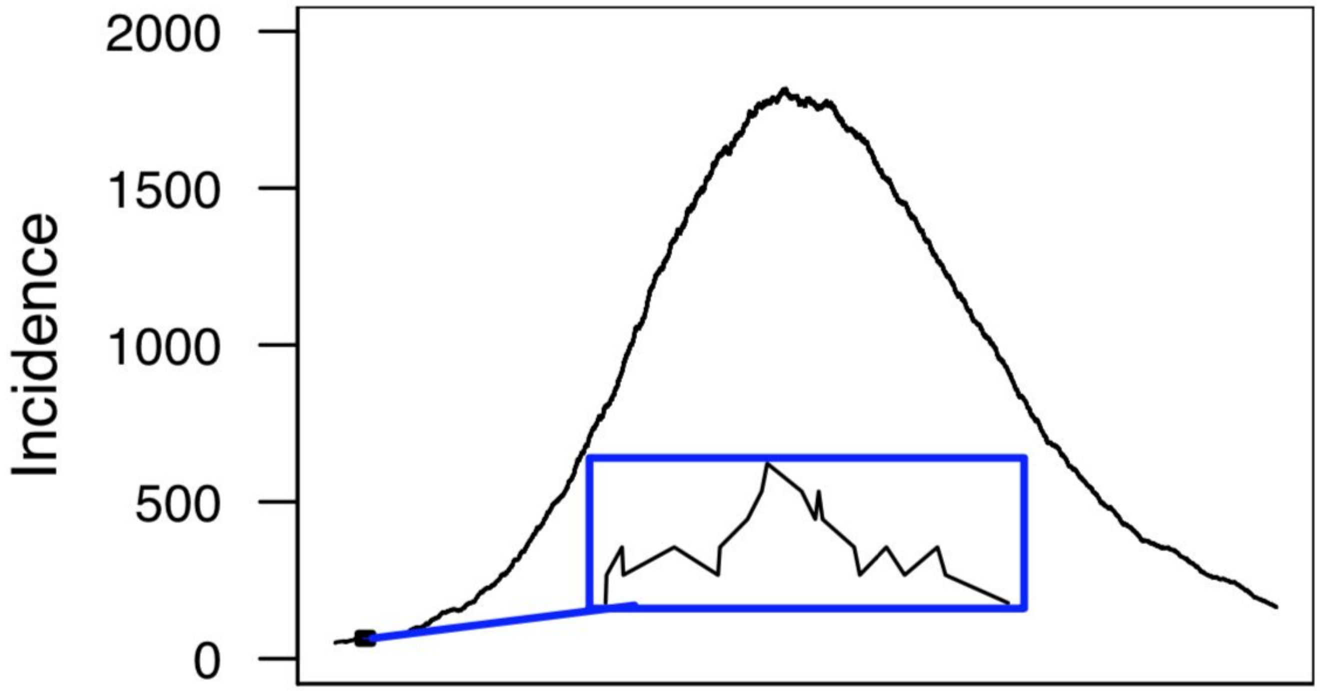
Two types of noise:

- Observation error: the data are probabilistically related to the true state of the system
- Process noise: the system progresses probabilistically
  - Environmental noise: some parameter is a random variable
  - Demographic noise: individual-level chance events



Noise is addressed using stochastic models

The SIR Model is an Approximation



The *SIR* model (e.g., $dY/dt = \beta XY/N - \gamma Y$) implies that changes in the states X, Y, and Z are continuous. But, in reality individuals are either susceptible, infected, or recovered so that X, Y, and Z are integer-valued and changes in the system state occur as discrete steps. The differential equation is an idealization.

