

II-325 - ESTIMADOR DE DENSIDADE KERNEL EM APLICAÇÕES GIS APOIANDO DECISÃO EM SISTEMAS DE ESGOTAMENTO DA RMSP

Paulo Raimundo Volkov Ripardo⁽¹⁾

Engenheiro Civil pela Escola Politécnica da Universidade de São Paulo. Engenheiro da Divisão de Operação de Esgoto da Unidade de Negócio Oeste (MOEE) - SABESP

Paulo Levy de Souza Rodrigues⁽²⁾

Engenheiro Civil pela Universidade Nove de Julho. Especialista em Gestão de Negócios – MBA pela FIA – USP. Gerente da Divisão de Operação de Esgoto da Unidade de Negócio Oeste (MOEE) - SABESP

Endereço⁽¹⁾: Rua Major Paladino, 300 prédio 11 – Vila Leopoldina – São Paulo - SP - CEP: 05307-000 - Brasil - Tel: +55 (11) 3838-6118 - e-mail: pripardo@sabesp.com.br

RESUMO

Mostrar o uso do Estimador de Densidade Kernel na criação de um mapa temático, com foco no auxílio às equipes de campo para a resolução de problemas inerentes a rede coletora de esgoto.

O uso do mapa de calor ou “heat map”, muito conhecido em diversas áreas do conhecimento, nesse trabalho é apresentado para o estudo de obstruções da rede coletora de um município da Região Metropolitana de São Paulo (RMSP) através da criação do mapa de calor usando um software GIS (Geographic Information Systems – Sistemas de Informações Geográficas) para visualização da densidade de pontos, dentro do software a rotina utilizada para a obtenção de tal figura é o Estimador de Densidade Kernel (EDK).

Com ele foi possível localizar os pontos de maiores incidências de serviços de desobstrução de rede, o EDK mostra uma tendência de incidência de um evento cuja a amostra é descontínua tornando-a contínua no espaço e, dessa forma, aplica-se níveis de cores de acordo com o nível de incidência em dado ponto no espaço. Para o dado município estudado, criou-se o mapa de calor das desobstruções da rede coletora.

PALAVRAS-CHAVE: GIS, Estimador Kernel, Mapa de Calor, rede coletora de esgoto.

OBJETIVO

O presente trabalho focou no estudo do Estimador de Densidade Kernel, que é uma das ferramentas presentes nos softwares de GIS, como aliada a análise de dados espaciais aplicada ao setor de saneamento.

INTRODUÇÃO

• HISTÓRIA DA APLICAÇÃO GIS NO SANEAMENTO

Uma das primeiras aplicações de GIS em saneamento foi o caso do mapeamento feito por John Snow em Londres no ano 1854, durante um surto de cólera. Nesse episódio o médico londrino identificou no mapa o local de residência de cada pessoa que falecia e no mesmo mapa comparou com o posicionamento dos poços de onde as pessoas consumiam a água. Esse trabalho de GIS realizado por Snow lhe proporcionou localizar o poço contaminado como mostrado na figura 1.

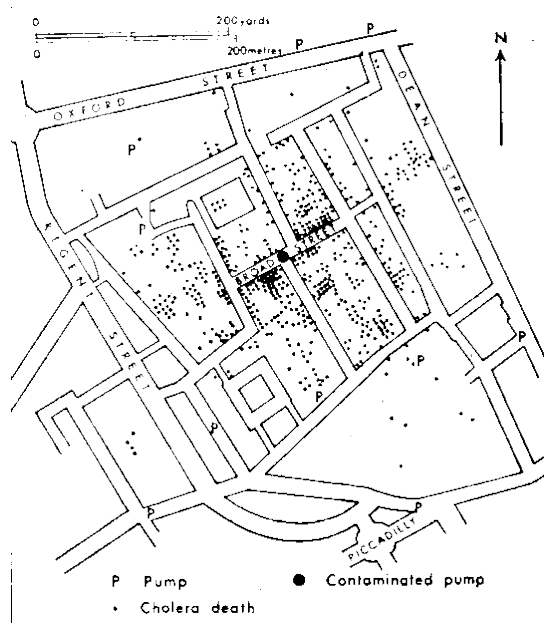


Figura 1- Mapa da Broad Street feito pelo médico John Snow em 1854, Londres, Inglaterra. Fonte: https://www.york.ac.uk/depts/maths/histstat/snow_map.htm

• APLICAÇÃO GIS NESTE ESTUDO DE CASO

Uma nova versão do mapa de Snow é apresentada na figura 2, onde já se utiliza as ferramentas atuais de GIS e uma das rotinas de criação de mapas temáticos é a utilização de cores para identificar um ponto de maior intensidade em relação aos demais locais, de modo a facilitar a visualização ou mesmo para criar um efeito de destaque para alguma posição que exija maior atenção do tomador de decisão.



