



## XVIII Seminário Nacional de Distribuição de Energia Elétrica

SENDI 2008 - 06 a 10 de outubro

Olinda - Pernambuco - Brasil

### Geoestatística na Prospecção de Perdas de Energia

<b>Eduardo de Rezende Francisco</b>	<b>Eduardo Bortotti Fagundes</b>
<b>AES Eletropaulo</b>	<b>AES Eletropaulo</b>
eduardo.francisco@aes.com	eduardo.bortotti@aes.com

#### Palavras-chave

Análise Geográfica; Geoestatística; Modelos Preditivos; Perdas de Energia; Variáveis Socioeconômicas

#### Resumo

Um dos grandes desafios da AES Eletropaulo é buscar a redução contínua do índice de perdas de energia, particularmente aquele correspondente às perdas comerciais. Diante desse desafio, muito esforço se faz no entendimento dos reais fatores geradores dessa perda, e também na busca de ações mais eficientes de combate. Para a efetiva redução de perdas, a utilização de técnicas sofisticadas para apoio à gestão tática e operacional contribui para um melhor retorno financeiro, e, por esse motivo, é muito bem-vinda. As ferramentas de análise geográfica e a geoestatística trazem essa contribuição, através da geração de modelos preditivos que consideram as relações espaciais e a vizinhança na compreensão do fenômeno. Este trabalho apresenta uma metodologia geoestatística de identificação de áreas da AES Eletropaulo com maior potencial de incidência de clientes com perdas de energia, que também confirma a relação estatística entre variáveis sócio-econômicas e as perdas de energia. Utilizado de forma sistemática, esta metodologia potencializa o desempenho das equipes de campo, priorizando clientes de alta propensão à perda para serem inspecionados.

#### 1. INTRODUÇÃO

Anualmente, a AES Eletropaulo deixa de faturar 2.700 GWh em decorrência de perdas não-técnicas, um montante de energia maior do que o consumo residencial anual nos municípios de São Bernardo, São Caetano, Santo André e Diadema, que juntos têm 2 milhões de habitantes. Segundo a ANEEL, a perda comercial anual do setor brasileiro de distribuição de energia elétrica foi de cerca de R\$ 5,1 bilhões em 2005.

Em função da relevância, a gestão de perdas se constitui em área estratégica para as empresas distribuidoras de energia elétrica e requer a concentração de esforços para seu entendimento, cujas origens e motivações podem ser classificadas em fraudes, anomalias de medição, falhas nos processos internos de faturamento, ligações clandestinas, clientes cortados ligados diretamente à rede ou técnicas, bem como, para a definição de estratégias de atuação que sejam eficazes na mitigação de seus impactos nos resultados financeiros da organização.

Estudos anteriores (FRANCISCO, 2006) destacam que o consumo de energia elétrica é um bom substituto da renda domiciliar, analisados para áreas de ponderação do IBGE (conjunto de setores censitários, que são cerca de 20.000 na área de concessão da AES Eletropaulo). Por outro lado, as

perdas de energia podem ser mensuradas através da diferença entre energia comprada e vendida, ao nível das Estações Transformadores de Distribuição (ETDs).

A utilização de técnicas sofisticadas para apoio à gestão tática e operacional, contribui para um melhor retorno financeiro com ações direcionadas de combate às perdas, e, por esse motivo, é muito bem-vinda. As ferramentas de análise geográfica trazem essa contribuição, através da manipulação de diversos níveis de informação. A geoestatística contribui com a avaliação e quantificação da influência geográfica do fenômeno. Esta é a inovação que este trabalho permite.

Em linhas gerais, os objetivos deste trabalho são:

- a) Desenvolver um indicador geográfico de potencial de propensão à perda, em especial considerando a perda comercial, através da combinação geográfica entre informações internas à empresa e variáveis socioeconômicas;
- b) Desenvolver modelos de regressão da perda comercial, com o objetivo de explicar e prever, sistematicamente, os fatores que afetam essa perda. Em especial, utilizar técnicas geoestatística de predição (modelos globais, como o SAR [*Spatial Auto-Regression*] e modelos locais, como o GWR [*Geographically Weighted Regression*]), que consideram a vizinhança e outras propriedades espaciais no entendimento do fenômeno.

