Assessing the use of phenomenological transmission functions to model disease spread on contact networks

Application to time series analysis

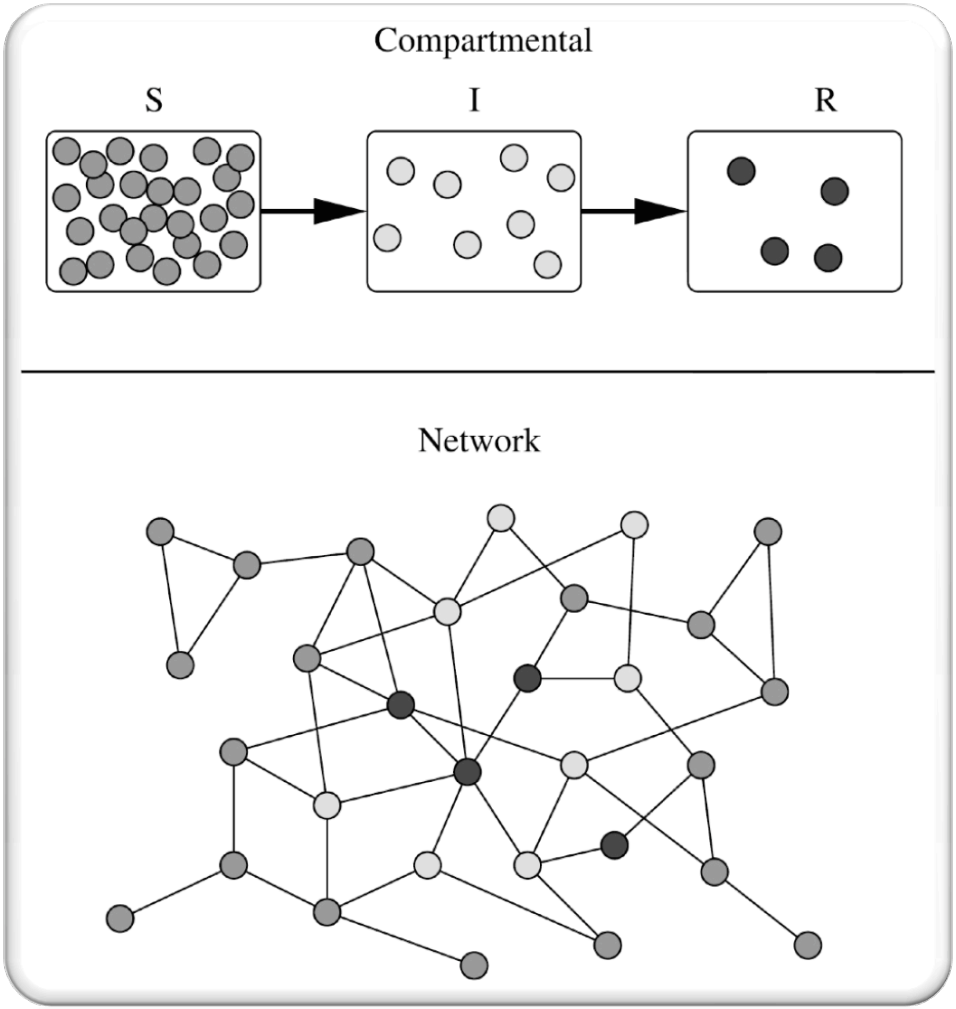
Anton Camacho, Quentin Herreros, Bernard Cazelles

CNRS – UMR 7625 Ecologie & Evolution
Equipe Eco-Evolution Mathématique
ENS Ulm -UPMC

Mass Action *vs* Contact Structure

$$\Delta I \propto SI$$

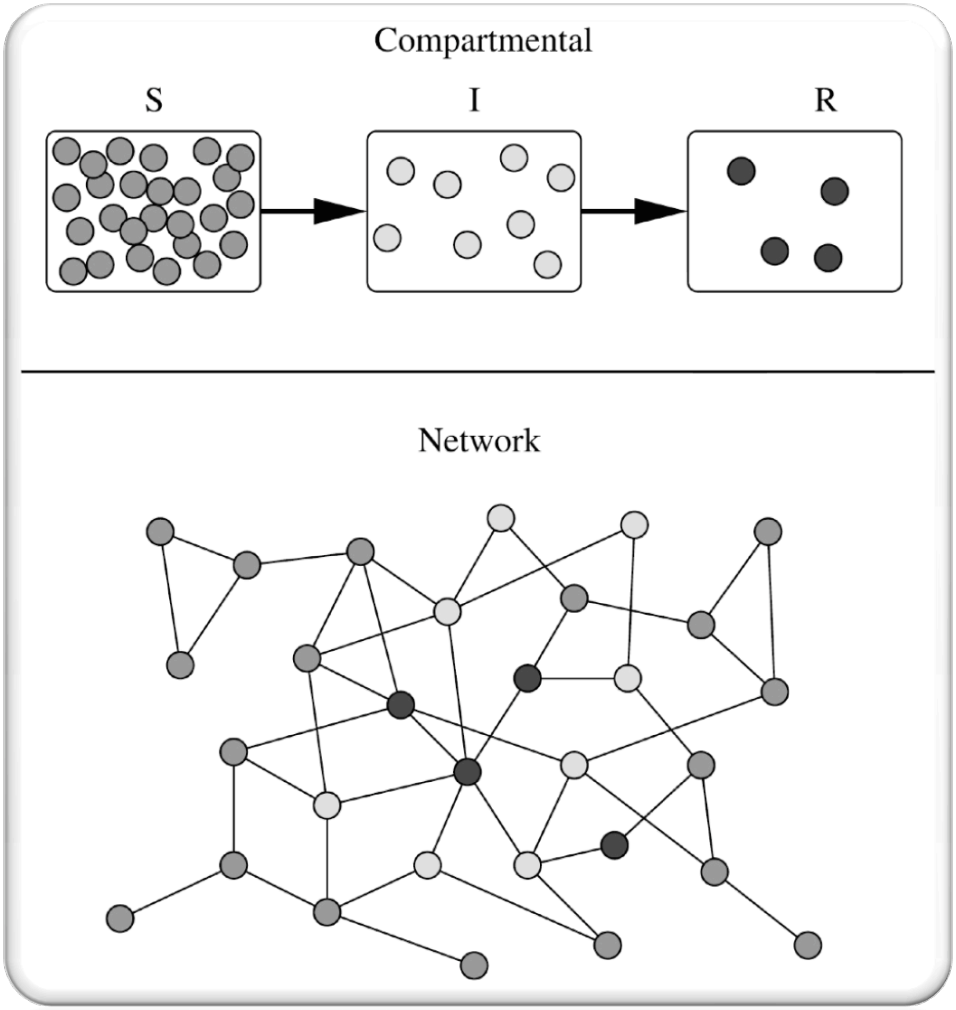
$$\Delta I \propto \sum_{POP} [SI]$$

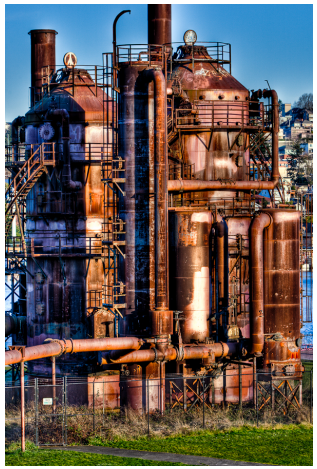


Mass Action *vs* Contact Structure

$$\Delta I \propto SI$$

$$\Delta I \propto \sum_{POP} [SI]$$





Incorporating (simply) Heterogeneity

Network-based analytical approaches incorporate:

- The degree distribution: percolation theory (Newman) , PGF-based differential equations (Volz).
- The local contact structure (clustering) : pair approximation methods (Keeling).

Incorporating (simply) Heterogeneity

Network-based analytical approaches incorporate:

- The degree distribution: percolation theory (Newman) , PGF-based differential equations (Volz).
- The local contact structure (clustering) : pair approximation methods (Keeling).

Phenomenological approaches to compartmental models:

Negative binomial

$$kS\ln(1+\beta\frac{I}{k})$$

Small *k* corresponds to highly clustered infection

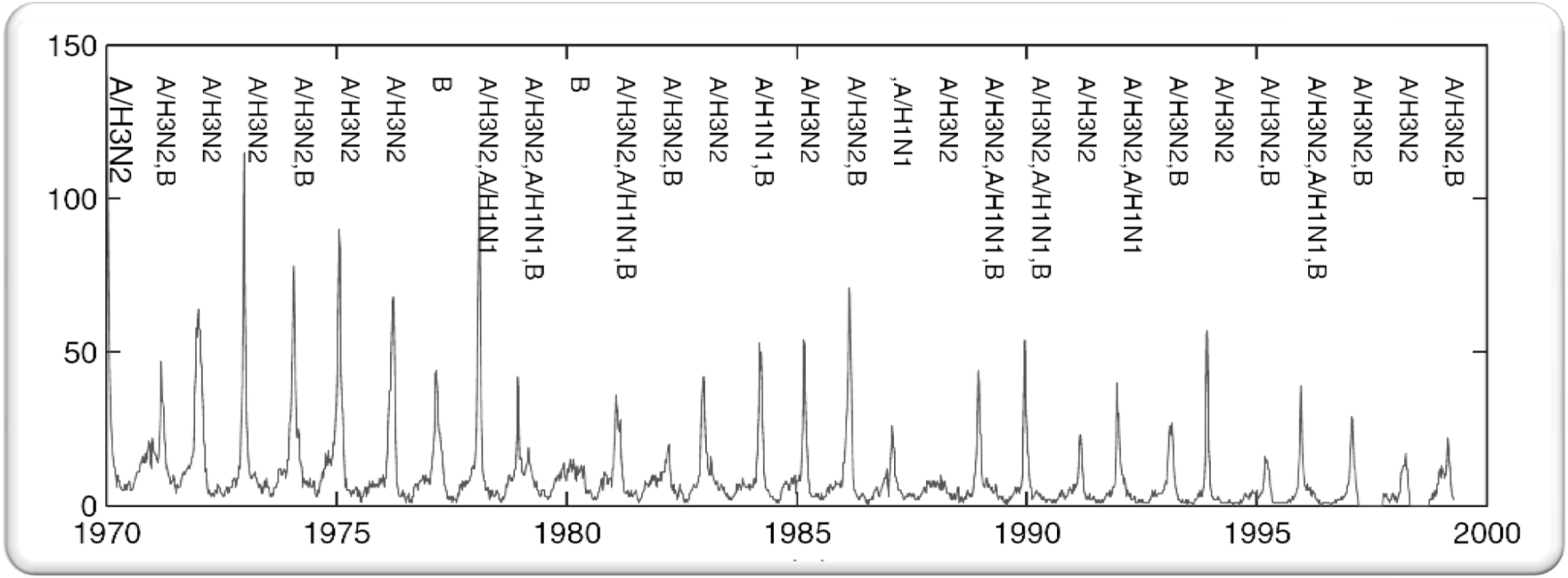
$$\beta SI$$

Power relationship

$$\beta S^p I^q$$

Modify the density of contacts

Time series analysis



- Contact patterns depend on: social structure, age, spatial structure etc
- Many other sources of heterogeneities

