SYNTHETIC LIKELIHOOD APPROACH FOR QUANTIFYING MULTI-SCALE EPIDEMIC PROCESSES FROM LARGE AND COMPLEX DATA SETS

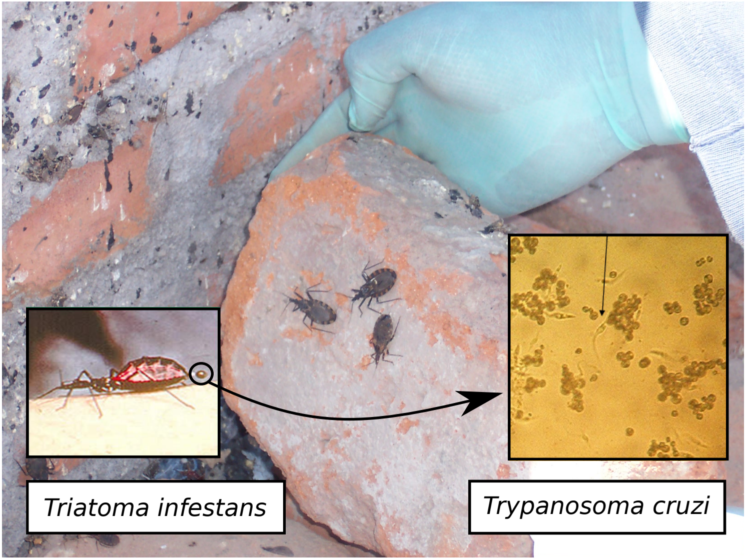
APPLICATION TO A VECTOR OF CHAGAS DISEASE

Corentin M. Barbu, Karthik Sethuraman & Michael Z. Levy

University of Pennsylvania

November 6, 2013

Triatoma Infestans & Trypanosoma cruzi



CHAGAS DISEASE CONTROL

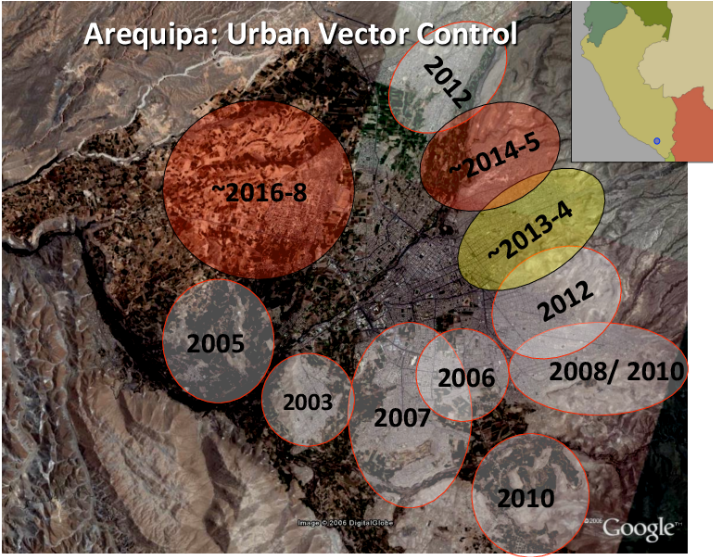
AVAILABLE TREATMENT AND PREVENTION METHODS

- No vaccine
- Imperfect treatment
- Vector control

TYPICAL VECTOR CONTROL SEQUENCE

- Initial treatment (blanket indoor spraying)
- Surveillance
    - Community reports → inspections
    - Infestation → retreatment

THE CONTROL CAMPAIGN IN AREQUIPA



HALF-WAY EVALUATION

ATTACK PHASE EVALUATION

- 10 years to treat half the city
- 60-85% participation in initial treatment
- very efficient treatment ($>$99% households cleared)

SURVEILLANCE

- redispersal from some non-treated households
- infected insects

CAN A REACTIVE SURVEILLANCE BE ENOUGH?