## **2. DESENVOLVIMENTO**

### **2.1. Contexto Geral e Motivação Metodológica**

Compreender a distribuição espacial de dados oriundos de fenômenos ocorridos no espaço constitui hoje um grande desafio para a elucidação de questões centrais em diversas áreas da administração pública e de empresas, como saúde, dinâmica urbana, meio-ambiente, marketing e gestão de ativos. Tais estudos vêm se tornando cada vez mais comuns, devido à disponibilidade de Sistemas de Informação Geográfica (GIS), e também à necessidade de explicação da distribuição espacial de problemas e variáveis de interação socioeconômica que modelos tradicionais e clássicos geralmente não endereçam. A geoestatística é um campo da ciência regional que aborda o tema da dependência e da heterogeneidade espacial, aspectos críticos em qualquer estudo de economia regional. Estas características podem invalidar o uso de técnicas econométricas convencionais. Os métodos quantitativos espaciais são um caso especial do enfoque estatístico geral e, portanto, necessitam de um conjunto de métodos e técnicas diferentes daqueles utilizados na estatística convencional. Empresas que possuem na geografia informação essencial para sua gestão se beneficiam tremendamente do uso destas técnicas.

Renda é o indicador tradicionalmente adotado em estudos sobre condições de vida e pobreza, uma vez que é através dela que se dá o acesso aos bens e serviços necessários à sobrevivência. Entende-se renda como a soma do rendimento mensal de trabalho com o proveniente de outras fontes (IBGE, 2003). No entanto, a dificuldade em obter informações precisas sobre essa variável, freqüentemente alterada por subdeclaração, superdeclaração, esquecimento, sazonalidade da fonte dos rendimentos e recusa, torna difícil o uso deste indicador em pesquisas de mercado.

Em conseqüência disso, os Institutos de Pesquisa optam por captar a Classe Econômica ou o Poder de Consumo dos indivíduos através de indicadores baseados na posse de bens duráveis da família e no grau de instrução do chefe. Tais indicadores podem ser usados com certa precisão para substituir a renda. Além de ser uma medida indireta da renda familiar, o estoque de bens possuídos por uma família indica o nível de conforto por ela alcançado ao longo do tempo.

O mais recente indicador deste tipo é o Critério de Classificação Econômica Brasil (CCEB), ou simplesmente Critério Brasil, definido em 1996 pela Associação Nacional de Empresas de Pesquisa (ANEP). Este indicador é baseado em uma escala de pontos que pode variar de 0 (zero) a 34 (trinta e quatro) pontos, segmentados em 7 classes econômicas (ABEP, 2004).

Porém, o critério apresenta importantes diferenças regionais e não é adequado para a caracterização de famílias posicionadas nos extremos da distribuição de. O critério se presta a segmentar grandes massas e se adequa bem a estudos de abrangência nacional. Para regiões ou segmentos específicos, estudos aprofundados necessitam de especializações ou adaptações do Critério Brasil para a caracterização inicial da população em análise, muitas vezes envolvendo variáveis que melhor caracterizem a aptidão para o consumo. O uso de indicadores de consumo que tenham abrangência e utilidade gerais pode mostrar-se útil nesse processo de caracterização dos consumidores.

Dentre os indicadores com essa natureza está o consumo de energia elétrica. Nacionalmente, o serviço de fornecimento de energia elétrica abrange 97,0% dos domicílios brasileiros, índice que aumenta para 99,6% na área urbana, para 99,4% na região Sudeste e 99,9% no município de São Paulo [3]. Tem mais capilaridade e cobertura que serviços de outras empresas de utilidades, como telefonia fixa e móvel, água encanada e gás. Isso porque, basicamente, todos têm acesso à luz, principalmente quando consideradas as ligações clandestinas e fraudes, mas nem todos têm água encanada, ou mesmo telefone. Além disso, as bases de dados das distribuidoras de energia elétrica contêm a informação de consumo de cada um de seus clientes, das classes mais às menos favorecidas.

Por ser um serviço essencial, abrangente e relativamente democrático em comparação a outros serviços de utilidade pública, suas informações cadastrais e comerciais podem oferecer subsídio para um conhecimento comparativo de características socioeconômicas e demográficas das famílias em estudo. Ademais, estudos aprofundados em determinadas classes como, por exemplo, as classes mais pobres, requerem mecanismos que permitam detalhamento de faixas específicas e o melhor entendimento do matiz de subníveis de classificação.

