



Trabalho Final

Análise espacial da ocorrência de ligações clandestinas em rede de distribuição de água

Diogo Reis

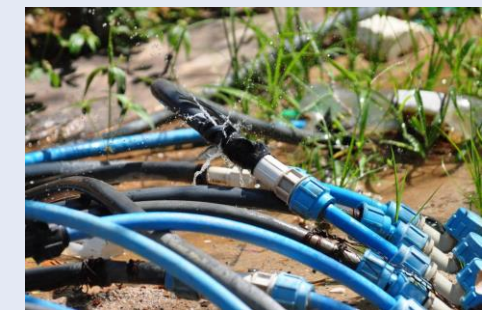
SER 301 – Análise Espacial de Dados Geográficos

Dezembro/2017

Contexto



- Sistemas de Distribuição de Água
 - Tubulações, Registros, Ramais, Tanques
- Elevados índices de perdas no país
- Causas
 - Falta de manutenção apropriada
 - Deterioração dos ativos
 - Ineficiência na fiscalização
 - Monitoramento inadequado
- Impactos ambientais e econômicos



Classificação de Perdas de Água em Sistemas de Distribuição

Volume de Entrada no Sistema	Consumo Autorizado	Consumo Autorizado Faturado	Consumo Medido Faturado	Receita		
			Consumo Não Medido Faturado			
	Perdas de Água	Consumo Autorizado Não Faturado		Consumo Medido Não Faturado	Perda de Receita	
				Consumo Não Medido Não Faturado		
		Perdas Reais	Perdas Aparentes	Consumo Não Autorizado		
				Imprecisões no Medidor		
Vazamentos na Adução e Distribuição						
		Vazamentos e Transbordamentos em Tanques de Armazenamento				
		Vazamentos em Ramais de Ligação e Conexões				

(Fonte: Lambert, 2000)

Contexto



- Consumo não autorizado: Ligações clandestinas
 - Furto de água do sistema de distribuição
 - Consumo ou comercialização
 - Perda comercial para a companhia
- Consequências
 - Impactos na eficiência operacional dos sistemas
 - Desperdício do recurso natural
 - Prejuízo ao abastecimento da população
 - Perda de receita



Objetivo

- A. Existe evidência de agrupamento espacial na ocorrência de ligações clandestinas?
- B. Existe dependência espacial ao analisar os padrões de ocorrência de furto de água?
- C. Como é possível melhor modelar a relação entre a ocorrência de ligações clandestinas e o perfil socioeconômico?



1. Área de Estudo



Recife (PE)
Região Centro-Sul
Fonte: Compesa



2. Eventos (V.D.)



Locais de ocorrência de ligações clandestinas
Pontos (Coordenadas Geográficas)
Agregação Espacial




3. Análise: Agrupamento Espacial



- Caracterização da distribuição de pontos
- **Efeitos de 1ª ordem**
 - Estimador de Kernel
- **Efeitos de 2ª ordem**
 - Função do Vizinho mais Próximo
 - Função K




4. Análise: Dependência Espacial




Índice de Moran (I)
Getis e Ord (G)




5. Regressão



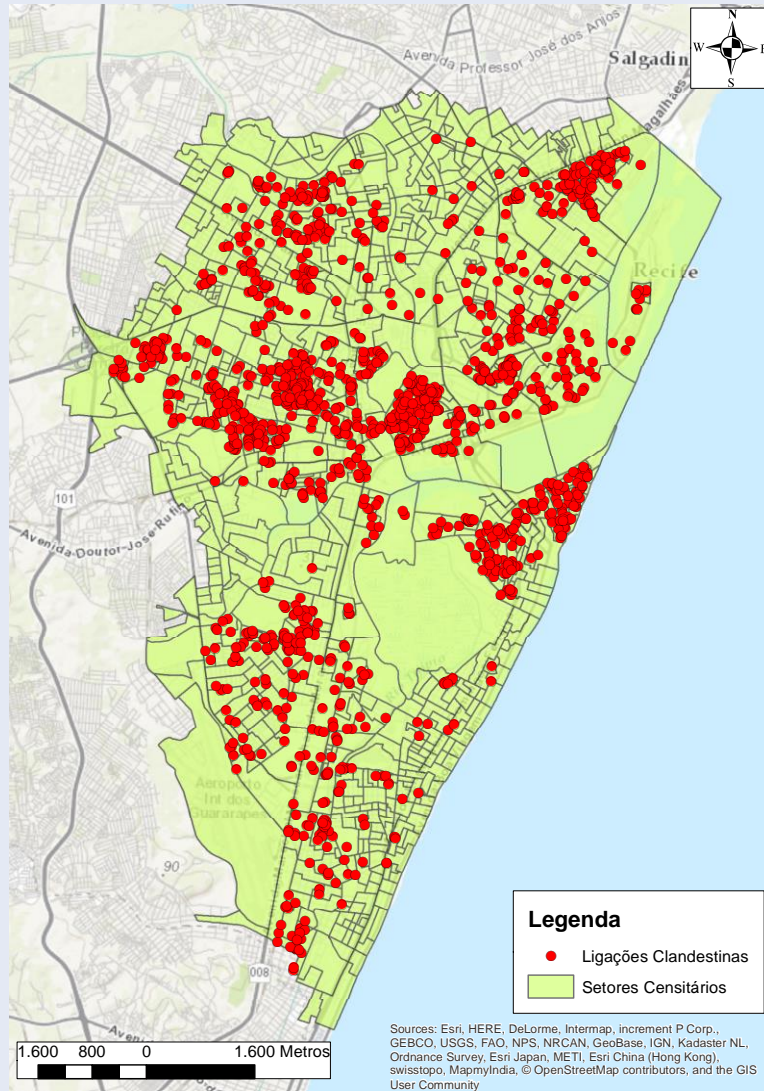
Renda
Setores Censitários (IBGE)