NEW YORK STATE | Department of Health

NYSDOH COVID-19 Tracker

Testing data as of: 7/13/2020
Testing data last updated on:
(Updated daily bef

Statewide

Total Persons Tested
4,784,927

Total Tested 7/13
60,045

Total Tested Positive
403,175

Sex Distribution of Positive Cases

Female	Male	Unknown
48.7%	50.7%	0.6%

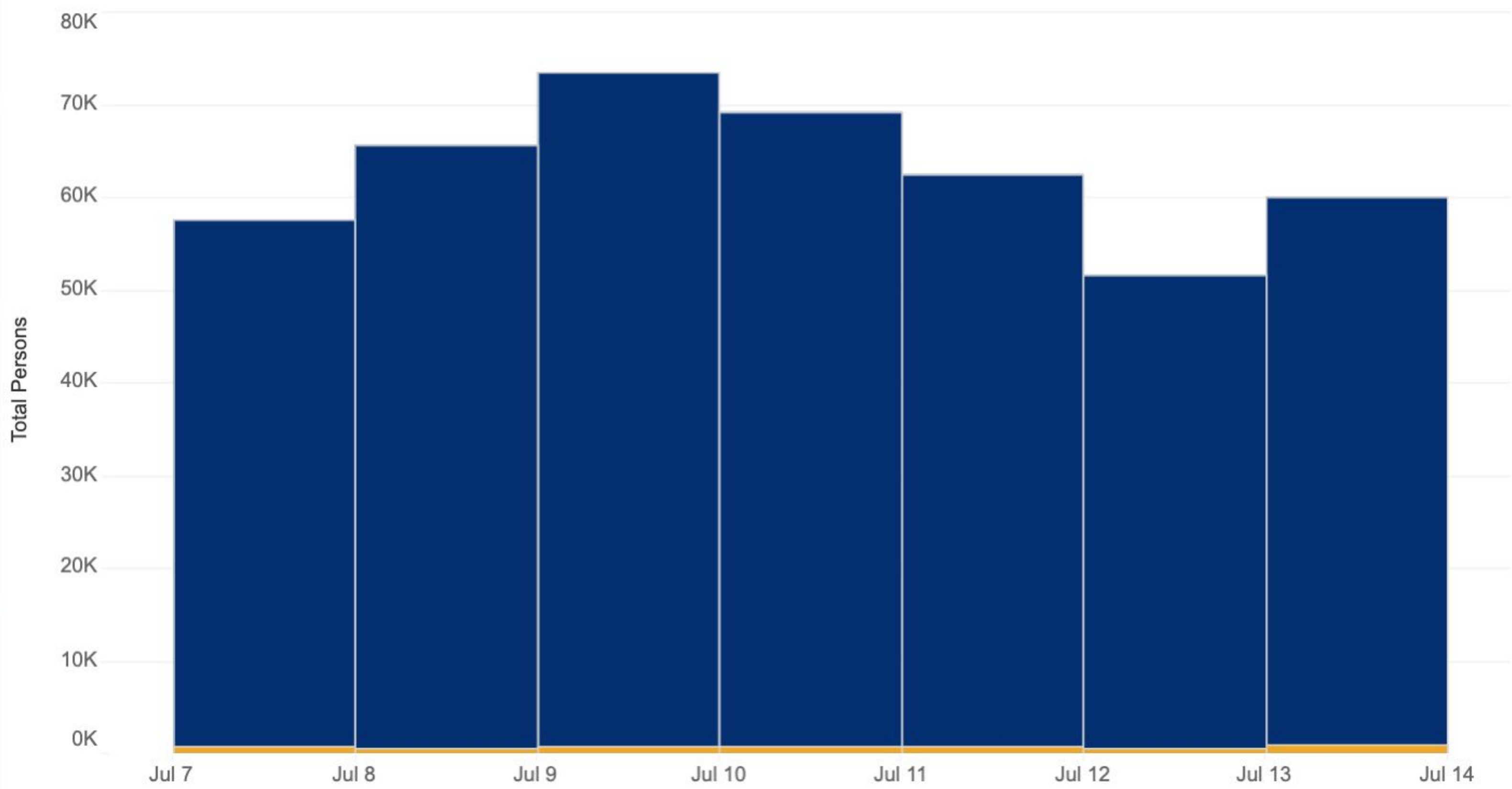
New Positives 7/13
912

Daily Totals: Persons Tested and Persons Tested Positive

Total Persons Tested   Total Tested Positive

Time Period
Last 7 Days

Hover over a bar to see details



Click County to See Detail
Click Again for Statewide

- Albany
- Allegany
- Bronx
- Broome
- Cattaraugus
- Cayuga
- Chautauqua
- Chemung
- Chenango
- Clinton
- Columbia
- Cortland
- Delaware
- Dutchess
- Erie
- Essex
- Franklin

Click for Map View

Click for Table View

Click for Fatality Data

Statewide

Total Persons Tested
4,784,927

Total Tested 7/13
60,045

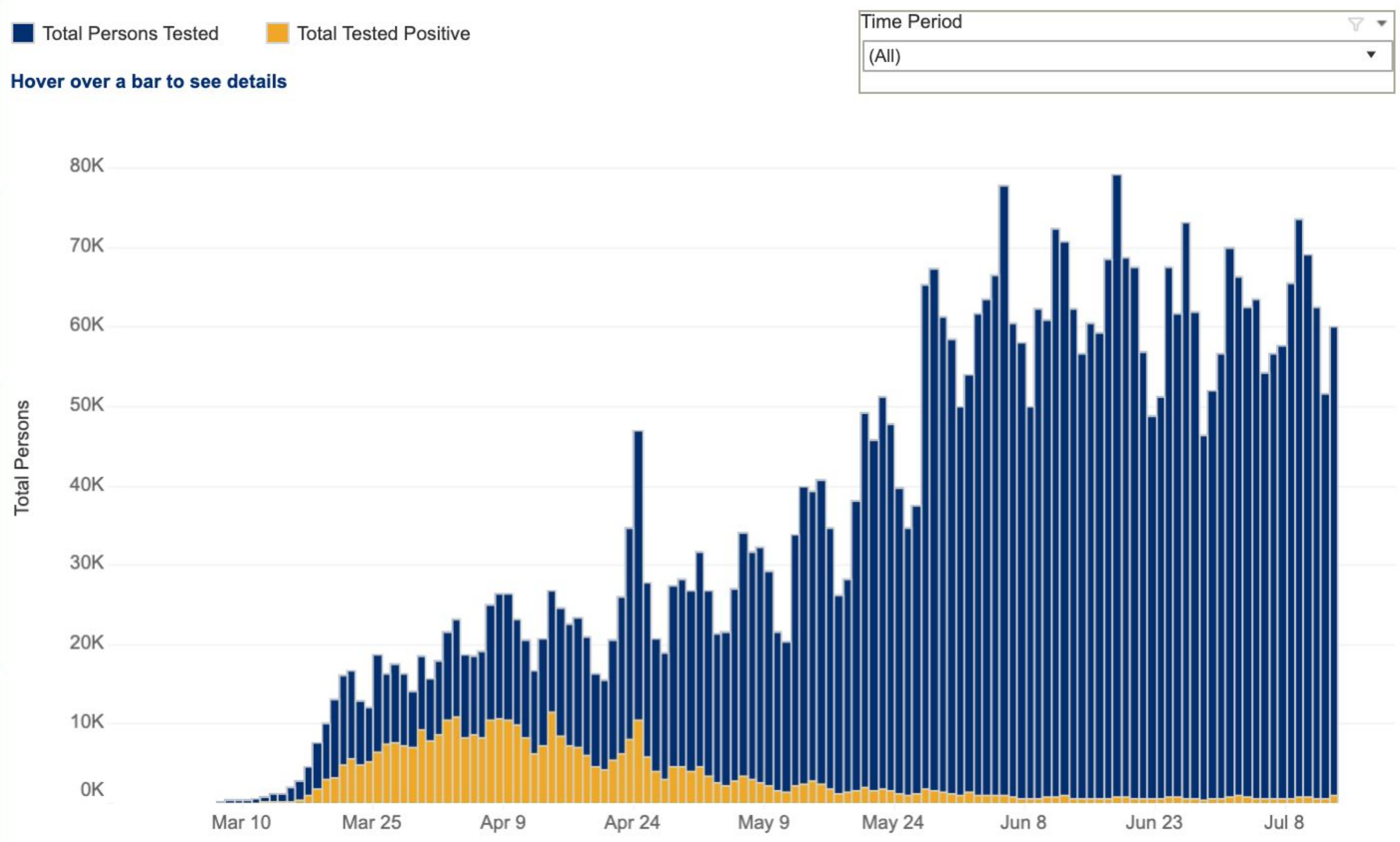
Total Tested Positive
403,175

Sex Distribution of Positive Cases

Female	Male	Unknown
48.7%	50.7%	0.6%

New Positives 7/13
912

Daily Totals: Persons Tested and Persons Tested Positive



Click County to See Detail
Click Again for Statewide

Albany	2,247
Allegany	68
Bronx	48,433
Broome	841
Cattaraugus	139
Cayuga	127
Chautauqua	180
Chemung	149
Chenango	171
Clinton	110
Columbia	489
Cortland	59
Delaware	92
Dutchess	4,318
Erie	7,833
Essex	51
Franklin	36

