

# Introduction to spatial point process and its Fourier analysis

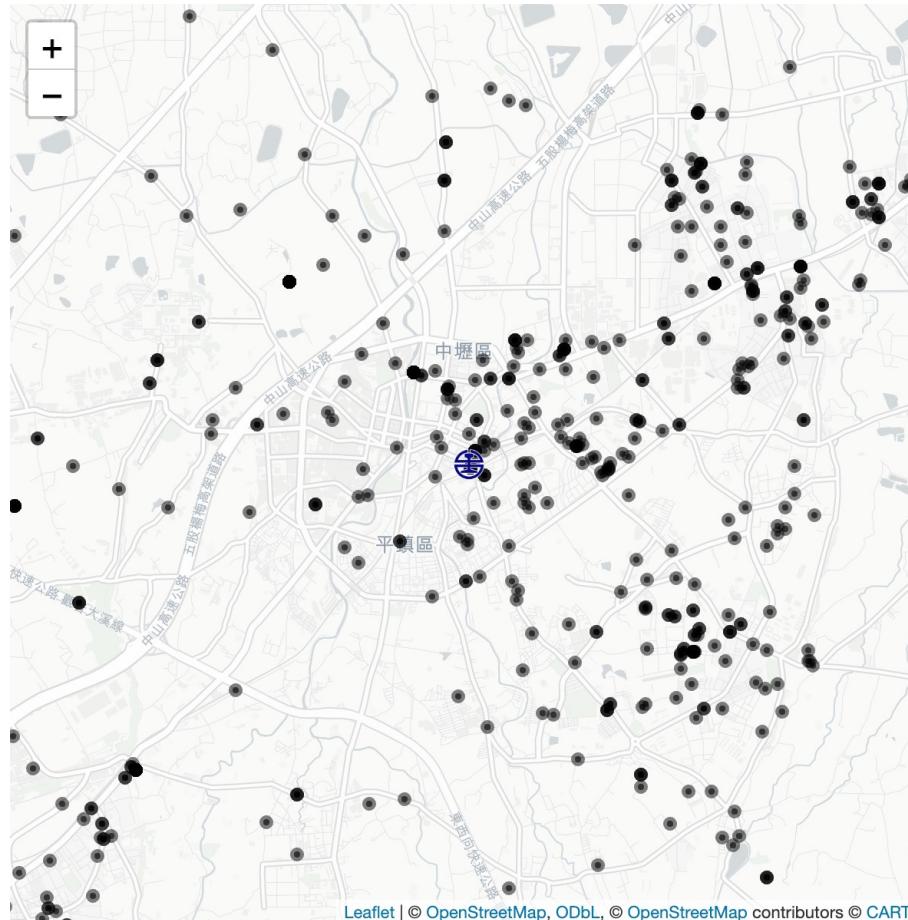
Qi-Wen Ding, Academia Sinica, Taiwan

Joint work with Junho Yang (AS) and Joonho Shin (Sungshin Women's U.)

# Fundamentals

# Point pattern data

Spatial **locations** of things or events



For a study area, **all locations are observable!**

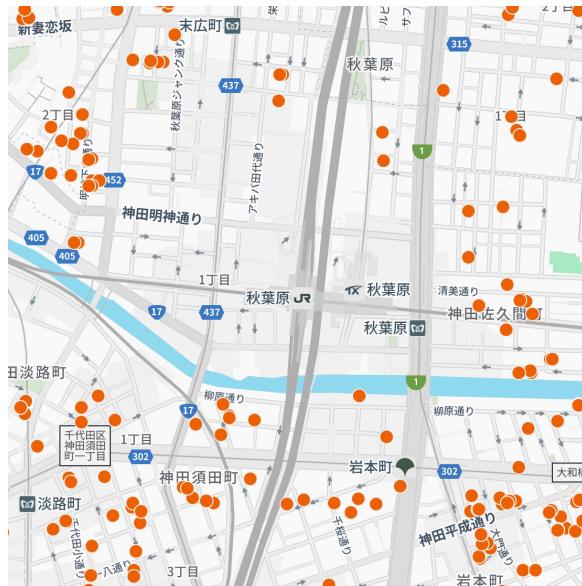
Locations without a point = observed but no event happened!

# Point pattern data

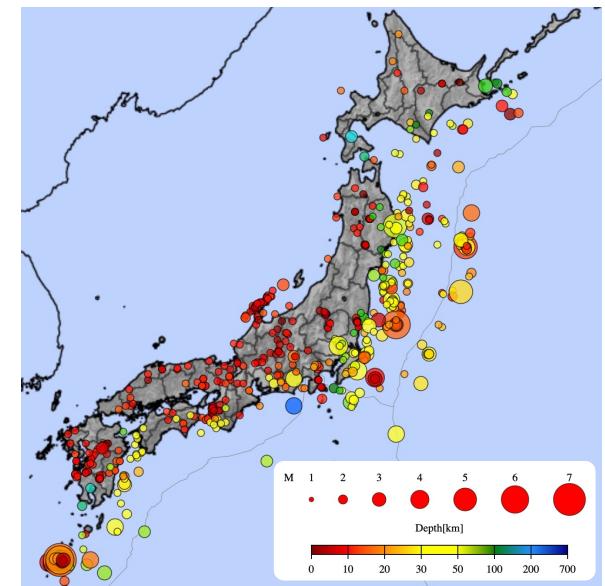
**Spatial locations** of things or events



Traffic injuries near Tokyo station (2021)



Rental housing spots near Akihabara (2025/6/7)

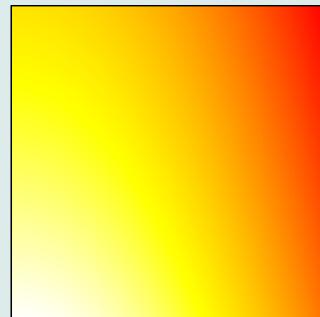
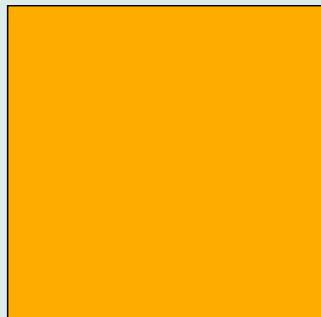
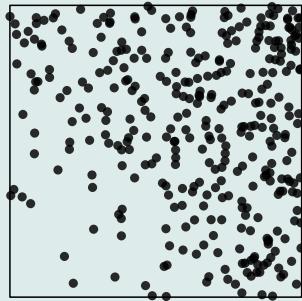
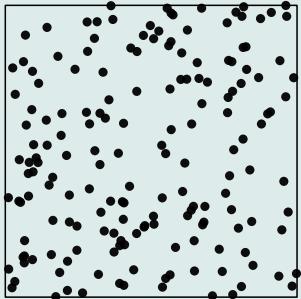


Earthquakes in Japan (2025/7/04-05)

# Characteristics of point process

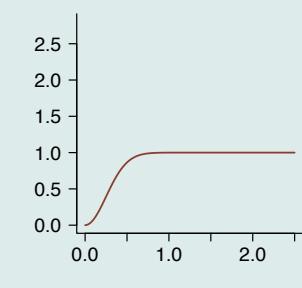
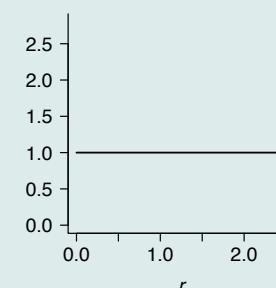
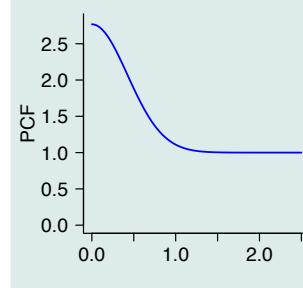
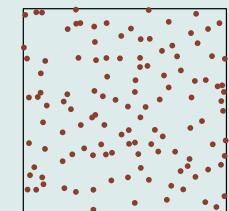
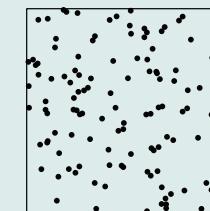
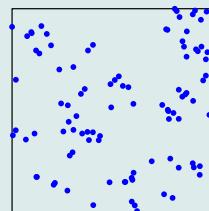
Intensity function:  $\lambda(x)$

