

Big data e geointeligência sistêmica: inovação na otimização de cenários para *utilities*ⁱ

Eduardo de Rezende Francisco
Doutor em Administração pela FGV.
Professor do Mestrado em Comportamento do Consumidor da ESPM-SP
E-mails: eduardo.francisco@espm.br.

André Insardi
Tecnólogo pela FATEC Praia Grande
MBA em Gestão Empresarial pela FGV
E-mail: andre.insardi@meiabandeirada.com.br.

Rubens de Almeida
Bacharel em Engenharia Civil pelo Mackenzie
e em Jornalismo pela PUC-SP.
E-mail: ra@gisbi.com.br.

Ricardo Maciel Gazoni
Engenheiro Químico pela USP
Doutorando em Tec. da Intelig. e Design Digital pela PUC-SP.
E-mail: gazoni@semiotic.com.br.

Fábio Franco da Silva
Bacharel em Administração pelo Mackenzie,
MBA em Finanças pelo INSPER.
E-mail: fabiofcos@yahoo.com.br.

Recebido: 02 mai. 2017

Aprovado: 29 out. 2017

Resumo: Esta pesquisa estuda a influência de variáveis geográficas e de modelos de otimização nos KPIs da operação da AES Eletropaulo no atendimento de emergência. Os resultados permitem que a realização de avaliações suporte melhores decisões para as empresas de serviços públicos.

Palavras-chave: *Big data*. Geointeligência. *Utilities*.

Abstract: This research studies the influence of geographic variables and optimization models in the KPIs of the AES Eletropaulo operation in emergency care. The results enable the performance of evaluations to support better decisions for utility companies.

Keywords: Big data. Geointelligence. Utilities.

Resumen: Esta investigación estudia la influencia de variables geográficas y de modelos de optimización en los KPI de la operación de AES Eletropaulo en la atención de emergencia. Los resultados permiten que la realización de evaluaciones soporte mejores decisiones para las empresas de servicios públicos.

Palabras clave: Big data. Geointeligencia. Utilidades.

Introdução

O presente trabalho, resultado do projeto de P&D ANEEL nº 0390-1076/2014 – “Estudo para Proposição e Análise de Cenários de Composição de Bases Operacionais através de Plataforma de GeoInteligência Sistêmica” estuda a influência de variáveis geográficas nos indicadores-chave de desempenho (KPIs – *Key Performance Indicators*) da área operacional da AES Eletropaulo no atendimento de chamadas de emergência na região de abrangência de sua concessão de distribuição de energia na Região Metropolitana de São Paulo.

O foco do estudo é identificar a influência da localização das bases operacionais na operação de atendimento de emergências e verificar as consequências de eventuais alterações de endereço dessas bases ocasionam em seus indicadores de desempenho. É importante destacar que a localização atual das bases é consequência natural da própria evolução histórica da prestação de serviços de distribuição de energia na cidade de São Paulo, e, evidentemente, não levou em conta a situação geográfica metropolitana hodierna, mas o cruzamento das oportunidades de existir patrimônio imobiliário (terreno) disponível e a sensibilidade da área de planejamento da operação.

As análises e conclusões deste trabalho levaram em conta a extensa base de dados disponibilizada pela AES Eletropaulo e contaram com a visão dos profissionais da concessionária na orientação quanto ao modus operandi da empresa. Foram criados modelos estatísticos de predição, avaliação, análise e validação, que serviram de base para a construção de um *software* que estará à disposição da AES Eletropaulo para futuras consultas e suporte a decisões estratégicas. A inovação que após a disponibilização desse *software*, a área de planejamento da concessionária conta agora com uma inédita visão geoespacial das variáveis relevantes na área de concessão, para verificar as consequências das alterações urbanas no desempenho de suas equipes.

As conclusões mais surpreendentes do trabalho estão na influência decisiva das variáveis de tráfego na performance operacional e a relação entre a localização geográfica das bases e a malha viária. Em particular, notam-se possíveis ganhos imediatos na significativa melhora na performance do tempo de atendimento da primeira ocorrência do dia, que é a ocorrência impactada pela localização da base, para cada equipe, através da simples redefinição da área de cobertura de cada base atual.

