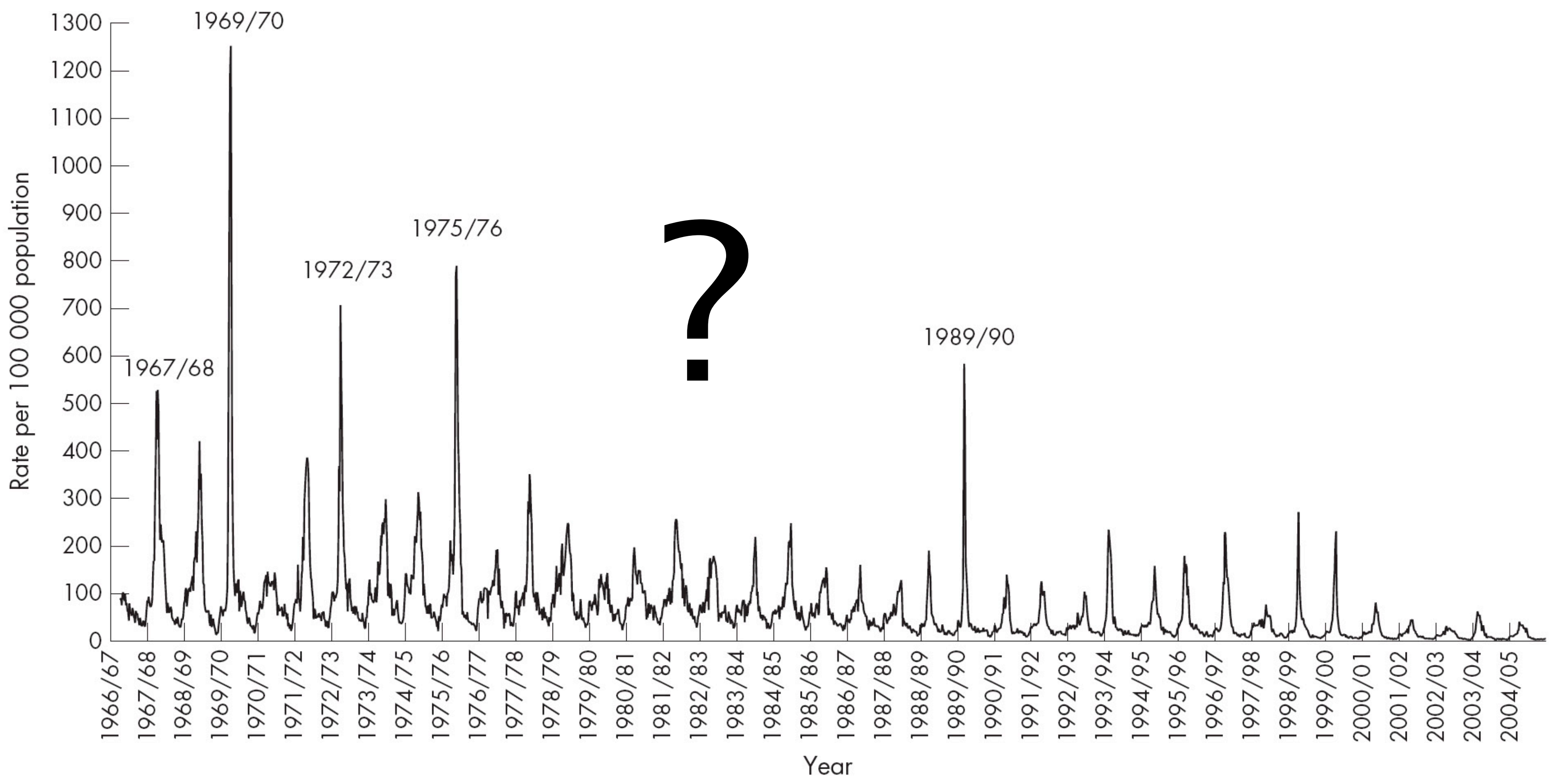
Time series analysis via maximum likelihood