Aliados às informações de localização, histórico e sazonalidade, os indicadores elétricos podem ajudar a inferir uma melhor classificação socioeconômica a partir de faixas de consumo, e contemplar uma melhor definição do consumidor de baixa renda em regiões de menor acesso e maior dificuldade de levantamento de bens de consumo.

Para as empresas de varejo, a incorporação do consumo de energia elétrica poderá melhorar a identificação das famílias potenciais consumidoras de ofertas específicas. Consumidores de mesma classificação econômica segundo o Critério Brasil poderão ser diferenciados através do consumo de energia elétrica de seus domicílios. Uma vez que o consumo de energia elétrica, em certa medida, reflete a posse e o uso de bens duráveis de natureza elétrica, associado ao total de cômodos do domicílio e ao número de pessoas que o habitam, as empresas poderão melhorar sua taxa de retorno em materiais promocionais, malas diretas e custos afins, através de uma melhor segmentação de seu mercado, do aumento da reatividade dos consumidores em função de seu maior poder aquisitivo, e da melhor identificação de seu mercado alvo.

Além disso, para as empresas distribuidoras de energia elétrica, a aplicação de um modelo de classificação socioeconômica parcialmente ou totalmente baseado no consumo de energia elétrica propicia maior eficácia em estudos de identificação, segmentação e previsão de mercado, análises de fraude e inadimplência e na elaboração de estratégias de relacionamento com seus clientes, entre outros benefícios.

A relação entre Renda e Consumo de Energia Elétrica é muito alta, com coeficientes de determinação  $R^2$  variando de 0,912 e 0,960. Este resultado (corroborado pela alta similaridade entre os dois mapas da Figura 1) suporta o uso do consumo médio de energia elétrica, em nível territorial, como um excelente indicador regional de concentração de renda na cidade de São Paulo, extensível a toda área de concessão da AES Eletropaulo.

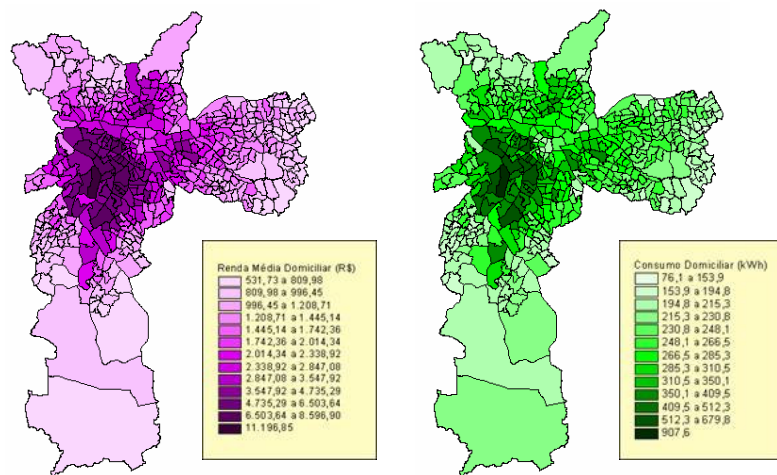


FIGURA 1. Mapas da Cidade de São Paulo representando: (i) Renda Domiciliar Total Média, e (ii) Consumo de Energia Elétrica Médio por Área de Ponderação (conjunto de Setores Censitários)  
Fonte: FRANCISCO, 2006.

## 2.2. Descrição da Metodologia – Indicador de Propensão à Perda

Este trabalho apresenta o desenvolvimento de uma metodologia de identificação de áreas com maior potencial de incidência de clientes com perdas de energia, que também confirma a relação estatística entre variáveis sócio-econômicas e as perdas de energia. Tal metodologia utiliza técnicas geoestatísticas e integra, para cada região geográfica em estudo, quatro informações básicas: (I) perda percentual de energia (total), (II) renda domiciliar média, (III) indicador de vulnerabilidade social (substituível por percentual da área ocupado por favelas), e (IV) percentual de clientes cortados.