A longo prazo, conforme o *software* for utilizado pelas equipes de planejamento, a ferramenta desenvolvida possibilitará a drástica adequação do número de bases operacionais, tornando disponível importante ativo imobilizado da empresa, podendo, inclusive, modificar procedimentos logísticos de posicionamento de veículos e equipamentos da companhia.

Como a dinâmica urbana das cidades envolve fenômenos evolutivos – perfil de ocupação dos bairros, volume de novos empreendimentos imobiliários, aumento do número de veículos, abertura de novas vias, entre outras – o *software* desenvolvido é capaz de incorporar em seus cálculos outras grandezas que representem a complexidade urbana e pode, portanto, ser utilizado para, ao longo de tempo, verificar se tais mudanças são capazes de influenciar a fluidez do tráfego, gerando dificuldades não previstas para o cumprimento dos tempos de atendimento.

Os resultados alcançados permitem também verificar se o número de centrais de atendimento e o seu posicionamento na malha urbana são os mais adequados em cada momento da evolução das características urbanas da região Metropolitana de São Paulo. Tal equação é complexa o suficiente para sugerir que as conclusões sejam expostas não só pela simples identificação dos melhores endereços para posicionar as Bases de Atendimento, mas pela geração de estimativas dos tempos de atendimento, de modo a oferecer elementos para a tomada de decisão e em que momento a concessionária deverá rever esse posicionamento, em função de eventuais alterações.

Para que isso seja possível, todo o trabalho de análise de variáveis, desenvolvimento de critérios de avaliação e de exposição de resultados utiliza uma plataforma de mapas digitais, introduzindo os conceitos de inteligência geográfica e Estatística Espacial no processo de decisão. Os mapas digitais ajudam os analistas a compreenderem melhor a evolução das mudanças nas condições operacionais no espaço urbano e oferecem a possibilidade de realizar rapidamente cálculos espaciais complexos, que consideram situações de arruamentos e vias com diferentes capacidades de tráfego, velocidades de deslocamentos em diversas condições de horários e dias de semana. Combinados com alterações climáticas, densidade de arborização local e valor imobiliário das regiões, o *software* utiliza algoritmos geográficos que indicam, a cada situação, as melhores localizações para se posicionar as bases operacionais, hoje e no futuro.

O *software* desenvolvido permite que a realização de avaliações e tomada de decisão para reposicionamento das bases operacionais tornem-se imediatas e

corriqueiras, desde que sejam mantidos os acessos às bases de dados de acompanhamento do trânsito, clima, características de arborização e valor imobiliário utilizadas para a construção do modelo geográfico-digital de análise.

Desenvolvimento da pesquisa

Em linhas gerais, o objetivo e o produto do projeto estão destacados na Figura 1, a seguir.

A integração de KPIs existentes e tradicionalmente acompanhados pela AES Eletropaulo, bem como a adaptação e eventual criação de novos indicadores, ligados ao desempenho imobiliário e suas idiossincrasias, é parte intrínseca deste trabalho. A utilização dessa plataforma destinar-se-á à análise de Modelos Preditivos dos KPIs em um modelo flexível e passível de inserção dinâmica e contínua de variáveis de acompanhamento do crescimento da demanda de ativos e da necessidade de bases operacionais que satisfaçam as diversas condições definidas nas regras da distribuição de energia.

Objetivo	Desenvolver uma metodologia com base em modelos estatísticos, pesquisa operacional e técnicas de análise de dados georreferenciados que permita propor uma configuração otimizada dos bens da AES sem interferir no desempenho operacional de serviços técnicos e comerciais em campo
Produto do Projeto	Plataforma Analítica dotada de ferramental de exploração de vários cenários compostos pelas possíveis estruturas de bases operacionais (quantidade/local/dimensão) e da oferta de serviços em campo, que permitirá a concessionária replicar as análises realizadas sistematicamente otimizando a performance geográfica da localização de suas bases operacionais

Tabela 1. Objetivos e Produto do Projeto

1. Metodologia

A Figura 1 apresenta as macro-etapas envolvidas no desenvolvimento da metodologia. Os subtópicos seguintes descrevem cada etapa em maior nível de detalhe.