Regressão Linear Múltipla:
Método dos Mínimos Quadrados



Regressão Espacial:
Spatial Lag Model (LAG)
Spatial Error Model (CAR)



Área de Estudo

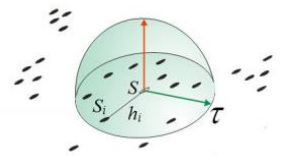
- Recife (PE)
 - Região Centro-Sul
- Compesa
 - Distribuição de água na RMR
- Ligações Clandestinas
 - 1414 registros de ocorrências
 - Período: 2014 a 2016
 - Coordenadas Geográficas
- Setores Censitários
 - IBGE

Agrupamento Espacial

Efeitos de 1ª Ordem

Estimador de Kernel

$$\hat{\lambda}_{\tau}(s) = \sum_{i=1}^n \frac{3}{\pi\tau^2} \left(1 - \frac{h_i^2}{\tau^2}\right)^2$$



I : função escolhida para construir uma superfície contínua sobre os eventos;
 τ : largura de faixa (ou largura de banda);
 S : localização qualquer na área de estudo a ser estimada;
 S_i : localizações dos eventos observados;
 n : representa o número de eventos dentro da largura de faixa (τ).

Efeitos de 2ª Ordem

Função K

$$\hat{K}(h) = \frac{R}{n^2} \sum_{i=1}^n \sum_{j=1}^n I_h(d_{ij})$$

n : número de eventos;
 I_h : matriz de pesos;
 d : distância entre os eventos.

Método do vizinho mais próximo

$$\hat{G}(w) = \frac{\#(w_i \leq w)}{n}$$

w : distância de entrada;
 w_i : distância entre eventos;
 n : número de eventos.

Dependência Espacial

Índice de Moran (I)

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

n: número de áreas;
 y_i : valor do atributo na área i ;
 \bar{y} : valor médio do atributo na região de estudo;
 w_{ij} : pesos atribuídos conforme a conexão entre as áreas.

Getis e Ord (G)

$$G = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{i,j} y_i y_j}{\sum_{i=1}^n \sum_{j=1}^n y_i y_j}, \forall j \neq i$$

y_i e y_j : valores de atributo associados às i -ésima e j -ésima áreas;
 w_{ij} : peso espacial entre as áreas i e j ;
n: número de áreas;
 $\forall j \neq i$: áreas não podem ser as mesmas.

Regressão Espacial

Spatial Lag Model (SAR)

$$Y = \lambda W_y + X\beta + \varepsilon$$

λ : coeficiente espacial autoregressivo - medida de correlação espacial;
 W_y : matriz de proximidade espacial em y ;
 β : coeficiente de regressão;
 ε : erro aleatório

Spatial Error Model (CAR)