From theory to practice

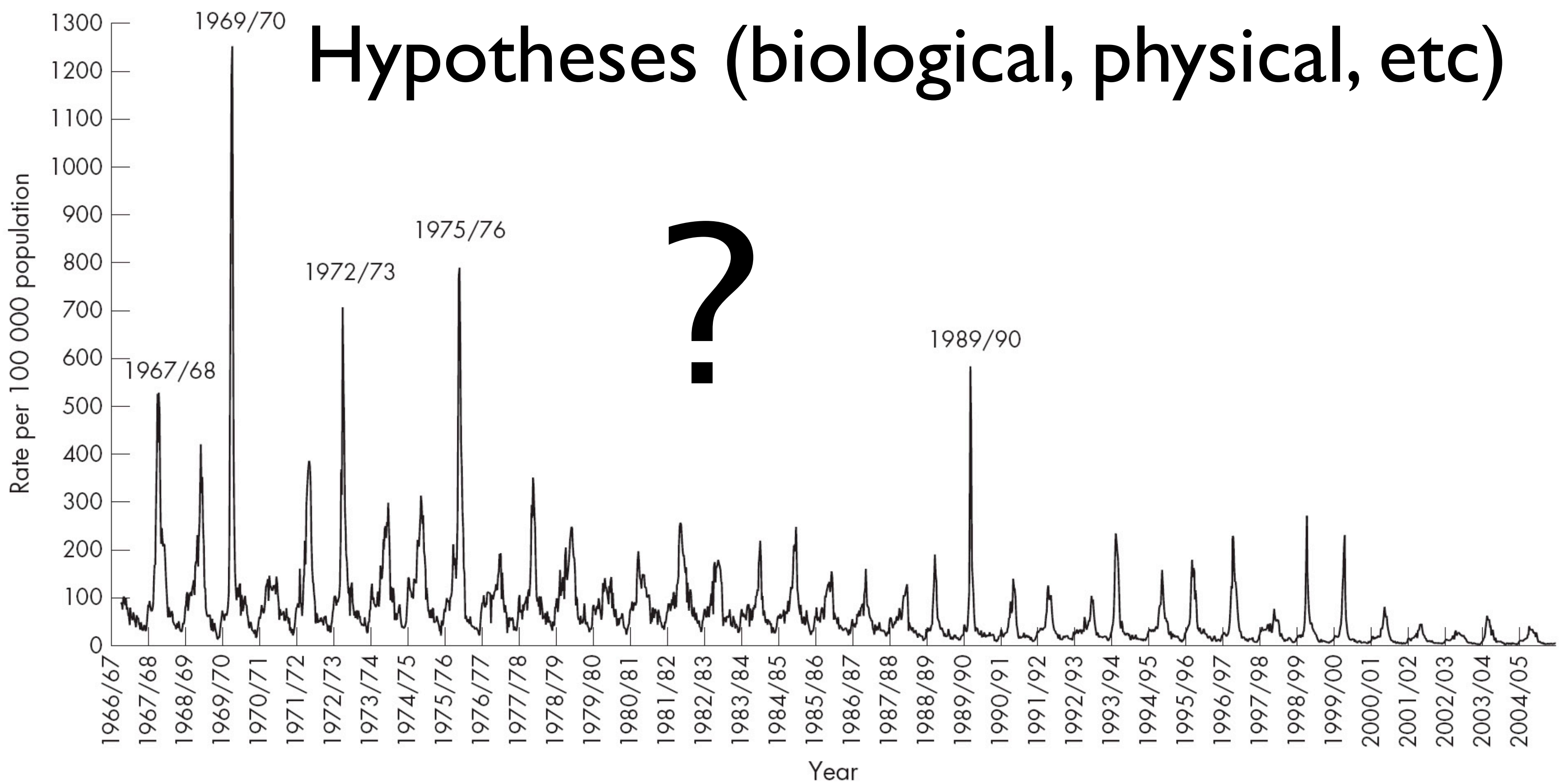
Anton Camacho & Bernard Cazelles

Department of Biology
UMR 7625, UPMC-CNRS-ENS
camacho@biologie.ens.fr

Time series analysis



Time series analysis

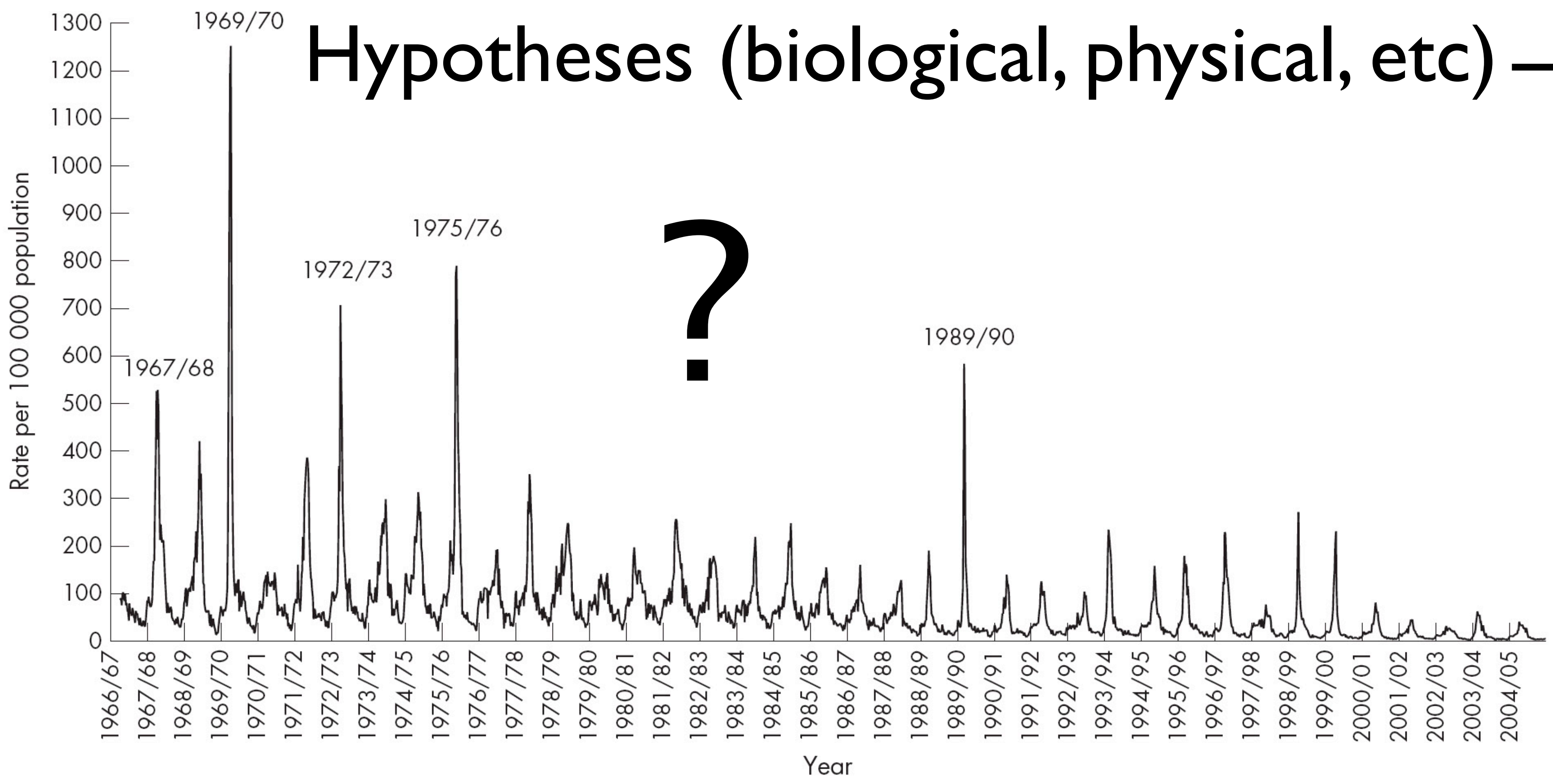


Hypotheses (biological, physical, etc)

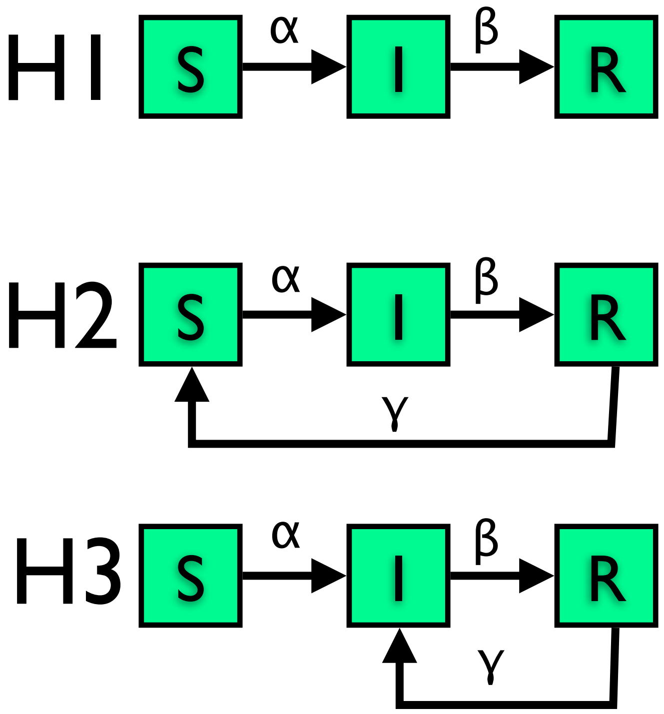
?

Time series analysis

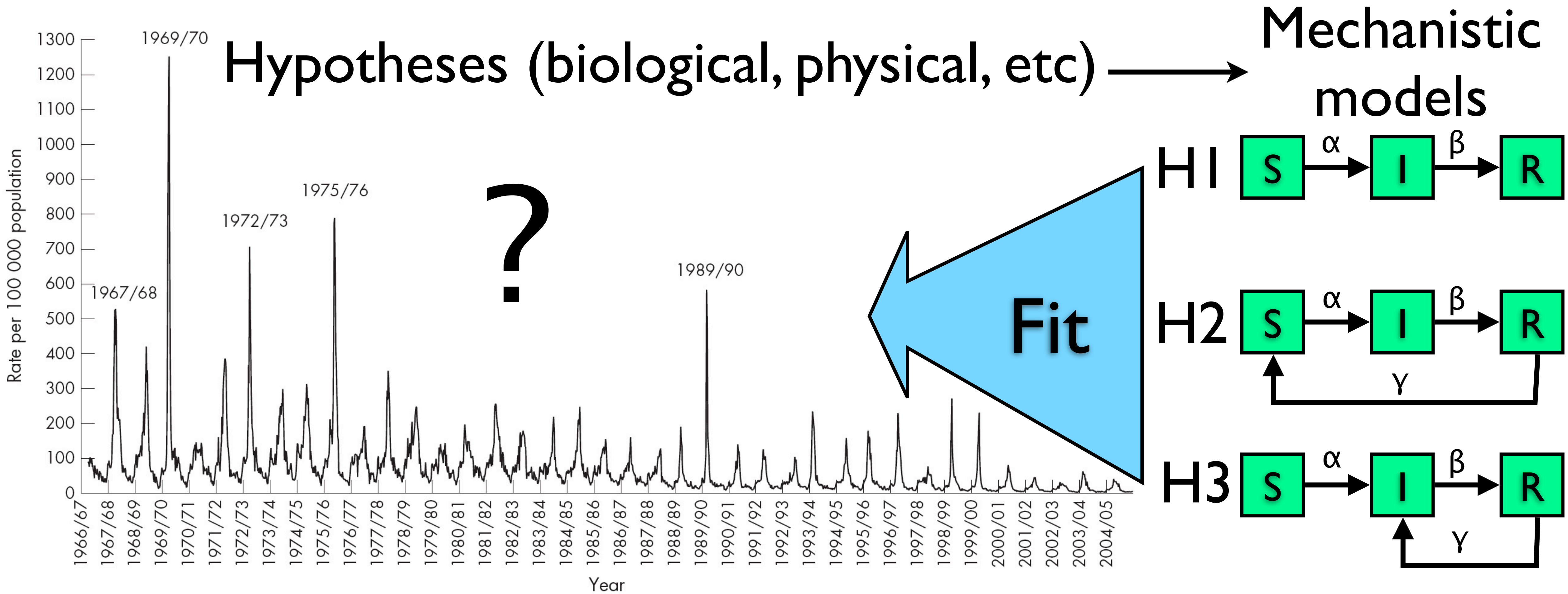
Hypotheses (biological, physical, etc) → Mechanistic models



?



Time series analysis



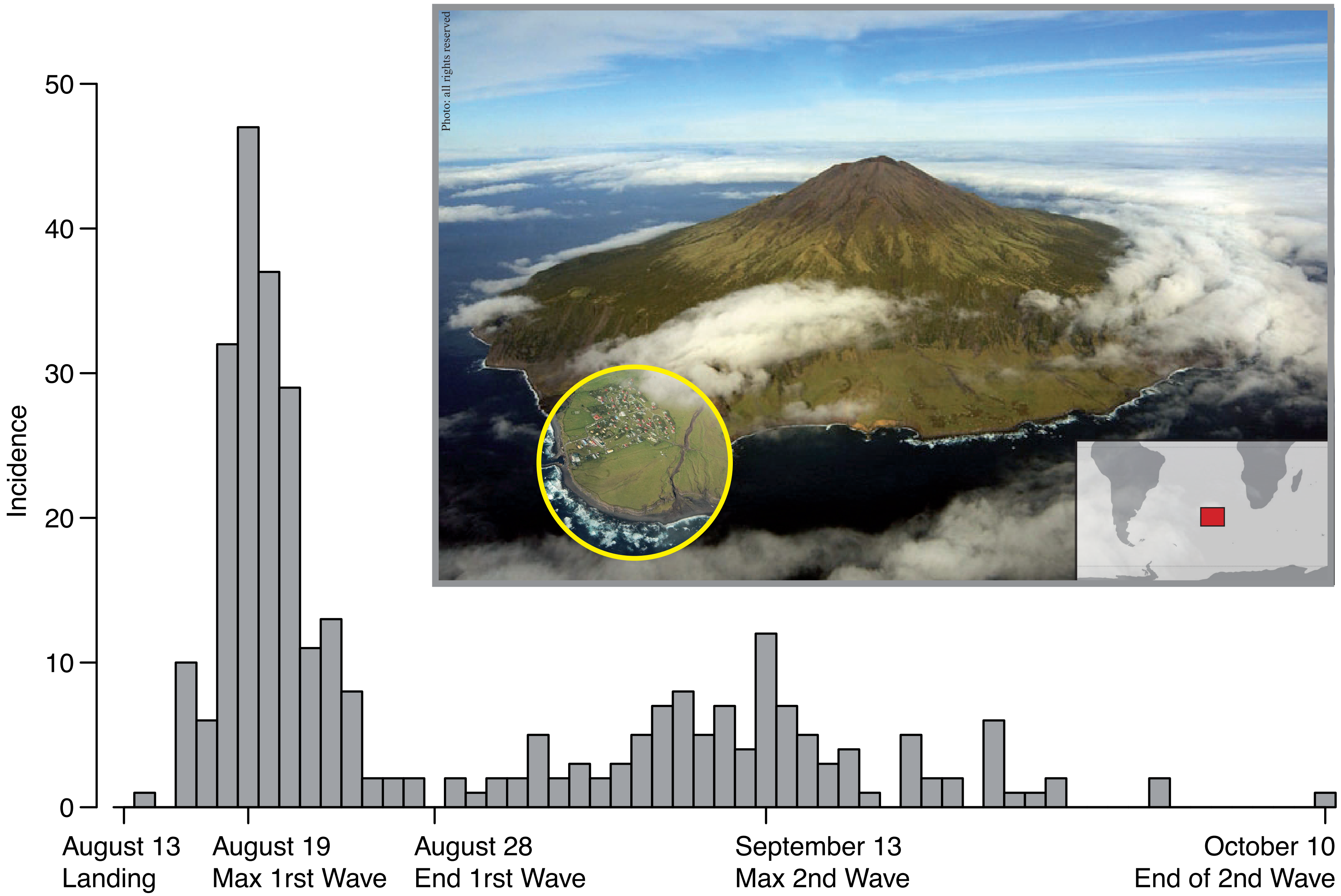
Parameter inference:

Identifiability, maximum likelihood estimates, confidence intervals

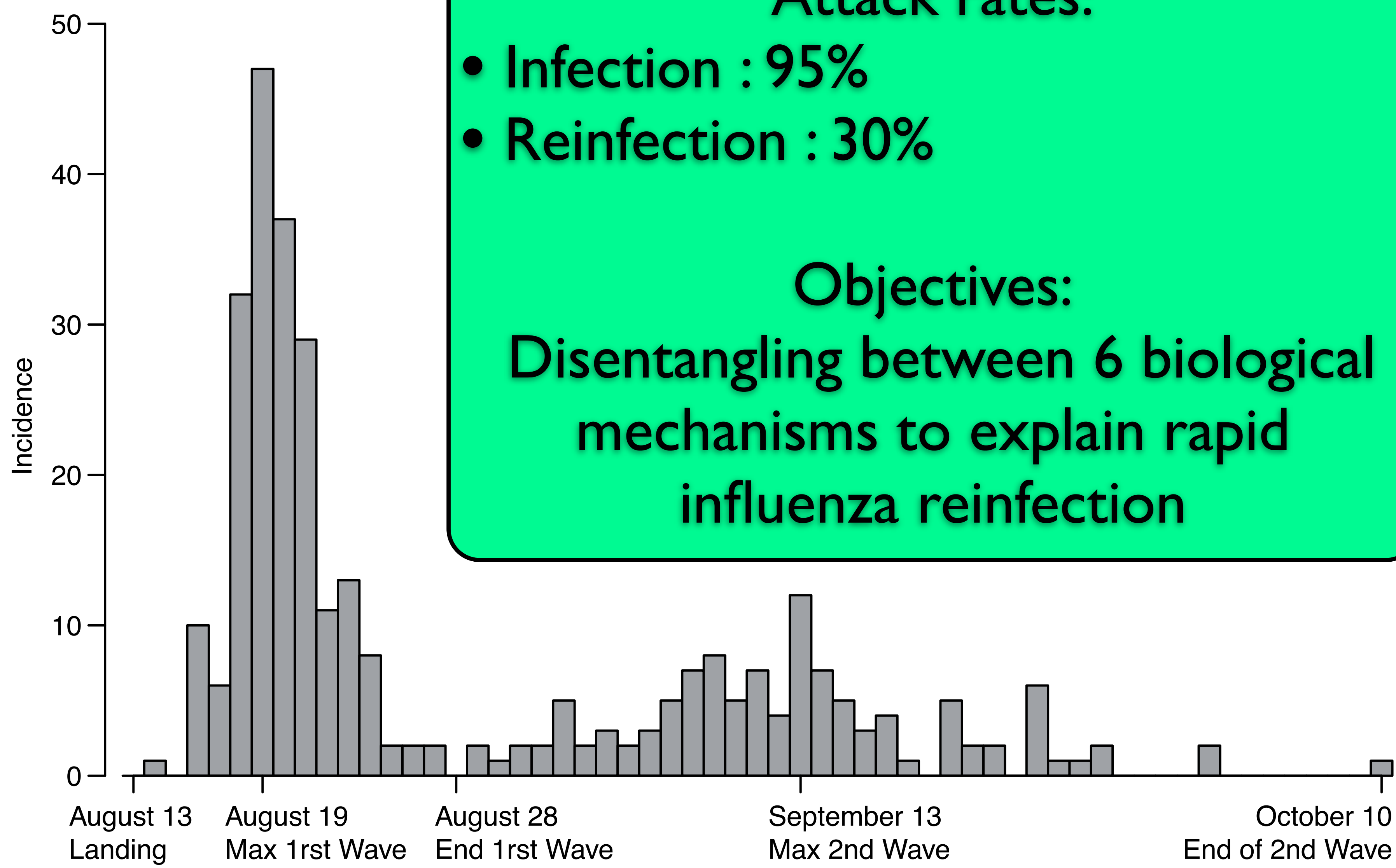
Model selection:

objective ranking of models, which hypothesis best explains the data?

Tristan da Cunha (1971) a two-wave flu epidemic



Tristan da Cunha (1971) a two-wave flu epidemic



Attack rates:

- Infection : 95%
- Reinfection : 30%

Objectives:

Disentangling between 6 biological mechanisms to explain rapid influenza reinfection

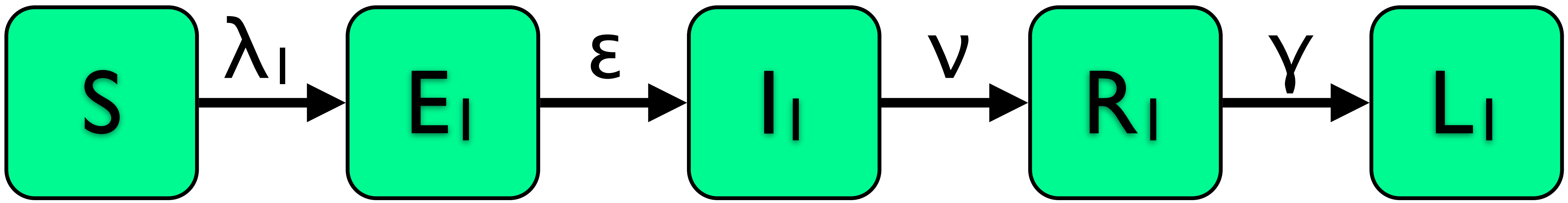
A simple mechanistic approach

$$S \xrightarrow{\lambda} E \xrightarrow{\varepsilon} I \xrightarrow{\nu} R \xrightarrow{\gamma} L$$

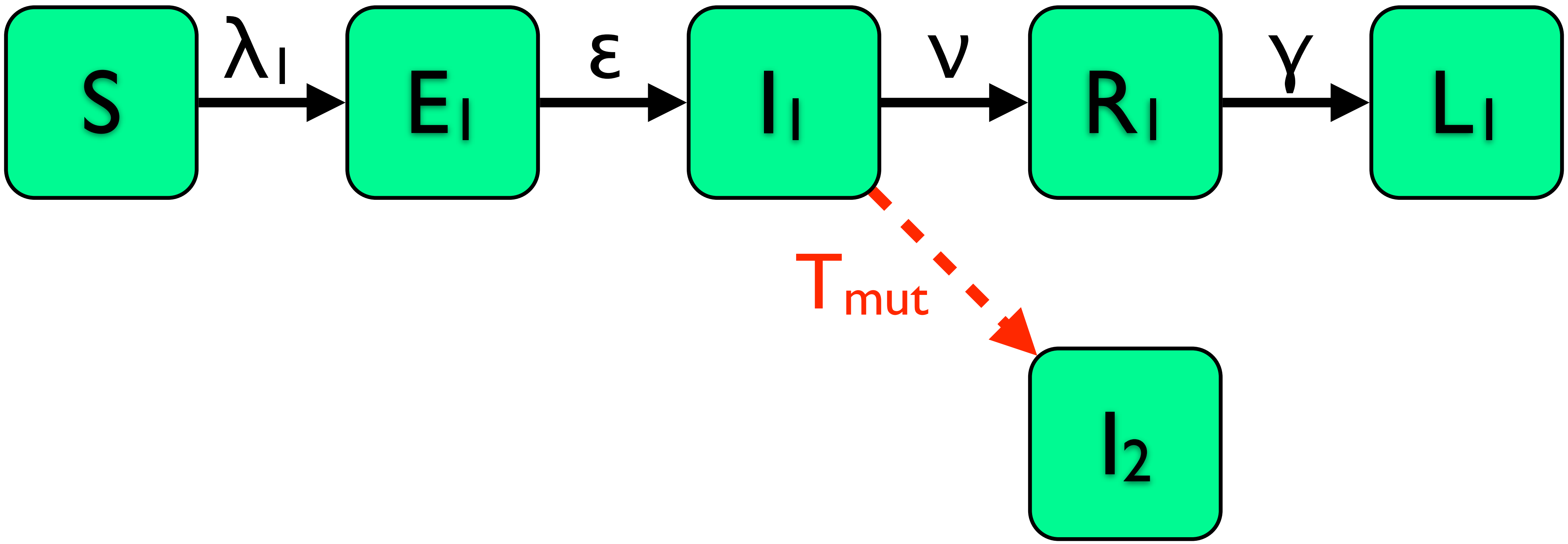
Long-term immunity

- $\lambda = \beta\, I/N$ mass-action
- $1/\varepsilon$: mean latent period
- $1/\nu$: mean infectious period
- $1/\gamma$: mean removed period