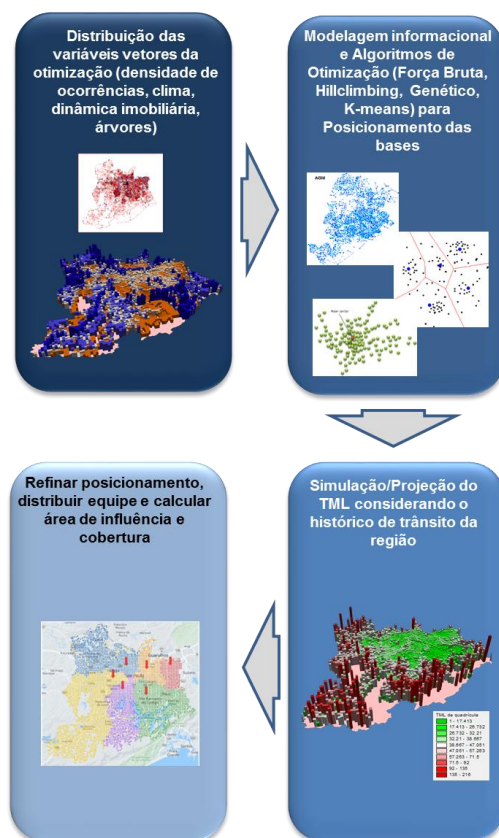


Figura 1. Etapas do Desenvolvimento da Metodologia de Big Data e Geointeligência Sistêmica

1.1 Levantamento de variáveis, análise e aproveitamento dos dados

Todo o desenvolvimento de ideias e soluções vinculadas partiu da disponibilização, pela AES Eletropaulo, de uma base de dados operacionais com mais de 1,2 milhão registros de atendimentos de emergência efetuados pelas equipes da Eletropaulo em sua região de atendimento, entre o período de dezembro de 2012 a novembro de 2015, fornecidos à equipe de P&D em tabelas de dados que configuraram um total de 12 milhões de dados (variáveis e observações) a serem analisados, com o seguinte detalhamento: Localidade da emergência, Hora de registro da solicitação de atendimento, Horário de saída da equipe, Horário de chegada da equipe ao local de atendimento, Equipe responsável, Equipamento, e Tipo de ocorrência.

A esses dados foram incorporados diversas outros processos de atendimento, em uma fase inicial denominada “Prova de Conceito”, em que ocorrências de outros serviços, em especial Leitura e Entrega, Inspeção de Fraude e Notas Técnicas, foram analisadas.

Os dados registrados das ocorrências em campo, em especial os horários de registro, saída e chegada, foram comparados com dados de trânsito consistentes da base externa GisBI/Here. A verificação da consistência dos dados coletados/ anotados pelas

equipes de campo foi então realizada pela leitura dos tempos de deslocamento de veículos (leves e pesados) na região urbana, consolidadas por sistemas de acompanhamento de trânsito na região metropolitana de São Paulo. O trabalho usou como dados referenciais as informações processadas e trabalhadas pela GisBI a partir dos dados da Here North America, LLC, em seu serviço denominado “Traffic Patterns”, originados do acompanhamento sistemático de veículos equipados com GPS e outros tipos de controles de deslocamentos, cuja licença de uso durante o período de desenvolvimento do P&D foi integrada aos valores contratados.

A base de trânsito utilizada neste projeto apresenta histórico mínimo de 3 anos, constantemente atualizados (de 15 em 15 minutos), como insumo para a identificação (e atualização) do trânsito típico horário de cada via (de e para), com médias horárias para cada dia da semana típico configuram 13.171 padrões diferentes de deslocamento médio horário por dia da semana e categorias de vias funcionais

Para a comparação preliminar, foram considerados os deslocamentos entre as bases operacionais e as primeiras ocorrências registradas para cada equipe num período de um ano (de dezembro de 2012 a novembro de 2013), em um total de sucesso em 111.912 ocorrências, sendo que para cada comparação calculou-se por dois critérios (caminho mais curto e caminho mais rápido), disponíveis na base de dados da GisBI/Here.