$\propto \mathbb{P}(\text{Having a point at location } x)$

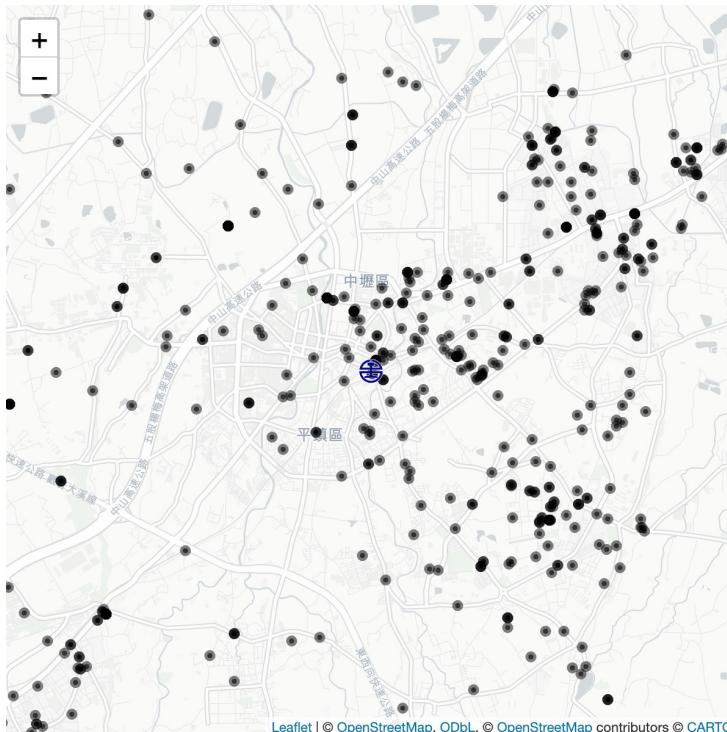


Pair correlation function:  $g(r)$

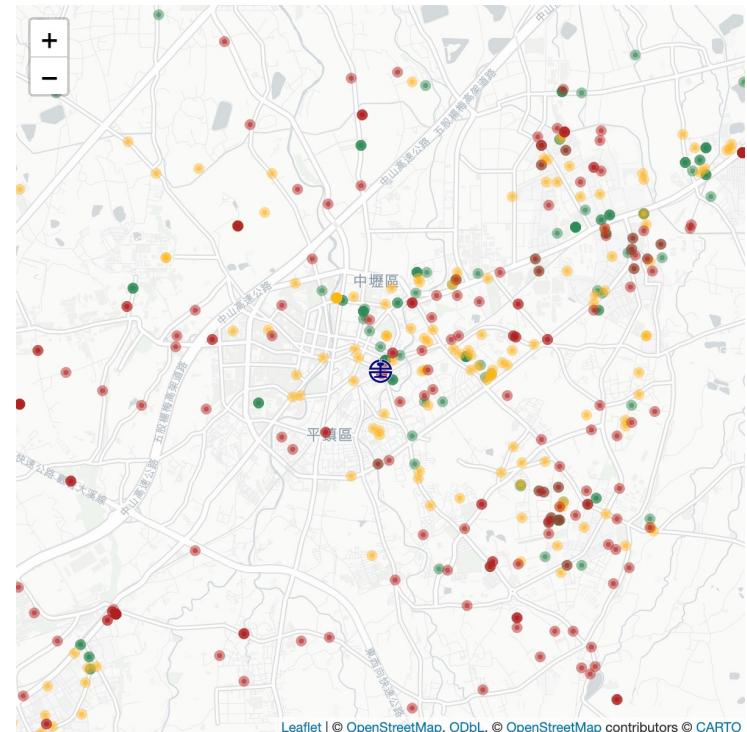
How events interact at radius  $r$



# From univariate to multivariate



● Vehicle burglary incident (2020-2023)



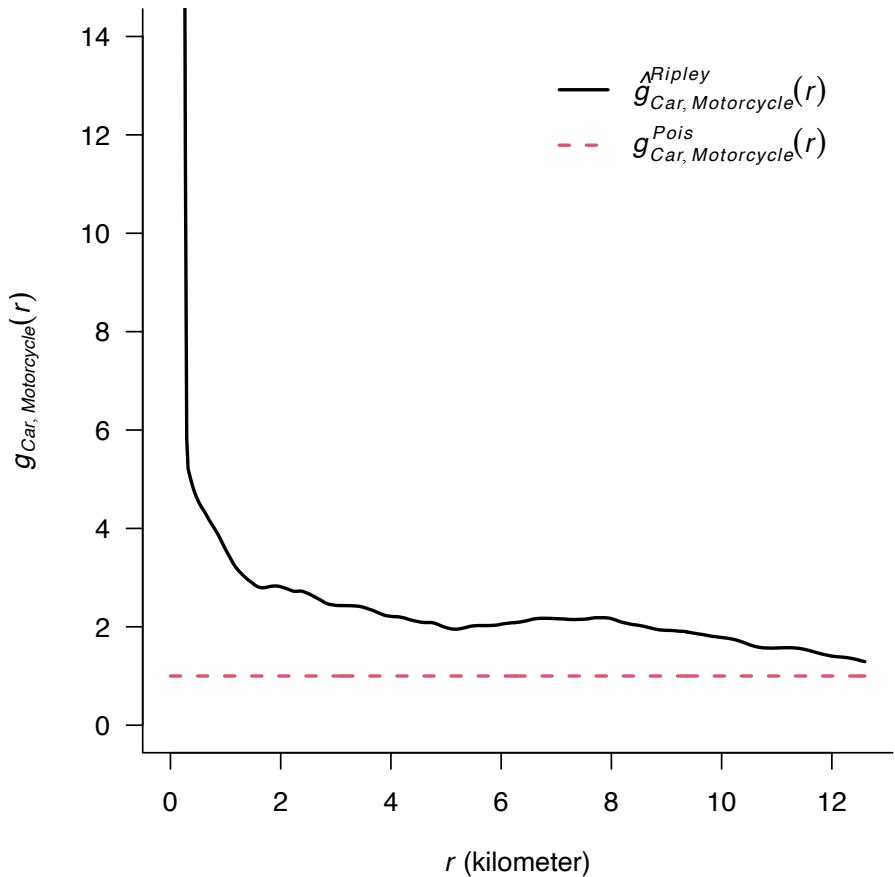
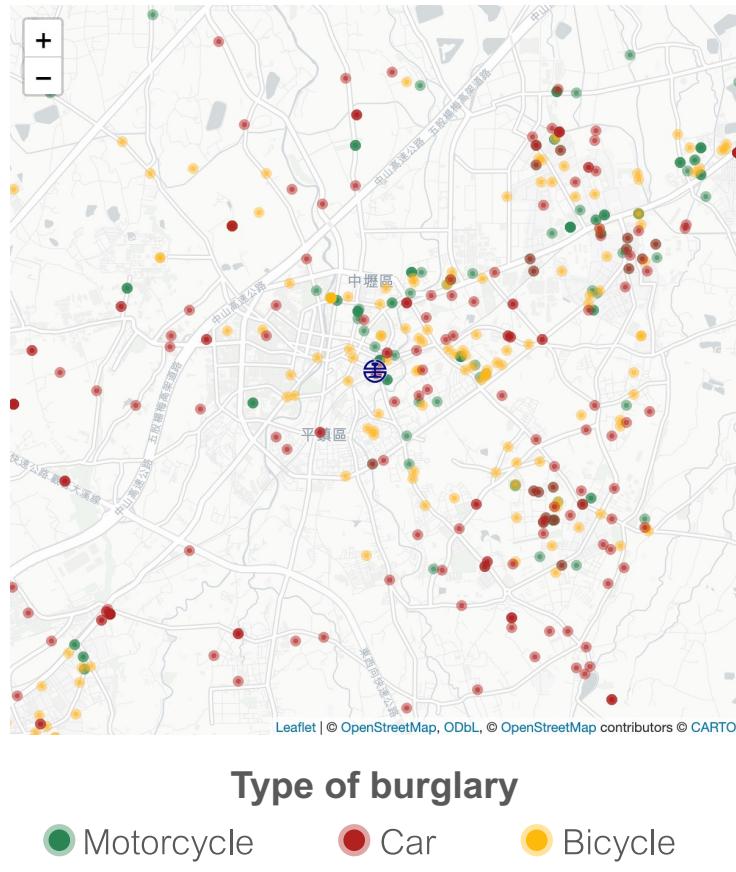
Type of burglary