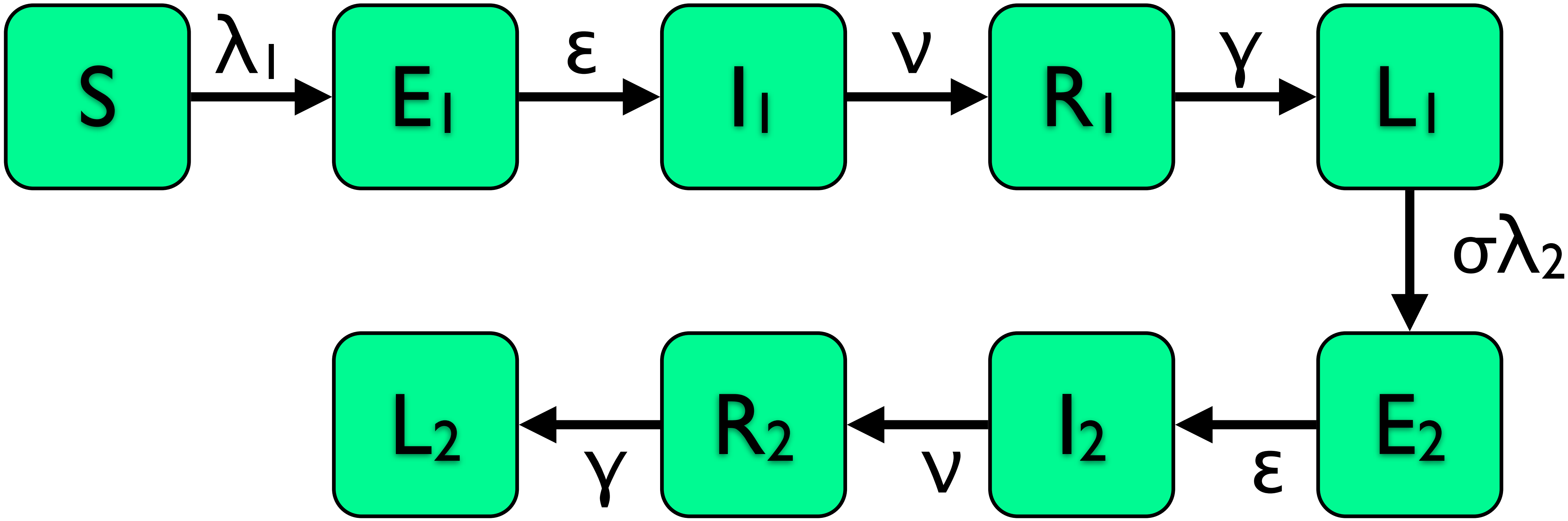
H1: the virus mutated during the first epidemic-wave (Mut)



H1: the virus mutated during the first epidemic-wave (Mut)

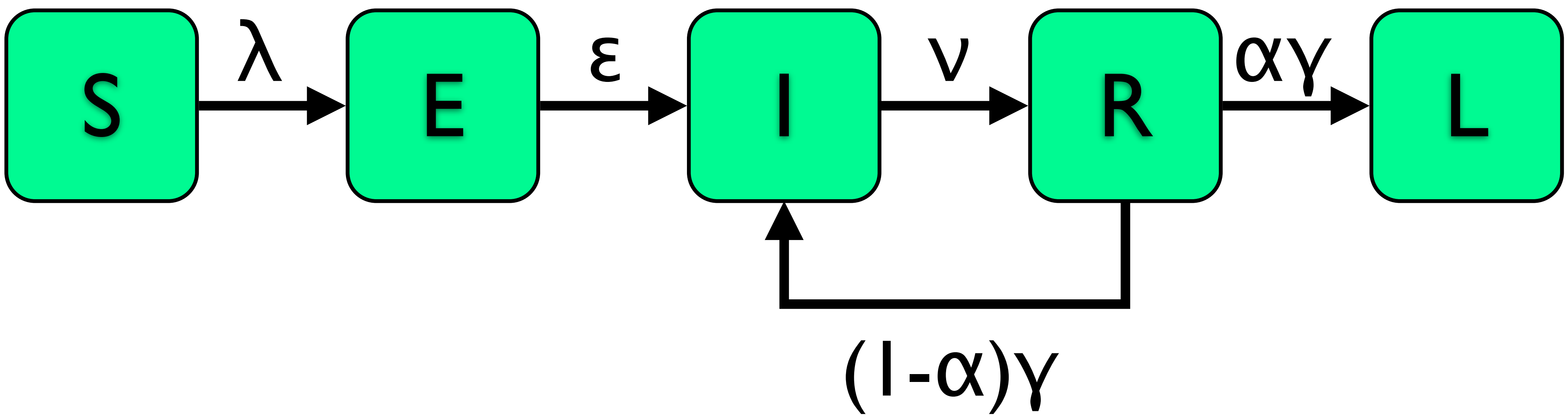


H1: the virus mutated during the first epidemic-wave (Mut)



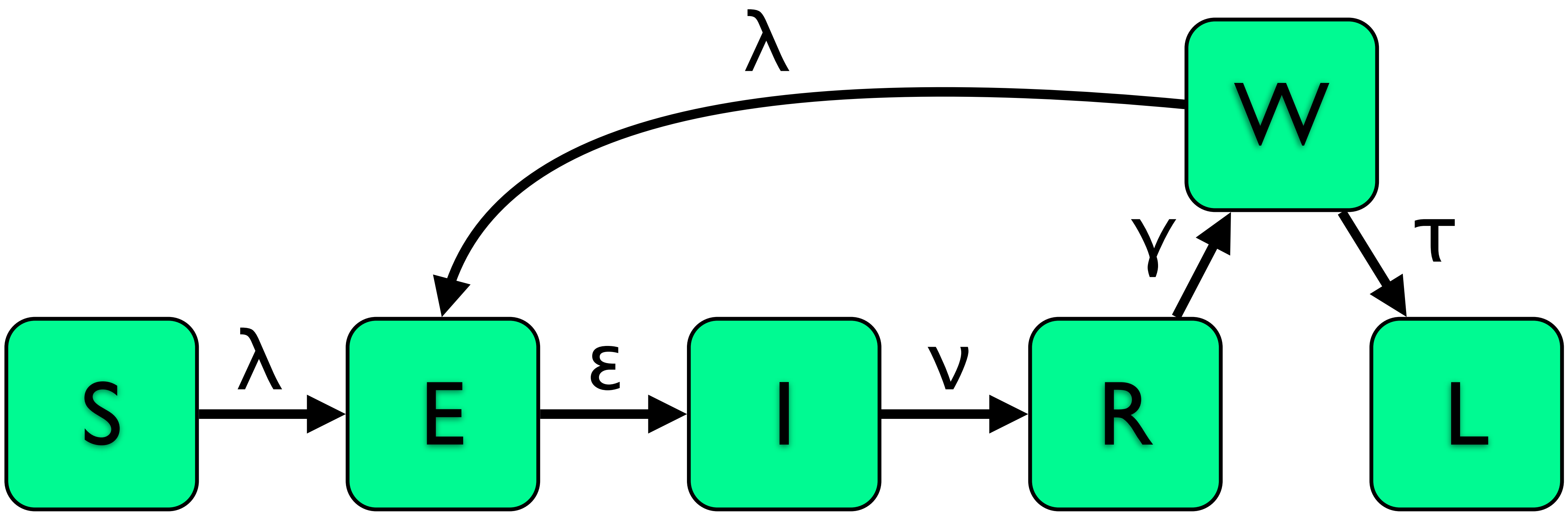
- $\sigma \in [0,1]$ cross-immunity
- 2-strain history-based model (Rios-Doria & Chowell 2009)

H2: intra-host recrudescence of infection (InH)



α: the probability to clear the viral load

H3: window-of-reinfection (Win)



$1/\tau$: the mean duration of the window of susceptibility before developing immunity

Likelihood-based inference

For a given time series: $y_{1:T} = (y_1, y_2, ..., y_T)$
and a state space model completely specified by:

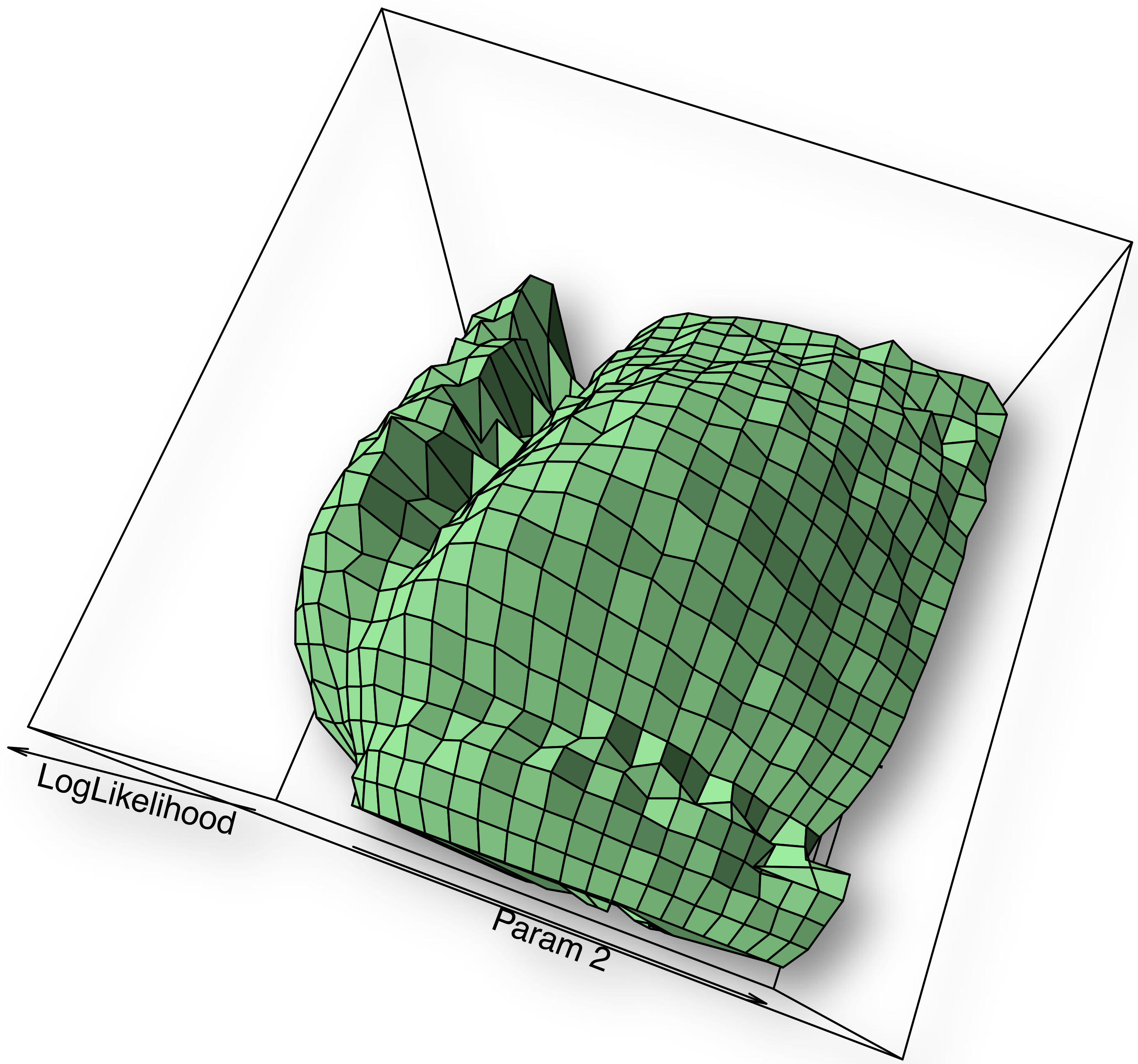
$$M : \begin{cases} f(x_t|x_{t-1}, \theta) & \text{the conditional transition density} \\ f(y_t|x_t, \theta) & \text{the conditional distribution} \\ & \text{of the observation process} \\ f(x_0|\theta) & \text{the initial density} \end{cases}$$

the likelihood is given by the identity:

$$f(y_{1:T}|\theta) = \prod_{t=1}^{T} f(y_t|y_{1:t-1}, \theta)$$

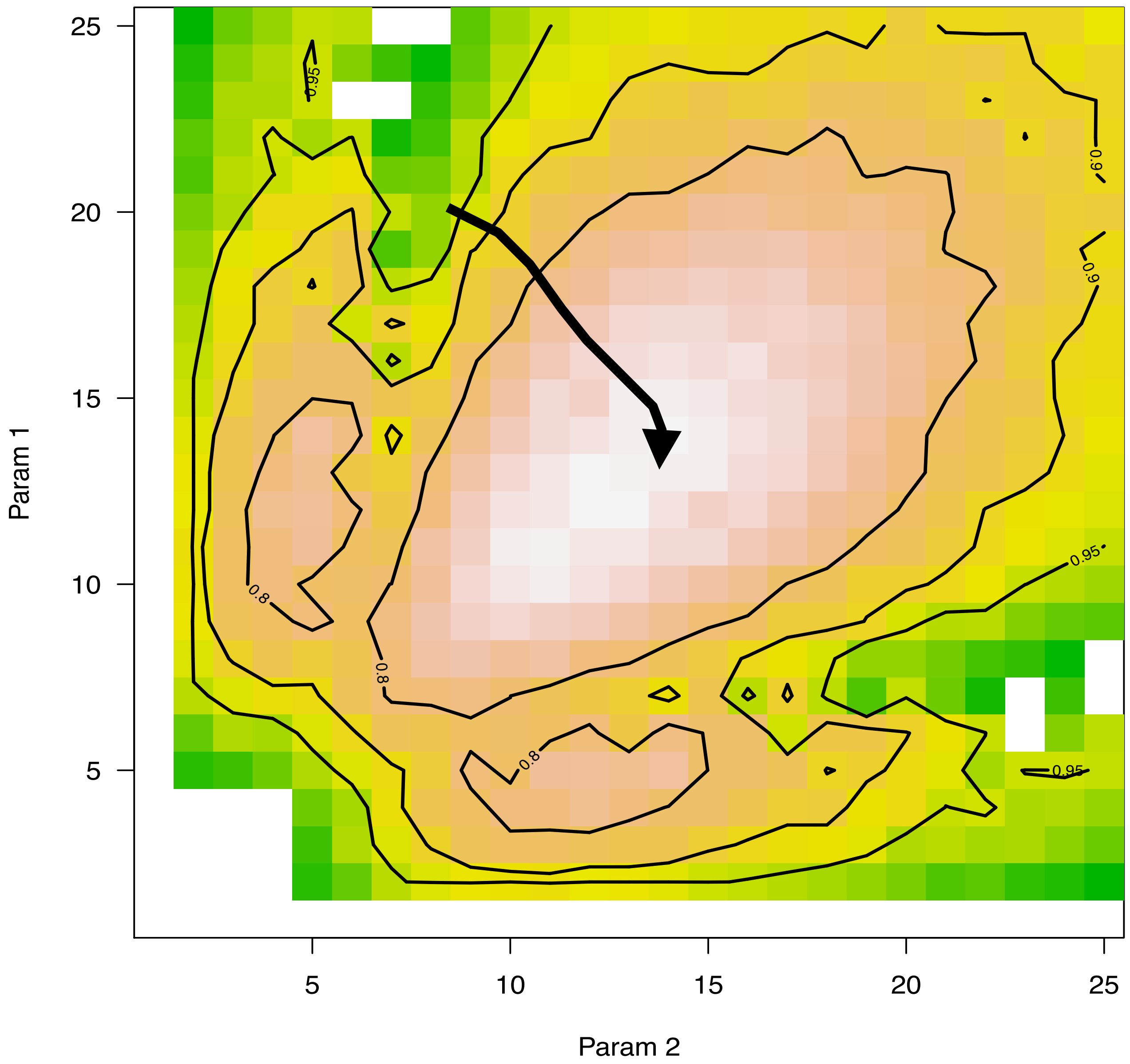
where x_t is the unobserved Markov process, θ is the unknown vector of parameters and $f(.|.)$ is a generic density specified by its arguments

Exploring the likelihood surface



Exploring the likelihood surface with MIF (Ionides *et al.* 2006)