1.2 Enriquecimento das variáveis e modelagem conceitual

Com base nas conclusões dos debates e definições metodológicas da etapa anterior, e para tornar possível a agregação das diversas variáveis possíveis ao modelo a ser proposto para a localização das bases operacionais, a equipe passou a se dedicar ao detalhamento e enriquecimento das bases de dados, trabalho que exigiu uma dedicação especial, para que todas as bases a serem integradas na metodologia guardassem coerência geográfica espacial com os dados originais e, ao mesmo tempo, permitissem a sua integração matemática de diferentes grandezas, descritas convenientemente na etapa do efetivo desenvolvimento do modelo computacional de análise geoespacial.

A partir de estudos anteriores ([1], [2]), verificou-se a viabilidade de concentração de grande volume de dados em quadrículas como uma possível simplificação da distribuição geográfica de bases de dados de diferentes grandezas e significados.

Com essa possibilidade de simplificação em tela, os pesquisadores consideraram,

no entanto, que o problema de localização das bases operacionais da AES Eletropaulo na região de concessão definida, envolvia não só a questão da distribuição das variáveis pelo território (e sua geografia), mas pela observação dos tempos de deslocamento pelas vias (ruas e avenidas) reais, caracterizadas pelos dados do trânsito na região, advindos da base da GisBI/Here. Ou seja, exigia a caracterização de cada uma das (e todas) quadrículas como origem e destino de deslocamentos a partir de todas as demais quadrículas.

Essa conjugação de informações e necessária iteratividade dos dados de diferentes origens e métricas, perceberam, tornava a resolução do problema como um exemplo de "NP – Completo", que na teoria da complexidade computacional significa um nível de dificuldade especial para solucionar o problema em um tempo polinomial.

Isso motivou a iniciativa do grupo em buscar representações agregadas e algoritmos computáveis para solucionar a busca do posicionamento ótimo das bases, dentro da malha de quadrículas, de modo a gerar condições para a tomada de decisão em condições operacionais, pelas equipes de planejamento da AES Eletropaulo.

1.3 Desenvolvimento do algoritmo de análise espacial

O título adotado para este P&D propõe a expressão “geointeligência sistêmica” (Estudo para Proposição e Análise de Cenários de Composição de Bases Operacionais através de Plataforma de Geointeligência Sistêmica) como uma inovação do processo de análise e planejamento operacional dentro de uma concessionária de distribuição de energia [3].

A inovação não visa apenas modificar e/ou incrementar o instrumental teórico e prático dos profissionais envolvidos nessas tarefas dentro das companhias de distribuição de energia com apenas mais uma tecnologia geográfica. Mas transformar toda a complexidade das análises espaciais em um instrumento simples o suficiente para ser adotado na prática cotidiana do planejamento operacional desse tipo de serviço público, incorporando as possibilidades analíticas das plataformas geográficas, tornadas acessíveis muito recentemente, graças ao rápido desenvolvimento dos mapas digitais e a sua disseminação de uso por todas as pessoas, notadamente por meio de aplicativos vinculados ao planejamento de rotas no trânsito das cidades, entre outros ([4], [5], [6], [3]).

As companhias que já utilizavam mapas e recursos gráficos em suas análises estratégicas mais sofisticadas têm a chance, agora, de desenvolverem novas rotinas de

planejamento a partir da utilização intensiva de raciocínios espaciais em plataformas simplificadas, que não exijam treinamento específico em *software* GIS. O pressuposto é de que a partir deste trabalho torne-se possível disseminar a cultura geoespacial entre os profissionais de análise e planejamento operacional, com interfaces de fácil manipulação ([7], [8]).