A metodologia consistiu, em descrição geral, nos seguintes passos:

- (i) Extração da rede primária de todos os circuitos da AES Eletropaulo e associação de cada trecho de rede à Estação Transformadora de Distribuição (ETD) que o alimenta (a Figura 2 apresenta a rede primária – cada cor representa a cobertura de cada ETD);

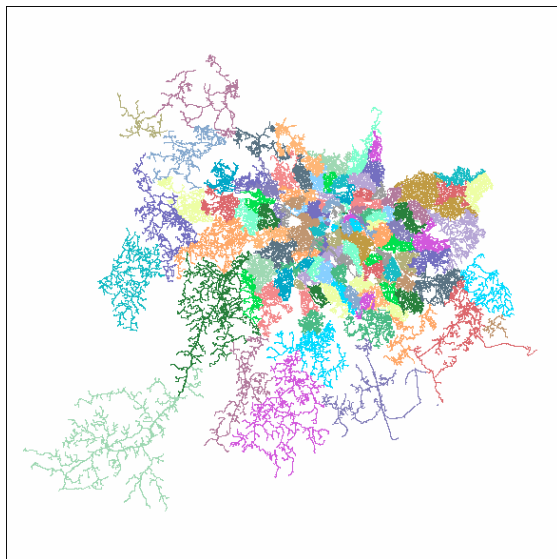


FIGURA 2. Mapeamento da Rede Primária  
Fonte: Sistema de Cadastro Técnico GIS da AES Eletropaulo

- (ii) Determinação da área de cobertura de cada ETD, através da técnica geoestatística de delimitação de polígonos de Thiessen, a partir da proximidade de cada ponto geográfico aos trechos de rede (a Figura 3 utiliza a mesma legenda para cada área de cobertura das ETDs);

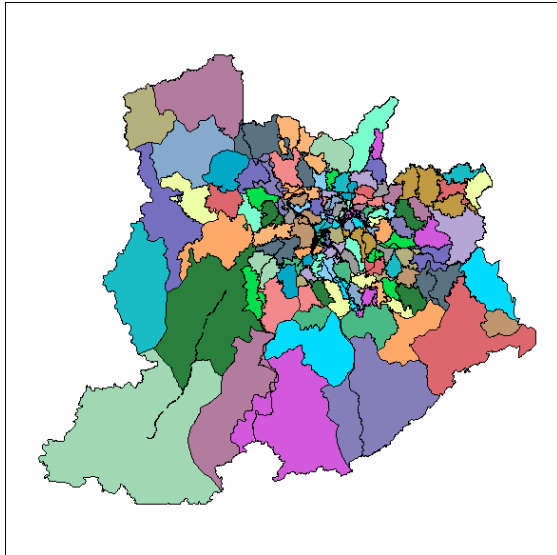


FIGURA 3. Determinação da área de Cobertura de cada ETD

Fonte: Os autores

- (iii) Mapeamento da Densidade Geográfica de Clientes Cortados (utilizamos uma malha quadrada de um quilômetro de lado) – a Figura 4 mostra o percentual de clientes cortados;

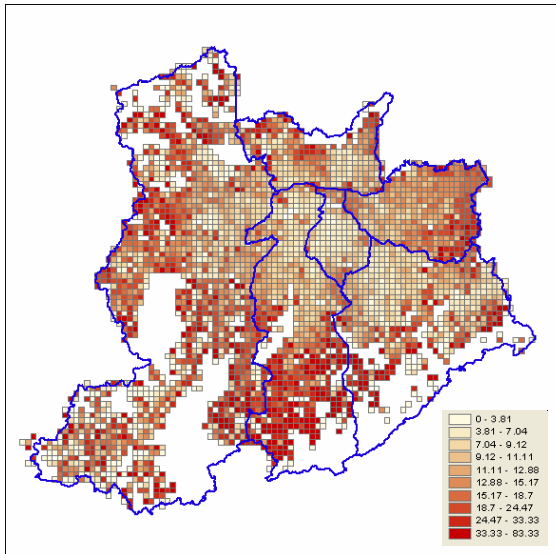


FIGURA 4. Concentração de Clientes Cortados

Fonte: Sistema Comercial SICON-B da AES Eletropaulo

- (iv) Mapeamento da Concentração de Consumo Residencial de Energia Elétrica, como *proxy* de Renda Domiciliar e associação às mesmas unidades de área da densidade de clientes cortados (malha quadrada de um quilômetro de lado) – a Figura 5 utiliza cores escuras para rendas baixas e cores claras para rendas altas;