Click for Map View

Click for Table View

Click for Fatality Data

Deterministic Models

Deterministic models run like “clockwork”, given the same starting conditions, exactly-the-same trajectory will always be observed

The SIR Model is an Approximation

- Transmission is obscured by three sources of noise: observation error, environmental variability, and intrinsic demographic noise

- Demographic noise is especially important in systems where $R0 \approx 1$

The SIR Model is an Approximation

- Transmission is obscured by three sources of noise: <u>observation error</u>, environmental variability, and intrinsic demographic noise

- Not all infections are symptomatic

- Not all symptomatic infections reported

BUSINESS INSIDER

Subscribe

40% of people infected with COVID-19 are asymptomatic, a new CDC estimate says

Ellen Cranley Jul 12, 2020, 2:36 PM

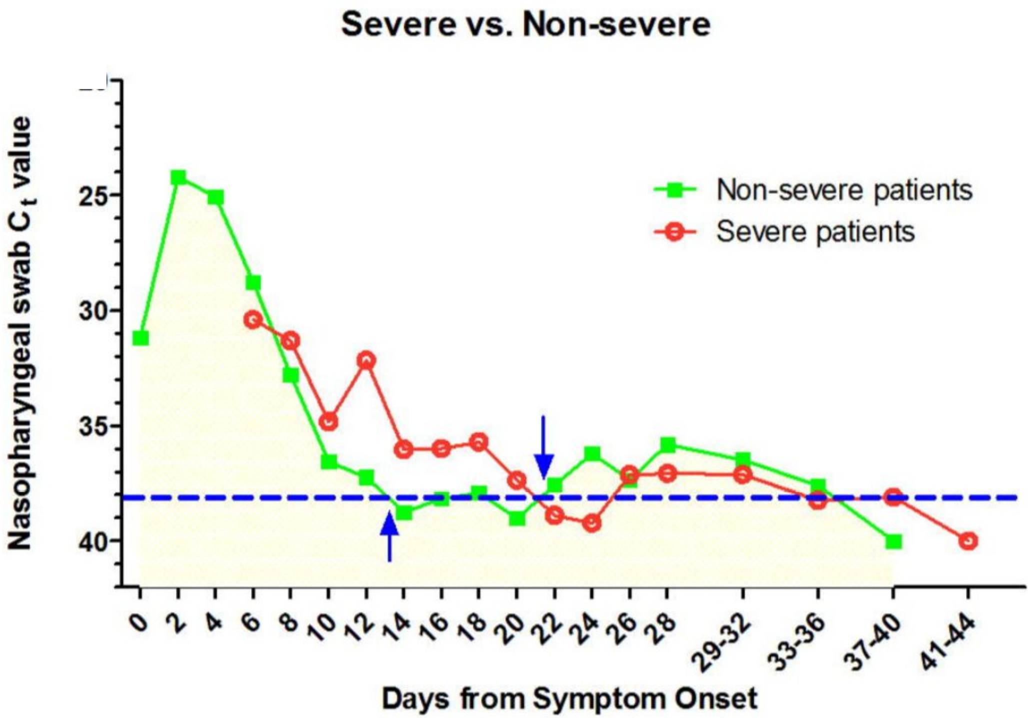
A nurse prepares to swab a patient at a COVID19 testing center on July 7, 2020 in Austin, Texas. Sergio Flores/Getty Images

The SIR Model is an Approximation

- Transmission is obscured by three sources of noise: observation error, environmental variability, and intrinsic demographic noise

- Variation among individuals can impact parameters

- Variation in environment can impact parameters



Tan et al. MedRXIV 2020

The SIR Model is an Approximation

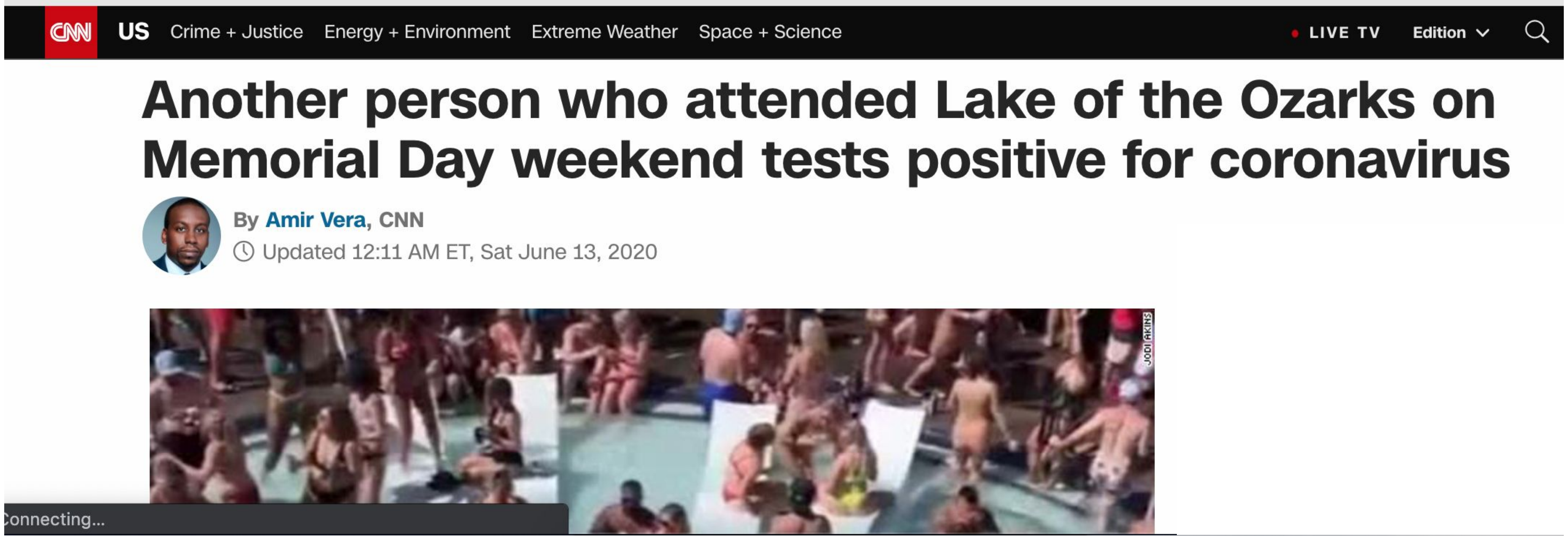
- Transmission is obscured by three sources of noise: observation error, environmental variability, and intrinsic demographic noise