1.4 Condicionantes urbanos

A hipótese inicial era a de que algumas características urbanas precisariam ser consideradas na formulação de um algoritmo de análise matemático-estatístico capaz de reconhecer, a qualquer tempo, a influência das modificações ocorridas na área de concessão, em especial, das condições de:

- Trânsito nas rotas de deslocamentos das viaturas de atendimento a emergências
- Tempos de identificação e acesso às ocorrências
- Padrão de arborização local (interferência no cabeamento aéreo)
- Intensidade dos fenômenos atmosféricos (Chuvas, ventos e descargas elétricas)
- Vetores imobiliários (valor do terreno das Bases Operacionais e tendências de ocupação urbana)

Em termos de trânsito era preciso observar com mais vagar os tempos anotados pelas equipes de campo para os deslocamentos de suas viaturas até os locais das ocorrências. Esse trabalho foi realizado na fase preliminar deste trabalho, quando foi detectada uma pequena parcela de inconsistências a partir de uma amostra de 111.912 deslocamentos de equipes, certamente derivadas do mau preenchimento dos formulários operacionais, o que é plenamente compreensível, embora indesejável para as verificações de desempenho, dadas as condições de emergência e às vezes urgência desses atendimentos.

1.5 Volume de dados versus capacidade de processamento

Após essa limpeza inicial da base de tempos de deslocamentos reais – entre o endereço da base o endereço da ocorrência, o desafio era promover a verificação dos tempos de deslocamentos, nessas mesmas rotas, oferecidos pela plataforma GisBI/Here, cujos dados disponibilizam padrões de trânsito que impactam nos tempos dos trajetos e que, em tese, poderiam ser aferidos de 15 em 15 minutos, nos 7 dias da semana.

Esse trabalho de comparações de ambos os tempos de deslocamento, para cada atendimento de emergência específico, foi realizado na fase preliminar e indicou uma forte correlação, oferecendo tranquilidade para se adotar as bases da GisBI/Here como

referência para a produção do algoritmo final de apoio à decisão e ao planejamento de melhor localização de bases operacionais.

Um dos maiores desafios identificado pela equipe de P&D foi o de encontrar uma maneira de simplificar a questão de geolocalização das eventuais ocorrências, de modo viabilizar que o volume de informações a serem consideradas nos cálculos de tempos de deslocamentos da base operacional até elas, pudessem, a um só tempo respeitar a força dessa correlação, sem, no entanto sobrecarregar o algoritmo a ponto de não permitir o seu processamento por microcomputadores comuns ou, na pior das hipóteses, supercomputadores.

1.6 Quadrículas de representação de endereços

A solução adotada partiu da premissa de que seria necessário reduzir o número de variáveis geográficas na matriz de possibilidades de endereçamentos das ocorrências e das bases, de modo a tornar possível uma solução computacional acessível à equipe de profissionais da concessionária. Ou seja, era preciso verificar a hipótese de representar uma série de endereços reais da cidade a partir de uma divisão aleatória do território da área de concessão em quadrículas de 1.000 metros de lado, convergindo-se todos os endereços daquela quadrícula para o endereço mais próximo de seu centróide ([1], [8], [10]).

Com isso, um número indefinido (e, certamente, enorme) de possibilidades de endereçamento de ocorrências passou a ser controlado, em uma matriz menor de quadrículas, cada uma delas representando uma parcela do território da área de concessão. Os endereços dos centroides, então, tornaram-se origem ou destino de deslocamentos, cujos tempos de trajeto seriam oferecidos pelas rotas de trânsito da GisBI/Here, disponibilizadas e padronizadas a cada 15 minutos do dia e da noite, para os 7 dias da semana na área de concessão.

Essa decisão implicou a verificação de outras possibilidades de proposição de quadrículas menores que 1.000 metros de aresta para tornar o cálculo ainda mais preciso e representativo, mas uma simulação do número de interações matemáticas em cada uma das hipóteses indicou que a quadrícula de 1.000 metros de aresta era uma boa escolha, em função da extensão do território da concessão, para a viabilização dos cálculos pelo algoritmo, já que todos os endereços disponibilizados (de qualquer centróide) deverão ser checados em termos de tempo de deslocamento para todos os demais endereços (demais centroides) [9].