Figura 2 Mapa de Calor do surto de cólera ocorrido em Londres no ano de 1854. Fonte: ESRI

A diferença entre os mapas, em primeira visualização, poderia tratar-se somente de uma questão gráfica, porém com toda a solução computacional, na qual está inserido o segundo mapa, há uma poderosa ferramenta da estatística, chamada de Estimador de Densidade Kernel, que neste trabalho será chamado de EDK.

O Estimador de Densidade Kernel (EDK), uma das ferramentas da estatística não paramétrica, ao longo dos anos vem se mostrando de grande utilidade em diversas áreas da ciência e tecnologia. Na área do saneamento, e.g, ela pode ser usada para localização de tendência de acúmulo de incidência de serviços executados de desobstrução de rede coletora de esgoto.

Neste estudo foi escolhido o município de Osasco operado pela SABESP e aplicamos o modelo EDK, a fim de gerar um mapa de calor para indicação de obstruções de rede para o citado município, cujo período de estudo foi o ano de 2016, por meio do uso do software comercial GIS, utilizado pela empresa. O trabalho se inicia

com a espacialização dos pontos dentro da área de interesse, em um município da região metropolitana de São Paulo. A base de dados utilizada vem do banco intermediário presente no *Plugin* de dados da empresa, presente no GIS, de onde é feita a coleta de informações de desobstruções ocorridas no município dentro de um dado período. Foram analisados o primeiro e segundo semestres de 2016, cada serviço executado com o código 2.46010 é associada ao código de cliente, já os serviços sem indicação do código de cliente ou coordenadas XY dentro do sistema de referência foram desprezados pela dificuldade de especializar o dado sem essas informações. O processo de espacialização das informações ocorre pela ligação de tabelas entre serviço executado no município de cada período com o arquivo Shapefile (SHP), que gera os pontos plotados no mapa.

MÉTODO

• FORMULAÇÃO/DESCRIÇÃO DA COMPOSIÇÃO DO EDK

Para uso no EDK, as informações que se fazem necessárias são largura de banda e o tamanho do pixel. A largura de banda, chamado de raio de pesquisa ou search radius nas versões em inglês, é um dado é opcional no GIS, mas a rotina de cálculo para a definição da largura de banda evita o aparecimento de anéis ao redor de pontos muito esparsos no espaço.

O tamanho do pixel, também é uma informação tratada como opcional pelo software, mas a sua utilização ajuda a definir o arredondamento que há no mapa. Para uma melhor visualização, se faz necessário observar qual o tamanho do elemento da amostra para que o “heat map” seja o mais fiel possível à realidade que se deseja representar ou prever.

A formulação apresentada por Wasserman (2006 apud Silverman, 1986) indica como outro item importante para a criação de um mapa de densidade a quantidade de pontos utilizados da população.

$$\hat{f}_n(\mathbf{x}) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h}\right) \quad \text{equação (1)}$$

Onde n é a população, h é a largura de banda, $\mathbf{X}_i$ é o ponto da população que está sendo analisado, $K(\mathbf{x})$ é a função kernel e $\hat{f}_n(\mathbf{x})$ é média dos kernels centrados em $\mathbf{X}_i$.

O kernel não possui uma função estritamente específica que o represente. Em Wassermann (2006), são apresentados diferentes tipos de kernels que podem ser aplicados à função estimador. Segundo Wasserman (2006), a escolha da função kernel não é importante. A escolha do kernel não é crucial, mas a escolha da largura de banda h é importante.

A determinação da largura de banda seguindo a metodologia recomendada pela ESRI é a seguinte:

$$h = 0,9 \cdot \min\left(DP, \sqrt{\frac{1}{\ln 2}} \cdot Dm\right) \quad \text{equação (2)}$$

Onde DP é distância padrão a qual é calculada por meio de uma função do software comercial GIS e Dm é mediana das distâncias médias calculadas a partir de um centro médio obtido de toda a população de dados.