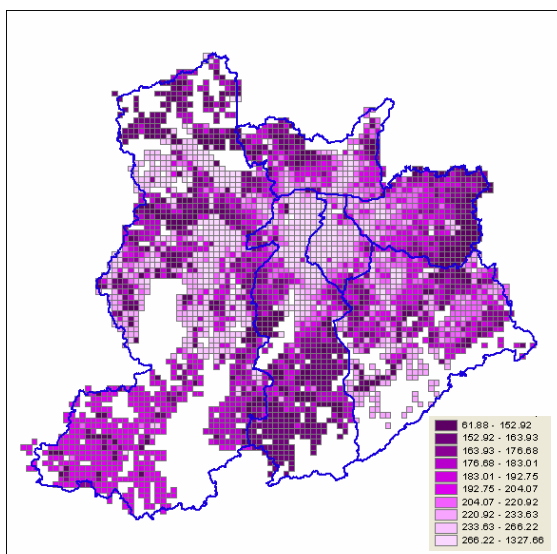


FIGURA 5. Concentração de Renda (Consumo de Energia Elétrica)

Fonte: Os autores, baseado em informações dos sistemas técnicos e comerciais da AES Eletropaulo, e em estudo anterior de Francisco (2006)

- (v) Identificação das áreas de Favelas e concentrações de aglomerados subnormais para a determinação do percentual de áreas de favelas de cada unidade de área de densidade de clientes cortados – a Figura 6 apresenta estas áreas;

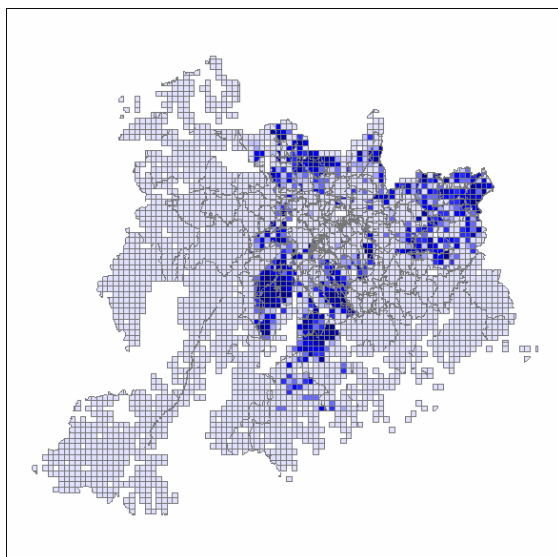


FIGURA 6. Concentrações de aglomerados subnormais (favelas)

Fonte: Prefeitura do município de São Paulo

- (vi) Junção Espacial (*Spatial Overlay*) entre os layers geográficos obtidos nos passos (ii), (iii), (iv) e (v) – usando esta técnica nós obtivemos um *layer* espacial em que cada entidade poligonal contém informação de quatro dimensões (respectivamente, os quatro passos mencionados acima) utilizados neste estudo do fenômeno de perdas comerciais;
- (vii) Ponderação entre as quatro informações associadas a cada unidade de área, para a geração do Indicador de Potencial de Perdas Comerciais.

As Figuras 7 e 8 resumem o propósito inicial do projeto, e a integração entre os domínios utilizados no projeto para a investigação da prospecção de perdas comerciais: (i) perdas por cobertura de ETD, (ii) densidade de renda (utilizando o consumo de energia elétrica como *proxy*), (iii) proximidade e concentração de áreas de favelas, e (iv) percentual de clientes cortados.

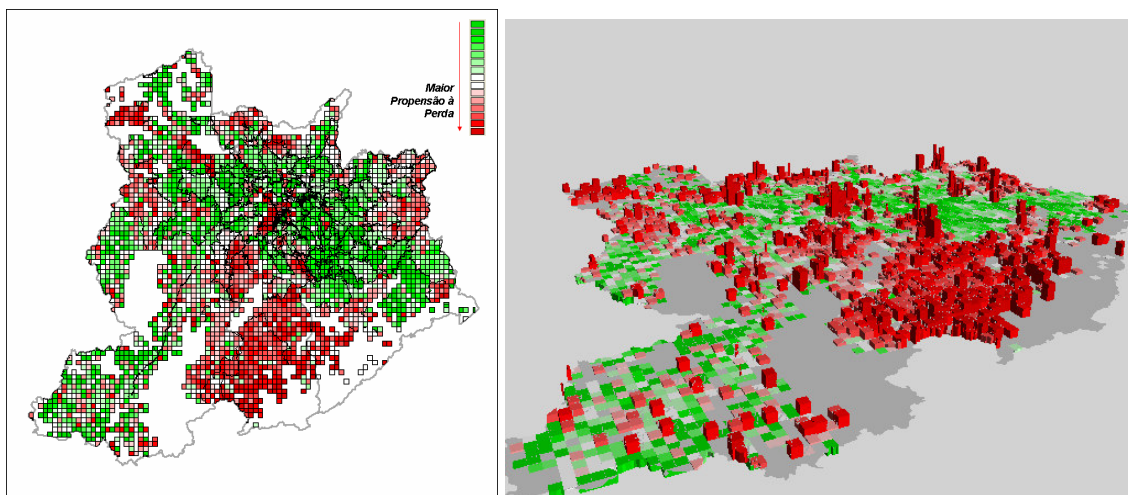


FIGURA 7. Ponderação entre as quatro dimensões de informação (mapa e Visualização 3D)

Fonte: Os autores.