- Think of coin flips to conceptualize demographic stochasticity. The more flips, the more precision you have to approximate the mean

The Real World

If it were possible to “re-run” an epidemic in the real-world, we would not expect to have exactly-the-same people become infected at exactly-the-same time.

Why is this?



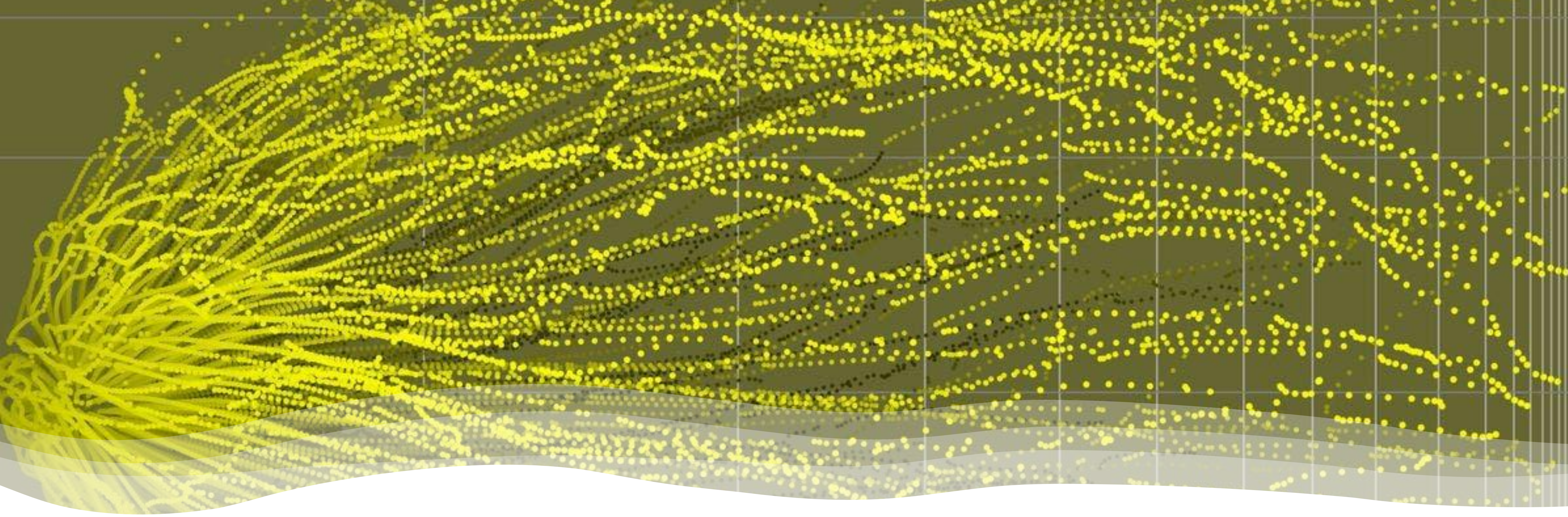
CNN US Crime + Justice Energy + Environment Extreme Weather Space + Science LIVE TV Edition

Another person who attended Lake of the Ozarks on Memorial Day weekend tests positive for coronavirus

By Amir Vera, CNN

Updated 12:11 AM ET, Sat June 13, 2020

Connecting...