● Motorcycle

● Car

● Bicycle

# Example 1: Type interaction

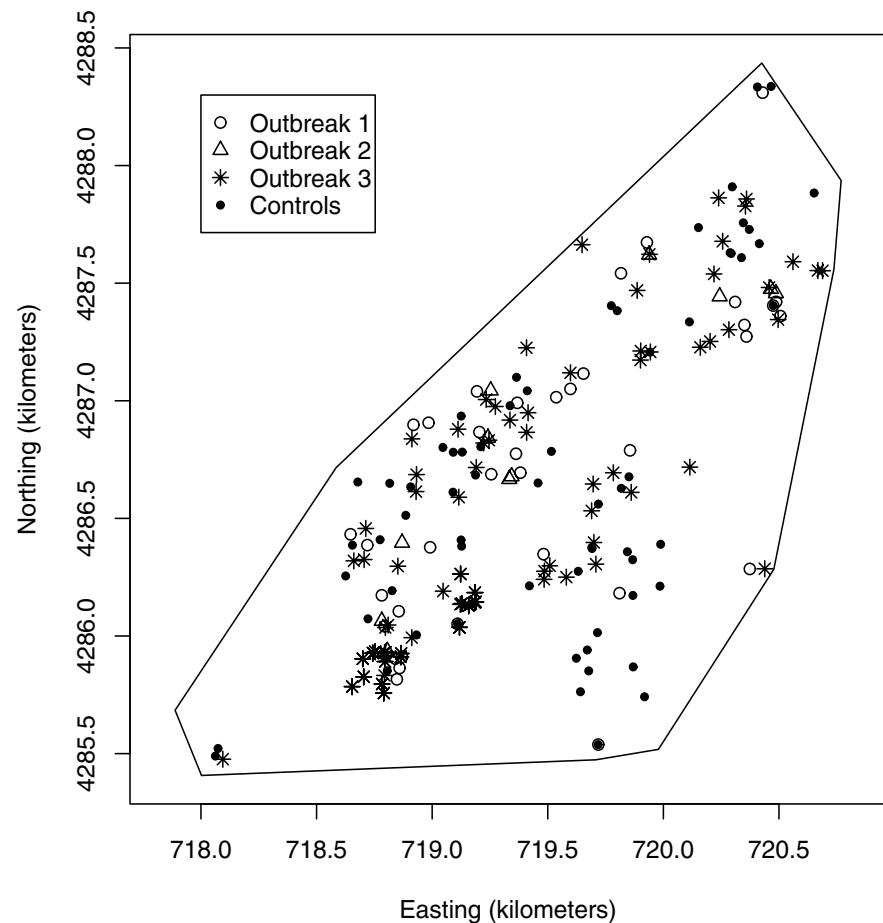


# Example 2: Source detection of infectious disease<sup>9</sup>

## Legionella disease (退伍軍人病)



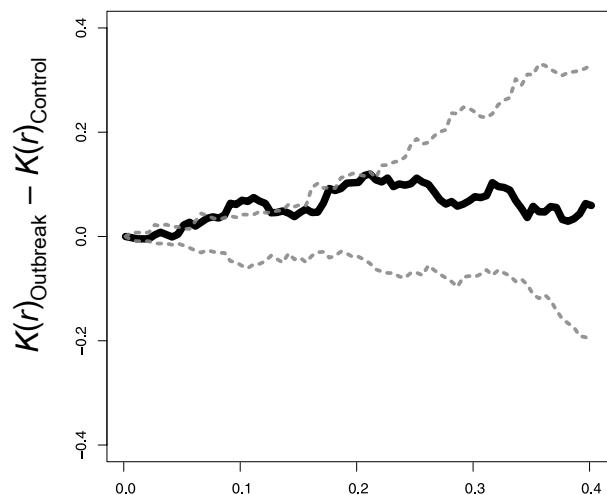
Category	Pneumonia (肺炎)
Transmission	Aerosols (氣溶膠)
Symptoms	High fever, cough, muscle aches, headaches, shortness of breath
Incubation period	2 ~ 10 days



# Example 2: Source detection of infectious disease<sup>10</sup>

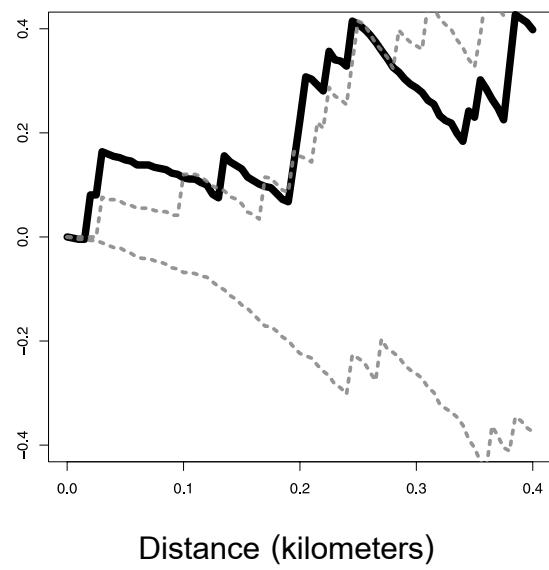
## Outbreak 1 vs. control

1999/9/20 - 2000/2/27  
36 people affected



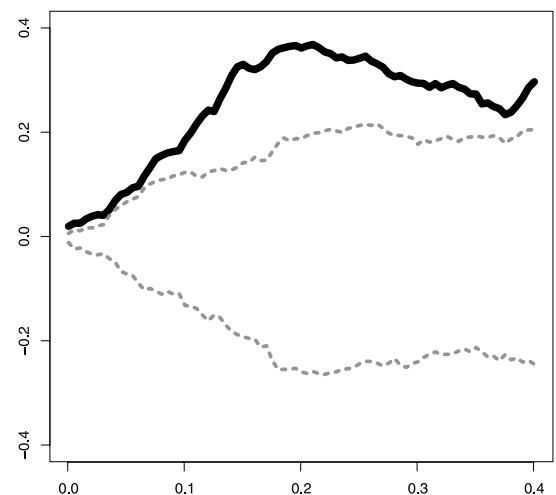
## Outbreak 2 vs. control

2000/4/9 - 2000/7/30  
11 people affected



## Outbreak 3 vs. control

2000/9/9 - 2000/12/1  
97 people affected



# Fourier analysis of point processes

# Time series vs. spatial point processes

## Time series model

- White noise :  $e_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$
- MA(1) :  $X_t = e_t + \theta e_{t-1}$
- AR(1) :  $X_t = \alpha X_{t-1} + e_t$

## ACF

- $\gamma(k) = \sigma^2 \mathbb{1}\{k = 0\}$
- $\gamma(k) = \theta \sigma^2, k = \pm 1$
- $\gamma(k) = \alpha^{|k|} \sigma_X^2$

