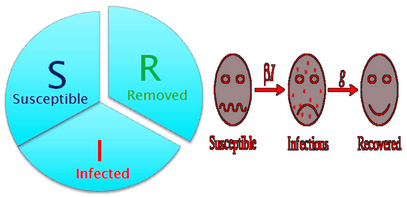


EMULATING A GRAVITY MODEL TO INFER THE SPATIOTEMPORAL DYNAMICS OF AN INFECTIOUS DISEASE

Roman Jandarov, **Murali Haran***, Ottar Bjørnstad and Bryan Grenfell

Department of Statistics, The Pennsylvania State University; Department of Entomology and Center for Infectious Disease Dynamics, The Pennsylvania State University; Ecology and Evolutionary Biology, Princeton University.

PENNSTATE 1855

BILL & MELINDA GATES foundation

CIDD CENTER FOR INFECTIOUS DISEASE DYNAMICS

OVERVIEW

- Our goal is to fit a gravity model (the Gravity T-SIR model) that describes the space-time dynamics of measles.
- Challenges:
  - Likelihood evaluations are expensive so computationally challenging.
  - Likelihood-based fitted model does not match biologically important characteristics of the data.
- Idea: Why not fit the model by trying to directly match the important characteristics of the data?
- Contributions:
  - General simulation-based approximate approach for inference using Gaussian processes
  - Avoids having to write down/evaluate the likelihood function
  - Only requires a relatively small number of simulations from the model

GRAVITY TIME SERIES SIR MODEL

- SIR = Susceptible-Infected-Recovered.
- Gravity Time Series SIR Model = SIR model for local dynamics + explicit formulation for the spatial transmission between different host communities. (Xia et al., 2004)
- **Number of incidents (I)**:

$I_{k(t+1)} = \text{Poisson}(\lambda_{t+1})$, where $\lambda_{t+1} = \beta_t S_{kt}(I_{kt} + L_{kt})^{\alpha}$

- **Number of susceptibles (S)**:

$$S_{k(t+1)} = S_{kt} + B_{kt} - I_{k(t+1)}$$

- **Unobserved number of infected immigrants (L)**:

$$L_{kt} = \text{Gamma}(m_{kt}, 1),$$

where $m_{kt} = \theta N_{kt}^{\tau_1} \sum_{j=1, j\neq k}^{K} \frac{(Ijt)^{\tau_2}}{d_{kj}^{\rho}}$.

PROBLEM STATEMENT

Goal: infer the unknown gravity parameters θ, τ_1, τ_2 and ρ
Data: 952 cities in England and Wales.

CHALLENGES

- **Computational challenge**: data dimensions are 546x952 = 519,792 $\Rightarrow$ expensive likelihood evaluations.
- **Poor inference/model fit**: Even if computational problem solved, likelihood-based inference results in biases in parameter estimates, model that fits data poorly.
- **Want to avoid heavy simulations:** Simulations take time making it inconvenient to use approximate Bayesian computation (ABC) methods.
- **Parameter identifiability** issues: cannot learn about all the parameters from the data.

LIKELIHOOD - BASED APPROACH

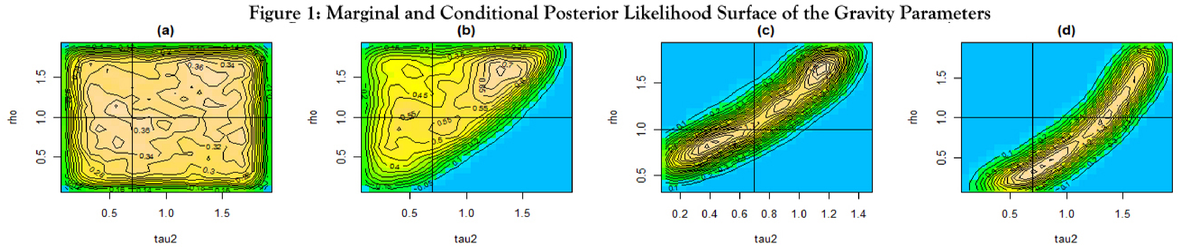
- Discretize part of the parameter space
- Instead of assuming that $L_{kt} \sim \text{Gamma}(m_{kt}, 1)$, we fix $L_{kt} \equiv m_{kt}$ to avoid integration.
- Select a grid on the range of possible values for τ_2 and ρ. For each point of the grid, we calculate and save matrices $\{M_{tk}\}$, where $M_{tk} = \sum_{j=1, j\neq k}^{K} \frac{(Ijt)^{\tau_2}}{d_{kj}^{\rho}}$.
- When calculating the likelihood, $L(\theta, \tau_1, \tau_2, \rho)$, assume that θ and τ_1 are real numbers, and τ_2 and ρ are from the selected discrete grid.
- Use pre-calculated matrices $\{M_{tk}\}$.