FOCI R_0 THINKING

As we detect/treat easily neighbors of reporting households:

- Local dispersal is already dealt with
- Medium range dispersal may be handled
- Large scale dispersal is catastrophic

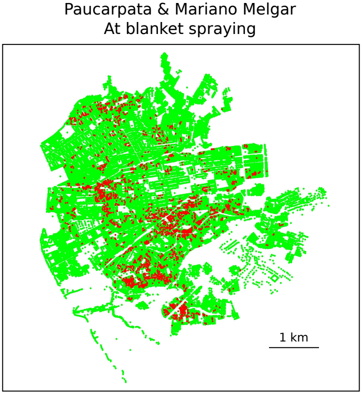
Can we on average control a foci before it generates infestation too far to be handled?

- Quantification of the dispersal at these three scales

LEVERAGING COMPLEX DATA

DECOUPLING SPATIAL AND TEMPORAL SCALE

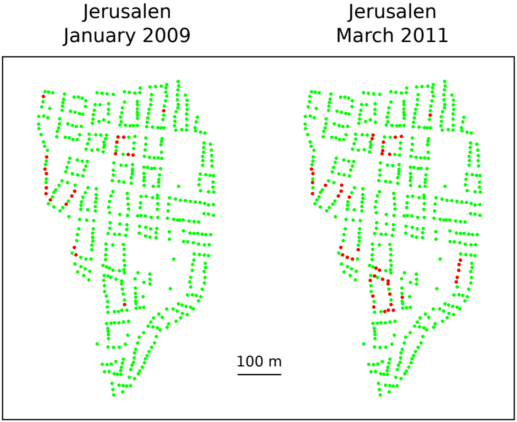
- cross-sectional → spatial patterns different scales



- longitudinal (400 households)→ time

DECOUPLING SPATIAL AND TEMPORAL SCALE

- cross-sectional → spatial patterns different scales
- longitudinal (400 households)→ time



COMBINE AND CONQUER

Cross-sectional study

- Unknown time
  → Impossible to fit the rate of movement
- Large scale
  → Relative rates at different scales

Longitudinal study

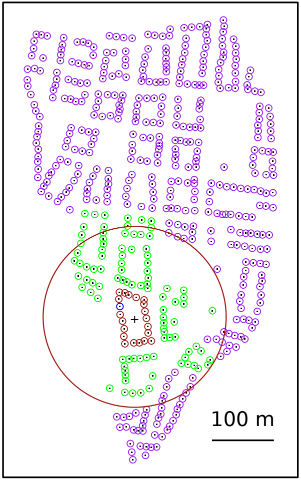
- Known time
  → Fit of the rate of movement
- Small scale
  → Vague fit of large scale movement

BAYESIAN COMBINATION:

The posterior from cross-sectional study becomes the prior for the longitudinal study.

⇒ The cross-sectional study gives the relative abundance of the dispersal at different spatial scales, the longitudinal a study gives the time-scale.

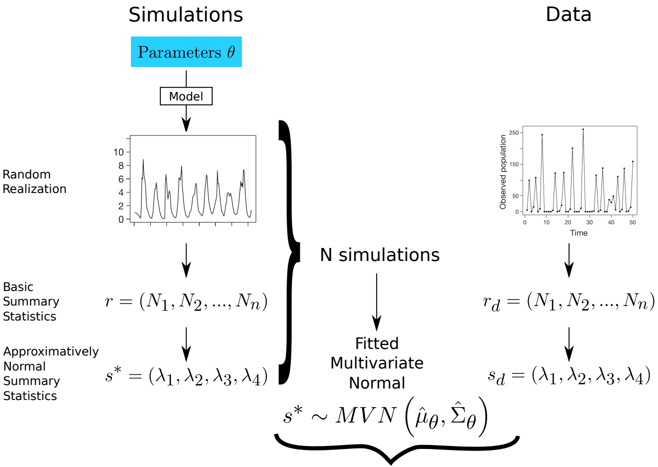
THE HOP/SKIP/JUMP MODEL



THE HOP/SKIP/JUMP MODEL OF DISPERSAL

- hop: within a city-block
- skip: in neighboring city-blocks
- Jumps: anywhere

THE SYNTHETIC LIKELIHOOD



Simulations

Parameters θ

Model

Random Realization

Basic Summary Statistics

$$r = (N_1, N_2, ..., N_n)$$

Approximatively Normal Summary Statistics

$$s^* = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$$

N simulations

Fitted Multivariate Normal

Data

$$r_d = (N_1, N_2, ..., N_n)$$

$$s_d = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$$

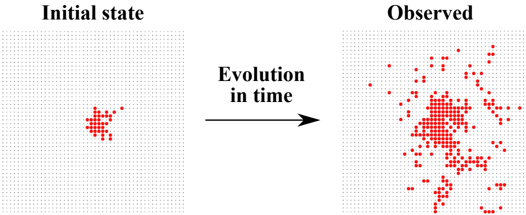
$$\underbrace{s^* \sim MVN\left(\hat{\mu}_\theta, \hat{\Sigma}_\theta\right)}$$