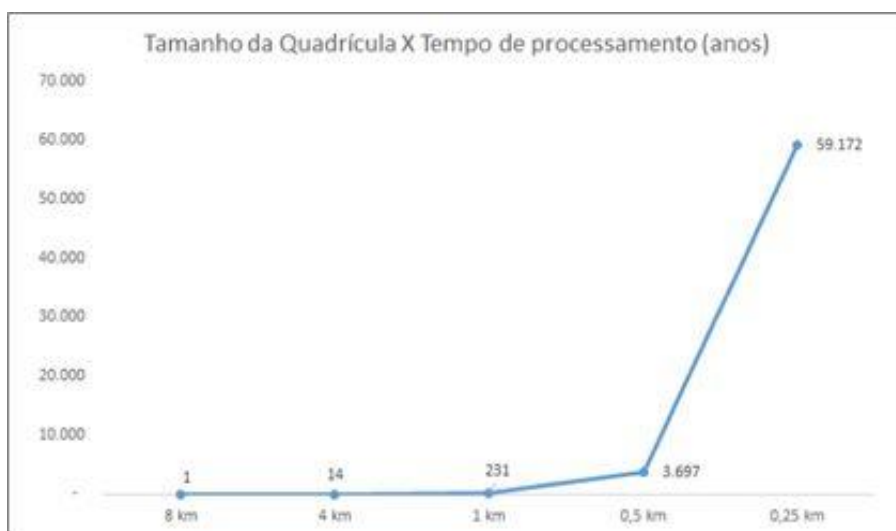


Gráfico 1 – Relação entre o Tamanho da Quadrícula e o Tempo de Processamento para a Solução Computacional adotada pelo estudo

O Gráfico 1 apresenta o tempo estimado de processamento – função do número de iterações entre os vários endereços (o mais próximo dos centroides das quadrículas) - conforme o número de quadrículas. O gráfico deixa evidente a limitação física do processamento de uma matriz de dados do tipo “todos os endereços para todos os endereços”. E demonstra que a simplificação para quadrículas de 1.000 metros de aresta com endereços representativos mais próximos dos centroides das quadrículas apresenta-se no limite da viabilidade computacional para a região de concessão da AES Eletropaulo.

1.7 Matriz de custo versus tempo

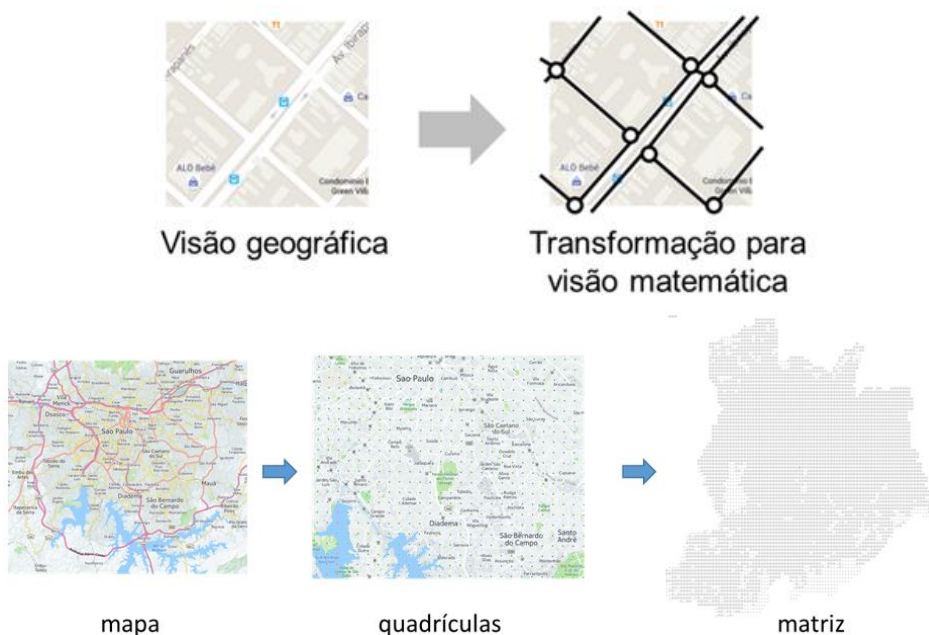
Após a simplificação da matriz de endereços, era preciso verificar o comportamento dos dados disponíveis nos vários bancos de dados escolhidos para o desenvolvimento do trabalho. Cada uma das quadrículas, representadas pelo endereço mais próximo de seu centroide recebeu, então, todas informações geolocalizadas pertencentes àquela quadrícula.