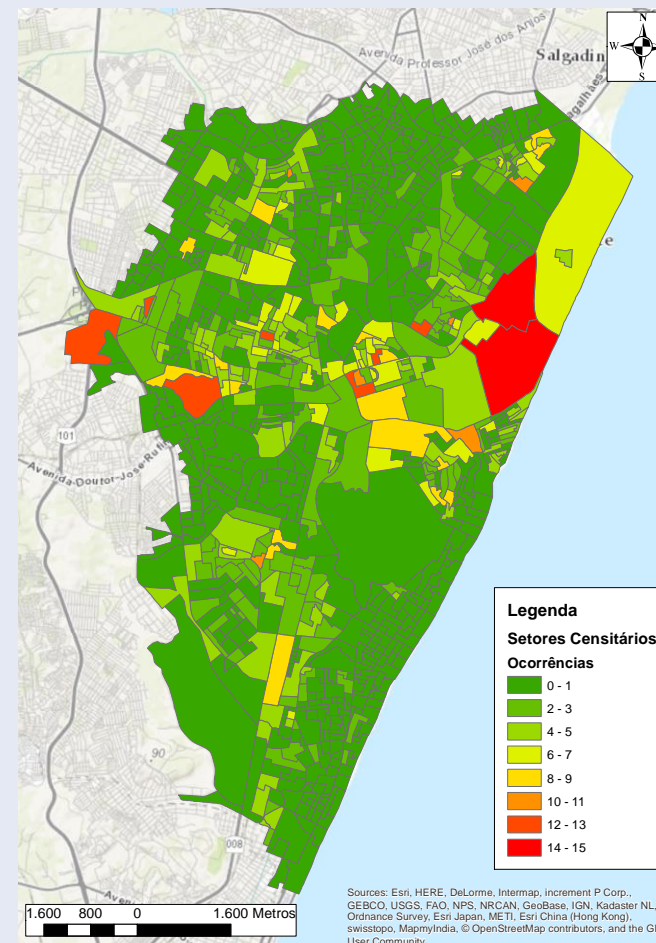
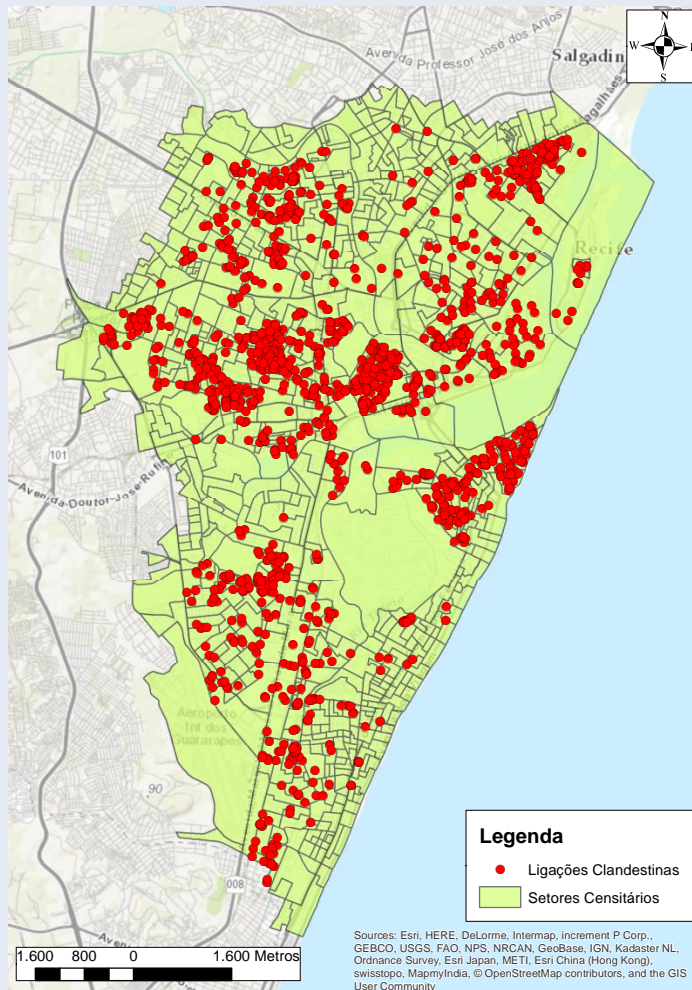
$$Y = X\beta + \varepsilon, \text{ sendo } \varepsilon = \lambda W\varepsilon + \xi$$

β : coeficiente de regressão;
 ε : erro aleatório;
 $W\varepsilon$: erro com efeitos espaciais;
 λ : coeficiente espacial autoregressivo - medida de correlação espacial;
 ξ : componente do erro (homocedástico e independente).