## Spectral density

- $f(\omega) = \sigma^2 / (2\pi)$
- $f(\omega) = \sigma^2 (1 + \theta^2 + 2\theta \cos \omega) / (2\pi)$
- $f(\omega) = \frac{\sigma^2}{2\pi(1 - 2\alpha \cos \omega + \alpha^2)}$

## Point process model

- Homogeneous Poisson process
- Thomas cluster process
- Determinantal point process

## PCF

- $g(r) = 1$
- $g(r) = 1 + \frac{e^{-r^2/4\sigma^2}}{4\pi\kappa\sigma^2}$
- $g(r) = 1 - \left(e^{-r^2/\sigma^2}\right)^2$

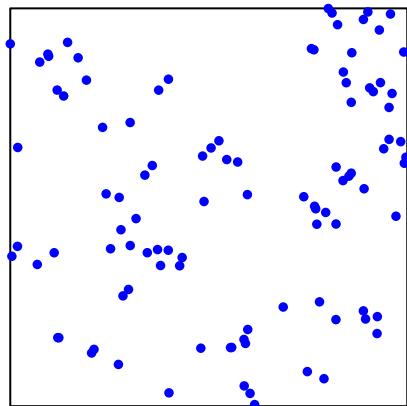
## Spectral density

- $f(\omega) = (2\pi)^{-d} \lambda$
- $f(\omega) = (2\pi)^{-d} \kappa \mu \left[ 1 + \mu e^{-\sigma^2 \|\omega\|^2} \right]$
- $f(\omega) = (2\pi)^{-d} \left[ \lambda - \lambda^2 \left( \frac{\pi\sigma^2}{2} \right)^{\frac{d}{2}} e^{-\frac{\sigma^2 \|\omega\|^2}{8}} \right]$

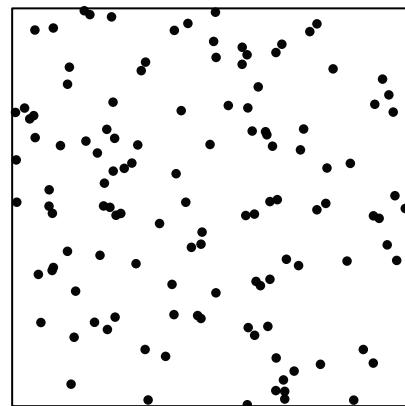
For stationary and isotropic case,  $g(\mathbf{x}_i, \mathbf{x}_j) = g(\|\mathbf{x}_i - \mathbf{x}_j\|) = g(r)$ .  
Parameter:  $\boldsymbol{\theta}_{\text{Thomas}} = (\kappa, \mu, \sigma^2)^T$ ,  $\boldsymbol{\theta}_{\text{Determinantal}} = (\lambda, \sigma^2)^T$ .

# PCF and spectral density

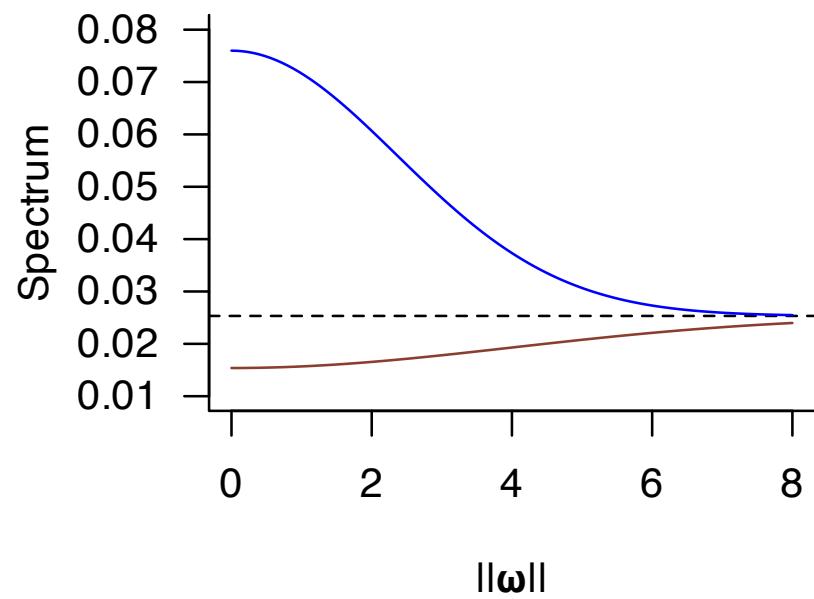
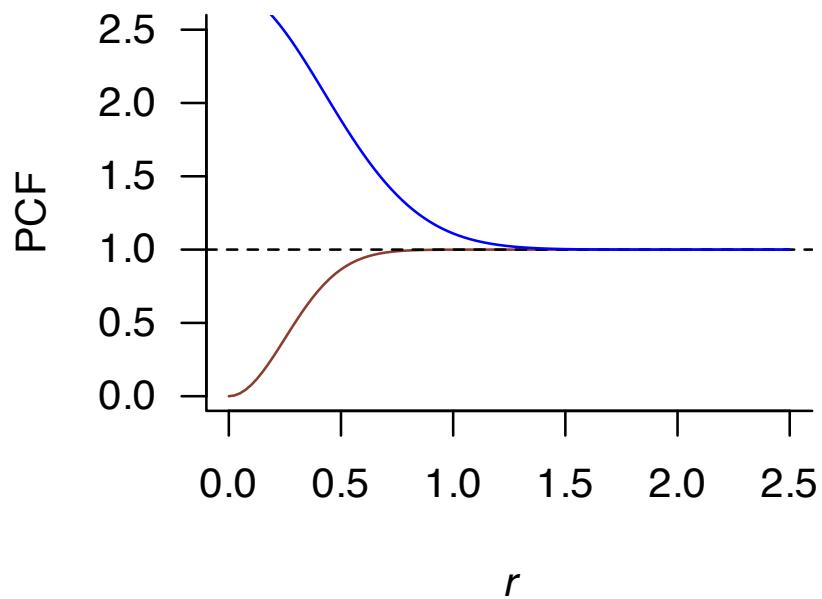
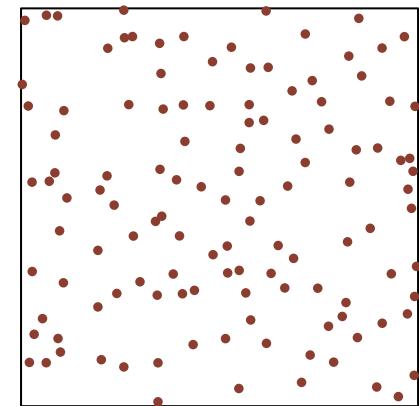
Thomas process  
 $(\kappa = 0.5, \sigma = 0.3, \mu = 2)$



Poisson process  
 $(\lambda = 1)$



DPP with Gaussian kernel  
 $(\lambda = 1, \sigma = 0.5)$



# Frequency domain parameter estimation

## Whittle-type likelihood

Let  $\{X_\theta\}$  be a family of 2<sup>nd</sup>-order stationary point processes with parameter  $\theta \in \Theta$ . The associated spectral density is denoted as  $f_\theta$ . Then, we fit the model using the pseudo-likelihood

$$L(\theta) = \sum_{\omega_k \in D} \left\{ \frac{\hat{I}(\omega_k)}{f_\theta(\omega_k)} + \log f_\theta(\omega_k) \right\}.$$

## Intuition

- $\hat{I}(\omega)$  can be regarded as the “truth” since  $\mathbb{E}[\hat{I}_n(\omega)] \rightarrow f(\omega)$ .
- $f_\theta(\omega)$  are the parameter family of spectral densities.
- $L(\theta)$  is the **spectral divergence between the truth and our guess. The smaller the better!**