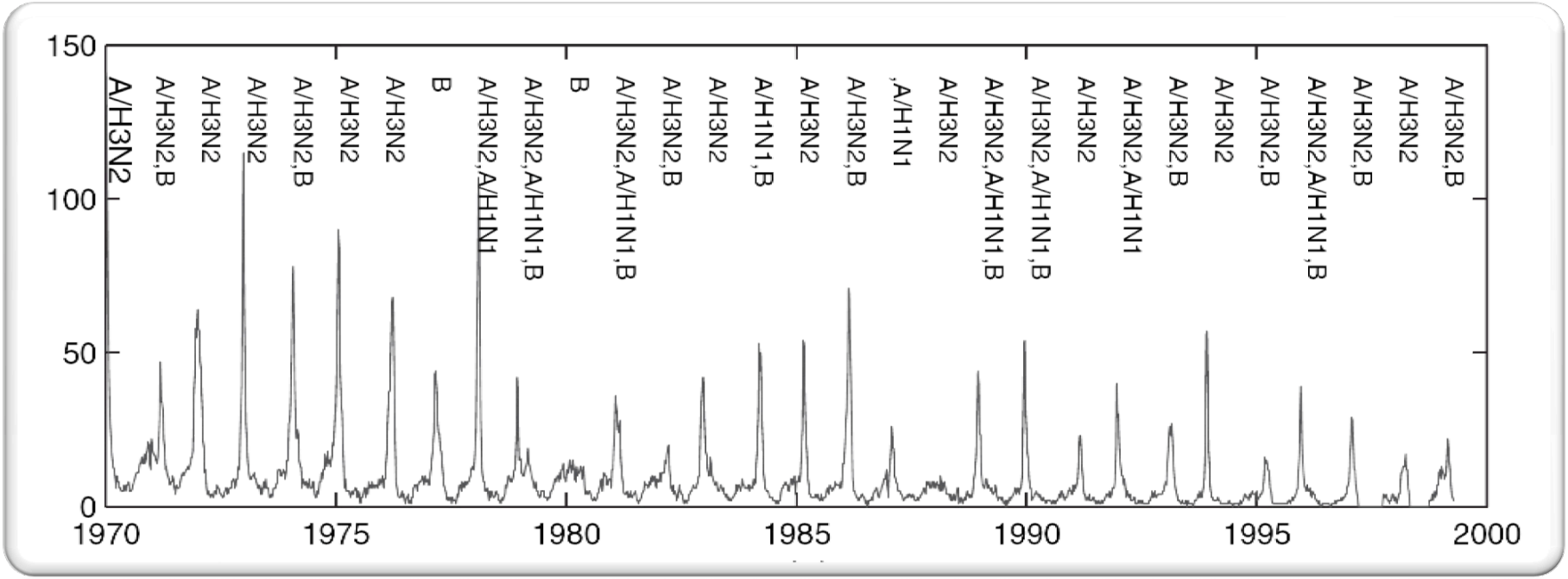
Caveats:

- Contact structure is unknown
- Reduced parameter space

In practice statisticians use phenomenological transmission functions

$$\beta S^p I^q$$

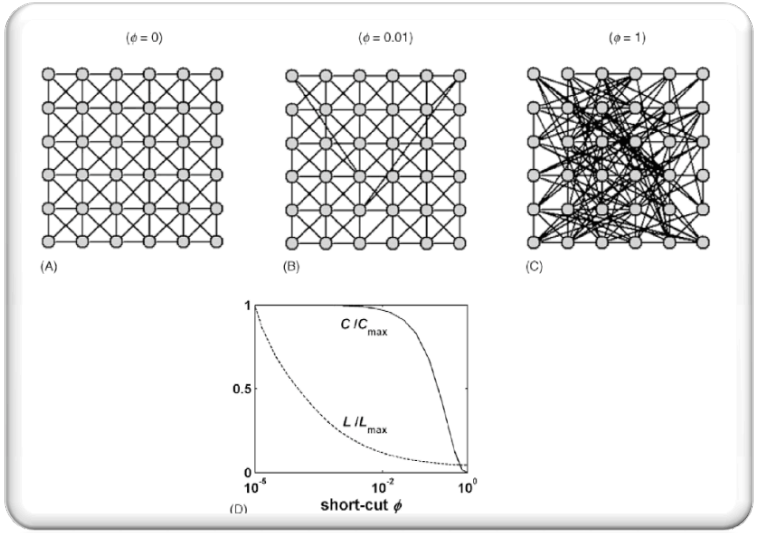
Time series analysis

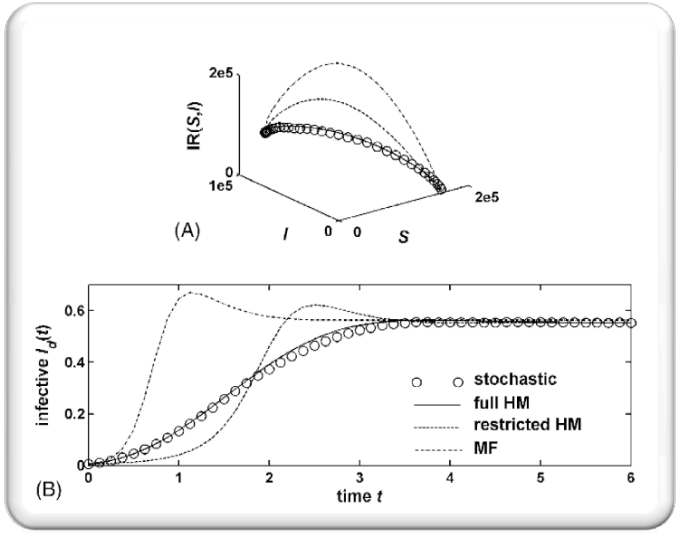


Time series analysis proposes to estimate several epidemiological relevant quantities
Public health policies like vaccine strategies depend on these estimates

DO PHENOMENOLOGICAL FUNCTIONS ALLOW FOR ACCURATE PARAMETER INFERENCE IN TIME SERIES ANALYSIS ?

Comparison with simulations on networks (Roy & Pascual, 2006)



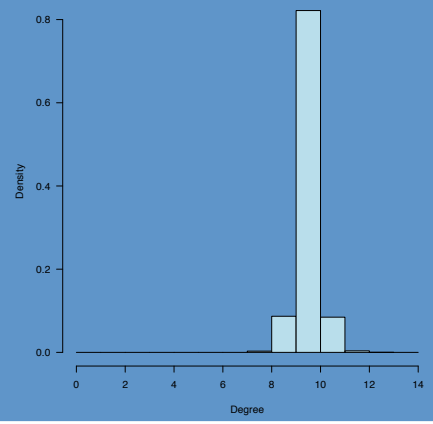


- On lattice, small-world and random networks the S^pI^q function can fit the time series
- What about other networks (exponential, scale-free)?
- The authors do not really fit the time series but instead the transmission rate "observed" in the output of the stochastic simulation (not available in practice).

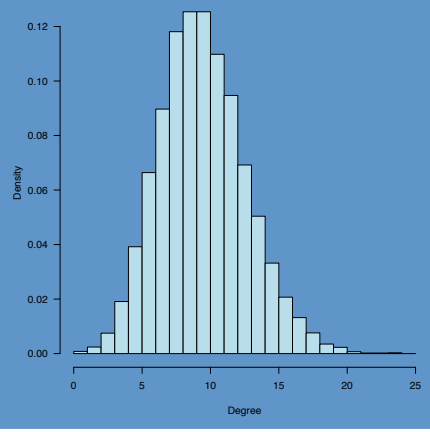
Networks: degree distribution

1. Mean degree = 10
2. Population size = 10 000
3. Undirected
4. Simple (no loop nor multiple edges between two nodes)
5. Connected

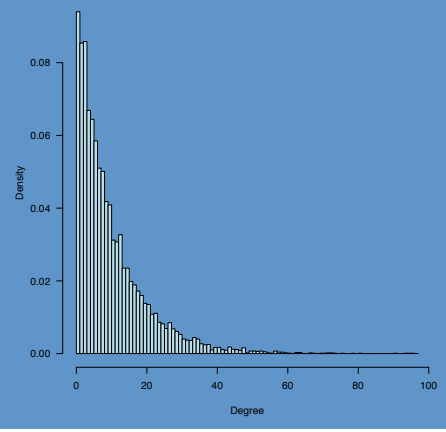
Small-World: [8 - 13] & $\sigma^2 = 0.2$



Poisson: [1 - 24] & $\sigma^2 = 10$

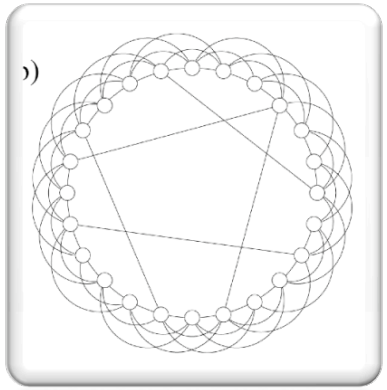


Exponential: [1 - 96] & $\sigma^2 = 100$

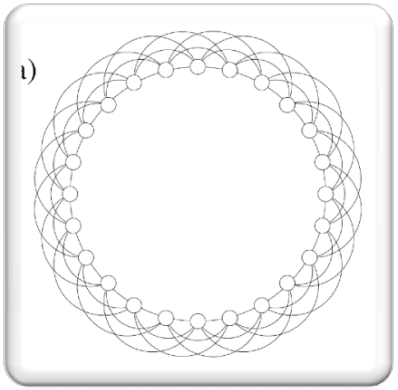


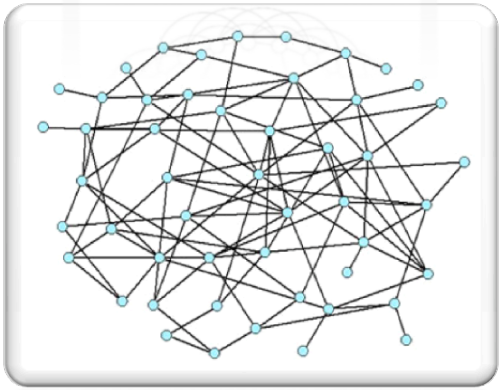
Networks: clustering

School and household transmission play a key role for influenza

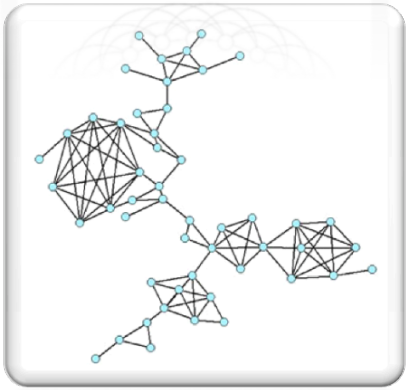


Small-World
Decrease clustering by increasing the rewiring probability





Exponential and Poisson
Increase clustering & keep the same degree distribution
(ClustRNet)



1. Small-World: 0.34 / 0.62 / 0.66 (lattice)
2. Exponential: 10^{-3} / 0.1 / 0.2 / 0.36
3. Poisson: 10^{-3} / 0.1 / 0.2 / 0.59

Epidemiological parameters

R_0: the expected number of secondary cases produced by an infected individual in a totally susceptible population

$$R_0 = \frac{\beta}{\nu} = \beta \times IP$$

β is the effective contact rate per time unit
IP is the infectious period

Epidemiological parameters

R_0: the expected number of secondary cases produced by an infected individual in a totally susceptible population

$$R_0 = \frac{r}{r+\nu}(n - 1 + \frac{\sigma^2}{n})$$

Disease-dependent parameters:

- *r* : transmission probability per contact per time unit
- *v* : 1/Infectious Period (IP)

Network-dependent parameters:

- *n*: mean of the degree distribution
- σ^2: variance of the degree distribution (Exponential > Poisson > Small-World)

Epidemiological parameters

R_0: the expected number of secondary cases produced by an infected individual in a totally susceptible population

$$R_0 = \frac{r}{r+\nu}(n - 1 + \frac{\sigma^2}{n})$$

Disease-dependent parameters:

- *r* : transmission probability per contact per time unit
- *v* : 1/Infectious Period (IP)

Network-dependent parameters:

- *n*: mean of the degree distribution
- σ^2: variance of the degree distribution

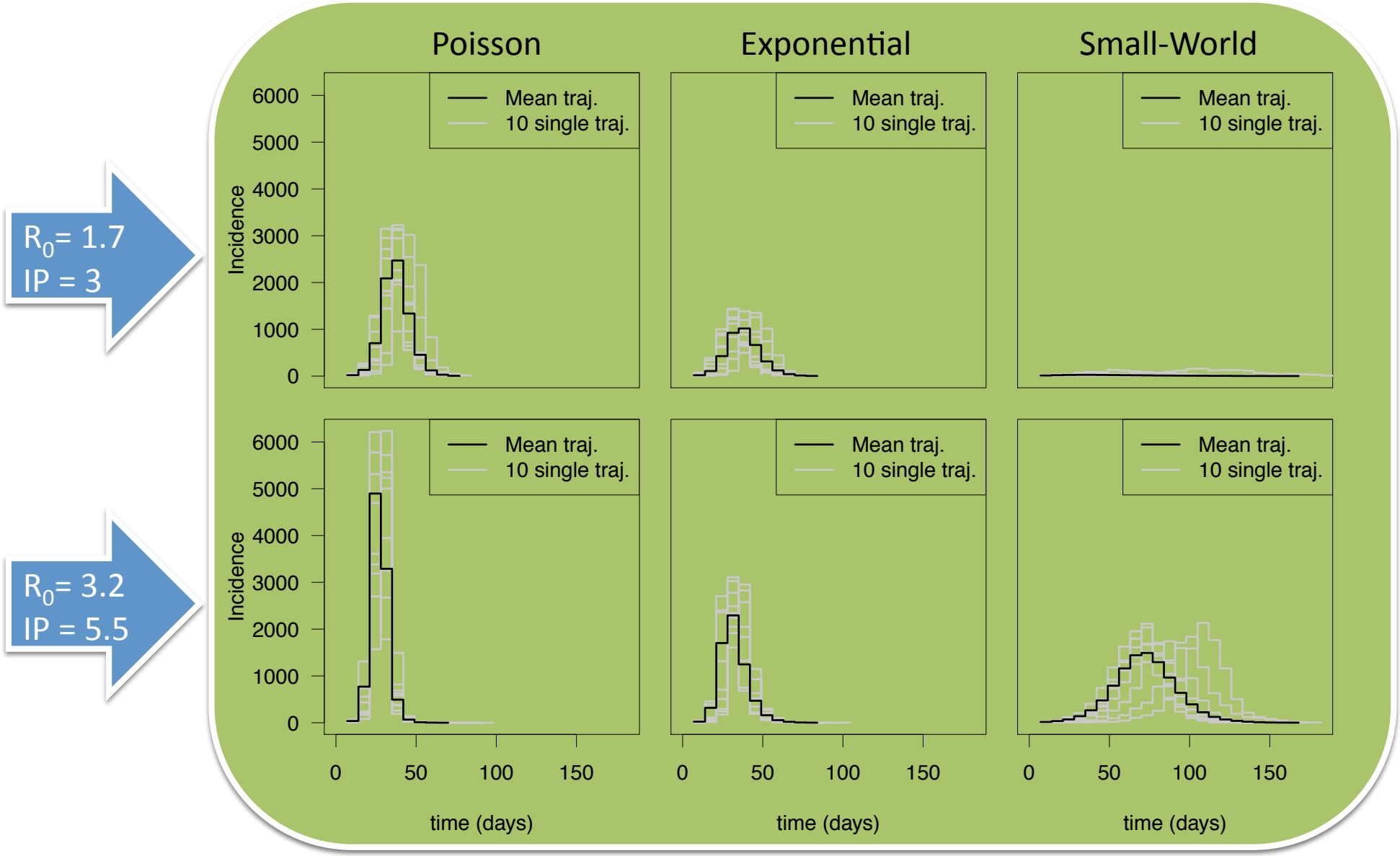
(Exponential > Poisson > Small-World)

Network comparison:

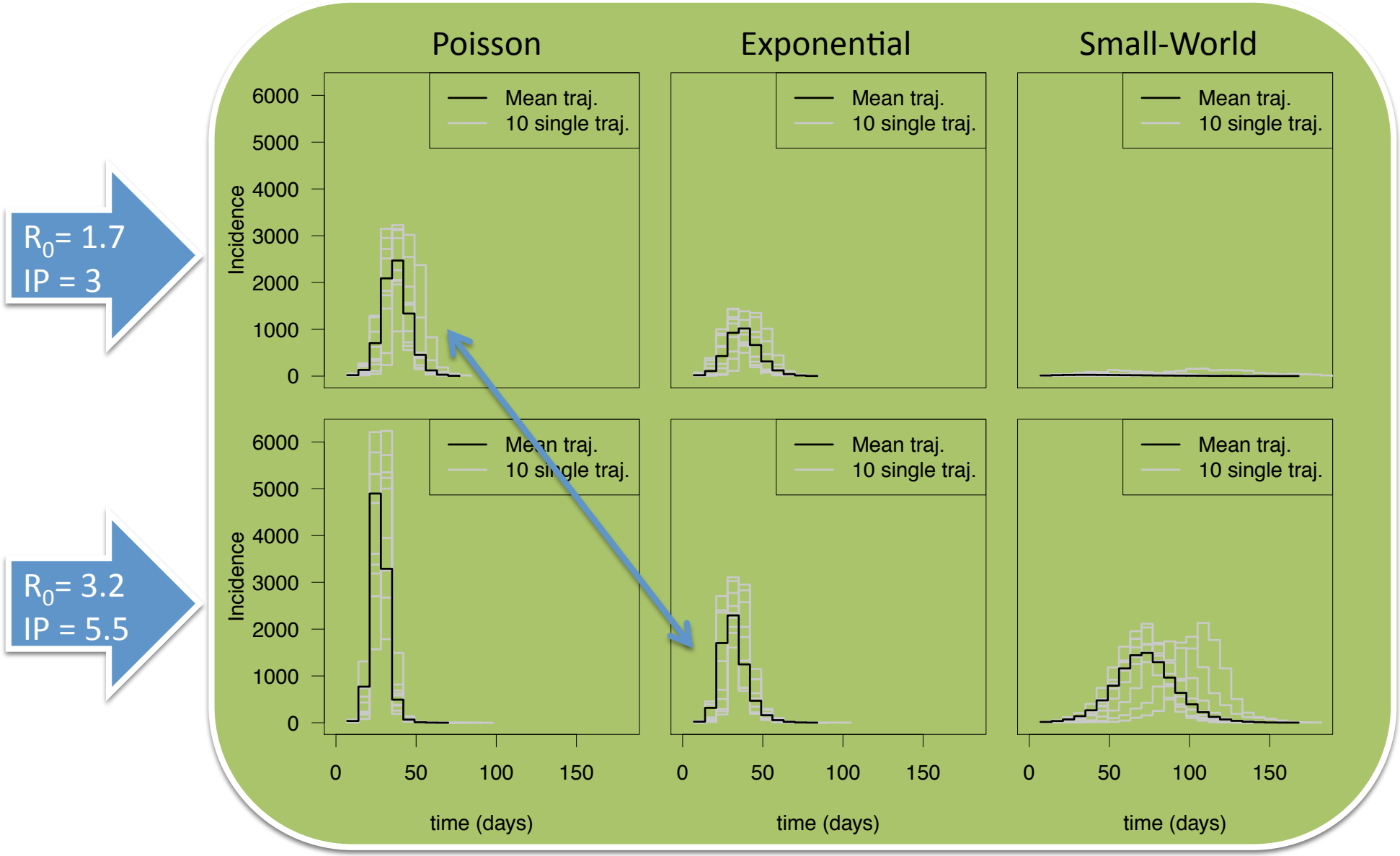
1. Same *r* & IP but $\neq R_0$ ☐
2. Same R_0 & IP but $\neq r$ ☑