$$l_s(\theta) = -\frac{1}{2}\left(s_d - \hat{\mu}_\theta\right)^T \hat{\Sigma}_\theta^{-1}\left(s_d - \hat{\mu}_\theta\right) - \frac{1}{2}\log\left|\hat{\Sigma}_\theta\right|$$

Wood, S. N. Statistical inference for noisy nonlinear ecological dynamic systems Nature, Nature Publishing Group, 2010, 466, 1102-1104

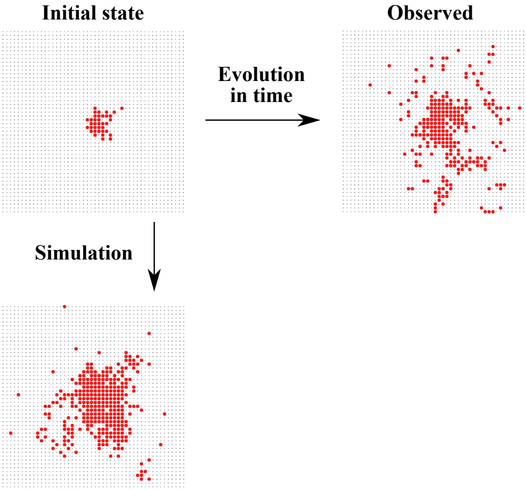
THE SYNTHETIC LIKELIHOOD

$$\textit{Syn.Lik.}(\theta) = P\left(\textbf{Observed Sum. Stats}|\text{Sum. Stats}(\theta)\right)$$



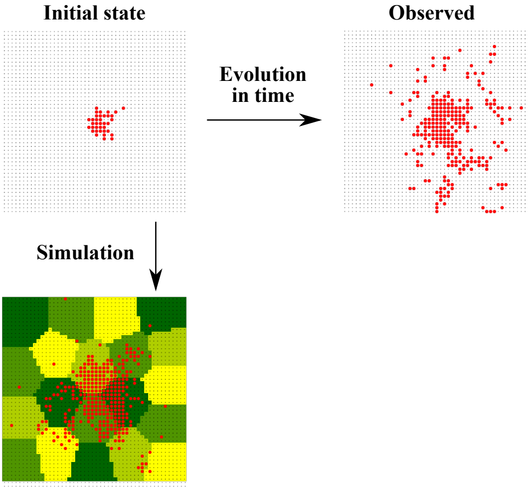
The synthetic likelihood

$$Syn.Lik.(\theta) = P\left(\textbf{Observed}\text{ Sum. Stats}|\text{Sum. Stats}(\theta)\right)$$



THE SYNTHETIC LIKELIHOOD

$$Syn.Lik.(\theta) = P\left(\textbf{Observed}\ \text{Sum. Stats}|\text{Sum. Stats}(\theta)\right)$$



Initial state

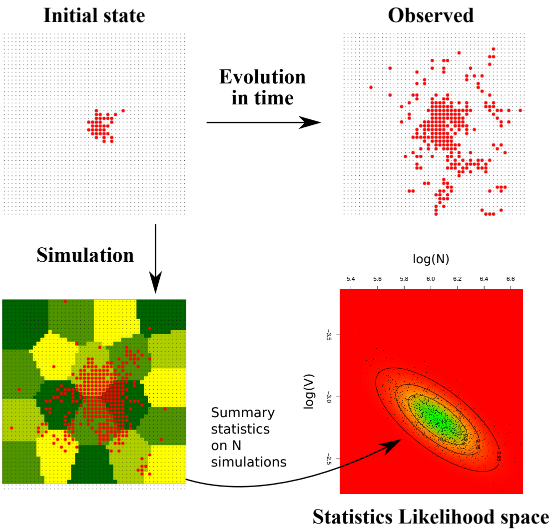
Evolution in time

Observed

Simulation

THE SYNTHETIC LIKELIHOOD

$$Syn.Lik.(\theta) = P\left(\textbf{Observed}\ \text{Sum. Stats}|\text{Sum. Stats}(\theta)\right)$$