## Proposed model parameter estimator

$$\hat{\theta} = \arg \min_{\theta \in \Theta} L(\theta)$$

# Frequency domain parameter estimation

<b>Model</b>	<b>Window</b>	<b>Parameter</b>	<b>Method</b>	
			<b>Ours</b>	<b>ML</b>
TCP	[−5, 5] <sup>2</sup> (≈ 200 pts)	κ	−0.04 (0.11)	−0.02 (0.11)
		μ	0.72 (3.52)	−0.24 (5.93)
		σ <sup>2</sup>	0.02 (0.07)	−0.04 (0.22)
	[−10, 10] <sup>2</sup> (≈ 800 pts)	<b>Time (sec)</b>	<b>0.74</b>	<b>0.38</b>
		κ	−0.02 (0.05)	−0.02 (0.05)
		μ	0.60 (1.77)	0.33 (2.79)
	[−20, 20] <sup>2</sup> (≈ 3200 pts)	σ <sup>2</sup>	0.01 (0.02)	−0.01 (0.11)
		<b>Time (sec)</b>	<b>2.38</b>	<b>5.67</b>
		κ	−0.01 (0.04)	0.00 (0.03)
	[−20, 20] <sup>2</sup> (≈ 3200 pts)	μ	0.25 (1.03)	0.15 (1.23)
		σ <sup>2</sup>	0.01 (0.02)	0.00 (0.03)
		<b>Time (sec)</b>	<b>9.15</b>	<b>173.66</b>

The bias and the standard errors (in parentheses) of the estimated parameters based on three different approaches for the Thomas clustered process (TCP)

# Spectral analysis

## Time series model

- White noise :  $e_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$
- MA(1) :  $X_t = e_t + \theta e_{t-1}$
- AR(1) :  $X_t = \alpha X_{t-1} + e_t$

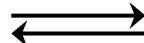
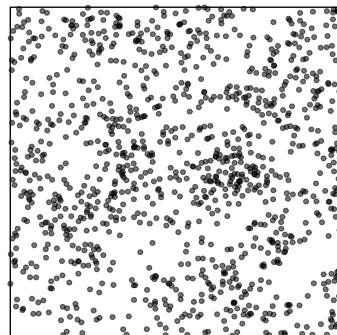
## ACF

- $\gamma(k) = \sigma^2 \mathbb{1}\{k = 0\}$
- $\gamma(k) = \theta \sigma^2, k = \pm 1$
- $\gamma(k) = \alpha^{|k|} \sigma_X^2$

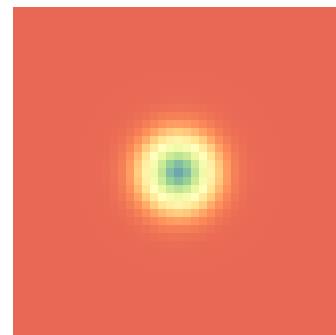
## Spectrum

- $f(\omega) = \sigma^2 / (2\pi)$
- $f(\omega) = \sigma^2 (1 + \theta^2 + 2\theta \cos \omega) / (2\pi)$
- $f(\omega) = \frac{\sigma^2}{2\pi(1 - 2\alpha \cos \omega + \alpha^2)}$

2<sup>nd</sup>-order stationary  
point process

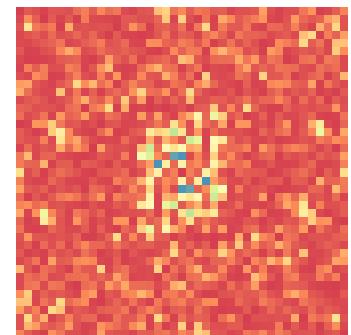


Spectrum  
 $f(\omega)$



Its estimator is

Periodogram  
 $\hat{I}(\omega)$



# Spectral analysis

## Time series model

- White noise :  $e_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$
- MA(1) :  $X_t = e_t + \theta e_{t-1}$
- AR(1) :  $X_t = \alpha X_{t-1} + e_t$

## ACF

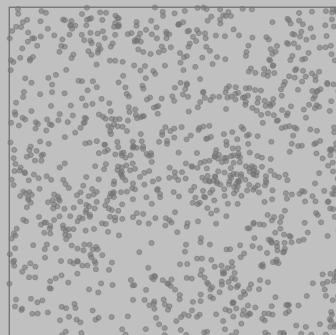
- $\gamma(k) = \sigma^2 \mathbb{1}\{k = 0\}$
- $\gamma(k) = \theta \sigma^2, k = \pm 1$
- $\gamma(k) = \alpha^{|k|} \sigma_X^2$

## Spectrum

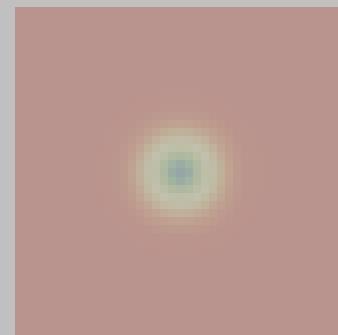
- $f(\omega) = \sigma^2 / (2\pi)$
- $f(\omega) = \sigma^2 (1 + \theta^2 + 2\theta \cos \omega) / (2\pi)$
- $f(\omega) = \frac{\sigma^2}{2\pi(1 - 2\alpha \cos \omega + \alpha^2)}$

**Does NOT work for inhomogeneous point processes!**

2<sup>nd</sup>-order stationary  
point process

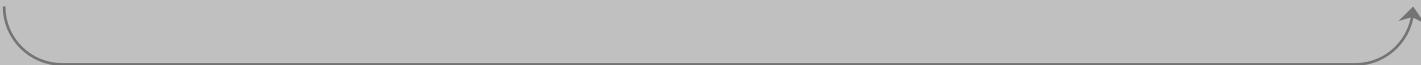
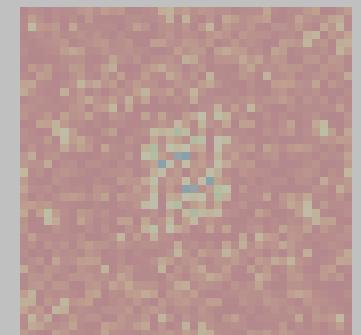