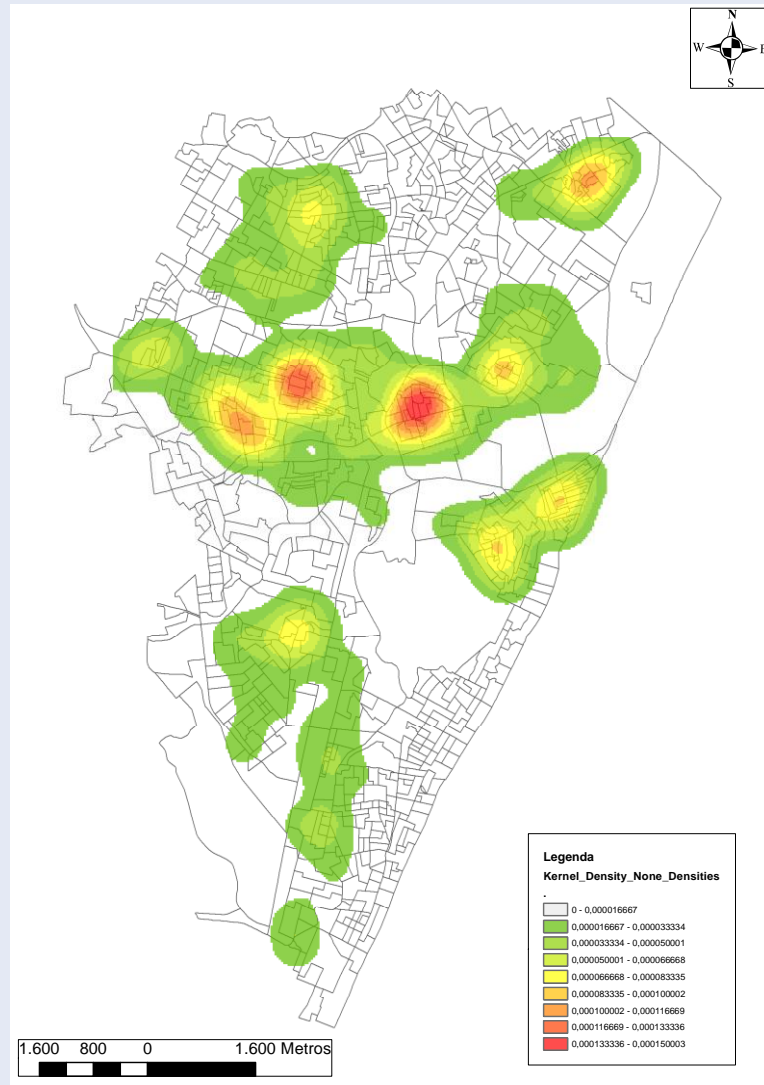
Resultados



Caracterização da distribuição de pontos



Agregados por setor censitário



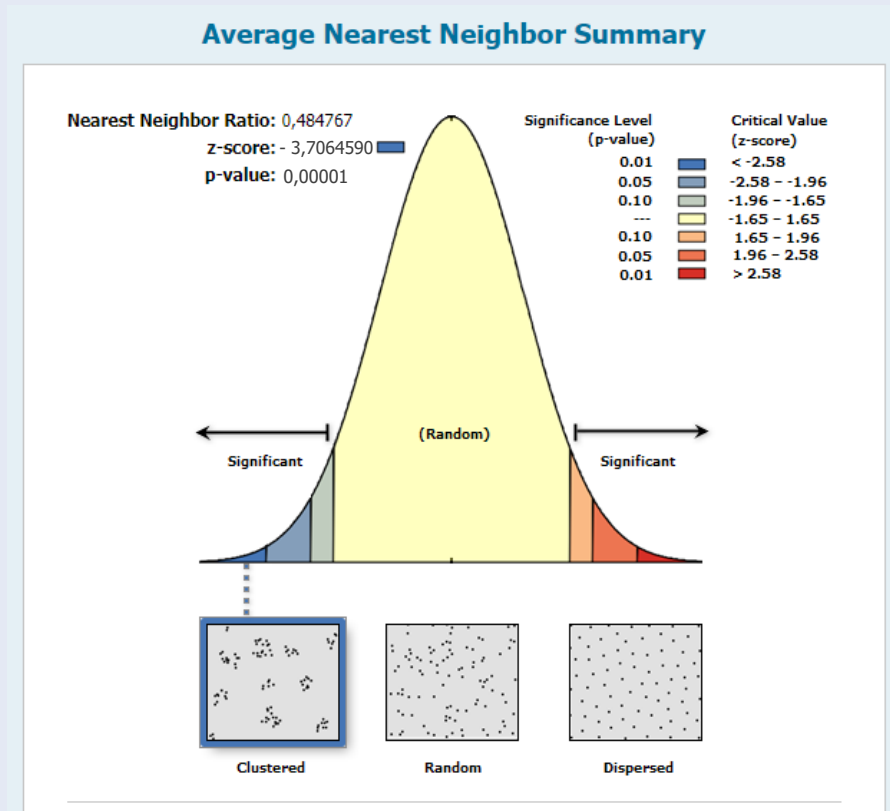
Avaliação do Agrupamento Espacial

- Estimador de Kernel
- Densidade de probabilidade de eventos na vizinhança
- Ponderada pela distância

- Identificação dos locais com maior probabilidade de ocorrência
- Avaliação da intensidade do processo (n° de eventos/uni. Área)