R_0 is the main estimated parameter whereas *r* is unknown

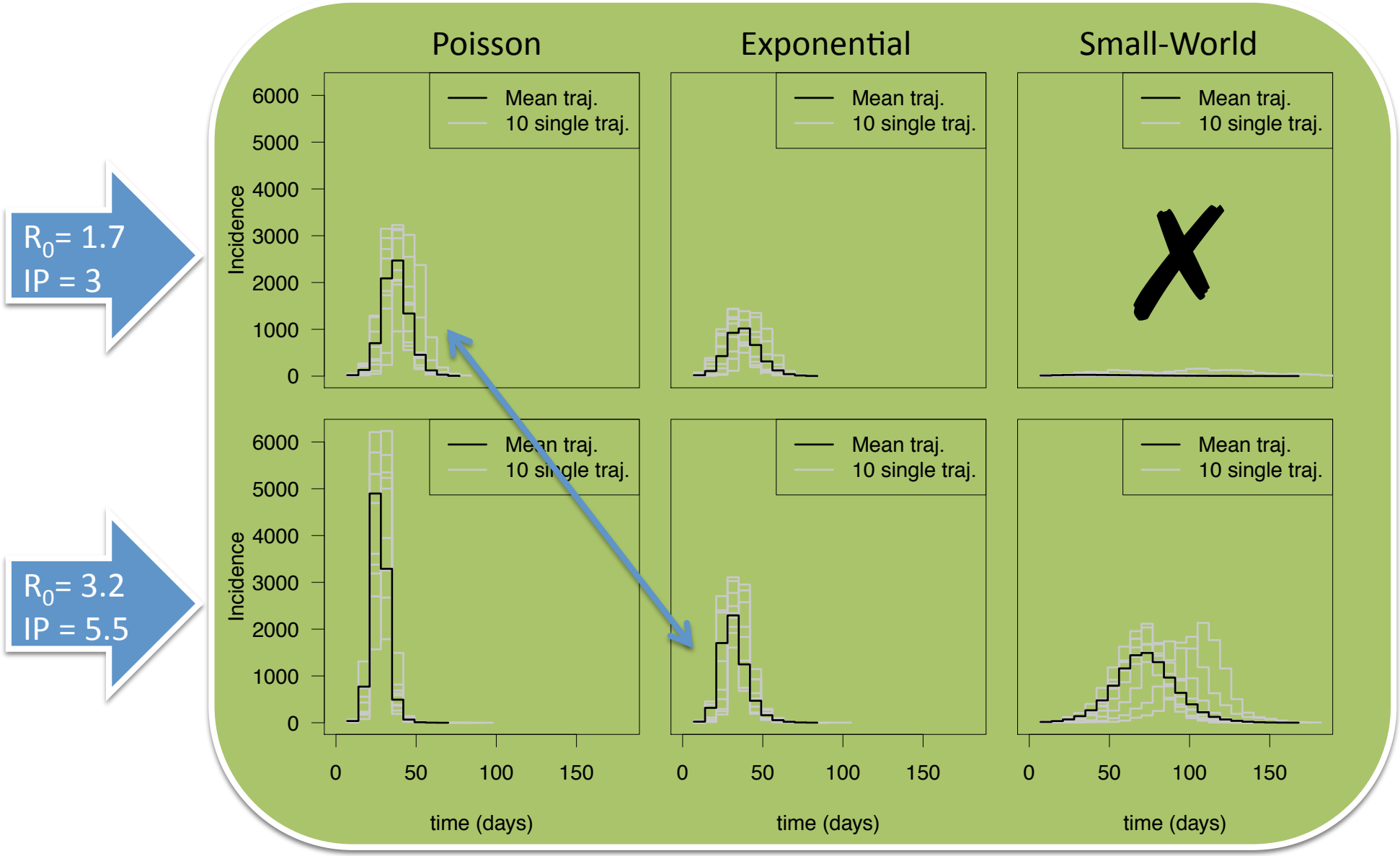
Networks comparison



Networks comparison



Networks comparison



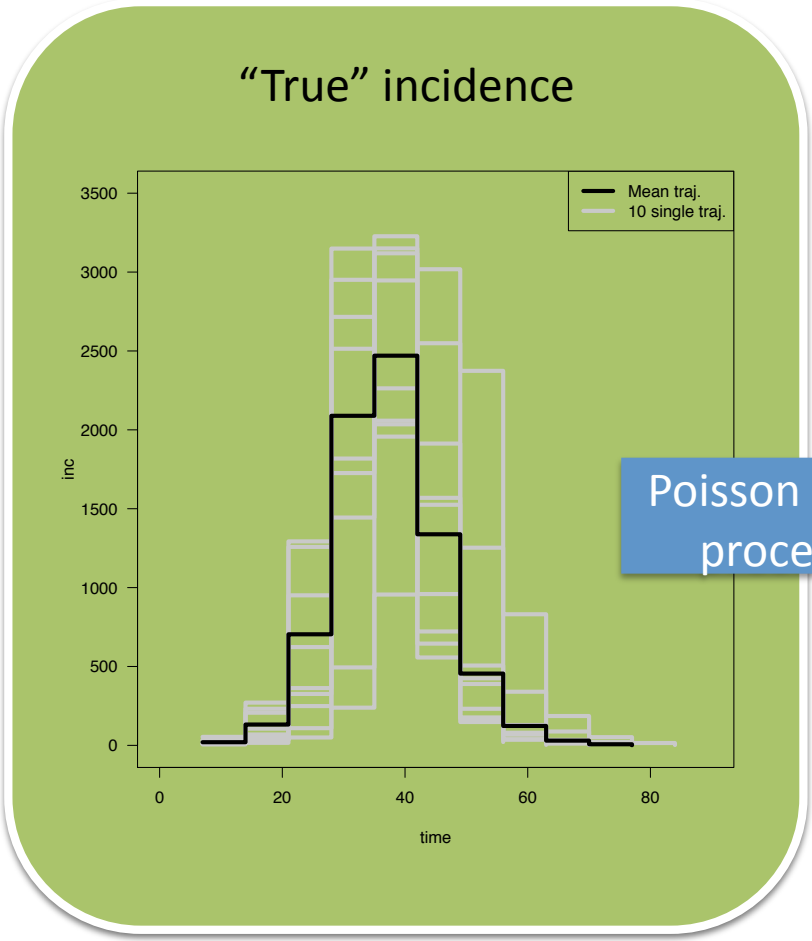
Real conditions for time series analysis

Incidence time series are weekly aggregated for influenza

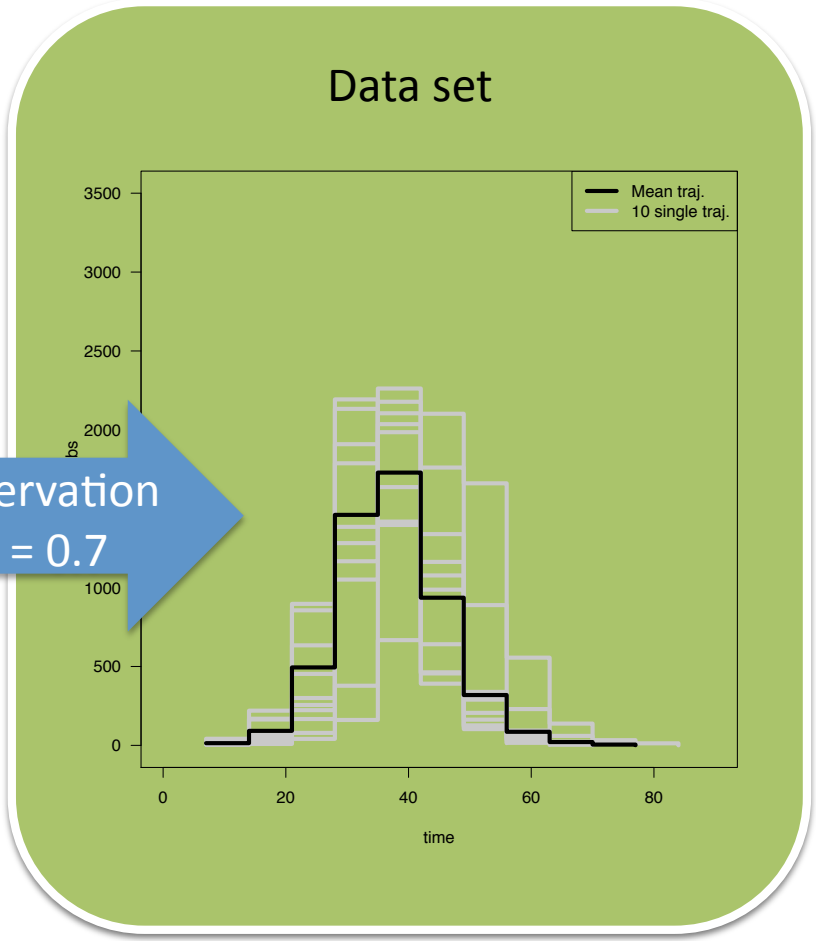
Incidence is only partially reported in the data

We use a statistical framework for parameter inference:
the Maximum likelihood via Iterated Filtering algorithm (Ionides, 2006)
Perfectly suited to non-linear stochastic compartmental models
Allow to explore the likelihood surface

Simulated data set



Poisson observation process ρ = 0.7



Real conditions for time series analysis

Stochastic compartmental model implemented by an Euler multinomial scheme with time step δt (tends to a continuous time Markov chain)

- For our simple SIR model:

$$Incidence(t, t+\delta t) \backsim Binomial(S, p_1)$$

$$Recovery(t, t+\delta t) \backsim Binomial(I, p_2)$$

$$p_1 = 1 - \exp(-\beta \frac{I^\alpha}{N} \delta t) \qquad p_2 = 1 - \exp(-\nu \delta t)$$

- Only one exponent on the number of infectious individuals
- δt << duration of infectiosity (2 < IP < 7 days for influenza): δt = 6 hrs
- β is the effective contact rate per time unit: $\beta = R_0 \nu$

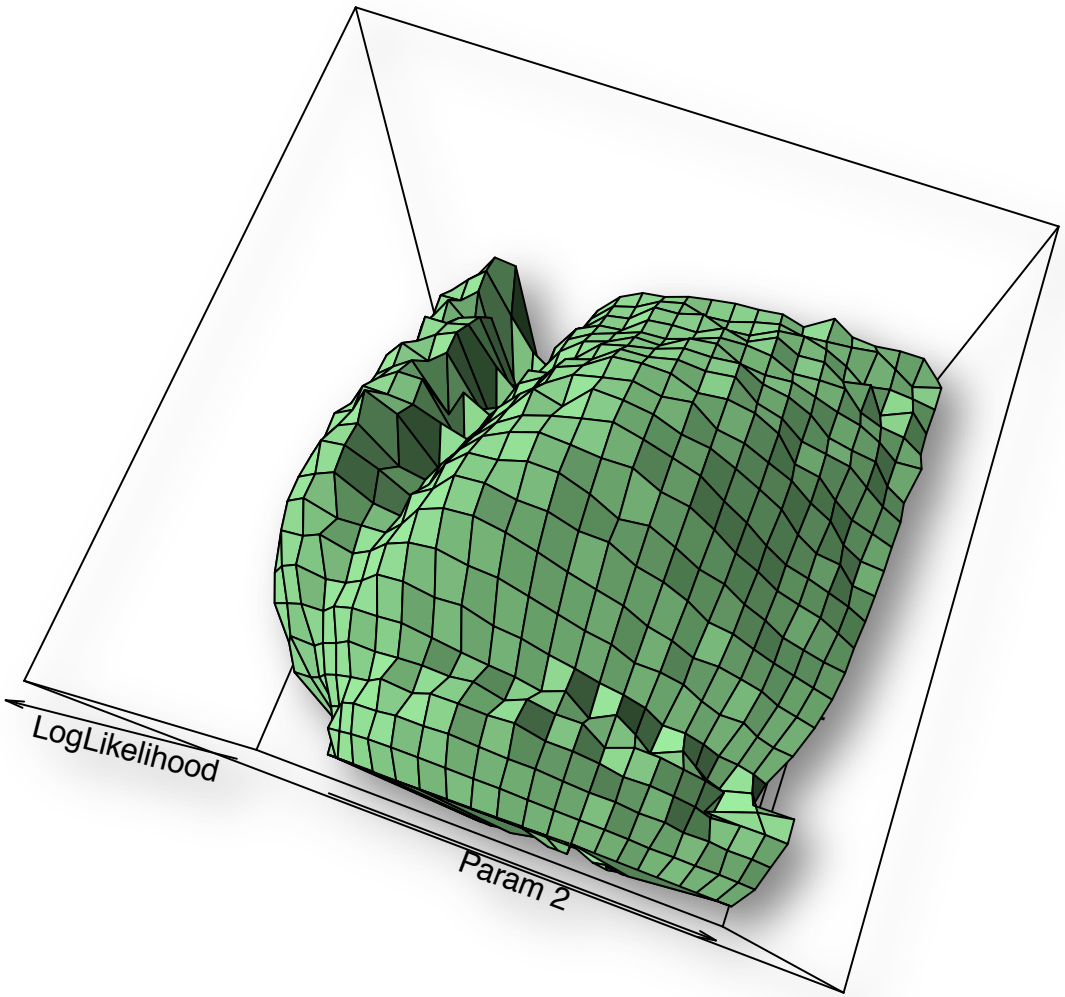
What do we estimate?