Stochastic Models

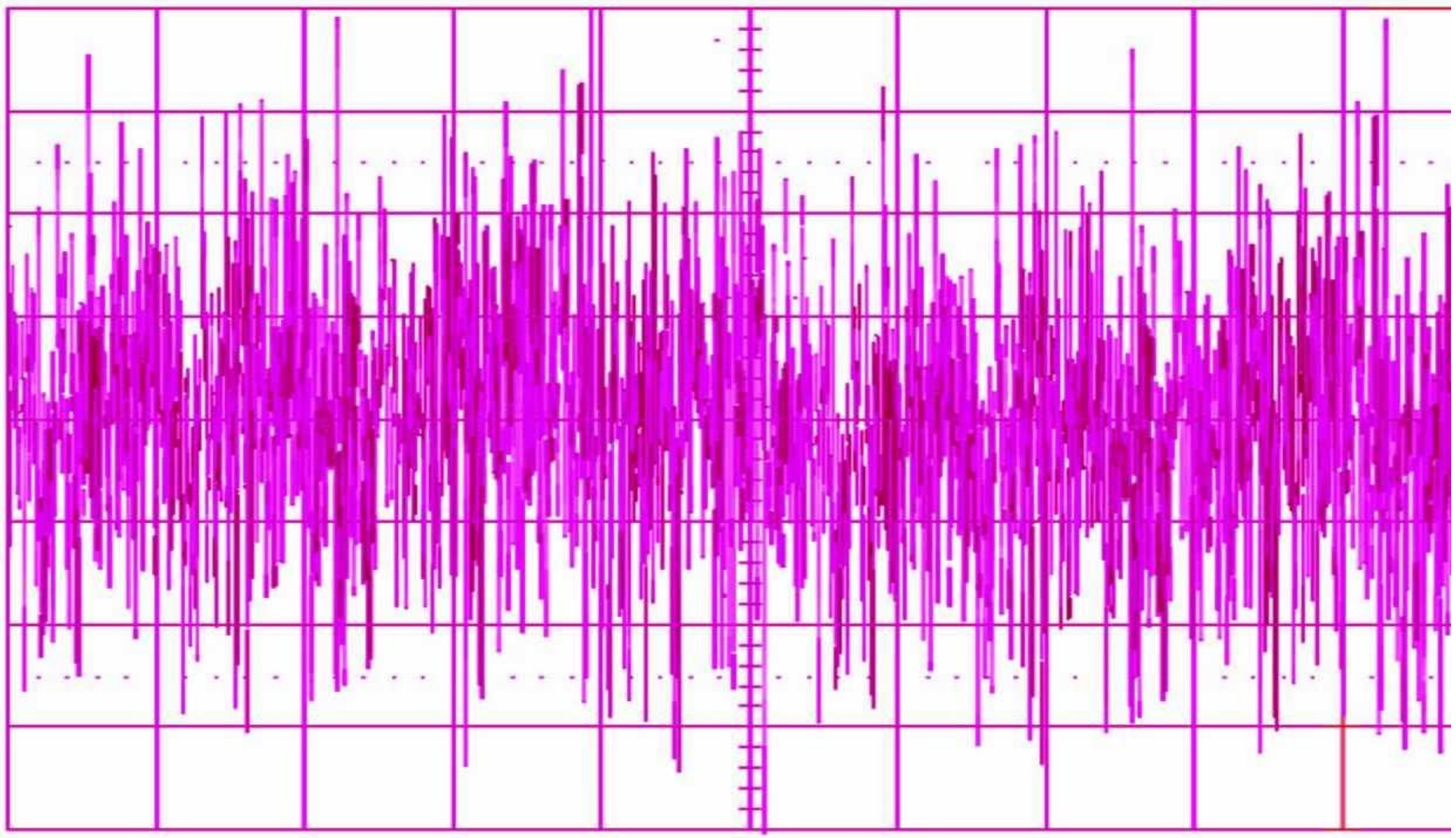
Stochastic models aim to capture some of the random and probabilistic features of the real-world.

Stochastic Models

Stochastic models aim to capture some of the random and probabilistic features of the real-world.

Stochasticity has the largest effect when:

- # infected is small
- population size is small
- when the infection has just invaded
- during the trough phase of an epidemic
- and when control measures are successfully applied



Stochastic Models

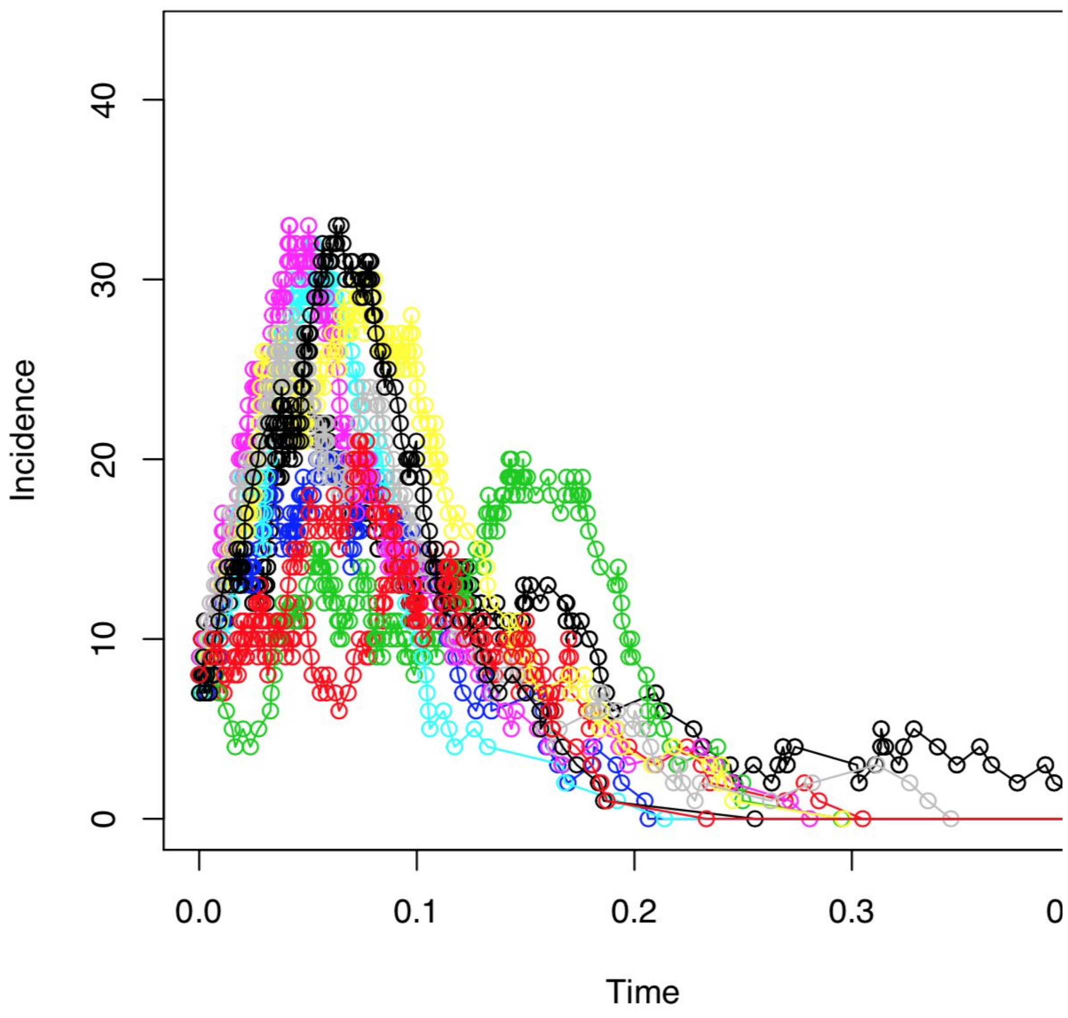
- Stochastic models aim to capture some of the random and probabilistic features of the real-world.