Spectrum  
 $f(\omega)$



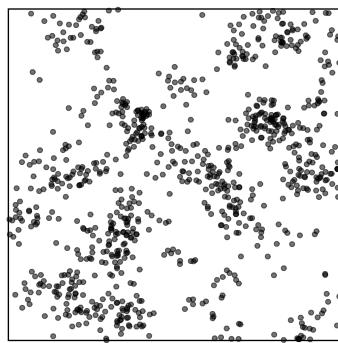
Periodogram  
 $\hat{I}(\omega)$

Its estimator is

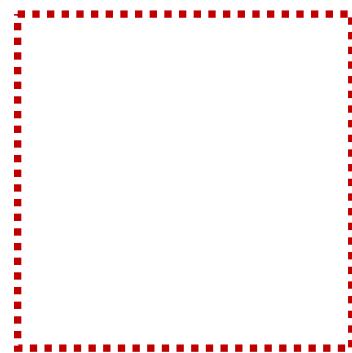


# Spectral analysis under inhomogeneity

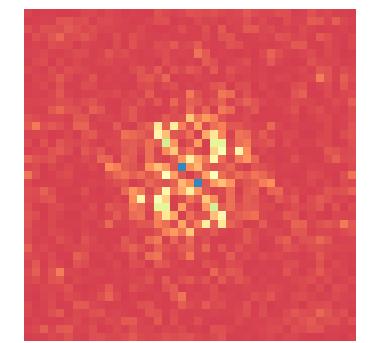
Inhomogeneous  
point process



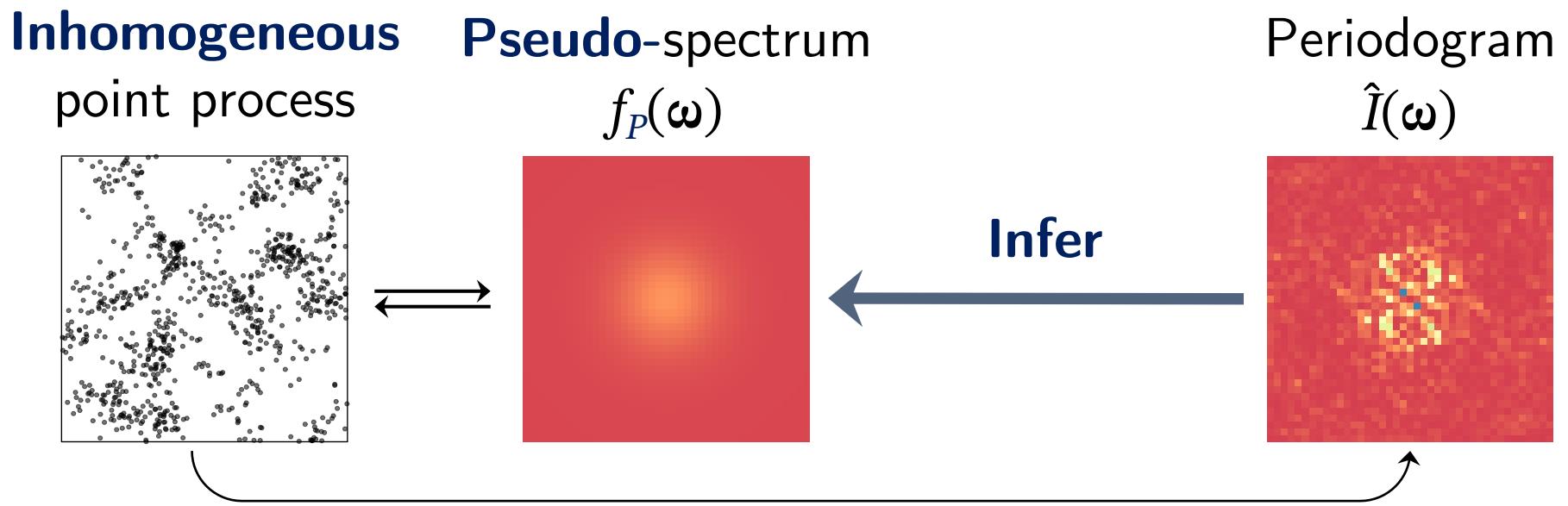
???



Periodogram  
 $\hat{I}(\omega)$



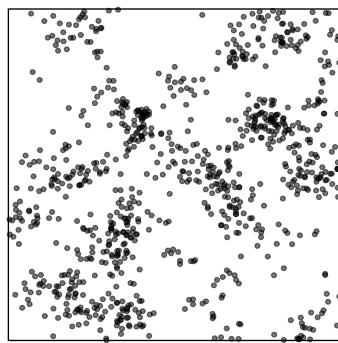
# Spectral analysis under inhomogeneity



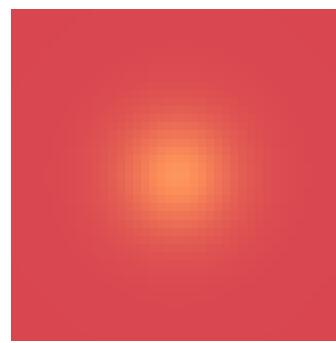
$$f_P(\omega) := \lim_{n \rightarrow \infty} \mathbb{E}[\hat{I}_n(\omega)], \quad \forall \omega \in \mathbb{R}^d.$$