6 dimensional space

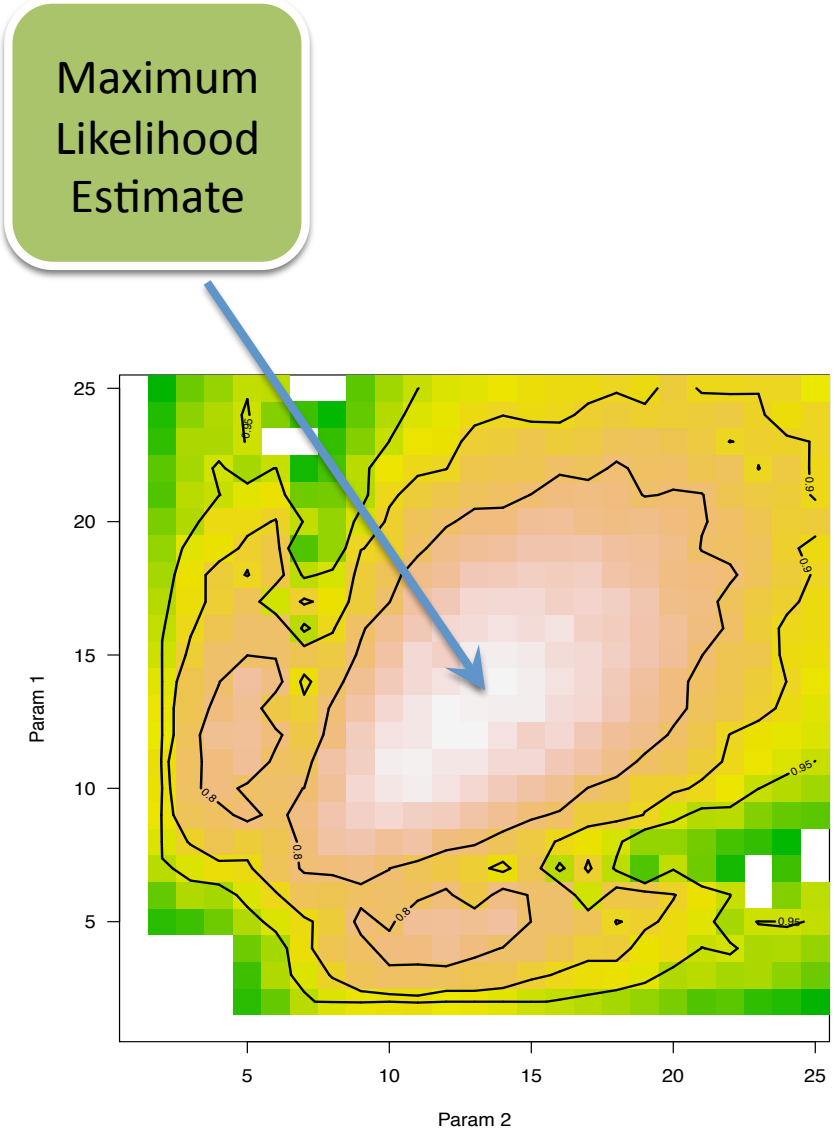
- 2 epidemiological parameters: R_0 & IP
- 1 heterogeneity parameter: α
- 1 observation parameter: ρ (reporting rate)
- 2 initial conditions: S_0 & I_0

The likelihood surface is thus 6 dimensional
How to explore it?
Is there a global maximum? Is it flat or sharp?

Exploring the likelihood surface

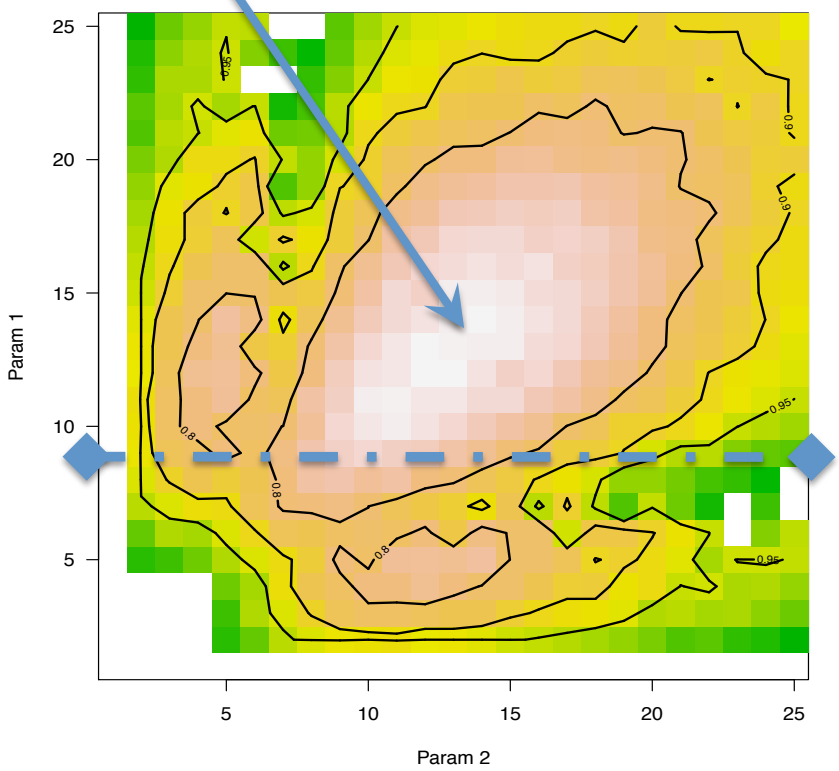


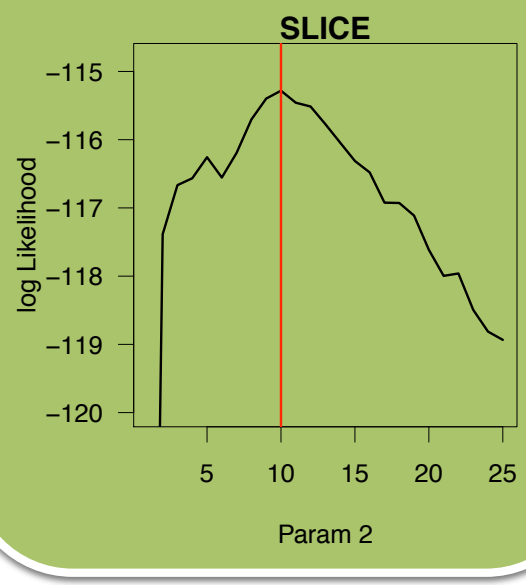
Exploring the likelihood surface



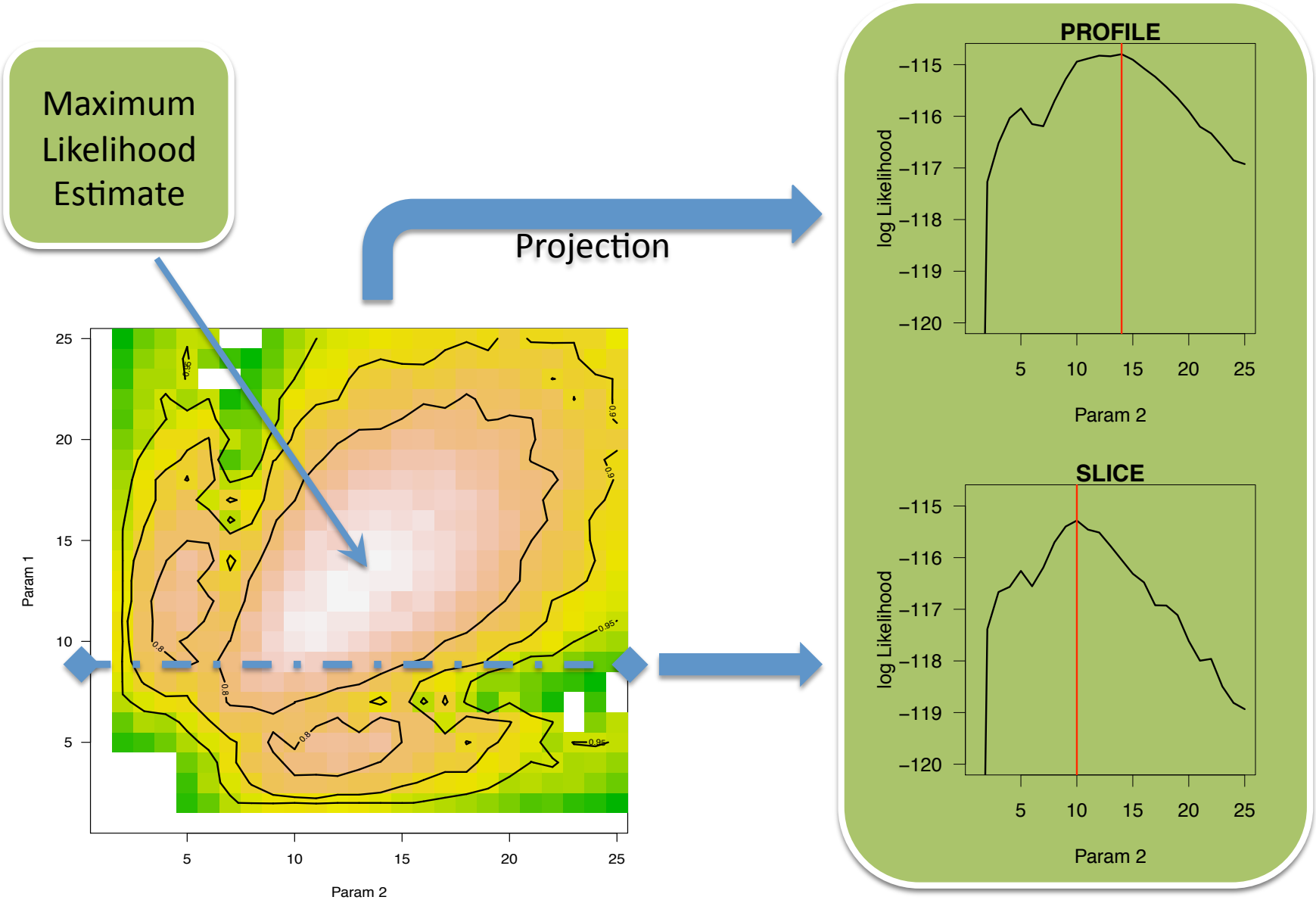
Exploring the likelihood surface

Maximum Likelihood Estimate





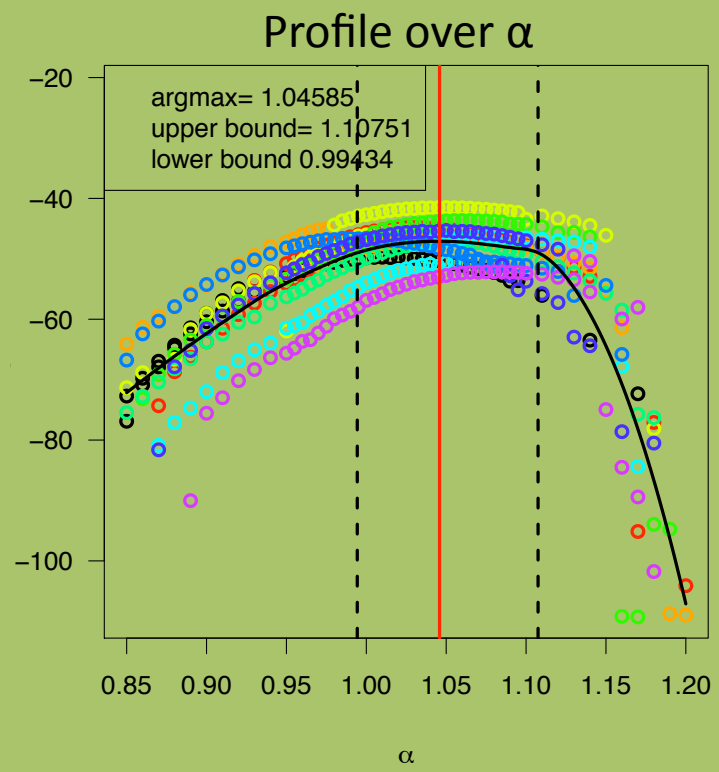
Exploring the likelihood surface

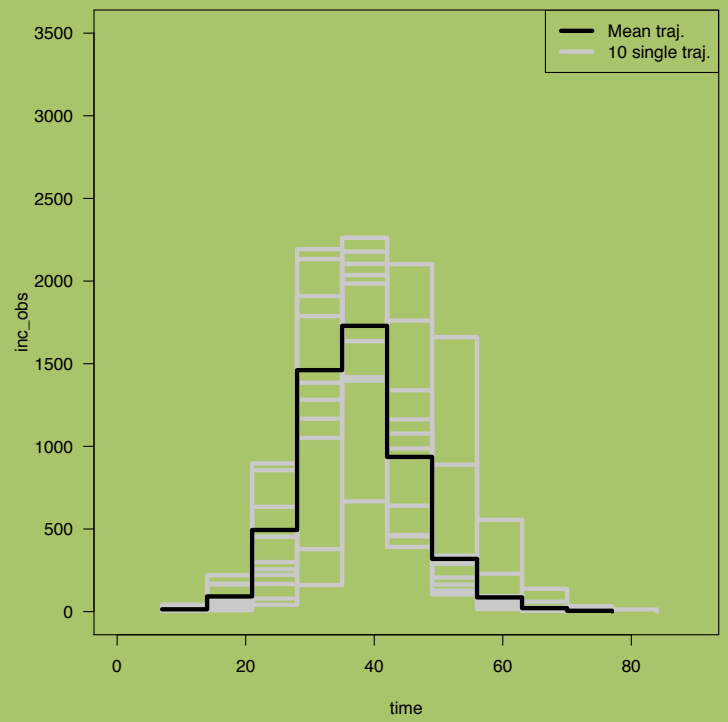


Profile: find the global maximum of the likelihood surface

R_0, IP, ρ, I_0 & S_0 are estimated for each trajectory and each fixed value of α