- Stochastic models use random number generators, for example:

  - random time-step length for events to occur
  - parameter values pulled randomly from distributions
  - reported cases pulled randomly from distributions with the mean being I_t*(mean report rate)*
  - Multiplicative white or pink noise on the force of infection (*beta** I_t)

Stochastic Models: Variability Between Simulations

Variability between simulations are the most obvious element of stochastic models. Mean and variance may accurately be predicted for simulations. However, since each simulation is different, it is generally impossible to predetermine the precise disease prevalence at any time in the future.



Goodness-of-Fit for Stochastic Models

- We focus on random process that (putatively) generated data
- A model is explicit, mathematical description of this random process
- "The likelihood" is probability that data were produced given model and its parameters:

$$L(\text{model} \mid \text{data}) = \Pr(\text{data} \mid \text{model})$$

- Likelihood quantifies (in some sense optimally) model goodness of fit

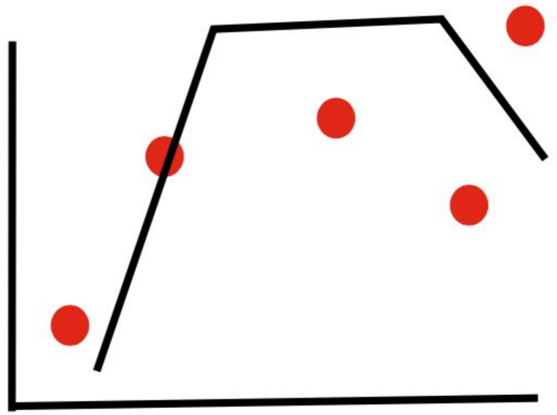
Likelihood Estimated for Stochastic Models

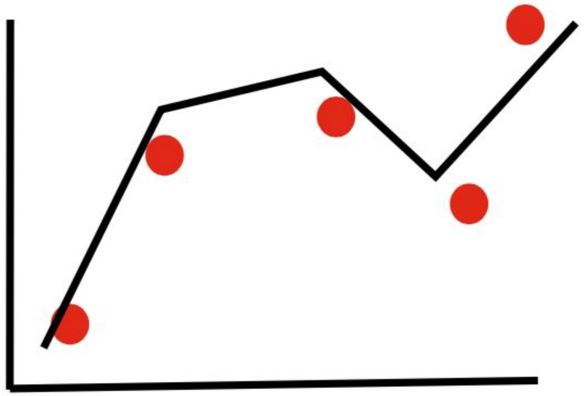
- Assume we have data, D, and model output, M (both are vectors containing state variables). Model predictions generated using set of parameters, θ

- Transmission dynamics subject to
  - "process noise": heterogeneity among individuals, random differences in timing of discrete events (environmental and demographic stochasticity)
  - "observation noise": random errors made in measurement process itself

Trajectory Matching

- If we ignore process noise, then model is deterministic and all variability attributed to measurement error
- Observation errors assumed to be sequentially independent
- Maximizing likelihood in this context is called 'trajectory matching'





Likelihood Estimation (with no process noise)

- Data, D
- Model output, M
- Parameters, θ

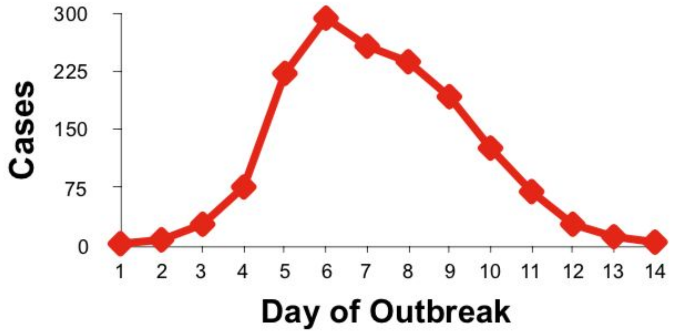
- If we assume measurement errors are normally distributed, with mean μ and variance σ^2 then

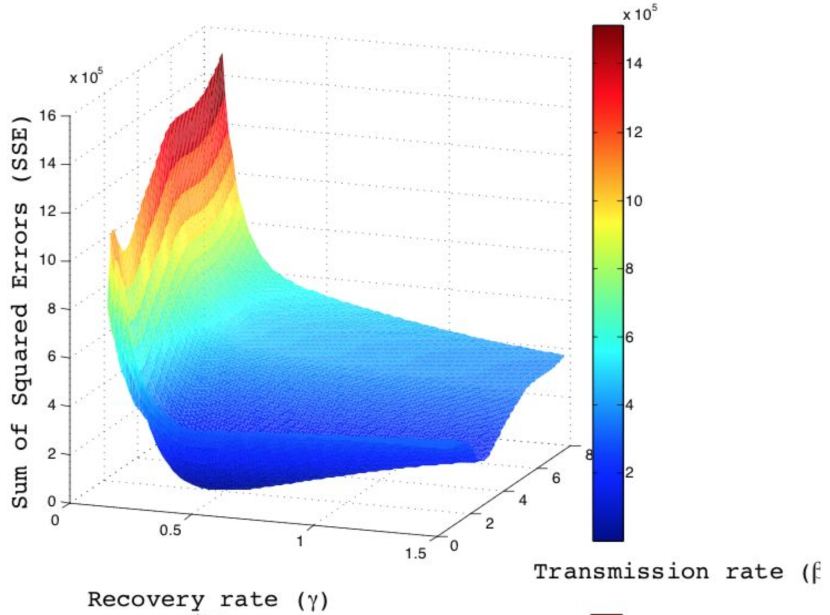
$$L(M(\theta) \mid D) = \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(D_i - M_i)^2}{2\sigma^2}}$$