Assim, foram transferidas para as quadrículas, as bases de dados de: ocorrências (AES Eletropaulo); valor imobiliário; clima e arborização (característica retirada da própria descrição das ocorrências AES Eletropaulo, por falta de cobertura territorial dessas informações das prefeituras). E, obviamente, os tempos de deslocamentos entre as quadrículas, duas a duas (trânsito pela base GisBI/Here), concentrando-as no endereço próximo ao centroide.

Qualquer tipo de informação que possa ser expressa por suas características geográficas de localização poderia ser atribuída às quadriculas da mesma forma, o que dá a esta investigação realizada pela equipe de P&D um caráter universal, certamente capaz de ser reproduzido em outras análises logísticas ou não, em qualquer recorte do território, em qualquer cidade brasileiras ou do mundo. A combinação dos diversos fatores (bases de dados) inseridos na quadrícula, ponderados conforme o interesse da análise, recebeu, neste trabalho, a nomenclatura de “matriz de custo” do sistema analisado, cujos valores podem ser avaliados a partir dos KPIs de referência.

Nas análises apresentadas a seguir, em função das exigências de performance da ANEEL, essa matriz de custo oferecerá sempre, como saída, os tempos de deslocamento entre cada quadrícula em particular e as quadriculas que representam a localização das bases operacionais mais bem colocadas sobre o território, a partir do quadro de ocorrências ponderado pelas variáveis da matriz de custo do sistema. A interpretação do território como quadriculas permite uma modelagem matemática computacional e considera, portanto:

- A adequação do posicionamento das bases operacionais da AES depende, em razoável grau, e entre outros fatores, do tempo de deslocamento entre as bases e as ocorrências.
- A malha viária é tratada como um grafo composto de nós e vértices que os ligam, conforme Figura 2 a seguir.
- Para viabilizar a computação dos resultados em equipamentos comerciais hodiernos foram feitas algumas simplificações, como a divisão da área em quadriculas de 1 km de lado.



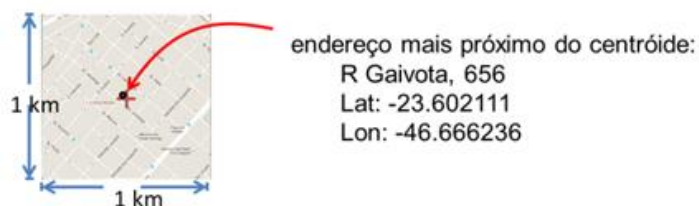


Figura 2 – Transformação da Visão Geográfica em Visão Matemática (representação matricial) para viabilização computacional dos algoritmos desenvolvidos no projeto

1.8 Cálculo da melhor localização das bases

A simplificação introduzida pelas quadrículas permitiu a compreensão dos processos de cálculo, mas não modificou a natureza do problema computacional a resolver. Ou seja, para encontrar a melhor localização das bases seriam necessários cálculos matemáticos intratáveis computacionalmente pelo seu volume de interações entre todas as quadrículas.

Para superar essa dificuldade, a modelagem teve que incorporar diferentes algoritmos de cálculos por aproximação, conhecidos para a solução de problemas NP-completos. Tais algoritmos foram, então, utilizados para se definir o melhor posicionamento das bases, quando o volume de cálculos tornava-se intratável computacionalmente. Para esses cálculos por aproximação, foram aplicados os algoritmos Força Bruta (tentativa de solução do problema NP-completo), Hillclimbing, Genético e K-means, explicados adiante.

Para os dados de tráfego utilizou-se a API da Here, que adota um padrão de tráfego – resultado de estudos que levam em conta dados históricos de alguns anos – e fornece o tempo e a distância entre dois pontos da malha viária na região de concessão, no caso, dos endereços mais próximos aos centroides de cada quadrícula.

Como o levantamento e preparação dos dados de tráfego toma vários dias de processamento, para as simulações apresentadas a seguir extraiu-se dados de tráfego referentes ao dia 15 de Dezembro de 2015 (uma terça-feira), no horário de 8:30 da manhã.