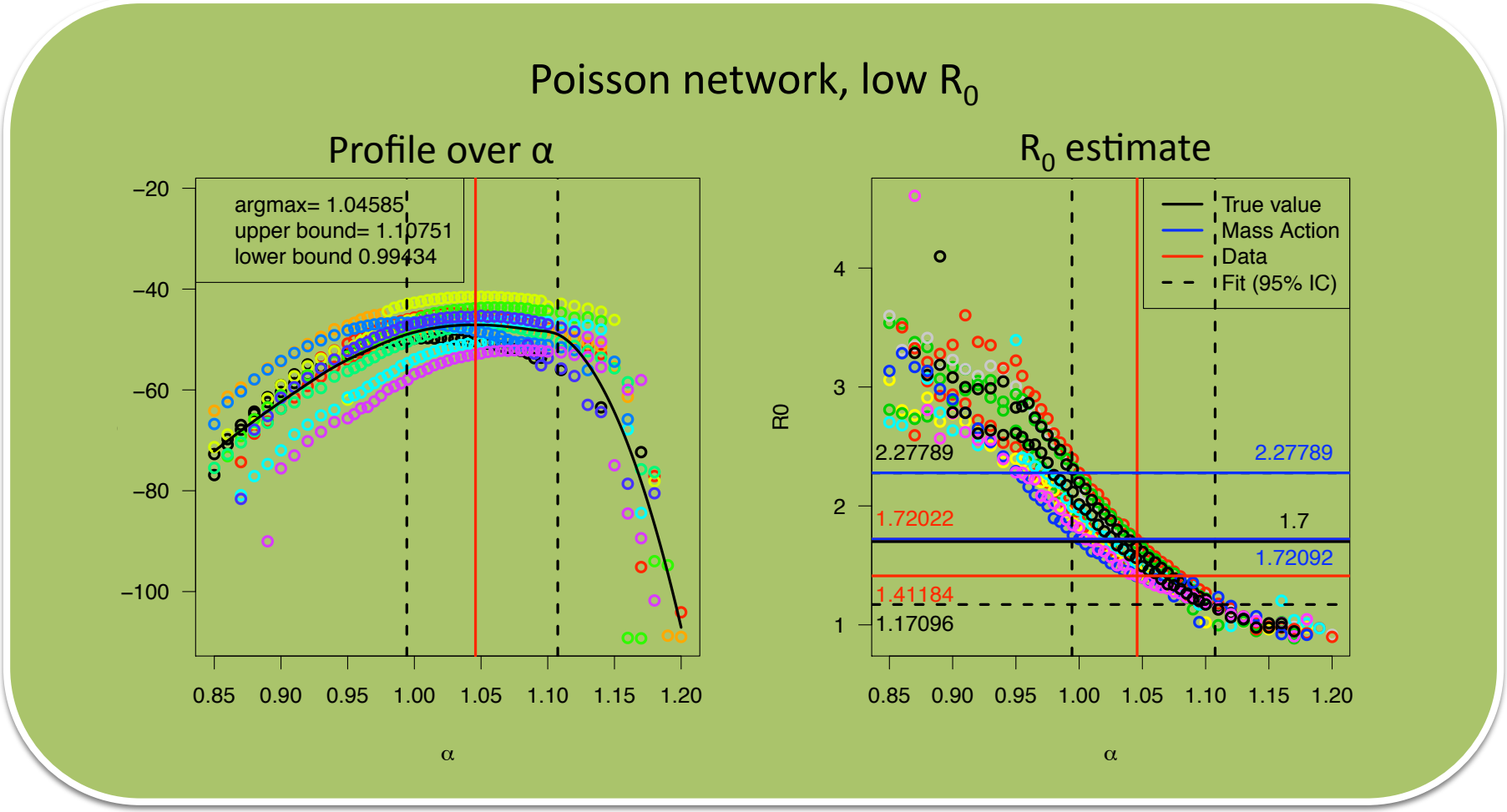
Poisson network, low R_0





Profile: find the global maximum of the likelihood surface

R_0, IP, ρ, I_0 & S_0 are estimated for each trajectory and each fixed value of α

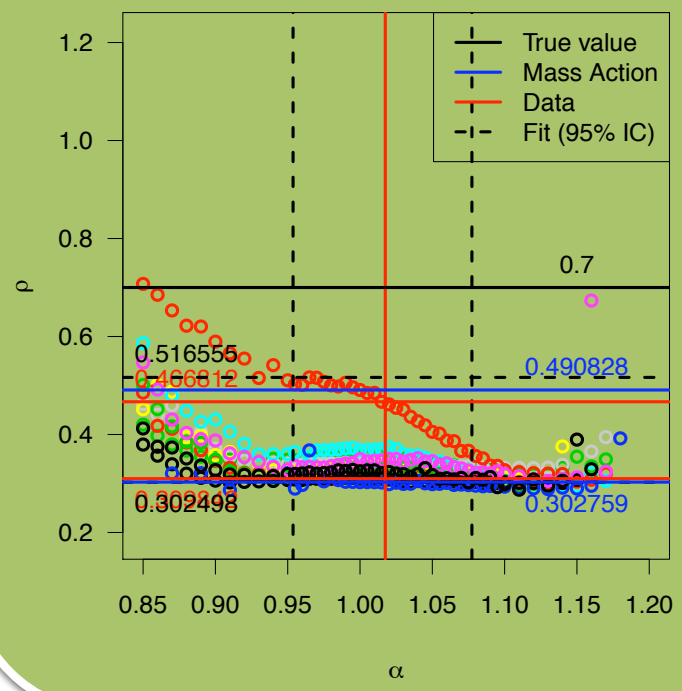


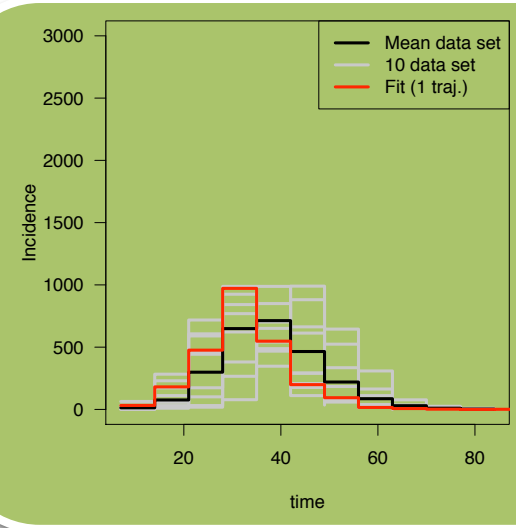
Results

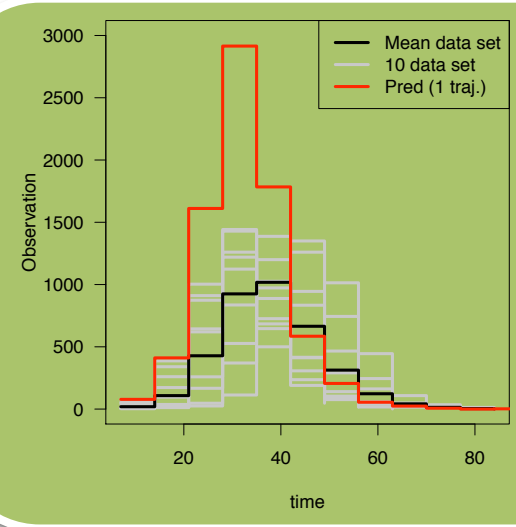
Network	Maximum Likelihood Estimate: α	"True" parameters well estimated	Mass-Action
Poisson (low R_0)	1.05 [0.99 – 1.10]	R_0, IP, ρ	R_0, IP, ρ
Exponential (low R_0)	1.02 [0.95 – 1.07]	R_0, IP	R_0, IP, I_0
Poisson (high R_0)	1.04 [0.99 – 1.05]	ρ, I_0	R_0, ρ, I_0
Exponential (high R_0)	0.97 [0.93 – 1.01]	R_0, I_0	I_0
Small-World (high R_0)	1.07 [1.00 – 1.08]	✗	✗

Exponential network (low R_0)

Both R_0 and IP are well estimated
The reporting rate of the observation process is well under the "true" value (0.35 vs 0.7)