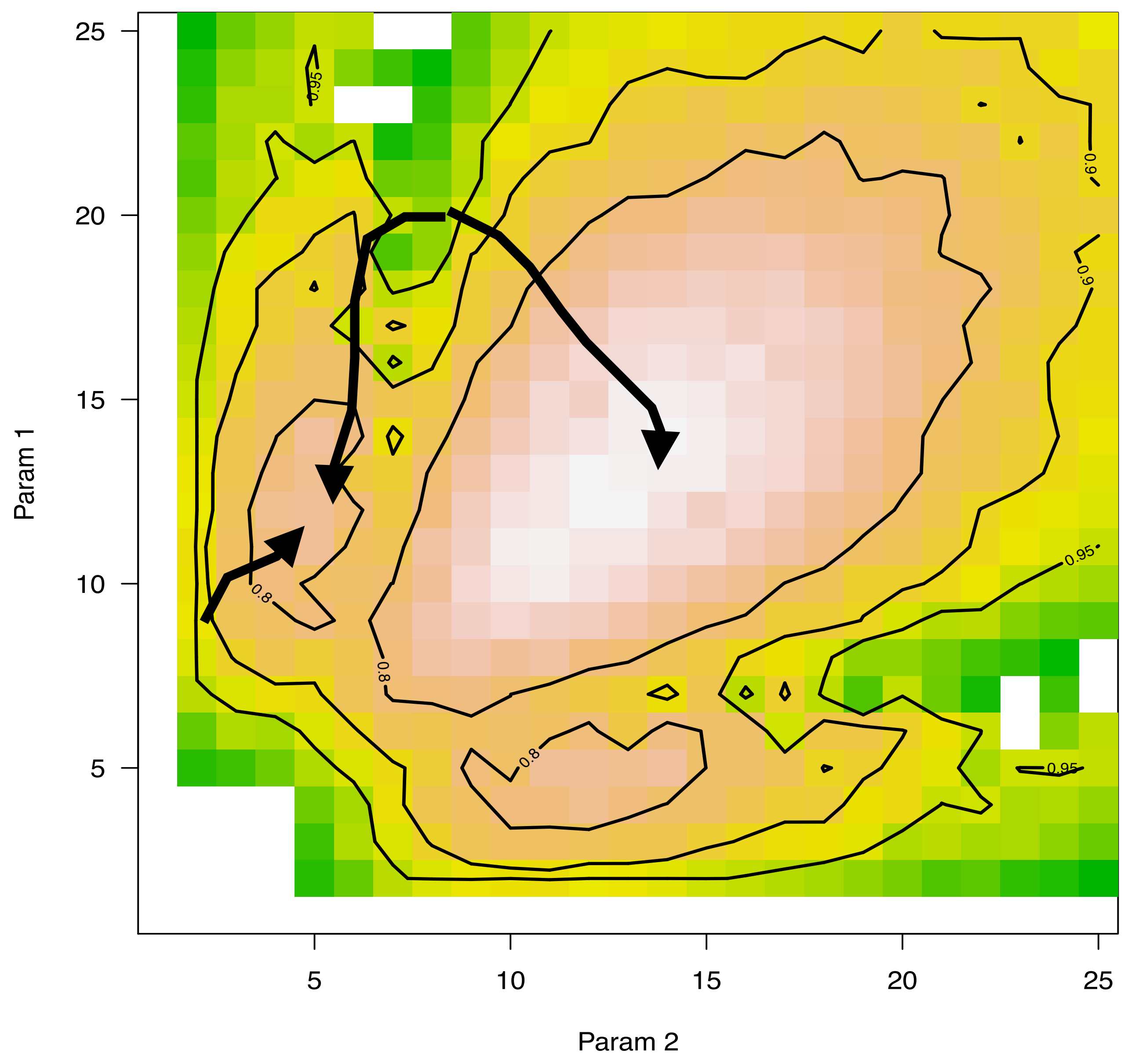
- 3 local maxima
- 1 global maximum



Global convergence

Exploring the likelihood surface with MIF (Ionides *et al.* 2006)

- 3 local maxima
- 1 global maximum

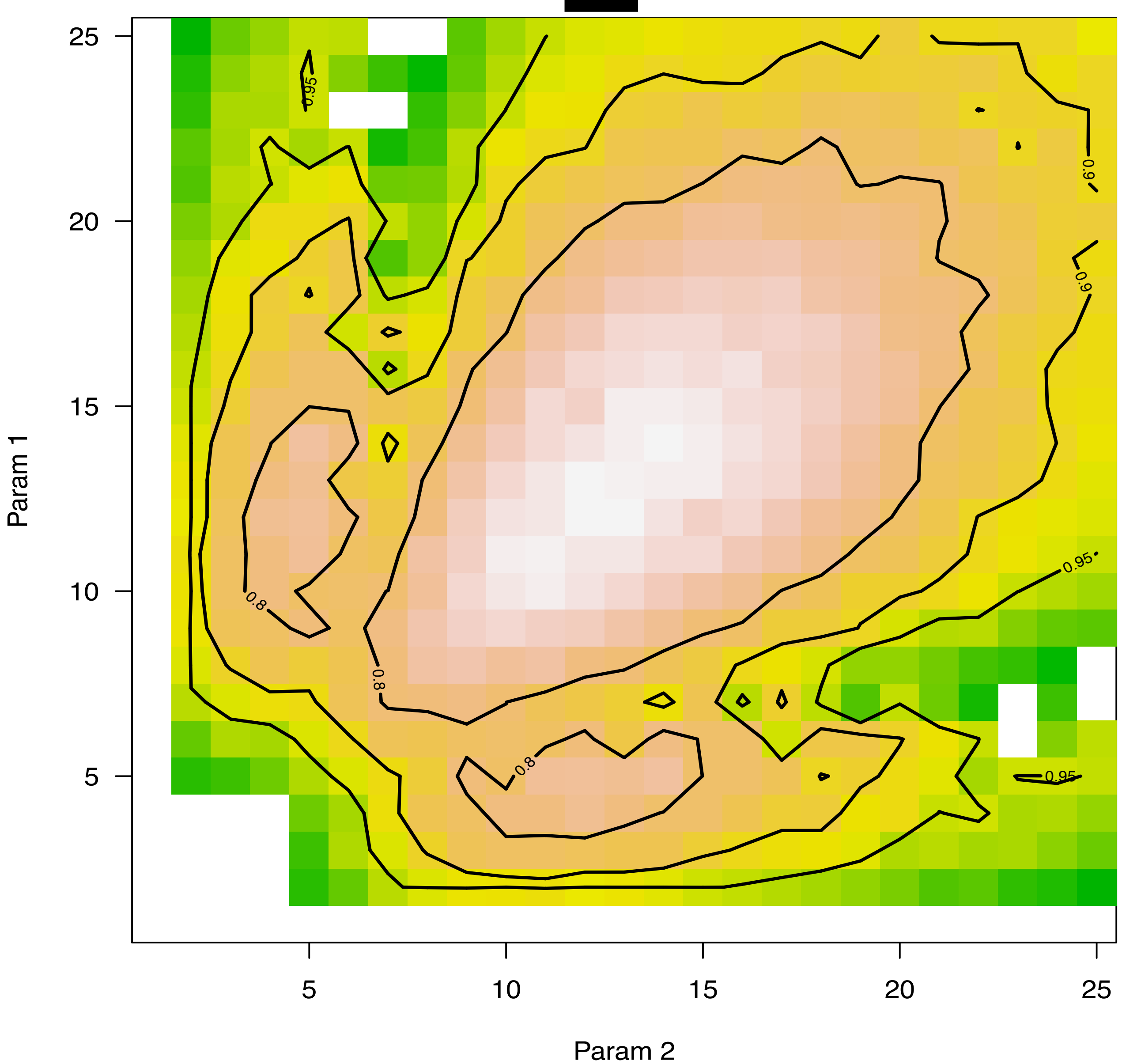


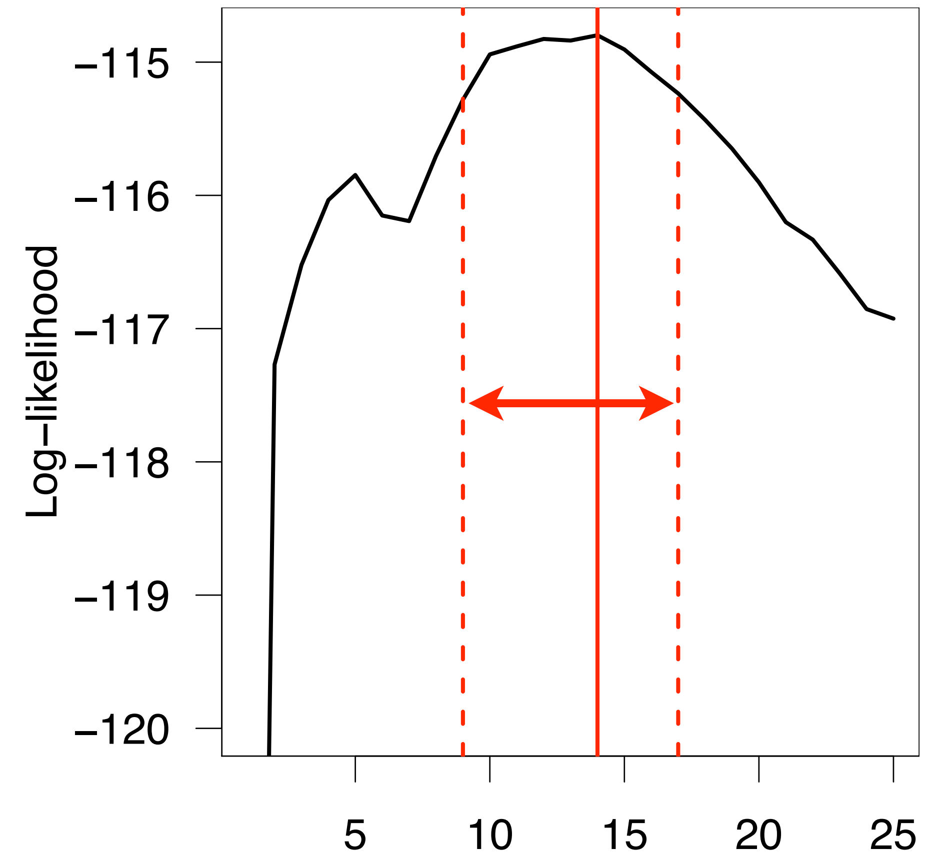
Local trap:

- initial θ
- MIF para-metrization

Log-likelihood profile

projection



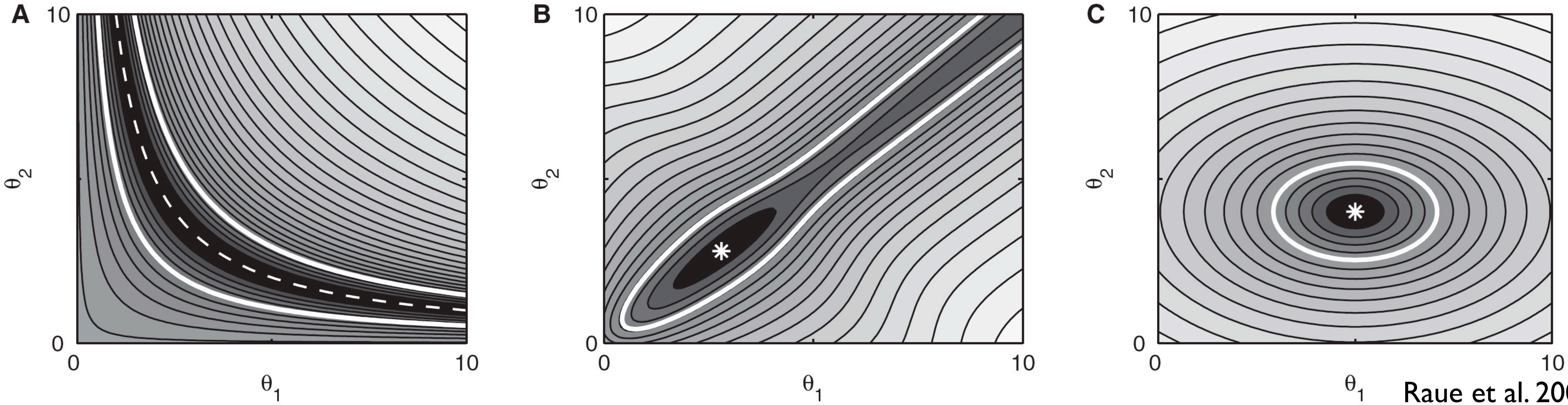


95% confidence interval:

$$\{\theta_2 : 2[l_{(p)}(\hat{\theta}_2) - l_{(p)}(\theta_2)] < \chi^2_{0.95}(1)\}$$

$$\text{with } \hat{\theta}_2 = \ \operatorname{argmax} l_{(p)}(\theta_2)$$

Parameter identifiability



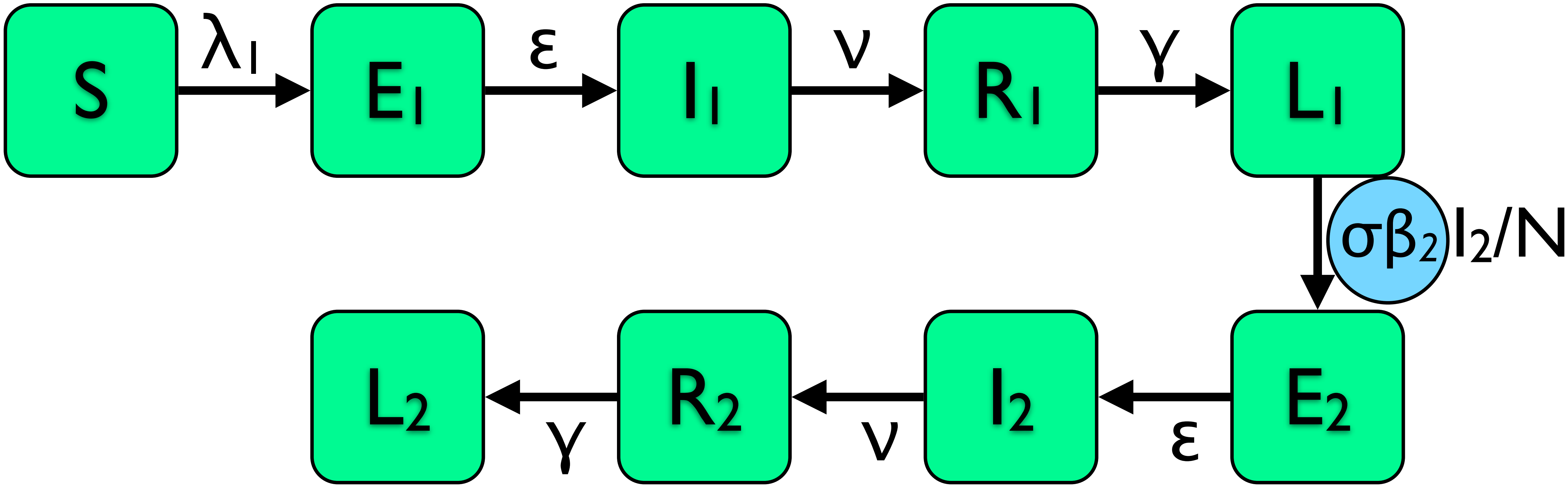
Raue et al. 2009

Structural non-identifiability

Practical non-identifiability

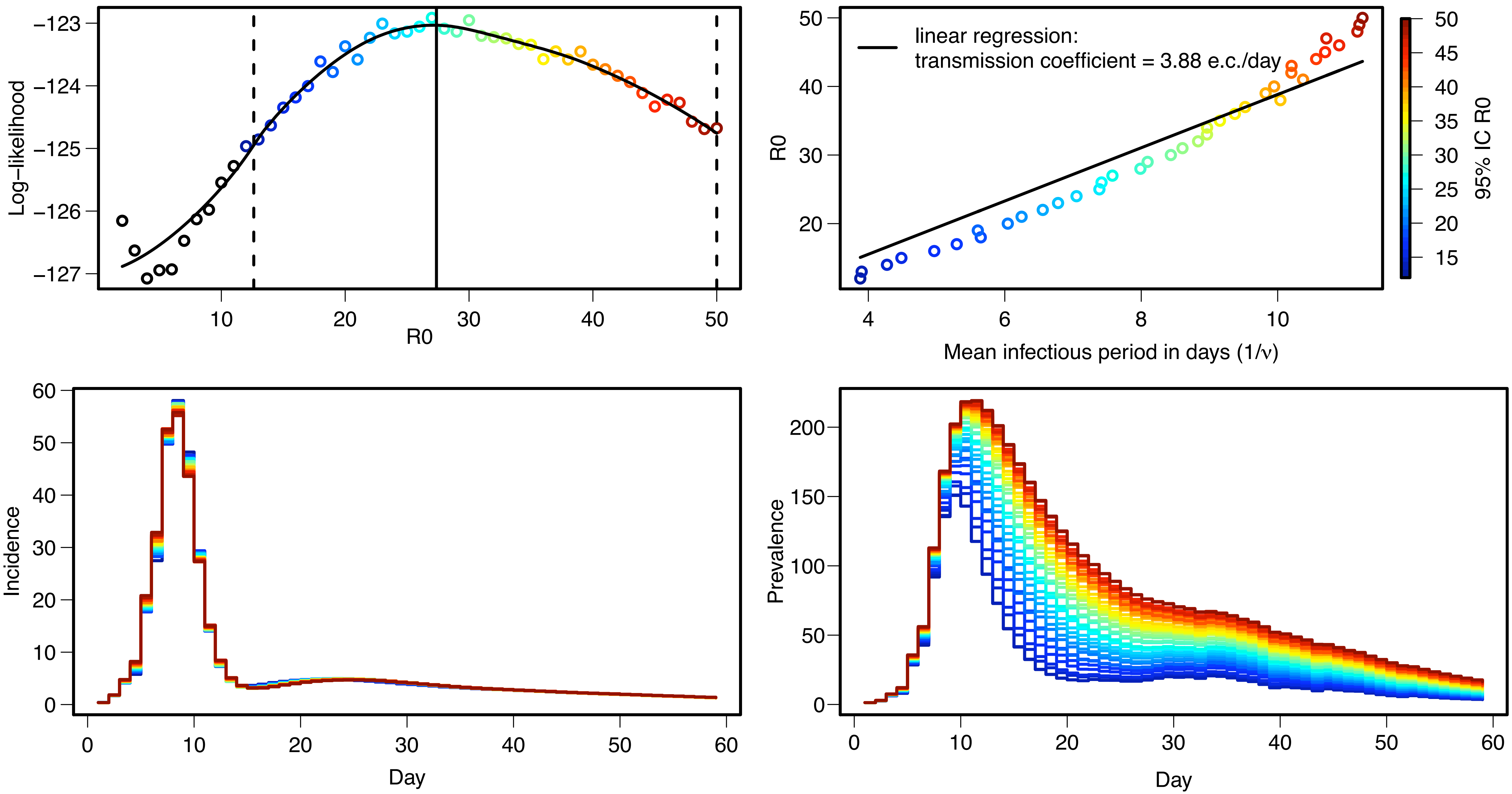
Identifiability

Structural non-identifiability (Mutation: H1)



Structural non-identifiability between σ and $\beta_2 \Rightarrow \beta_2 = \beta_1$

Practical non-identifiability (In-Host: H2)



Likelihood-based inference

For a given time series: $y_{1:T} = (y_1, y_2, ..., y_T)$
and a state space model completely specified by:

$$M : \begin{cases} f(x_t|x_{t-1},\theta) & \text{the conditional transition density} \\ f(y_t|x_t,\theta) & \text{the conditional distribution} \\ & \text{of the observation process} \\ f(x_0|\theta) & \text{the initial density} \end{cases}$$

the Log-likelihood is given by the identity:

$$\log(f(y_{1:T}|\theta)) = \sum_{t=1}^{T} \log(f(y_t|y_{1:t-1},\theta))$$

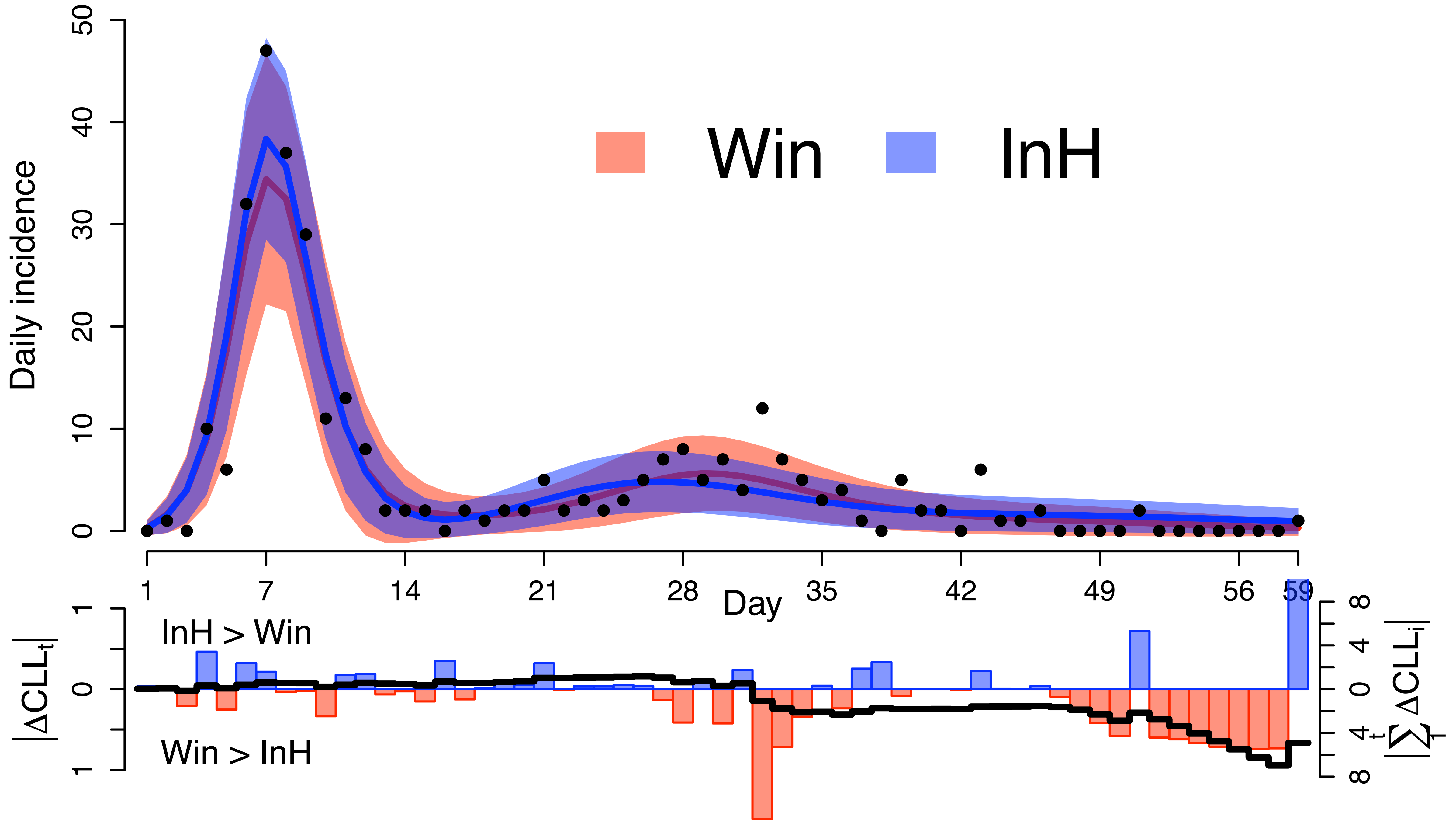
where x_t is the unobserved Markov process, θ is the unknown vector of parameters and $f(.|.)$ is a generic density specified by its arguments

Model selection: Akaike information criterion

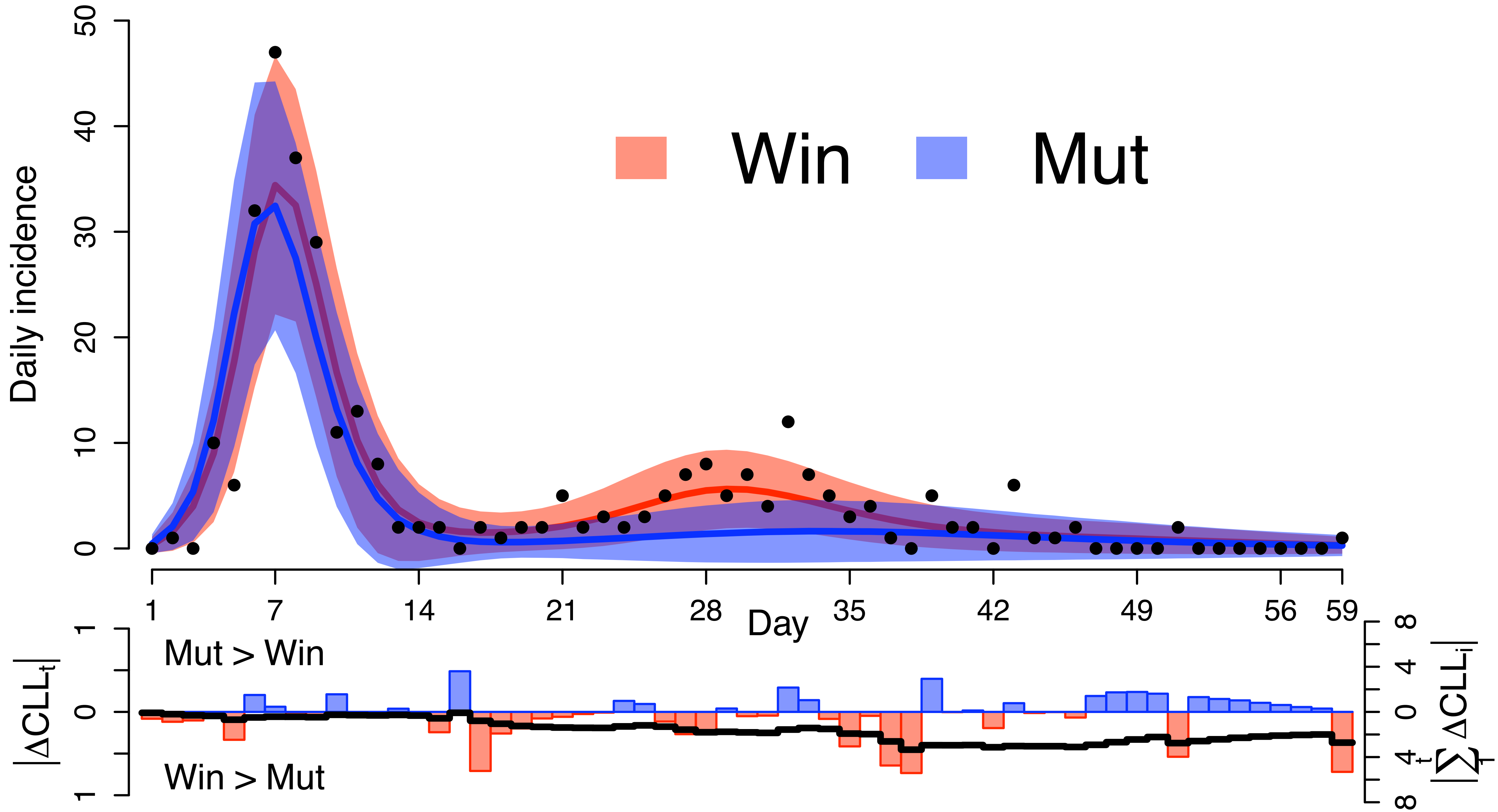
$$AIC_c = -2\mathcal{L}(\theta_{MLE}) + 2k + \frac{2k(k+1)}{T-k-1} \text{ with } k = ||\theta||$$

Model	Win	Mut	In-Host
k	9	10	9
Log-Like	-112.52	-115.20	-117.50
ΔAIC_c	0	8.27	9.96

Dynamics comparison



Dynamics comparison



Conclusion

- Maximum likelihood via Iterated Filtering (MIF, Ionides *et al*. 2006) is a rigorous statistical framework for parameter inference and selection based on AIC for mechanistic stochastic models.
- Identifiability analysis and 95% CI via log-likelihood profile
- Illustrate and compare the dynamics of the different models in a visual and intuitive manner

(1) Ionides EL, Breto CM, King AA. Inference for nonlinear dynamical systems. *PNAS*. 2006

(2) Raue A, Kreutz C, Maiwald T, et al. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics (Oxford, England)*. 2009

(3) Camacho A, Ballesteros S, Graham AL, Carrat F, Cazelles B. Explaining rapid reinfection in multiple-wave influenza outbreaks? Tristan da Cunha epidemic as a case study. *(in preparation)*

Likelihood-based inference

For a given time series: $y_{1:T} = (y_1, y_2, ..., y_T)$
and a state space model completely specified by:

$$M : \begin{cases} f(x_t|x_{t-1},\theta) & \text{the conditional transition density} \\ f(y_t|x_t,\theta) & \text{the conditional distribution} \\ & \text{of the observation process} \\ f(x_0|\theta) & \text{the initial density} \end{cases}$$

the Log-likelihood is given by the identity:

$$\log(f(y_{1:T}|\theta)) = \sum_{t=1}^{T} \log(f(y_t|y_{1:t-1},\theta))$$

where x_t is the unobserved Markov process, θ is the unknown vector of parameters and $f(.|.)$ is a generic density specified by its arguments

Maximum likelihood via Iterated Filtering (Ionides *et al.* 2006)

Global estimate: $\hat{\theta}^{(n)} = \hat{\theta}^{(n-1)} + V_1^{(n)} \sum_{t=1}^{T} \frac{\hat{\theta}_t^{(n)} - \hat{\theta}_{t-1}^{(n)}}{V_t^{(n)}}$

As shown by Ionides *et al.* (2006), under rather mild assumptions,

$$\lim_{\sigma \to 0} \sum_{t=1}^{T} \frac{\hat{\theta}_t - \hat{\theta}_{t-1}}{V_t} = \nabla \log f(y_{1:T} | \theta, \sigma = 0)$$

so that, for a sufficiently small σ_n, the algorithm iteratively updates $\hat{\theta}^{(n)}$ in the direction of increasing likelihood, with a fixed point at a local maximum of the likelihood surface.