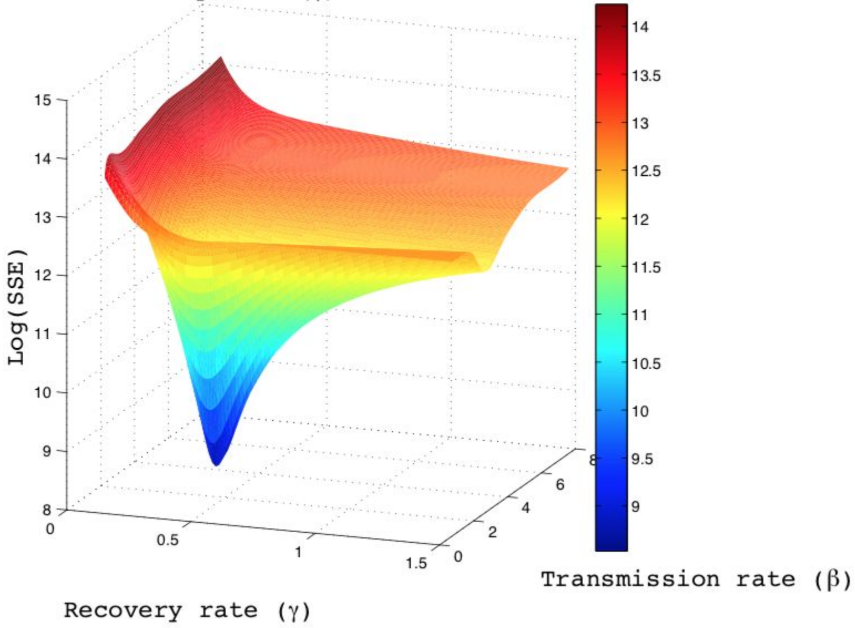
Likelihood Estimation

- Under such conditions, Maximum Likelihood Estimate, MLE, is simply parameter set with smallest deviation from data
- Equivalent to using least square errors, to decide on goodness of fit
  - Least Squares Statistic = SSE = $\Sigma(D_i - M_i)^2$
- Then, minimize SSE to arrive at MLE

Parameter Estimation: Influenza Outbreak



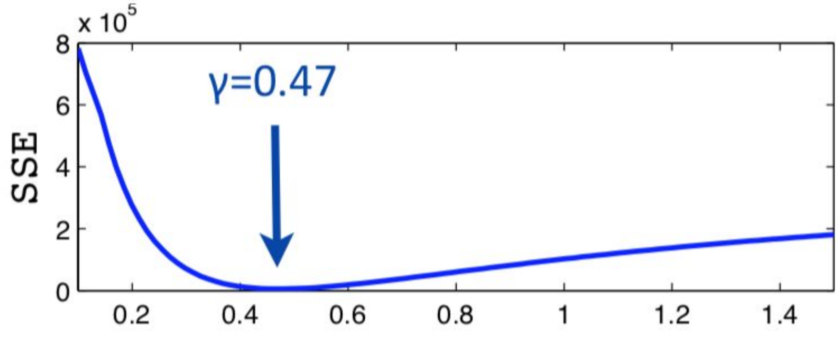




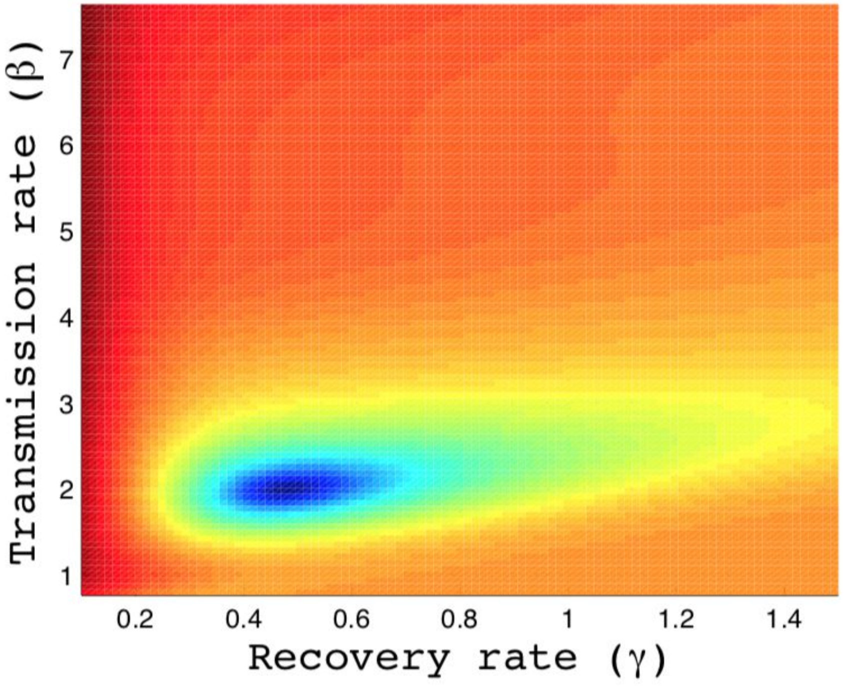
- Systematically vary β and γ, calculate SSE

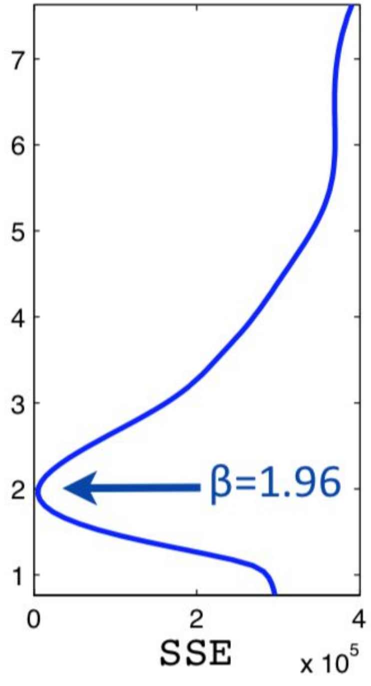
- Parameter combination with lowest SSE is 'best fit'

Parameter Estimation: Influenza Outbreak



Best fit parameter values:
β = 1.96 (per day)
$1/\gamma$ = 2.1 days
R_0 ~ 4.15





Generally, may have more parameters to fit, so grid search not efficient

Nonlinear optimization algorithms (eg Nelder-Mead) would be used

SSE vs. Log Likelihood

- How do we relate SSE to logLik?