### 2.3. Descrição da Metodologia – Modelos Geoestatísticos de Regressão da Perda Comercial

As técnicas geoestatísticas utilizadas foram: *Spatial Auto-Regression* (SAR) e *Geographically Weighted Regression* (GWR). São basicamente, técnicas exploratórias para análise de dados espaciais. A técnica SAR remonta aos anos 1950 e é tradicionalmente adotada em estatística espacial (DE SMITH, GOODCHILD & LONGLEY, 2007). Utiliza uma matriz de vizinhança ( $W$ ), associada ao próprio termo dependente  $y$  na estimação de  $y$ . Por isso, é denominada uma técnica auto-regressiva. As Equações 1 e 2 a seguir mostram como o termo auto-regressivo é incorporado, e a Figura 8 exemplifica a estrutura da matriz de vizinhança.

Equação 1. *Regressão Linear Multivariada*: 
$$y = \beta X + \varepsilon$$

Equação 2. *Regressão Linear Multivariada Espacial*: 
$$y = \rho W y + \beta X + \varepsilon$$

onde  $y$  é a variável dependente ;  $X$  é o vetor de variáveis independentes ;  $\beta$  é o vetor de coeficientes e  $\varepsilon$  é o erro da estimação (supostamente, normal).

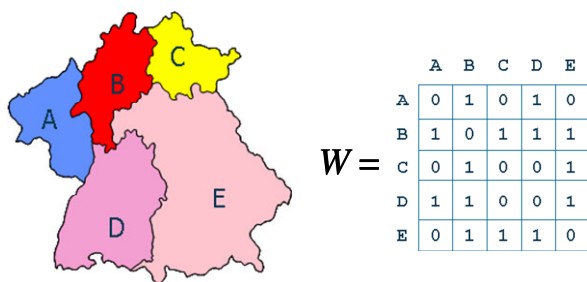


FIGURA 8. Exemplo de Matriz de Vizinhança

Fonte: Os autores.

A técnica GWR foi introduzida por Fotheringham, Charlton e Brunsdon (1997, 2002) para descrever uma família de modelos de regressão em que os coeficientes,  $\beta$ , podem variar espacialmente.

Em uma regressão linear tradicional, assumimos que a relação que está sendo modelada é estável em qualquer lugar da área estudada (isso vale inclusive para a técnica SAR). Em outras palavras, os parâmetros da regressão são iguais em todo o espaço estudado. A GWR, por sua vez, provê uma elegante maneira de modelar essa relação com variações locais dos coeficientes  $\beta$ , conforme apresentado na Equação 3.

Equação 3. *Geographically Weighted Regression*: 
$$y(g) = \beta(g)X + \varepsilon$$

onde  $y$  é a variável dependente ;  $X$  é o vetor de variáveis independentes ;  $\beta$  é o vetor de coeficientes,  $\varepsilon$  é o erro da estimação (supostamente, normal) e  $g$  é um ponto da amostra, localizado no espaço bidimensional (latitude, longitude).  $\beta(g)$  significa que o parâmetro está sendo estimado para a observação  $g$ , segundo uma amostra, local, específica para esse ponto.

Os coeficientes  $\beta(g)$  são determinados examinando-se um conjunto de pontos dentro de uma vizinhança definida para cada ponto  $g$  da amostra, utilizando um esquema de ponderação (*weighting scheme*), normalmente bi-quadrado ou gaussiano. A largura de banda que define essa vizinhança é um fator chave para esse modelo; ela pode ser definida manualmente ou através de um método adaptativo, como a minimização do Akaike Information Criterion (AIC) (FOTHERINGHAM, CHARLTON & BRUNSDON, 2002; DE SMITH, GOODCHILD & LONGLEY, 2007).