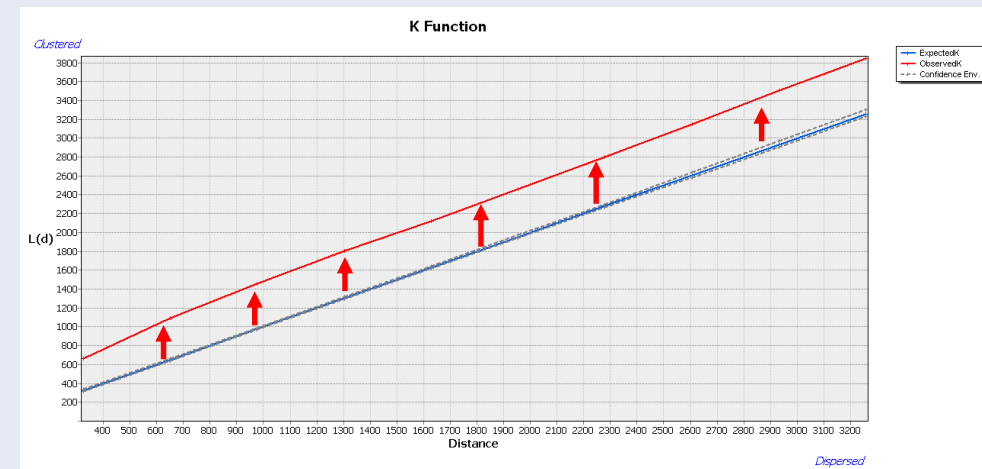
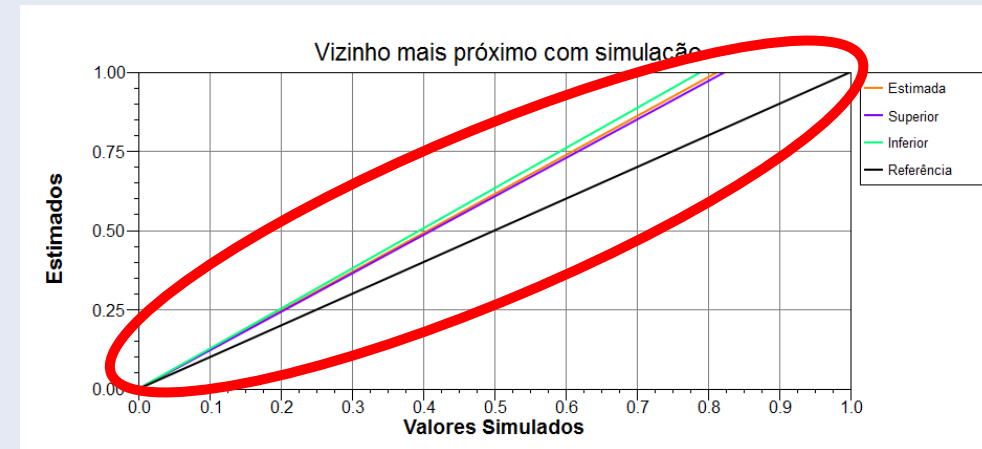
Resultados