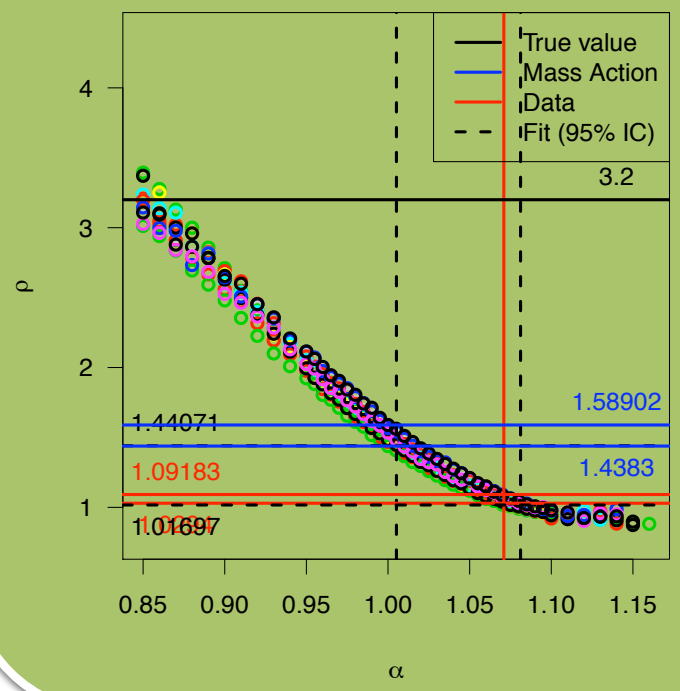


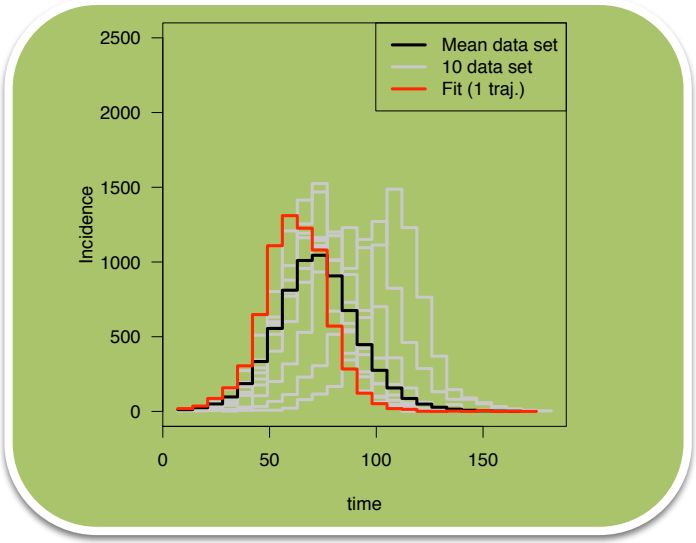


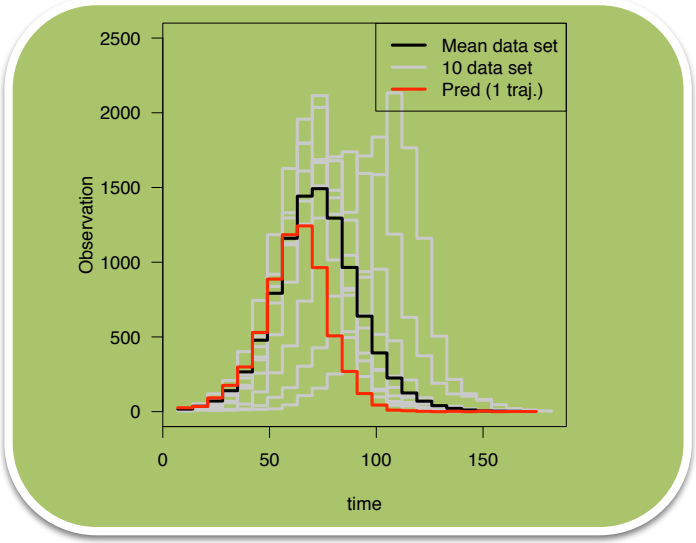


Small-World network (high R_0)

Surprisingly, all the parameters are badly estimated
The R_0 is well under the "true" value (1.1 vs 3.2)
Mass-action also fails







Conclusion

Statistical framework to test whether simple models are reliable in a realistic context of parameter inference via time series analysis:

- For small $R_0 \approx 1.5$ the mass-action law allows to accurate estimates for all networks tested
- For higher R_0, parameter inference is poorly reliable and α is not very helpful
- For exponential networks, epidemiological parameters can accurately be estimated but the resulting dynamics is false
- If the real contact structure of influenza transmission is small-world we are underestimating the R_0 even if the fit is very good

Ongoing work:

- Random networks with higher clustering
- Effect of the mean contact degree
- Inference with other phenomenological functions
- Dynamic networks

Mechanistic derivation via the theory of heterogeneous populations (Novozhilov)

Separable mixing (≠ static network):
n(t,ω) denotes the density of individuals in the population, which are making *ω* contacts in average. Each individual can be contacted by another individual proportionally to their average number of contacts.

$$\frac{\partial}{\partial t} i(t,\omega) = r\omega s(t,\omega) \frac{\int_\Omega \omega i(t,\omega) d\omega}{\int_\Omega \omega n_0(\omega) d\omega}$$

$$\frac{d}{dt} I(t) = rh(S)\left[1 - \frac{h(S)}{K}\right]$$

If ω follows a Gamma(k,v) distribution

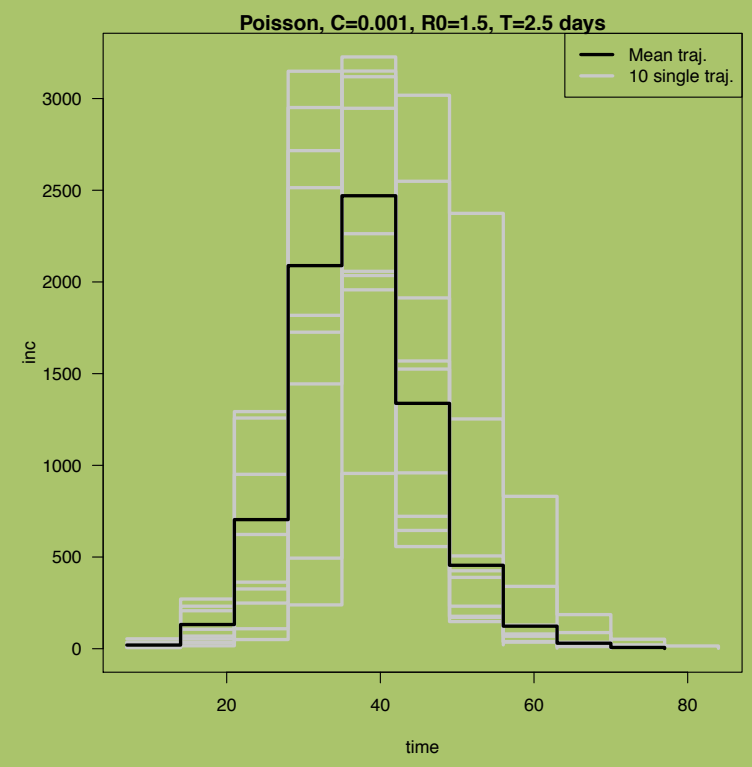
$$h(S) = \frac{kS}{\nu}\left[\frac{S}{S_0}\right]^{1/k}$$

The exponent (p=1+1/k) range from 1 (mass action) to 2 (exponential distribution) but usually estimates are below 1

Disease spread on Networks

Stochastic simultaneous SIR:

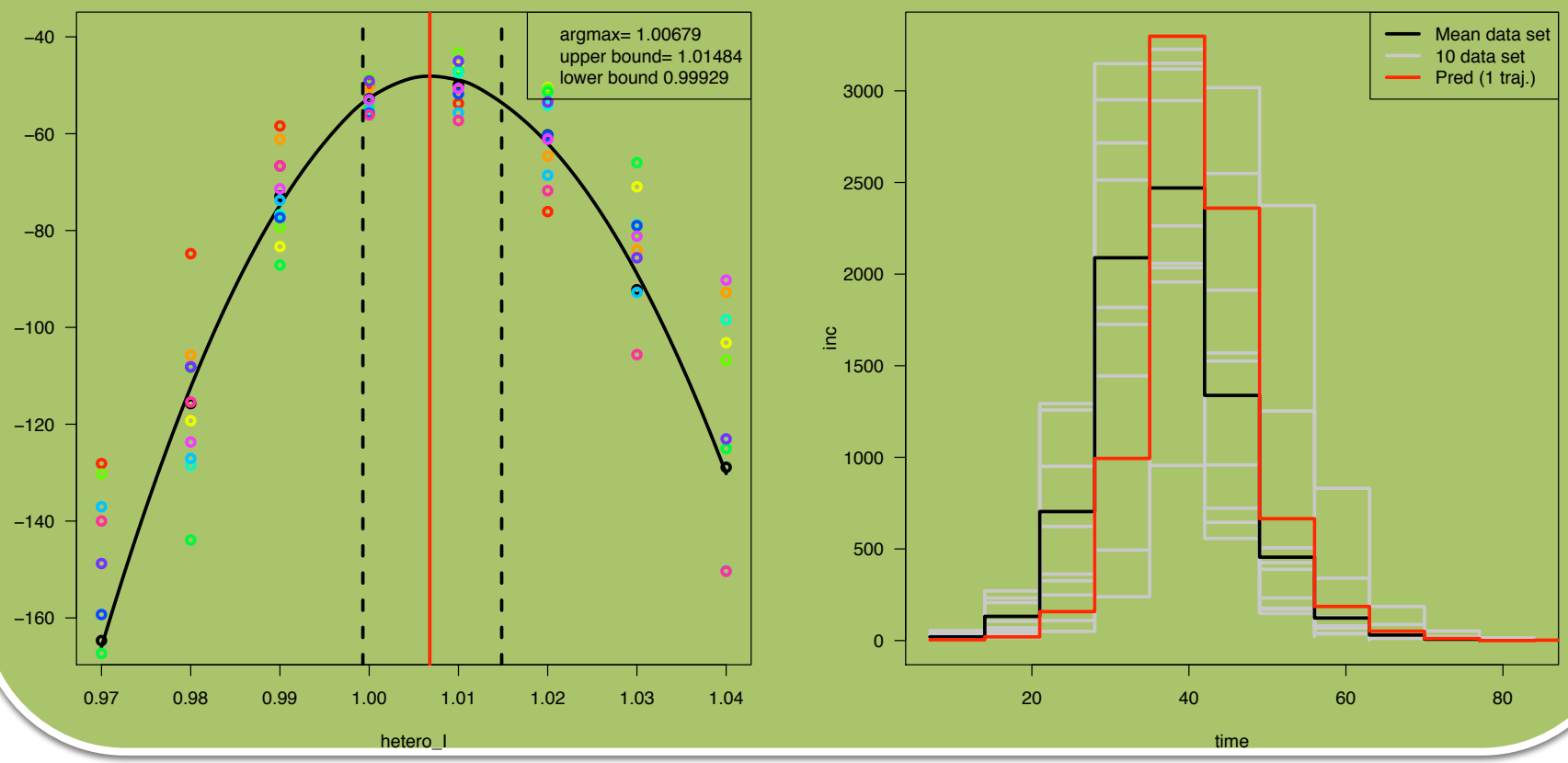
- At t=0, S(0)=9999 and I(0)=1 (chosen randomly)
- While t<t_end
  1. For each s in S:
     a) n_i: infectious neighbor
     b) s becomes infected with probability: $p=1-\exp(-rn_i\delta t)$
     c) If s -> I, he remains infectious until: T=t+IP (IP=Exp(v))
  2. For each i in I:
     a) If t>T: recover
  3. $t=t+\delta t$



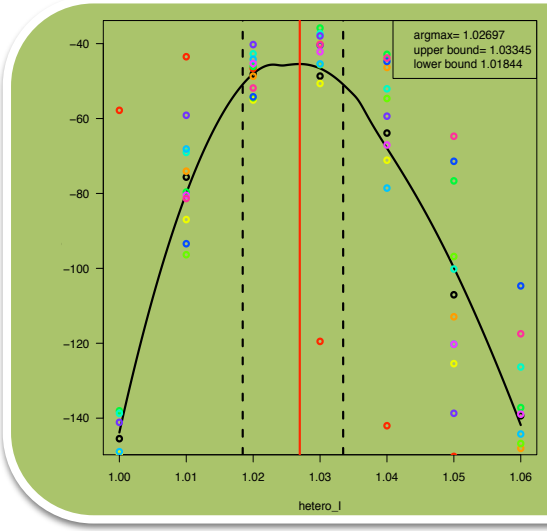
Slice: is there a "good" local maximum near the "true" parameter set?