A variável dependente utilizada neste modelo foi a Perda por cobertura da ETD, obtida no passo (ii) da Metodologia de criação do indicador de propensão à perda (seção 2.2).

### 3. RESULTADOS

Baseado no Indicador de Potencial de Perdas Comerciais, regressões multivariadas foram realizadas, utilizando o *layer* do passo 2 da metodologia apresentada (perda total associada a cada ETD) como variável dependente, e as outras três variáveis como independentes. Como um dos mais importantes resultados deste estudo, o uso de métodos de regressão espacial SAR aumentou em 251% o poder de compreensão do fenômeno (os coeficientes de determinação  $R^2$  variaram de 11% para 38,6%). A Figura 9 apresenta o *LISA Map* (*Local Indicator Spatial Association*) e o índice de Auto-correlação Espacial de Moran (uma medida de dependência espacial). Cerca de 43% da perda está correlacionado com a perda em áreas vizinhas.

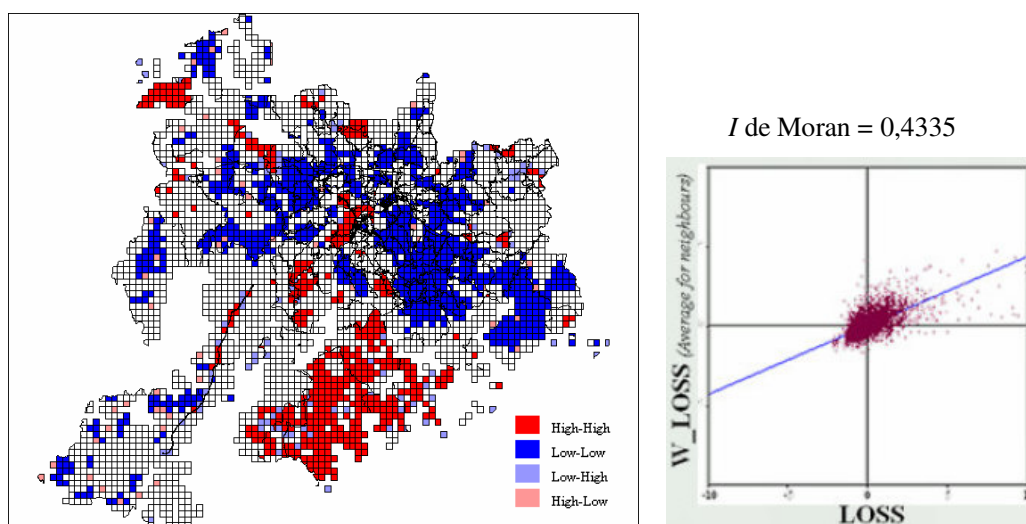


FIGURA 9. *LISA map*, *Scatter-plot* e Indicador I de Moran da perda das áreas de cobertura de ETDs  
Fonte: Os autores.

A Figura 10 apresenta os modelos de regressão tradicional e SAR utilizados nesta metodologia.



## Regressão Multivariada Tradicional

$$Perda = \beta_0 + \beta_1 Renda + \beta_2 ClientesCortados + \beta_3 Favelas$$

### REGRESSION

#### SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

Data set : fase1  
Dependent Variable : **Perda Percentual** Number of Observations: 7651  
Mean dependent var : 1.21553e-008 Number of Variables : 4  
S.D. dependent var : 0.999935 Degrees of Freedom : 7647

R-squared : **0.111343** F-statistic : 319.373  
Adjusted R-squared : 0.110994 Prob(F-statistic) : 0  
Sum squared residual: 6798.23 Log likelihood : -10404.2  
Sigma-square : 0.889006 Akaike info criterion : 20816.4  
S.E. of regression : 0.942871 Schwarz criterion : 20844.2  
Sigma-square ML : 0.888541  
S.E of regression ML: 0.942625

## Regressão Multivariada Espacial

$$Perda = \beta_0 + \beta_1 Renda + \beta_2 ClientesCortados + \beta_3 Favelas + \beta_5 W Perda$$

### REGRESSION

#### SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION

Data set : fase1  
Spatial Weight : fase1.GAL  
Dependent Variable : **Perda Percentual** Number of Observations: 7651  
Mean dependent var : 1.21553e-008 Number of Variables : 5  
S.D. dependent var : 0.999935 Degrees of Freedom : 7646  
Lag coeff. (Rho) : 0.545362

R-squared : **0.386082** Log likelihood : -9571.13  
Sq. Correlation : - Akaike info criterion : 19152.3  
Sigma-square : 0.67913 Schwarz criterion : 19187  
S.E of regression : 0.824093

FIGURA 11. Resultados dos Modelos de Regressão (Tradicional e SAR)

Fonte: Os autores.