# Estimating the pseudo-spectrum

Inhomogeneous  
point process

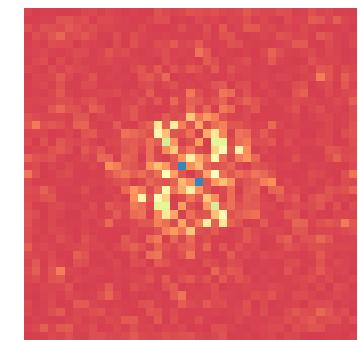


Pseudo-spectrum  
 $f_{\text{P}}(\omega)$



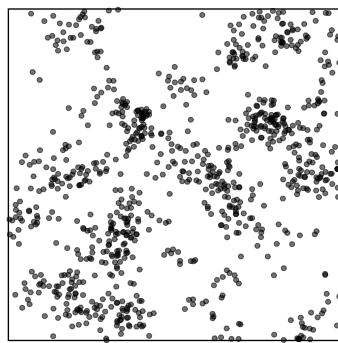
Estimators: ----- ↘

- ① Periodogram  
 $\hat{I}(\omega)$

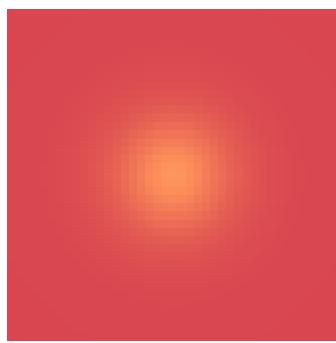


# Estimating the pseudo-spectrum

Inhomogeneous  
point process

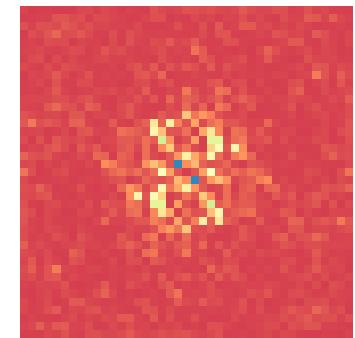


Pseudo-spectrum  
 $f_P(\omega)$

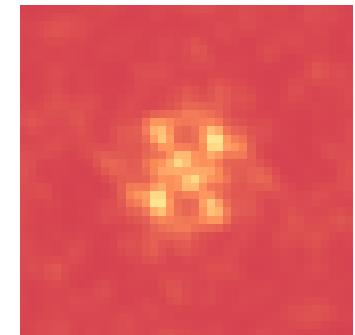


Estimators: ----- ↘

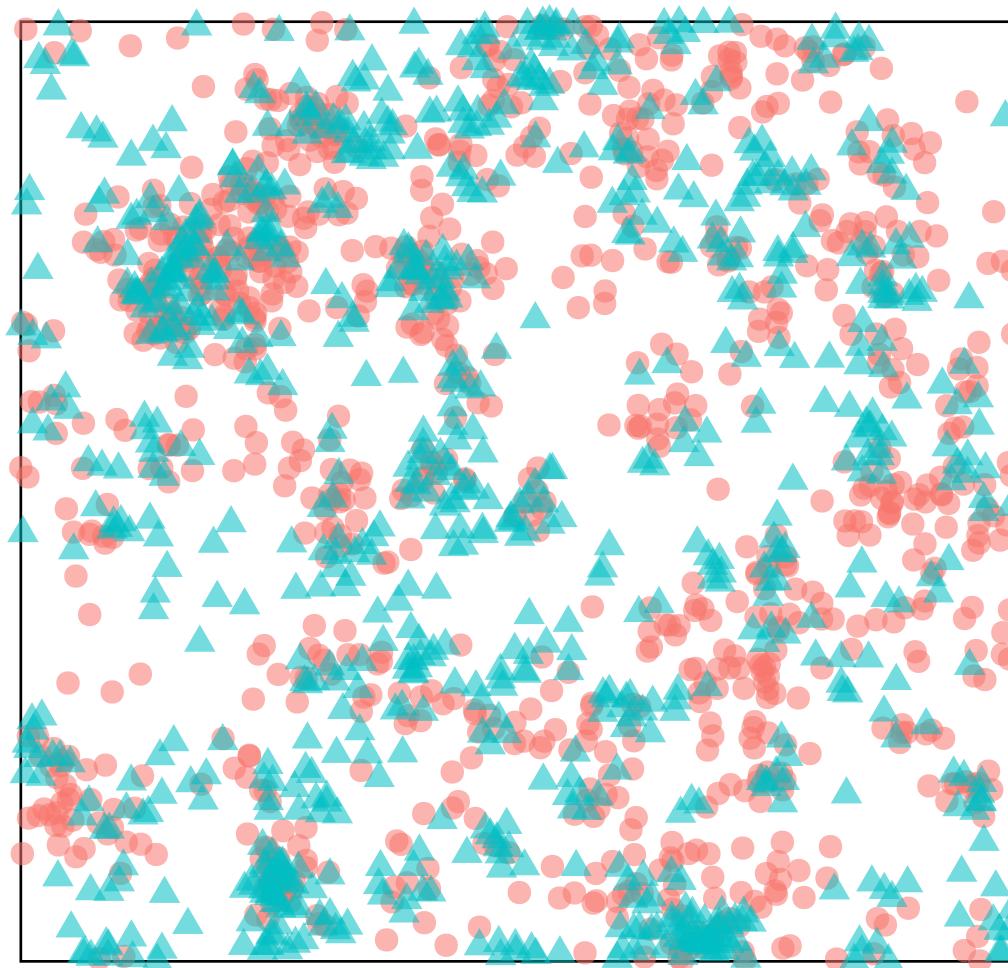
- ① Periodogram  
 $\hat{I}(\omega)$



- ② Kernel estimator  
 $\hat{f}_P(\omega)$



# Simulation: Data generating process

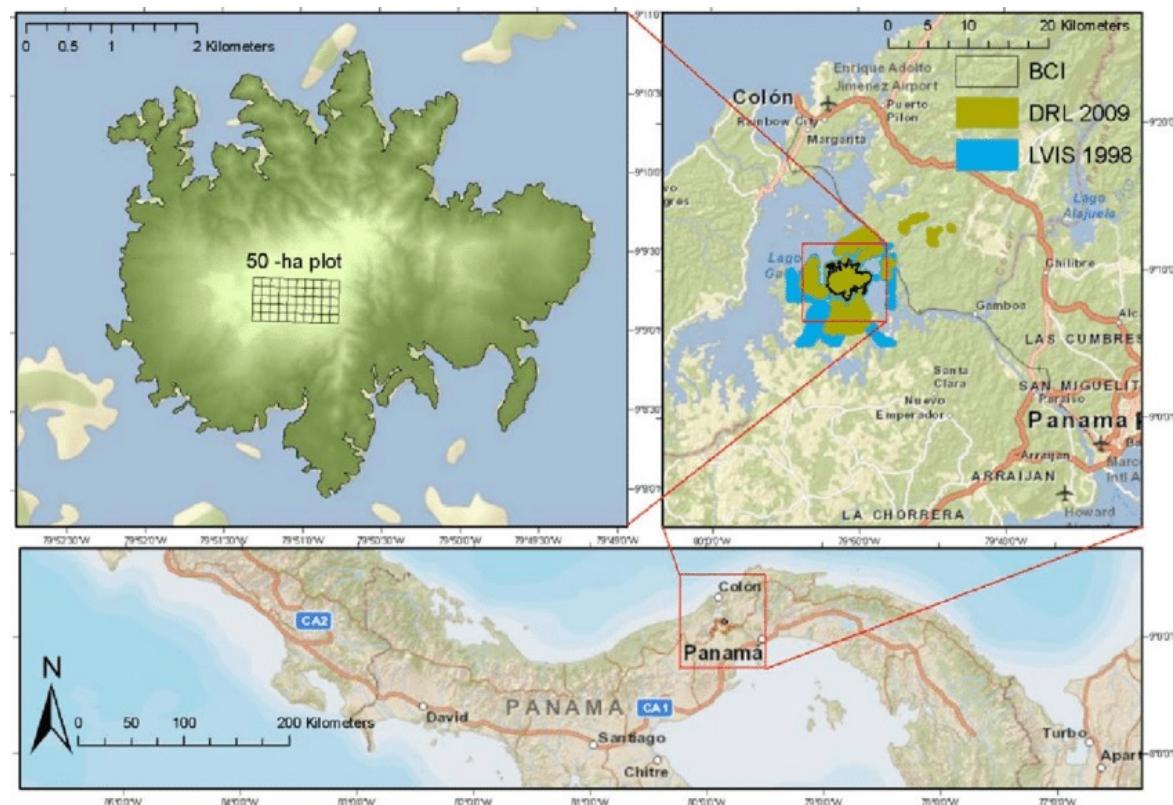


Bivariate point process with marginal and joint clustering interactions

# Simulation: Result

Pseudo-spectrum	Window	$\hat{I}(\omega)$		$\hat{f}_P(\omega)$	
		Bias	MSE	Bias	MSE
Marginal	$[-5, 5]^2$	0.03	1.49	0.02	0.49
	$[-10, 10]^2$	0.01	1.09	0.00	0.17
	$[-20, 20]^2$	0.00	1.02	0.00	0.13
Cross	$[-5, 5]^2$	0.06	28.14	0.01	3.76
	$[-10, 10]^2$	0.05	22.61	0.01	2.45
	$[-20, 20]^2$	0.05	20.18	0.01	2.11

# Barro Colorado Island (BCI) data



**207,718** alive trees

**310** species

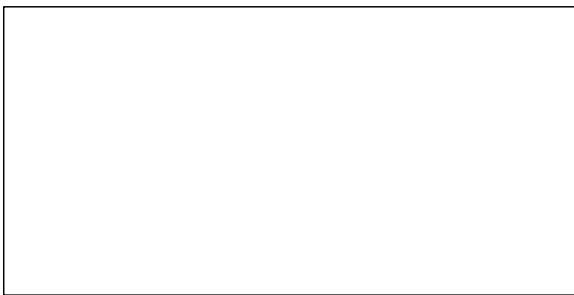
**7** censuses



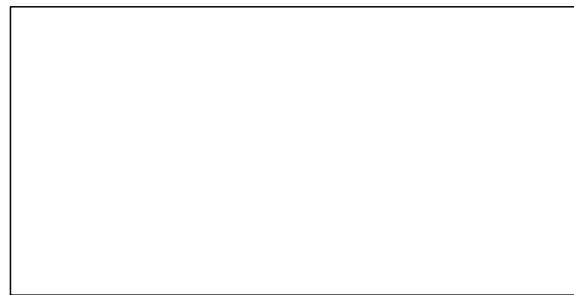
BCI Plot Census  
November 26, 2015 by Forest Global Earth Observatory (ForestGEO).