GP-EMULATOR APPROACH

- Idea: construct a **new likelihood function** using summary statistics that capture important biological characteristics of the disease dynamics.
- **STAGE 1:** Find approximate model using model simulations of summary statistics (e.g. proportions of zeros) for a pre-selected grid of parameters. Use Gaussian processes to approximate the model.
- **STAGE 2:** Fit approximate model to observations/data using Bayesian methods
- Detail: have a "discrepancy" term to capture model inadequacies.

RESULTS

Figure 1: Marginal and Conditional Posterior Likelihood Surface of the Gravity Parameters



Simulation using $\theta = 0.71$, $\tau_1 = 0.3$, $\tau_2 = 0.7$, $\rho = 1$
Figure 1 (a) inference (marginal) for (τ_2, ρ).
Figure 1 (b) inference for (τ_2, ρ) given $\theta = 0.71$.
Figure 1 (c) inference for (τ_2, ρ) conditioned on $\theta = 0.71)$, $\tau_1 = 0.3$
Figure 1 (d) inference for (τ_2, ρ) conditioned on $\theta = 1$, $\tau_1 = 1$

Figure 2: GP-Emulator vs Likelihood

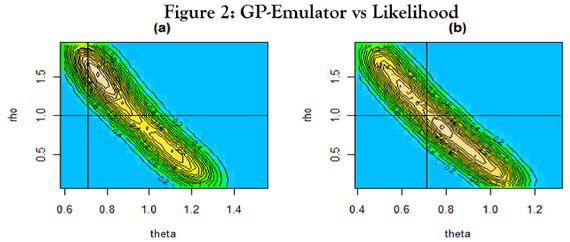


Figure 2 (a) posterior estimates using likelihood-based method is biased
Figure 2 (b) our method corrects the bias and produces better parameter inference

Figure 3: likelihood-based fitted model fails to captures important characteristics of the data

Figure3: Fitted Model from the Likelihood-Based Approach

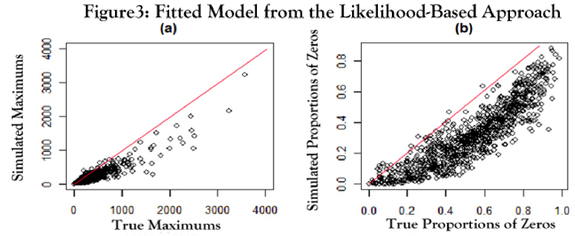
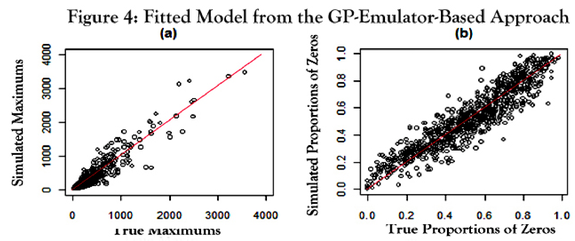


Figure 4 our method results in a fitted model that captures key features ("signature") of the disease

Figure 4: Fitted Model from the GP-Emulator-Based Approach



SUMMARY

- (i) The data have some information about the gravity parameters, but it is only possible to learn about two parameters. **Overparameterized model**
- (ii) There is no statistically significant change in these parameters during the periods of holidays vs non-holidays. **No holiday/non-holiday effect on dynamics**
- (iii) The gravity parameters do not seem to change from years 1944-55 to 1956-1966. **No vaccination effect on dynamics**
- (iv) **Can estimate movement for different cities**
- (v) **GENERAL**: The GP-based approach is computationally efficient and results in a better fitted model than standard likelihood-based methods

REFERENCES

[1] R. Jandarov, M. Haran, O. Bjørnstad, B. Grenfell (2014) Emulating a gravity model to infer the spatiotemporal dynamics of an infectious disease *Journal of the Royal Statistical Society, Series C*

[2] Y. Xia, O. Bjørnstad, B. Grenfell (2004) Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics In *American Naturalist*

ACKNOWLEDGMENT

This work was supported in part by the Bill and Melinda Gates Foundation.