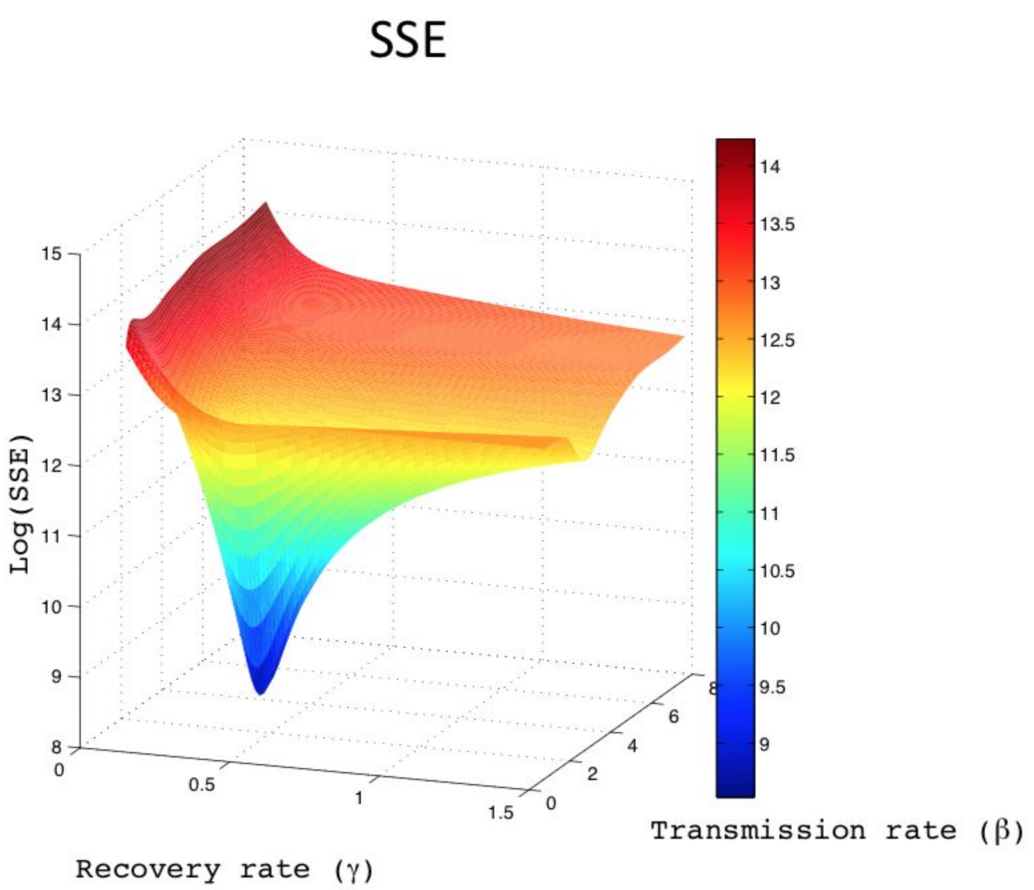
=length of data

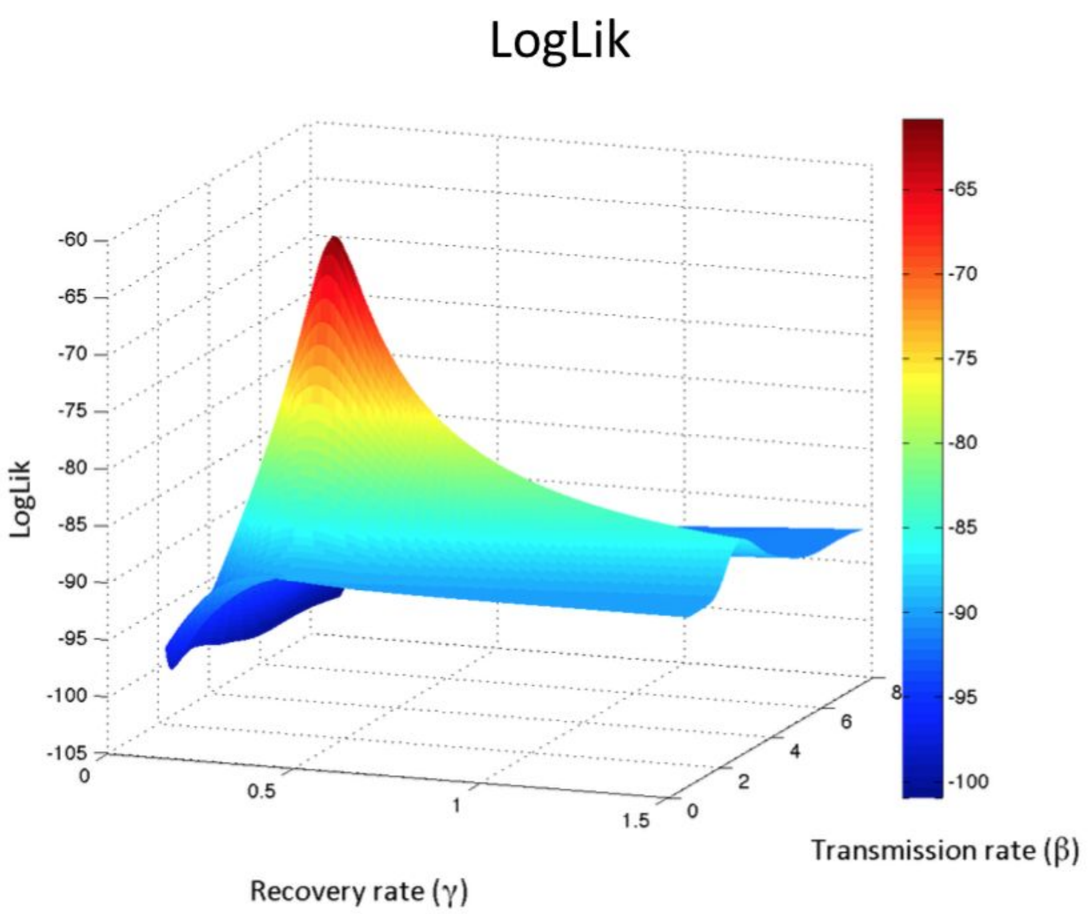
$$\log\left(L(M(\theta) \mid D)\right) = -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_i (D_i - M_i)^2$$

=SSE/n

=SSE

SSE vs. Log Likelihood





Surfaces often Complex