# Barro Colorado Island (BCI) data

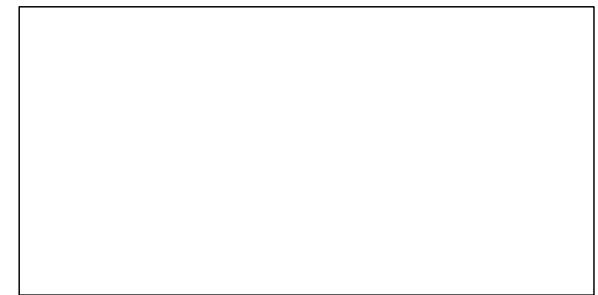
cappfr



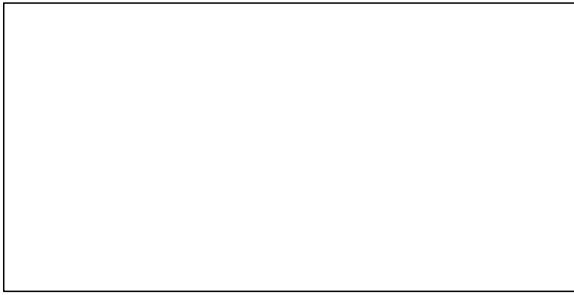
hirttr



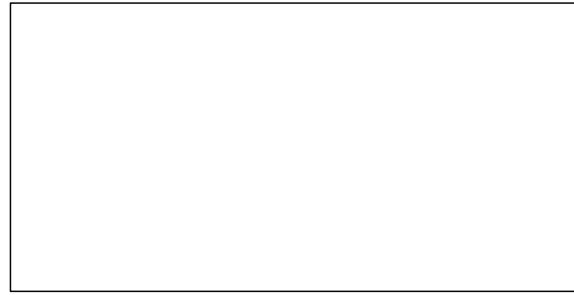
protpa



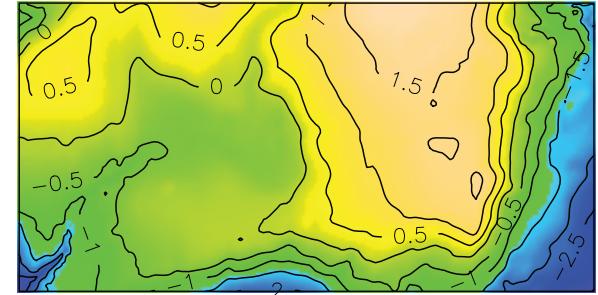
protte



tet2pa



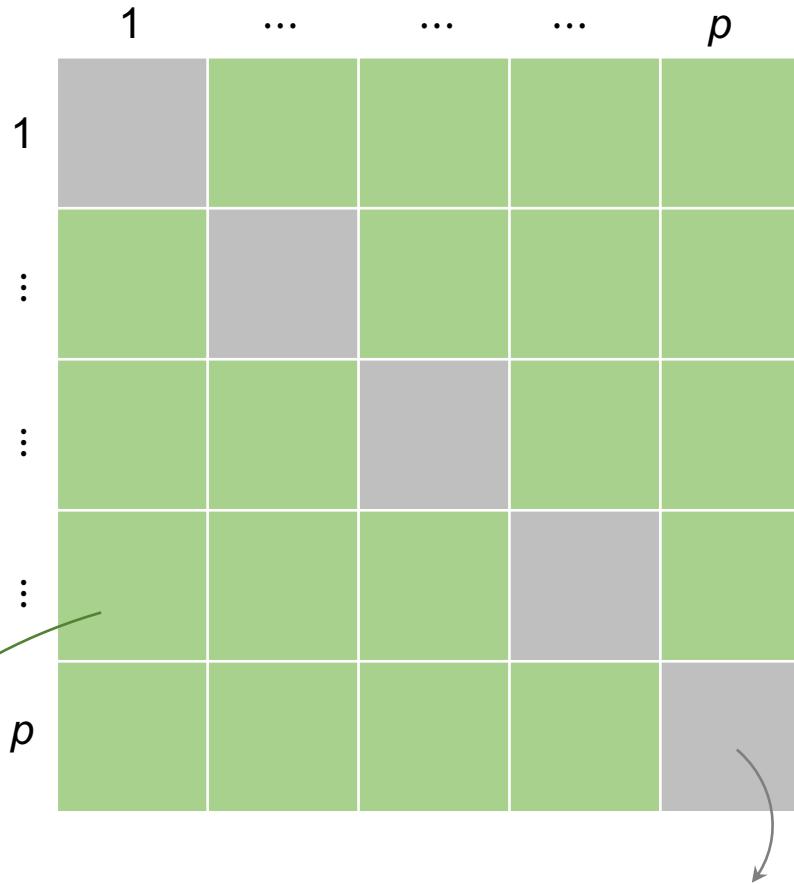
elevation



Point patterns of five species in the BCI dataset and image of elevation in the study region.

# Multivariate point pattern

Consider a  $p$ -variate process:  $i, j \in \mathcal{V} = \{1, 2, \dots, p\}$



Spectral coherence:

$$R_{ij}(\boldsymbol{\omega}) = \frac{f_{ij}(\boldsymbol{\omega})}{[f_{ii}(\boldsymbol{\omega})f_{jj}(\boldsymbol{\omega})]^{\frac{1}{2}}}$$

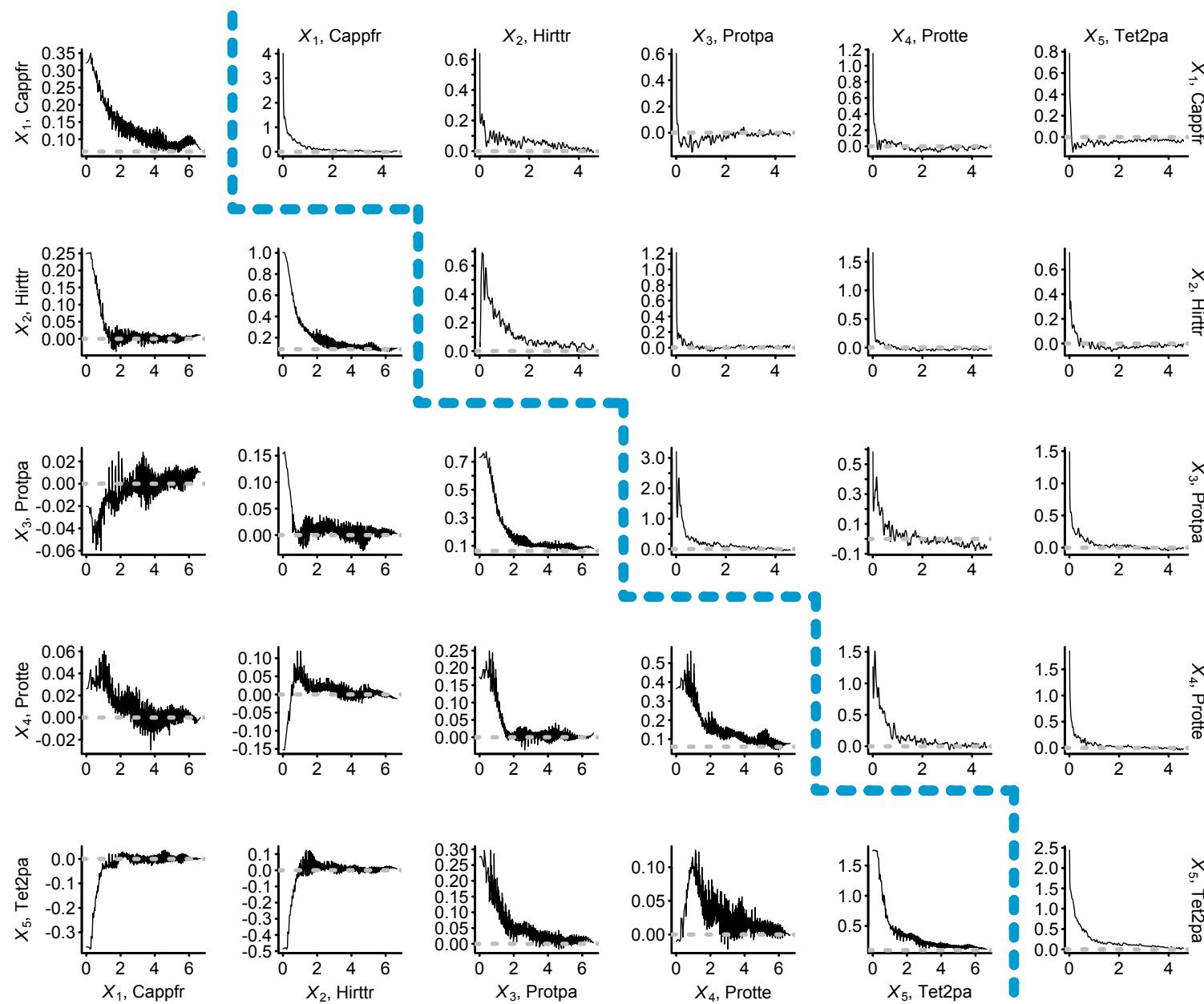
Cross-spectrum:

$$f_{ij}(\boldsymbol{\omega}) = (2\pi)^{-d} \int_{\mathbb{R}^d} C_{ij}(\mathbf{x}_1 - \mathbf{x}_2) e^{-i\mathbf{x}^\top \boldsymbol{\omega}} d\mathbf{x}$$

Marginal spectrum:

$$f_{ii}(\boldsymbol{\omega}) = (2\pi)^{-d} \int_{\mathbb{R}^d} C_{ii}(\mathbf{x}_1 - \mathbf{x}_2) e^{-i\mathbf{x}^\top \boldsymbol{\omega}} d\mathbf{x}$$

# Pseudo-spectra



# Coherence analysis