Avaliação do Agrupamento Espacial



Método do Vizinho Mais Próximo

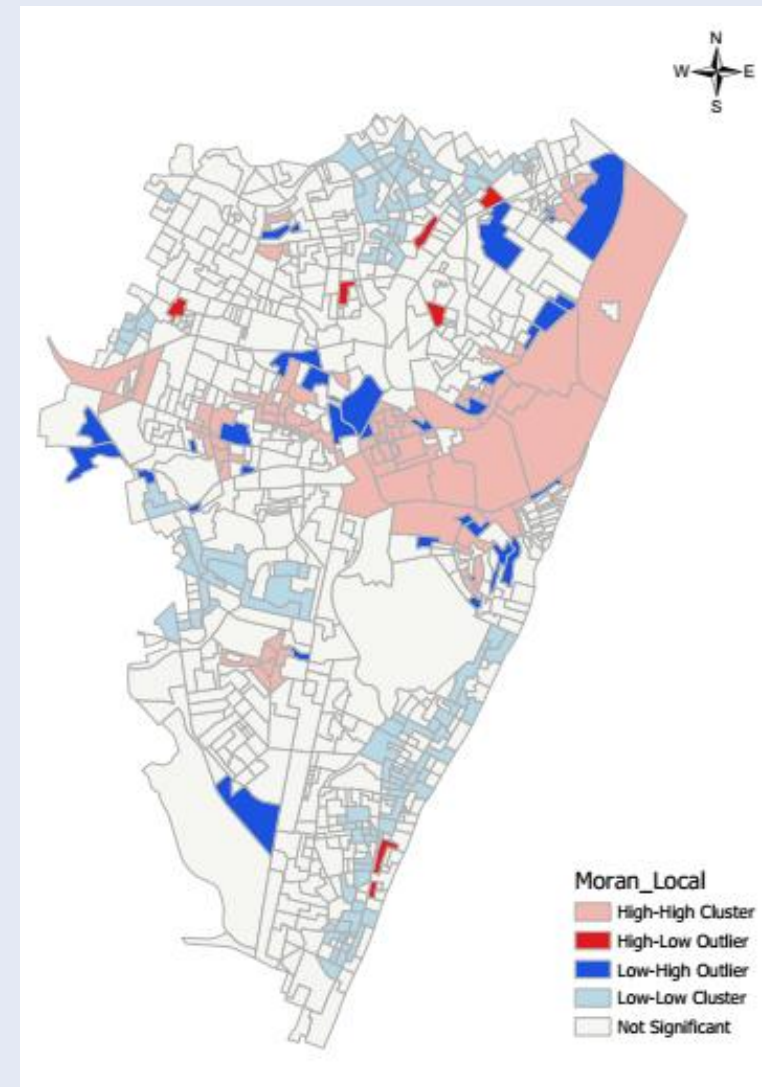
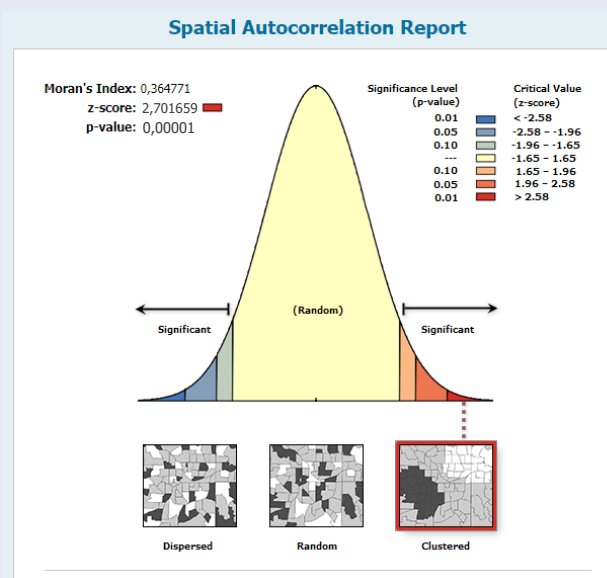
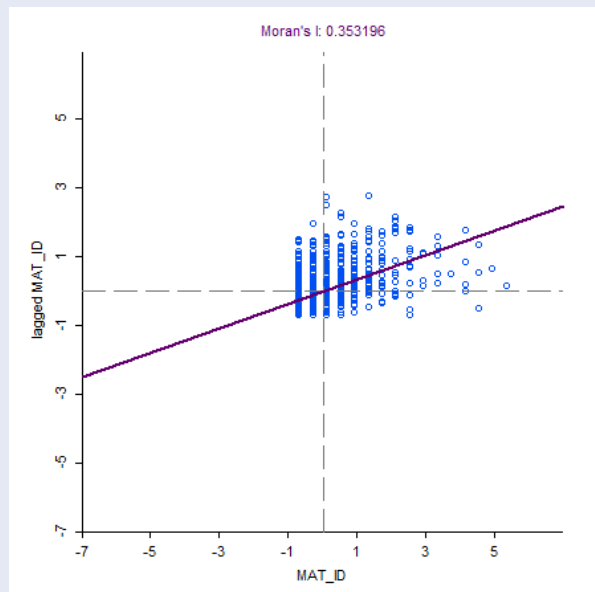
Tendência de agrupamento



Tendência de Agrupamento

Função K

Avaliação da Dependência Espacial



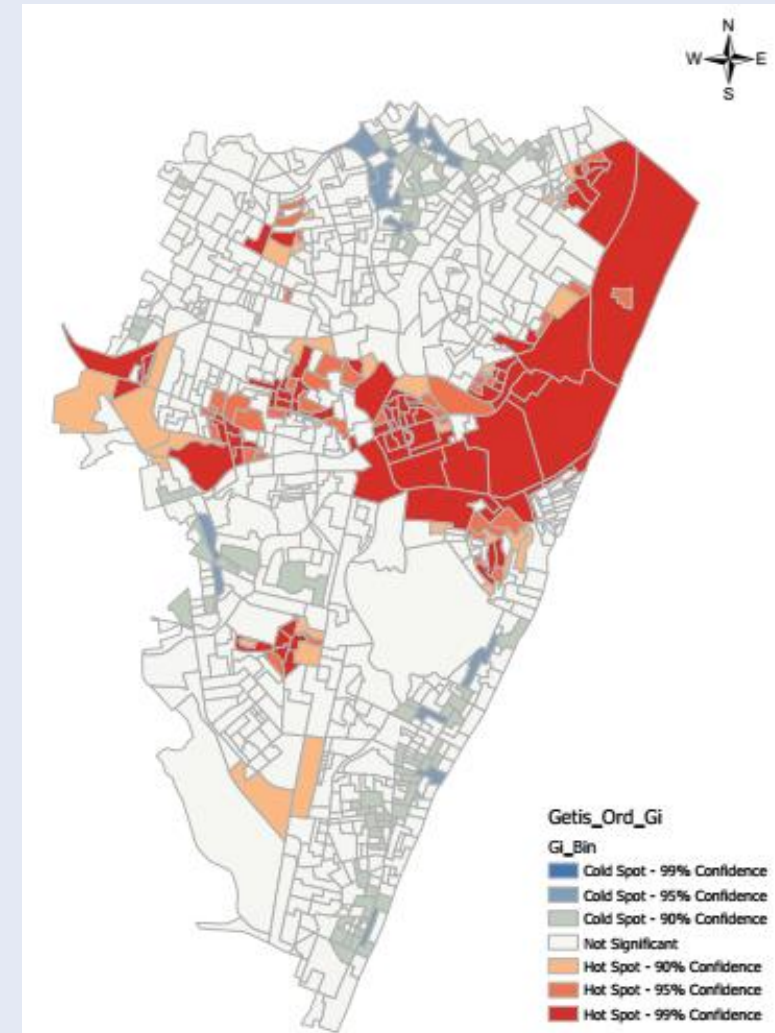
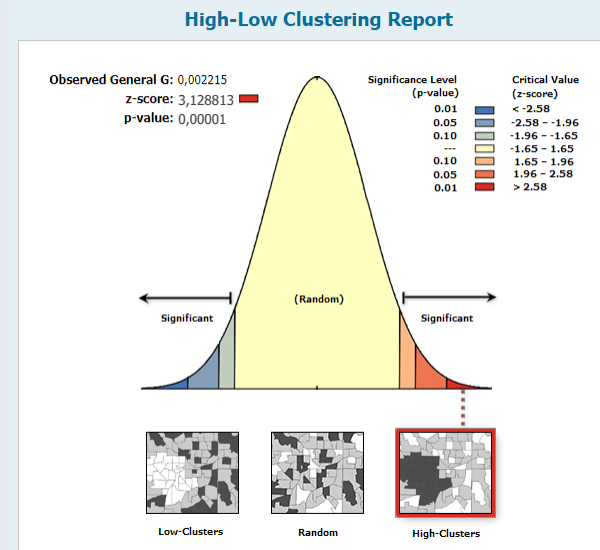
Índice de Moran