RESULTADOS OBTIDOS

O “heat map” do município de Osasco mostrou para a alta gerência a importância da realização de obras de interligação de rede coletora de esgoto aos coletores troncos na queda na demanda de serviço. A facilidade de análise provocada pela variação de cores, assim como mostrado na figura 2 do mapa de Snow atualizado, capacita aos colaboradores, que não trabalharam diretamente com a confecção do mapa, analisar as informações que são apresentadas.

Notou-se através da comparação dos mapas do primeiro e segundo semestre a diminuição do índice de obstrução da rede coletora, indo de 527 desobstruções/ 100 km.ano para 461 desobstruções/ 100 km.ano, que representa uma queda de 12,5%. É esperado ainda que haja impacto no índice de reclamações de clientes e de chamadas ao atendimento telefônico da empresa. Todos os fatores associados tendem a levar a uma melhoria da imagem da empresa perante aos clientes finais do serviço prestado, devido à diminuição do número de ocorrências de obstrução conforme o mapa da figura 3.

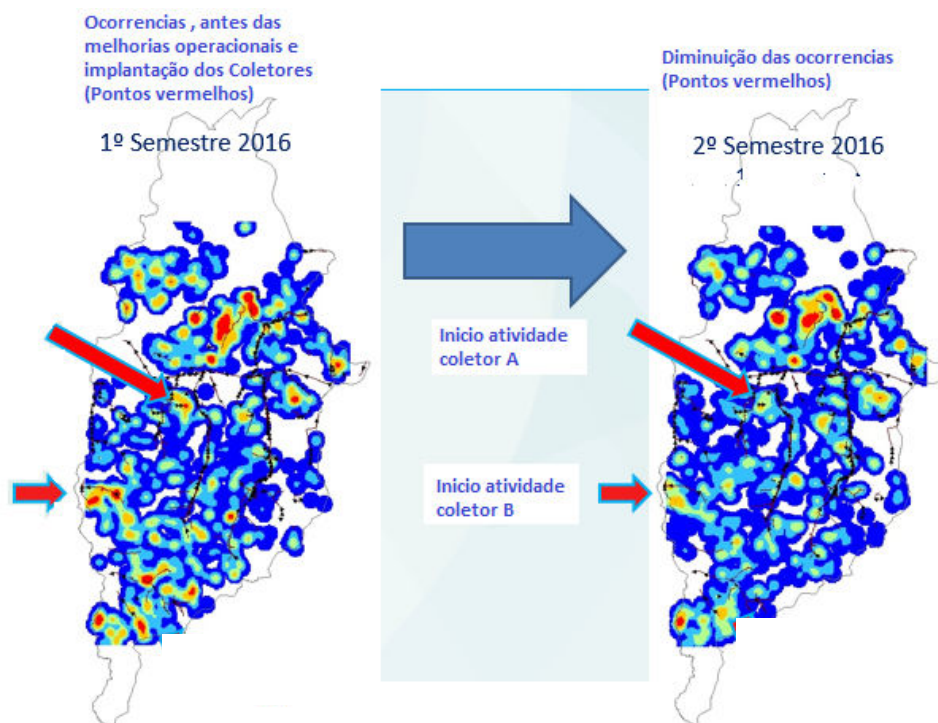


Figura 3 - Incidência de obstrução da rede coletora antes e depois da realização da interligação.

ANÁLISE DOS RESULTADOS

Durante a confecção do mapa e comparando-se com as tabelas de serviço executado ao longo do ano nota-se uma tendência de queda no indicador de serviços de desobstrução de rede coletora.

Observando os resultados obtidos nesse trabalho nota-se que a análise estatística deve ser cada vez mais introduzida no ambiente de trabalho da equipe de engenheiros das empresas de saneamento, pois a consciência da existência da ferramenta, que está presente nas rotinas e funções do software, e o entendimento do porquê de haver a função, capacitam o usuário de GIS ou engenheiro a compreender a saída (output) de cada informação entregue pelo software GIS.

A redução no valor do índice de ocorrências de desobstrução de rede coletora de esgoto é bem satisfatória. No nível operacional, as equipes começam a se dedicar aos serviços de manutenção preventiva que são entendidos como benefícios que a empresa de saneamento oferece à população, uma vez que os problemas de desobstrução deixam de ocorrer com frequência e, de acordo Garcia (2014 apud Rodrigues, 2012), “a manutenção preventiva tem como ponto positivo a facilidade por parte da produção em cumprir o planejado, assegura a continuidade do funcionamento das máquinas, aumento do tempo de vida das máquinas, diminui o número total de intervenções corretivas e reduz o custo da corretiva”. Assim, trazendo economia para o caixa da empresa, proporcionando capacidade de investimento em outras atividades entre elas dedicar-se a melhorias do sistema de saneamento.

