Bayesian Estimation of Modified ETAS Model for Invasive Species

Binhui Deng Dr. Earvin Balderama

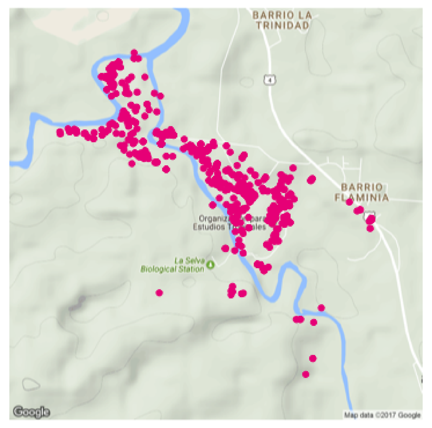
Department of Mathematics & Statistics, Loyola University Chicago, Chicago, IL, USA

Motivation

- **Invasive Plant:** Invasive alien species of plants can destroy the habitat of native species and disturb the natural evolution that takes place in the environment in which they spread.
- **ETAS Model:** Epidemic-Type Aftershock Sequence (ETAS) model, a spatial-temporal marked point process model, is the most popular stochastic model used to describe earthquake occurrences, to forecast earthquakes and to detect fluid/magma signals or induced seismicity.
- Adjusted ETAS model is applied onto the study of red banana tree to characterize its spatial-temporal spreading patterns.

Data

- **Red Banana Tree:** An invasive plant in Costa Rican rainforest have been observing in recent few years. Heights of 1008 red banana plants were observed and longitude and latitude coordinates were recorded by GPS. The age of plants was estimated according to the growth rate.
- 788 plants have complete location and origin times data, and were used for the subsequent analyses.



Epidemic-Type Aftershock Sequence (ETAS) Model

- The spatial-temporal ETAS model (Ogata, 1998) is a point process model where the rate of an earthquake of magnitude-M occurring at time t and location (x, y) depends on the histroy of occurrences H_t, characterized by the conditional intensity function:

$$\lambda(t,x,y|H_t) = \mu(x,y) + \sum_{\{i:t_i<t\}} \frac{K_0}{(t-t_i+c)^p} \cdot \frac{e^{\alpha(M-M_0)}}{((x-x_i)^2+(y-y_i)^2+d)^q}$$

$\mu(x,y)$: A non-homogeneous background rate.
$\sum$: Triggering function defined in the earthquake context.
K_0, α, c, p, d, q : Parameters to be estimated.

- The ETAS model for red banana plants by Balderama (2012) is

$$\lambda(t,x,y|H_t) = (1-\rho)\mu(x,y) + \frac{\rho\alpha\beta}{\pi} \sum_{\{i:t_i<t\}} e^{-\alpha(t-t_i)-\beta\{(x-x_i)^2+(y-y_i)^2\}}$$

$\frac{\alpha\beta}{\pi}$: A normalizing constant so that $\sum$ integrates to unity.
α : Exponential decay rate due to temporal distance apart.
β : Exponential decay rate due to spatial distance apart.
ρ : Proportion of events attributed to triggering by previous events.