Initial state

Evolution in time

Observed

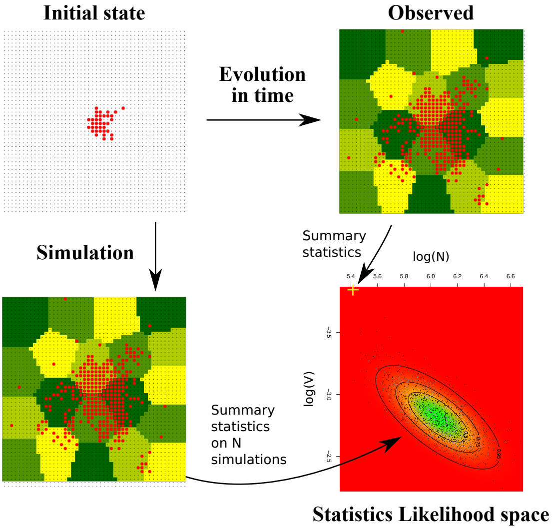
Simulation

Summary statistics on N simulations

Statistics Likelihood space

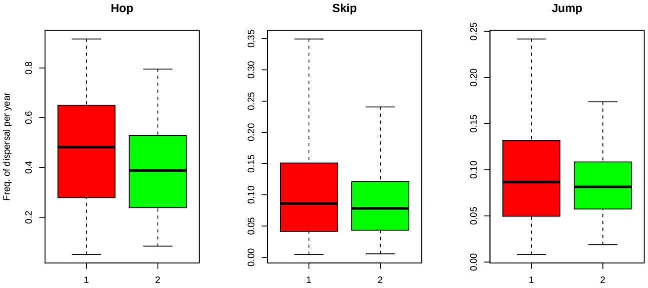
THE SYNTHETIC LIKELIHOOD

$$Syn.Lik.(\theta) = P\left(\textbf{Observed}\ \text{Sum. Stats}|\text{Sum. Stats}(\theta)\right)$$



DISPERSAL FREQUENCY

IMPACT OF THE COMBINED ANALYSIS



MODES OF DISPERSAL

- hop $\sim$ walk (68% [39-86])
- skip $\sim$ flight (16% [0.8-45])
- jump $\sim$ passive (15% [3-32])

$\Rightarrow$ Mainly active dispersal but significant passive dispersal

Re-infestation risk from one household

Average time to new infestation

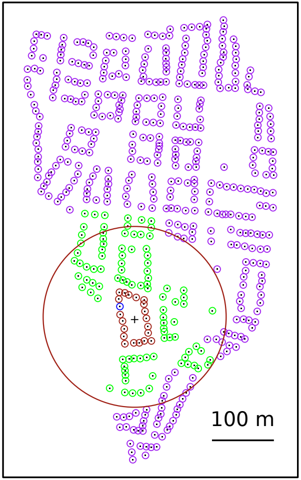
Total: 1.8 years [1.0-4.3]

Re-infestation from a single household

Time (years)	Same City Block	Diff. City Block	Neigh. City Blocks	Further away	Total
0.5	0.24	0.11	0.05	0.06	0.35
1.0	0.56	0.29	0.13	0.16	0.85
1.5	0.91	0.56	0.24	0.32	1.48
2.0	1.47	0.99	0.40	0.59	2.46
2.5	1.87	1.48	0.58	0.90	3.35
3.0	2.63	2.50	0.94	1.57	5.14

Average number of households infested, starting from a single household randomly chosen among the households in the cross-sectional study and using 1000 different parameter sets drawn from the posterior of the joint analysis.