R_0, IP, ρ, I_0 & S_0 are fixed to the values used to obtain the data set ➔ slice on α

Poisson network, low R_0: good local maximum near α ≈ 1

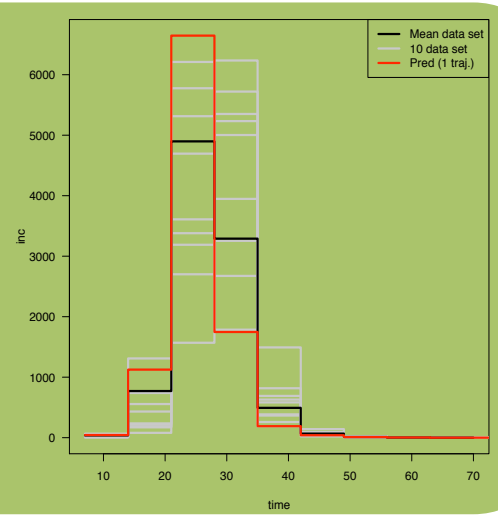


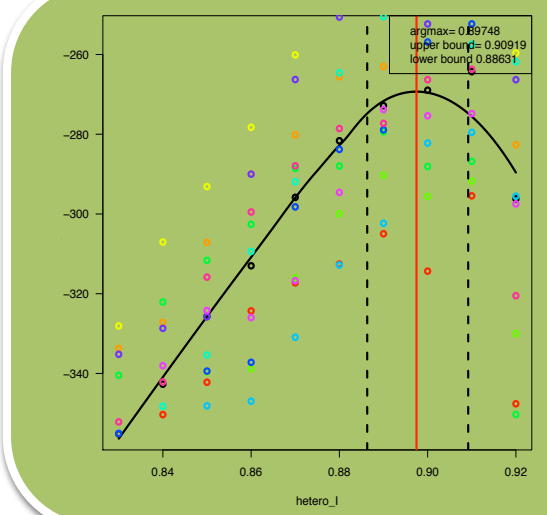
Slice: is there a “good” local maximum near the “true” parameter set?



Poisson network
High R_0
$\alpha \approx 1$

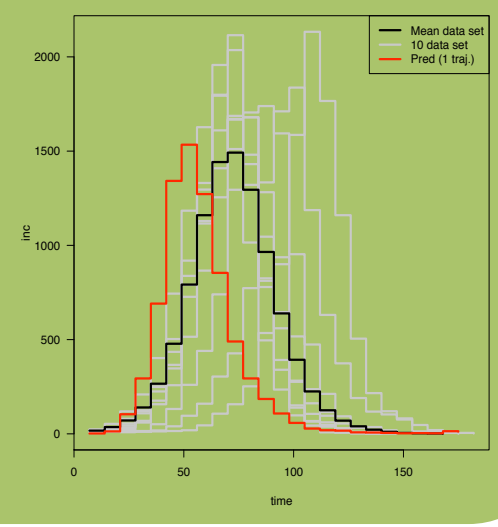
✔





Small-World network
High R_0
$\alpha \approx 0.9$

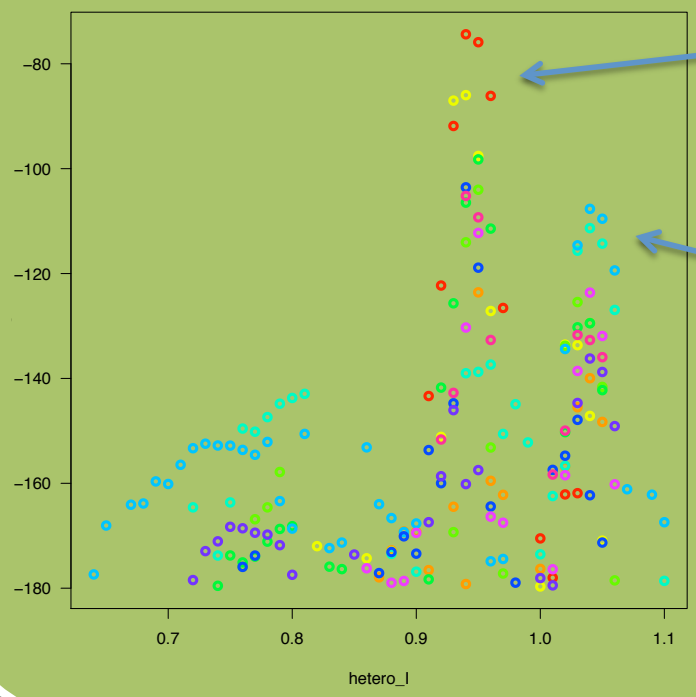
✔



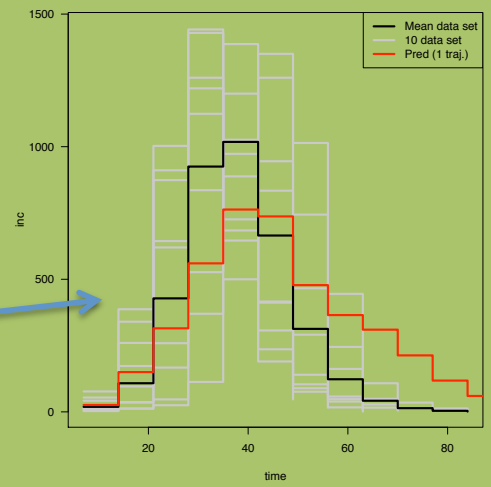
Slice: is there a “good” local maximum near the “true” parameter set?

Exponential network, low R_0
2 local maxima:
$\alpha \approx 0.95$ & $\alpha \approx 1.05$

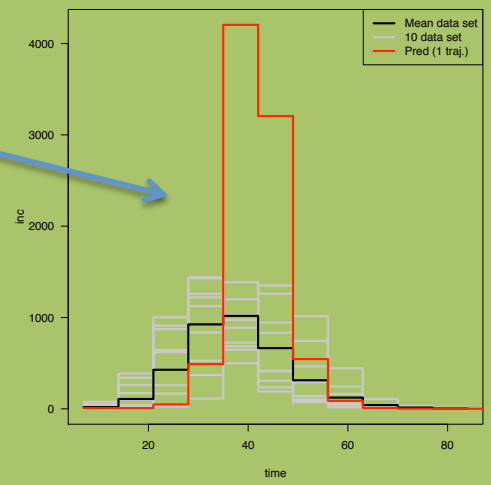
✗



Too long !



Too high!



Summary: slice

For the Poisson and Small-World networks there is a "good" local maximum for both low and high R_0 near the "true" parameter set:

- For Poisson network, $\alpha \approx 1$ (homogeneous mixing)
- For Small-World network, $\alpha < 1$ (as in Roy & Pascual's study)

BUT IS IT THE GLOBAL MAXIMUM OF THE LIKELIHOOD SURFACE ?

For the Exponential network there is no "good" local maximum maximum for neither low and high R_0 near the "true" parameter set

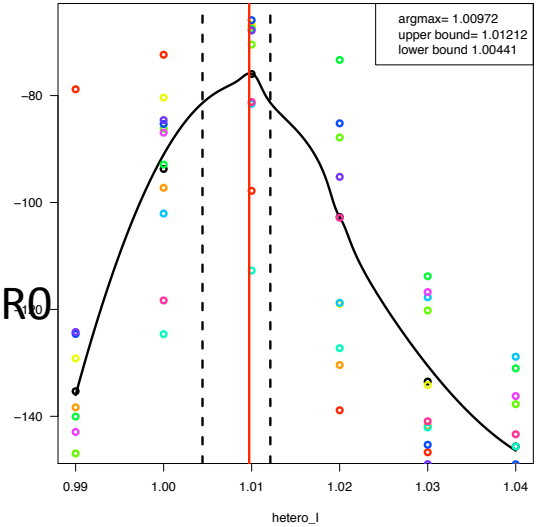
WHERE IS THE GLOBAL MAXIMUM ? IS IT FAR FROM THE "TRUE" PARAMETER SET ?

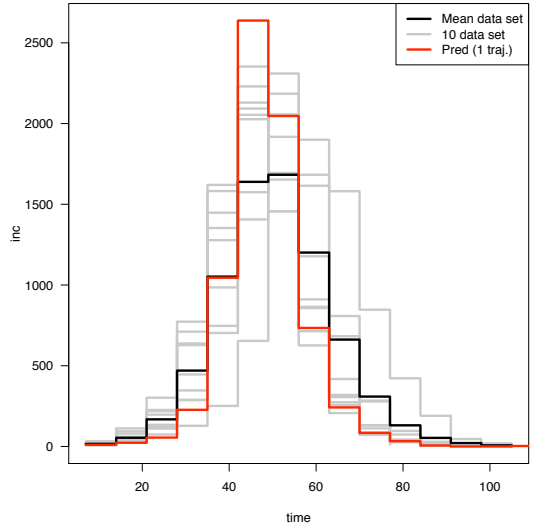
Results

Network	Slice: α	Good Local Max ?	Maximum Likelihood Estimate: α	"True" parameters recovered	Mass-Action
Poisson (low R_0)	1.007	✔	1.05 [0.99 – 1.10]	R_0, IP, ρ	R_0, IP, ρ
Exponential (low R_0)	0.95 & 1.05	✗	1.02 [0.95 – 1.07]	R_0, IP	R_0, IP, I_0
Poisson (high R_0)	1.027	✔	1.04 [0.99 – 1.05]	ρ, I_0	R_0, ρ, I_0
Exponential (high R_0)	1.01	✗	0.97 [0.93 – 1.01]	R_0, I_0	I_0
Small-World (high R_0)	0.9	✔	1.07 [1.00 – 1.08]	✗	✗

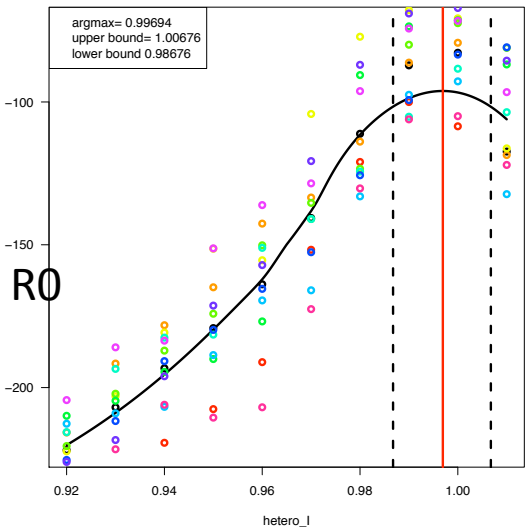
appendix: small world 0.1

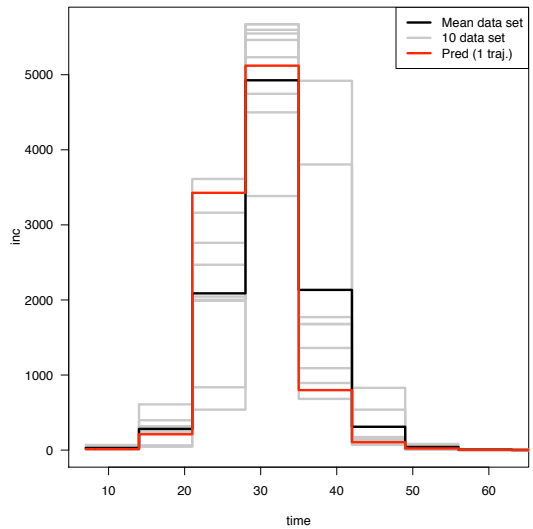
low R0





High R0





appendix (1)

- We apply a geometric correction for the duration of infectiosity: $T = \frac{\delta t}{1 - \exp(-\nu \delta t)}$