- The log-likelihood function for the occurrence of red banana plants is

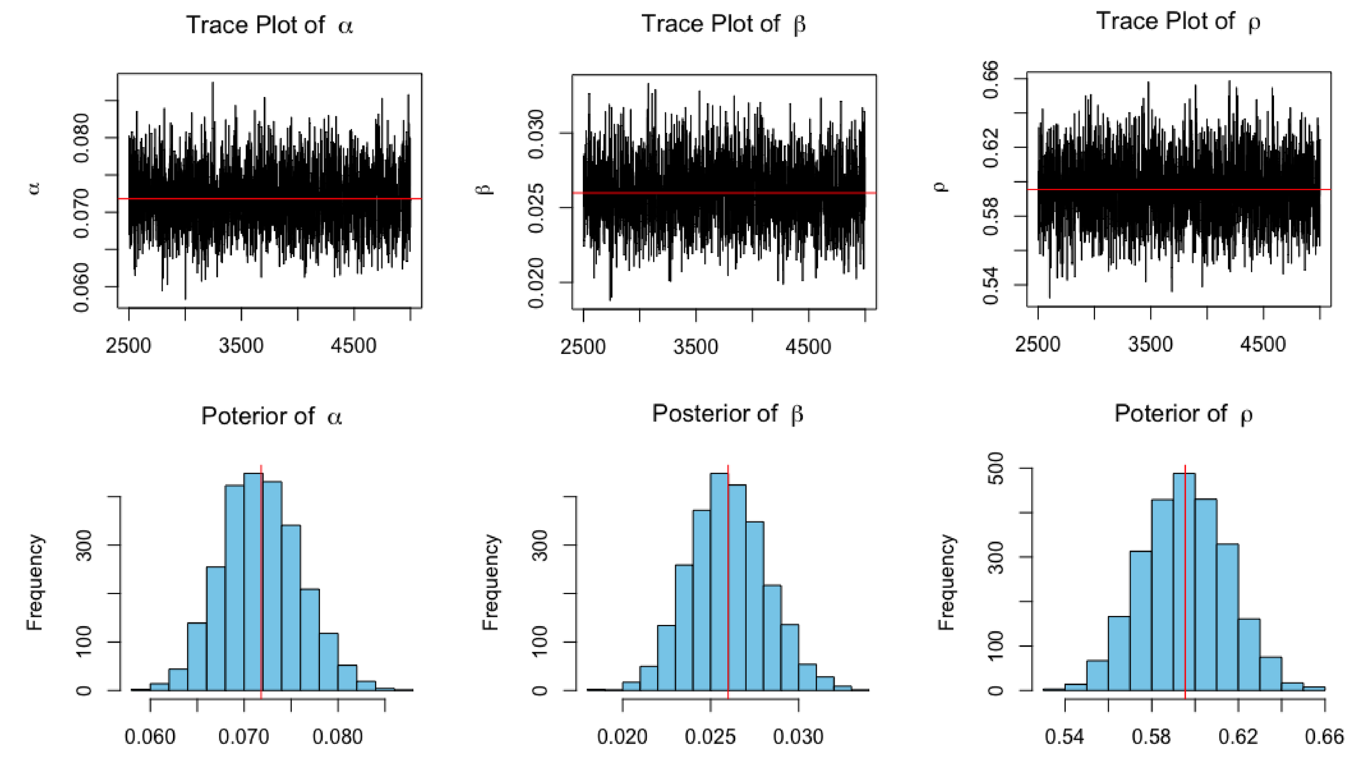
$$logL = \sum_{i=1}^{n} log\lambda(t_i,x_i,y_i) - \iint_A \int_0^\infty \lambda(t,x,y)dtdxdy$$
$$= \sum_{i=1}^{n} log\lambda(t_i,x_i,y_i) - (1-p)T\iint_A \mu(x,y)dxdy - pN$$

Bayesian Estimation

- **Bayesian Analysis:** A statistical paradigm that answers research questions about unknown parameters using probability statements.
- **Prior Distribution:** A probability distribution with respect to current knowledge about parameters, written as $p(\theta)$.
- **Likelihood:** The information regarding parameters contained by new data y, which is proportional to $p(y|\theta)$.
- **Posterior Distribution:** An updated probability distribution, on which all Bayesian inference is based, produced by the combination of likelihood and the prior, written as $p(\theta|y)$.
- The posterior is proportional to the prior times the likelihood:

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{\int p(\theta)p(y|\theta)d\theta}, \text{ where } \theta = \{\alpha, \beta, \rho\}$$

Results of Bayesian Estimation



Parameter	Posterior Mean	Posterior SD
α	0.0718	0.0042
β	0.0260	0.0022
ρ	0.5958	0.0195

- α: The intensity of triggering events decays at a rate of $1 - e^{-0.0718} = 6.93\%$ for every week that passes.
- β: The intensity of triggering events decays at a rate of $1 - e^{-0.026} = 2.57\%$ for every squared distance (in meters) away.
- ρ: 59.58% of events were triggered by previous events. Thus, 40.42% is the background rate, i.e., percentage of immigrant events.