CONTRIBUIÇÕES E LIMITAÇÕES DO ESTUDO

O trabalho aqui se dedicou a analisar somente o efeito para uma variável simples, todavia a estatística e os recursos computacionais fornecidos pela empresa possibilitam trabalhar uma análise multivariada do Estimador de Densidade Kernel. De acordo com Wasserman, o equacionamento do problema passa a ser:

$$\hat{f}_n(\mathbf{x}) = \frac{1}{n h_1 \cdots h_d} \sum_{i=1}^n \left\{ \prod_{j=1}^d K\left(\frac{x - X_i}{h_j}\right) \right\} \quad \text{equação (3)}$$

Onde d indica número de variáveis independentes existentes no problema de análise multivariada em software GIS. Neste caso, pode ser necessário conhecimentos de linguagem Python ou mesmo de R, por se tratar de um estudo com foco em dados de estatística espacial.

CONCLUSÃO

A apresentação do “heat map” para a alta direção da empresa corroborou com a comprovação técnica para respaldar o planejamento financeiro-econômico-ambiental, a fim de pleitear recursos para as obras na região de atuação dessa superintendência junto à diretoria e órgãos financiadores de obras de saneamento. Vale ressaltar que além das questões de planejamento há benefício imediato para as equipes operacionais na medida que visualizam o resultado de suas ações e a expressão gráfica dos pontos com maior ocorrência/incidência de obstruções

REFERÊNCIAS BIBLIOGRÁFICAS

1. WASSERMAN, L. All of Nonparametric. Springer, , p. 55-56,125-144, 2006
2. SILVERMAN, B. W. Density Estimation for Statistics and Data Analysis, Monographs on Statistics and Applied Probability, Chapman and Hall (1986)
3. SCOTT, D. W, Multivariate Density Estimation, John Wiley & Sons, Inc., New York (1992)
4. POTTER, L. SLUO Lectures on Statistics , 2006. Disponível em: < http://www-group.slac.stanford.edu/sluo/Lectures/stat_lecture_files/sluo2006lec4.pdf>. Acesso em: 15 de nov. 2016
5. SOUZA, N. P , SILVA, E. M. G. C , TEIXEIRA, M. D, LEITE, L. R , REIS, A. A , SOUZA, L. N , ACERBI JR, F. W RESENDE, T. A. Aplicação do Estimador de Densidade *kernel* em Unidades de Conservação na Bacia do Rio São Francisco para análise de focos de desmatamento e focos de calor. Anais XVI Simpósio Brasileiro de Sensoriamento Remoto – SBSR, Foz do Iguaçu, PR, Brasil, 13 a 18 de abril de 2013, INPE. p.4958.
6. PEREIRA, G. P , SANO, E. E. Mapa de densidade de Kernel como indicador de desmatamento futuro na Amazônia Legal. Anais XVI Simpósio Brasileiro de Sensoriamento Remoto – SBSR, Foz do Iguaçu, PR, Brasil, 13 a 18 de abril de 2013, INPE. p.4404.
7. ESRI, ArcGIS Pro, How Kernel Density works, 2016. Disponível em < <http://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/how-kernel-density-works.htm>> acesso em 15 nov de 2016
8. GARCIA, F . L , NUNES, F . L, Proposta de implantação de manutenção preventiva em um centro de usinagem vertical: um estudo de caso. Tecnologia e Tendências. Vol.10 nº2. 2º sem. 2014. Novo Hamburgo, RS. Feevale. Disponível em < <http://www.feevale.br/Comum/midias/d3e6b4f1-6778-4673-9904-78f1cd083801/5%20PROPOSTA%20DE%20IMPLANTA%C3%87%C3%83O%20DE%20MANUTEN%C3%87%C3%83O%20PREVENTIVA%20EM%20UM%20CENTRO%20DE%20USINAGEM%20VERTICAL%20UM%20ESTUDO%20DE%20CASO.pdf>> Acesso em 01. mar. 2016
9. CAMERON, D. & JONES, G. John Snow, the Broad street pump and modern epidemiology. Int. J. Epidemiol., 12: 396-3, 1983. Disponível em < <http://ije.oxfordjournals.org/content/12/4/393.full.pdf+html>> Acesso em 13 dez. 2016