A técnica GWR mostrou desempenho um pouco inferior ao da técnica SAR, mas seu resultado é também significativamente superior ao da regressão tradicional para a explicação do fenômeno da perda. A Tabela 1 apresenta os coeficientes de determinação ( $R^2$ ) dos três modelos utilizados.

TABELA 1 – Comparação entre os Coeficientes de Determinação ( $R^2$ ) dos Modelos Preditivos de Perda

Modelo de Regressão	Coefficiente de Determinação ( $R^2$ )
LR – Linear Regression	11,0 %
GWR – Geographically Weighted Regression	32,0 %
SAR – Spatial Auto-Regression	38,6 %

A AES Eletropaulo vêm sistematicamente incorporando o indicador de propensão à perda em sua estratégia de gestão de inspeções. O propósito é sistematizar essa metodologia para torná-la

incorporável ao modelo de gestão de priorização. A Geoestatística é uma ferramenta de grande potencial para a realização desse objetivo.

#### **4. CONCLUSÕES**

Desenvolvido em fase piloto na área correspondente à Diretoria Regional Leste da AES Eletropaulo, e já expandido para toda a área de concessão da empresa, este estudo está em fase inicial de apoio à definição de estratégias para cada região. Tal apoio se dá vis-a-vis a cada fonte de perda detectada, tais como: combate à fraude apenas em clientes ligados e em determinadas áreas geográficas, regularização de núcleos clandestinos, resgate de clientes cortados através de estratégias comerciais, e pode vir a ajudar no melhor direcionamento de futuras instalações de redes anti-fraude e equipamentos específicos, entre outras estratégias.

Uma vez que as informações utilizadas são todas internas e sob gestão da AES Eletropaulo, esta metodologia torna-se de fácil implementação e manutenção. A inovação está na integração entre informações técnicas, como a perda total por estação transformadora de distribuição, e informações comerciais, como o consumo médio faturado de energia elétrica e a situação (cortado versus não cortado) de cada unidade consumidora. Tal integração é possibilitada pelo GIS, e a geração do indicador se dá pelo estabelecimento de técnicas geoestatística, de uso muito incipiente e, por isso, pioneiro no setor de distribuição de energia elétrica.

Outro ponto de importante citação neste trabalho é a identificação de correlação representativa entre as variáveis analisadas: Renda, Inadimplência, Concentração de Favelas e as Perdas. Isso potencializa os benefícios da incorporação de variáveis externas aos processos de apoio à decisão e às análises de gestão comercial da companhia.

Mais ainda, este estudo está sendo aprimorado para identificar e entender o fator gerador das perdas de energia. Para os próximos passos, vislumbra-se a avaliação da relação entre as variáveis vistas neste estudo com a satisfação do cliente em relação à distribuidora de energia.

#### **5. REFERÊNCIAS BIBLIOGRÁFICAS**

- ABEP – Associação Brasileira de Empresas de Pesquisa (2004, dezembro 22). Critério de Classificação Econômica Brasil. São Paulo. Acesso em <http://www.anep.org.br>.
- DE SMITH, M. J., GOODCHILD, M. F. & LONGLEY, P. Geospatial analysis: a comprehensive guide to principles, techniques and software tools, Matador: Leicester, United Kingdom, 2007.
- FOTHERINGHAM, A. S., BRUNSDON, C. & CHARLTON, M. Geographically Weighted regression: the analysis of spatially varying relationships, Wiley: Hoboken, New Jersey, 2002.
- FRANCISCO, E.R., Relação entre o Consumo de Energia Elétrica, a Renda e a Caracterização Econômica de Famílias de Baixa Renda do Município de São Paulo. Tese de Mestrado, Escola de Administração de Empresas de São Paulo, Fundação Getúlio Vargas, São Paulo, Brasil, 2006.
- IBGE – Instituto Brasileiro de Geografia e Estatística. Censo Demográfico 2000. Documentação dos Microdados da Amostra. Rio de Janeiro, 2003.
- IBGE – Instituto Brasileiro de Geografia e Estatística (2005). Pesquisa Nacional por Amostra de Domicílios 2004. Rio de Janeiro, Brasil: Author.