- Probabilidade menor que 1% do padrão observado ser resultado de um processo aleatório (Rejeita H_0)
- Autocorrelação espacial positiva ou direta: Vizinhos com valores semelhantes (Q1 e Q2)

Avaliação da Dependência Espacial

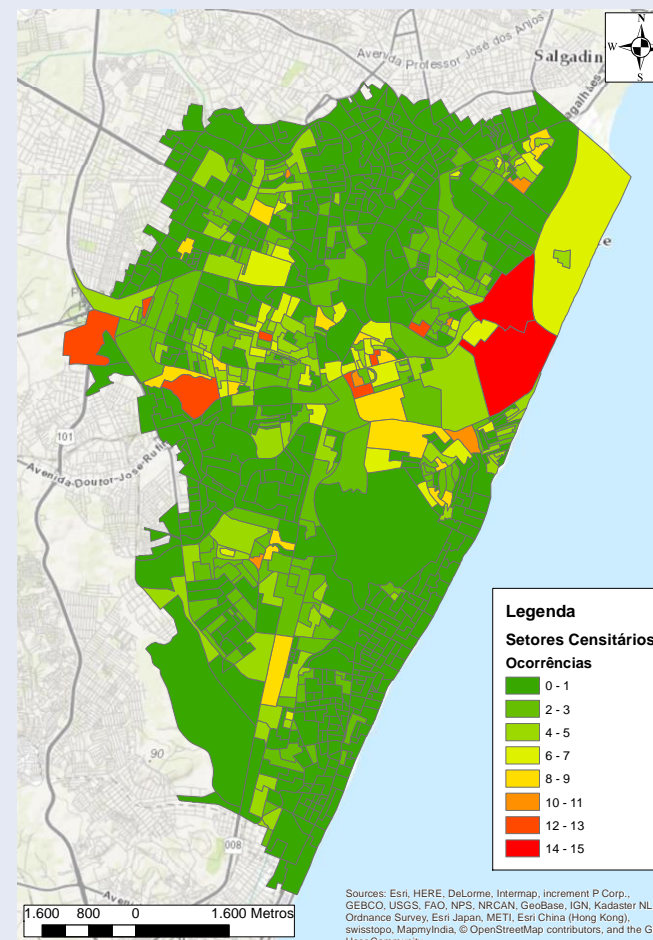
Getis and Ord (G)

- Probabilidade menor que 1% do padrão observado ser resultado de um processo aleatório (Rejeita H0)
- Identificação de Hotspots e Coldspots (90-99% confiança)



Regressão Espacial

- Variáveis Independentes:
 - Distribuição de População por Faixa de Renda
 - Setores Censitários (IBGE)
 - Todas as faixas de renda
 - DOM_REN_V0: 1 a 2 SM
 - DOM_REN__1: 2 a 3 SM
 - DOM_REN__2: 3 a 5 SM
 - DOM_REN__3: 5 a 10 SM
 - DOM_REN__4: Mais de 10 SM
 - DOMREN_AGP: Até 1 SM



Resultados

Regressão Espacial

- DOM_REN_V0: 1 a 2 SM
- DOM_REN__1: 2 a 3 SM
- DOM_REN__2: 3 a 5 SM
- DOM_REN__3: 5 a 10 SM
- DOM_REN__4: Mais de 10 SM
- DOMREN_AGP: Até 1 SM