PUBLIC HEALTH IMPACT



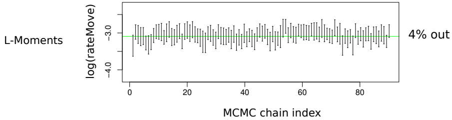
IMPLICATIONS FOR THE SURVEILLANCE CAMPAIGN

1. Same city-block inspection and treatment can stop epidemics.
2. Inspection of neighboring city-blocks is not efficient.
3. Infestation → treatment time must stay under 2 years.

VALIDATION OF THE ESTIMATES

INTERNAL VALIDATIONS

- Synthetic likelihood approximation: coverage analysis



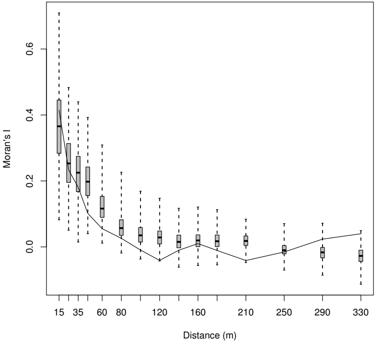
VALIDATION OF THE ESTIMATES

INTERNAL VALIDATIONS

- Synthetic likelihood approximation: coverage analysis

- Adequacy of the model: autocorrelation generated by the model

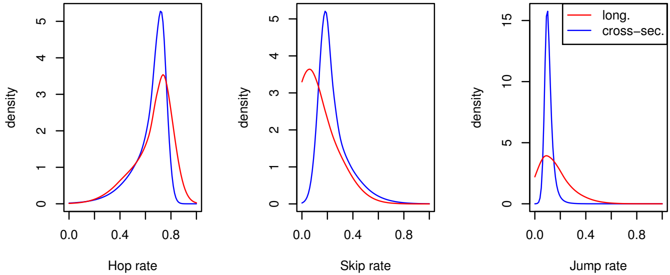
T. infestans, Jerusalen



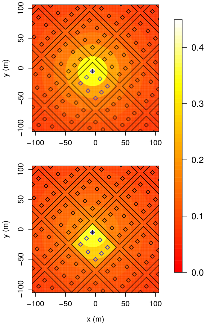
VALIDATION OF THE ESTIMATES

INTERNAL VALIDATIONS

- Synthetic likelihood approximation: coverage analysis
- Adequacy of the model: autocorrelation generated by the model
- Consistency of the data: similar posteriors for the rates in independent analysis.



CONNECTION WITH OTHER APPROACHES



EPIDEMIOLOGICAL DATA

- Average number of household found infested in close neighbors of infested, non-treated households, 6 months to 2 years after treatment: 0.27 [0.1-0.6].
- Autocorrelation analysis [a]

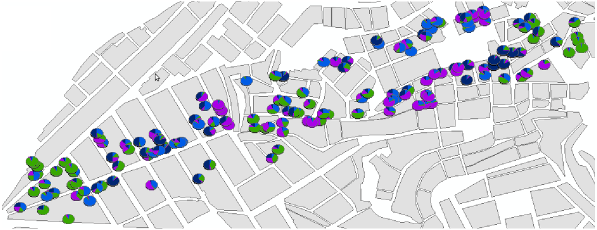
---

[a] Barbu, C. et al. The Effects of City Streets on an Urban Disease Vector PLOS Computational Biology, Public Library of Science, 2013, 9, e1002801

CONNECTION WITH OTHER APPROACHES

MICRO-SATELLITE ANALYSIS

- Micro-satellite analysis (Foley, in prep)



- Existence of migration at the scale of dozens of kilometers (Foley 2013).

Foley, E. A.; et al. 2013 Population structure of the Chagas disease vector, Triatoma infestans, at the urban-rural interface Molecular Ecology

STRENGTHS AND LIMITATIONS

EXISTING STRENGTHS

- Can handle arbitrary complex dispersal models
- Can combine cross-sectional and longitudinal datasets
- Can integrate large datasets by “small scale merging”

CURRENTLY WORKING ON

- Improving joint posteriors passing
- Improving usability (R package)
- Adding distance kernel handling

FUTURE DIRECTIONS

- Accounting for heterogeneity (households, observation)
- Using information from surveillance phase
- Integrating genetic information

THANK YOU

FIELD TEAM IN AREQUIPA