Bayesian Estimate Versus MLE

Maximum Likelihood Estimates obtained by Newton-Raphson algorithm:

Parameter	MLE	Standard Error
α	0.0761	0.0045
β	0.0292	0.0022
ρ	0.5767	0.0193

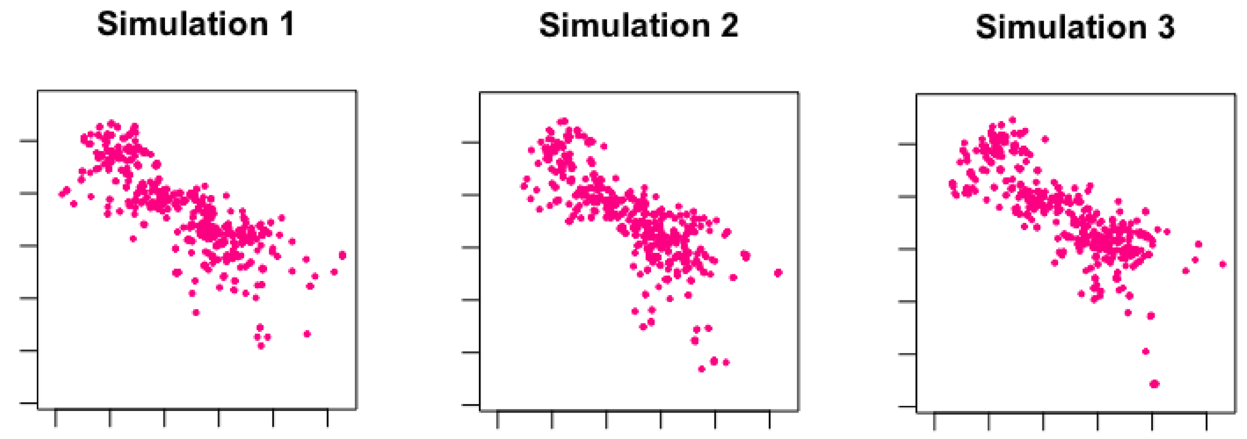
- Estimates result from Bayesian MCMC are very close to MLEs.
- Most of the time, the standard deviations in Bayesian MCMC are smaller than the standard errors of MLEs.

Markov Chain Monte Carlo (MCMC Algorithm)

- **Goal of MCMC:** Draw samples from a posterior distribution without having to know its exact height (probability) at any point.
- **MCMC Algorithm** was created to compute estimates of parameters. Procedures are as below:
  1. Set up prior distributions and proposal distributions:

Parameter	Prior Distribution	Proposal Distribution
α	Gamma(1,10)	Normal($\alpha^{(t-1)}$,0.1)
β	Gamma(1,10)	Normal($\beta^{(t-1)}$,0.1)
ρ	Uniform(0,1)	Beta($\rho^{(t-1)}$,$1-\rho^{(t-1)}$)

  2. Initialize $\theta^{(0)}$ by randomly picking an initial state from each prior.
  3. For t in (1:50000), update $\theta^{(t)}$:
     - Propose a candidate $\theta^{(c)}$ based on the proposal distributions.
     - Compute Metropolis-Hastings ratio $\mathbf{R} = \frac{p(\theta^{(c)}|y)}{p(\theta^{(t-1)}|y)}$.
     - Let $\theta^{(t)} = \theta^{(c)}$ with probability $\mathbf{R}$, otherwise $\theta^{(t)} = \theta^{(t-1)}$.
  4. The 50,000 iterations represent the Monte Carlo samples of posterior distributions.
  5. Compute posterior means to represent estimates of parameters.
- **Thinning:** Discard all but 10^{th} observations (5,000 left) to get rid of autocorrelation to guarantee the independecy of samples.
- **Burn-in:** Throwing away first 2,500 observations to minimize the effect of initial values on the posterior inference.
- Below are simulations of the adjusted ETAS model for red banana plants by using Bayesian estimates.



Future Work

- More environmental covariates should be considered into the model to characterize the spreading pattern of red banana plants, such as distance from rivers, amount of sunlight, elevation and temperature.
- The adjusted ETAS model should be applied on other invasive species or point-process natural events to assess its effectiveness, such as the patterns of volcanic eruption occurrences.

Reference

- Ogata, Y.(1988), "Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes," *Journal of the American Statistical Association*, 83, 9-27.
- Balderama, E.(2012), "Application of Branching Models in the Study of Invasive Species," *Journal of the American Statistical Association*, 107.498: 467-476.

bdeng@luc.edu