Resultados - OLS

R-squared	: 0.326479	F-statistic	: 65.5199
Adjusted R-squared	: 0.321496	Prob(F-statistic)	: 0
Sum squared residual:	3406.5	Log likelihood	: -1744.16
Sigma-square	: 4.20037	Akaike info criterion	: 3502.32
S.E. of regression	: 2.04948	Schwarz criterion	: 3535.27
Sigma-square ML	: 4.16443		
S.E of regression ML:	2.04069		

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	0.85364	0.200002	4.26816	0.00002
DOM_REN_V0	-0.00822816	0.00483868	-1.7005	0.08942
DOM_REN__1	0.00192492	0.0112083	0.171741	0.86370
DOM_REN__2	-0.00509142	0.00956337	-0.532388	0.59462
DOM_REN__3	-0.00415942	0.00769967	-0.540208	0.58920
DOM_REN__4	-0.00324715	0.00677569	-0.479235	0.63189
DOMREN_AGP	0.0124413	0.00100328	12.4006	0.00000

Resultados - LAG

R-squared	: 0.419990	R-squared (BUSE)	: -
Sq. Correlation	: -	Log likelihood	: -1697.942785
Sigma-square	: 3.58624	Akaike info criterion	: 3409.89
S.E of regression	: 1.89374	Schwarz criterion	: 3442.83

Variable	Coefficient	Std.Error	z-value	Probability
CONSTANT	0.539996	0.202401	2.66795	0.00763
DOM_REN_V0	-0.0067806	0.00479045	-1.41544	0.15694
DOM_REN__1	0.00291196	0.0104341	0.27908	0.78018
DOM_REN__2	-0.00327822	0.00881587	-0.371854	0.71000
DOM_REN__3	-0.00340199	0.00735819	-0.462341	0.64384
DOM_REN__4	-0.000880138	0.00681685	-0.129112	0.89727
DOMREN_AGP	0.0131677	0.000972419	13.5412	0.00000
LAMBDA	0.417777	0.0440919	9.47514	0.00000

Resultados - SAR

R-squared	: 0.404701	Log likelihood	: -1704.7
Sq. Correlation	: -	Akaike info criterion	: 3425.4
Sigma-square	: 3.68078	Schwarz criterion	: 3463.05
S.E of regression	: 1.91854		

Variable	Coefficient	Std.Error	z-value	Probability
W_MAT_ID	0.36316	0.0408388	8.89252	0.00000
CONSTANT	0.0470217	0.198774	0.236559	0.81300
DOM_REN_V0	-0.00841548	0.00453056	-1.85749	0.06324
DOM_REN__1	0.00676183	0.0105047	0.643697	0.51977
DOM_REN__2	-0.00326827	0.00895315	-0.365042	0.71508
DOM_REN__3	-0.00127853	0.00721603	-0.177179	0.85937
DOM_REN__4	0.000233906	0.00634582	0.0368599	0.97060
DOMREN_AGP	0.0113277	0.000965907	11.7275	0.00000

- Evidência de agrupamento e dependência espacial
- Spatial Error Model (LAG)
 - Inclusão da variável espacial
 - Multiplicadores de Lagrange
 - Análise dos Resíduos
- Inclusão de outras variáveis:
 - Categoria dos clientes
 - Demanda de água
 - Histórico de abastecimento
- Variabilidade temporal



Conclusão



- Evidências de agrupamento espacial na ocorrência de ligações clandestinas na região Centro-Sul de Recife, para o período estudado.
- Verificação de dependência espacial na ocorrência dos eventos
 - Hotspots e Coldspots
- Melhores resultados: Spatial Error Regression (CAR)
 - Dependência espacial e efeitos da vizinhança
- Estabelecimento de diretrizes para combate a perdas aparentes em sistemas de abastecimento de água.



Referências Bibliográficas

- Van den Berg, C. The Drivers of Non-Revenue Water : How Effective are Non-Revenue Water Reduction Programs?. Policy Research Working Paper. Nº 6997. World Bank Group, Washington, DC. 2014.
- Bailey, T. C., and A. C. Gatrell. Interactive Spatial Data Analysis. Longman Scientific & Technical, Harlow, U.K. pp. 395. 1995.
- Faria, L. T., et al. Análise Espacial de Pontos para Mapeamento de Perdas Comerciais. In V Simpósio Brasileiro de Sistemas Elétricos – V SBSE, Foz do Iguaçu-PR, pp. 1- 6. 2014.
- Getis, A. Interactive Modeling Using Second-Order Analysis. Environment and Planning A, 16: 173–183. 1984.
- Anselin, Luc. Local Indicators of Spatial Association — LISA Geographical Analysis 27(2): 93–115, 1995.



Obrigado!

Trabalho Final

**Análise espacial da ocorrência de ligações
clandestinas em rede de distribuição de água**

Diogo Reis

SER 301 – Análise Espacial de Dados Geográficos

Dezembro/2017