1.9 Problema geográfico transformado em matricial

O problema geográfico de determinar a melhor posição para as bases foi, então, convertido no problema matemático de determinar quais posições ocasionam o menor

custo ponderado de localização das bases numa matriz em que cada célula representa uma quadrícula, conforme apresentado na Figura 2, anterior.

Para o cálculo da melhor localização, levou-se em conta uma matriz de custos que determina o custo (tempo) de deslocamento entre cada origem e cada destino possível, e uma ponderação da quadrícula que determina o peso dessa quadrícula enquanto destino de deslocamento.

Assim, o algoritmo pode ser utilizado para determinar a melhor posição das bases de acordo com qualquer critério. Por exemplo, é possível adotar como matriz de custo o tempo de deslocamento entre as quadrículas e como ponderação o número de ocorrências de cada quadrícula.

Dessa forma, foi possível determinar a localização das bases operacionais que permitem o menor tempo médio de deslocamento até as ocorrências, embora essa abordagem matemática/matricial permita calcular o posicionamento de acordo com qualquer outro critério que se queira.

2. Desenvolvimento dos algoritmos de otimização

Como já mencionado, o trabalho analisou quatro metodologias de cálculo iterativo, cujas características serão apresentadas na sequência (itens II.3.1 e II.3.2). As metodologias de produzir iterações foram implantadas dentro do algoritmo por um programa em C++ (padrão C++11), a ser processado em microcomputadores comuns.

O cálculo básico da área de concessão, nestas condições, é desenvolvido considerando os custos de deslocamentos entre todas as quadrículas, duas a duas, sendo a quadrícula destinada a receber uma base operacional aquela que apresentar o menor custo na matriz de custo. O custo da configuração é dado pela soma das multiplicações da ponderação de cada quadrícula com o custo de deslocamento da base operacional até ela. O custo médio por ocorrência, em cada condição de localização da base operacional, é o custo total dividido pela ponderação:

$$Custo_{configuracao} = \sum_{i=1}^{N_{bases}} \sum_{j=1}^{M_{quadrículasDaBase_i}} Custo_{ij} * Ponderacao_j$$

$$Custo_{medio} = \frac{\sum_{i=1}^{N_{bases}} \sum_{j=1}^{M_{quadrículasDaBase_i}} Custo_{ij} * Ponderacao_j}{\sum_{k=1}^{P_{quadrículas}} Ponderacao_k}$$

2.1 O método força bruta – implementação (inviável a partir de 3 bases) do problema NP-completo

É o método mais intuitivo: nele são testados os custos de deslocamento a partir de bases operacionais colocadas sobre cada uma das quadrículas, até todas as demais quadrículas.

A base operacional será colocada sobre a quadrícula que apresentar o menor custo total.

- O método é inviável para configurações com muitas quadrículas;
- Por isso, para efeito deste trabalho, fixou-se o número máximo de iterações em 10 milhões, ou seja, pouco mais de uma hora de processamento em microcomputadores de bom desempenho, utilizados no projeto;
- Esse tempo corresponde ao cálculo de posição de duas bases, se levados em conta os dados da área de concessão da AES;
- Pode ser utilizado para aferir os resultados dos demais métodos;
- Pode ser utilizado em áreas menores que a de concessão.

Traz o resultado exato, embora o processamento massivo o torne inviável para o caso em estudo, que possui um número de bases operacionais maior. Para verificar a viabilidade de iterações (e processamento) conforme o número de bases, estimou-se o tempo de processamento a partir do número necessário de iterações considerando-se de uma até 10 bases operacionais simultâneas. O tempo de processamento torna-se inviável computacionalmente (anos de processamento de máquina) a partir da otimização de 4 bases operacionais.

2.2 Os métodos Hillclimbing, Genético e K-means

O quadro da Figura 3 destaca os conceitos, características e limitações de cada método de otimização desenvolvido no projeto.

	Descrição do Algoritmo	Características e Limitações
Hillclimbing	Esse método consiste em iterações de cálculo, considerando as seguintes etapas: É sorteada uma configuração inicial. Uma base por vez é deslocada para cada uma das quadriculas adjacentes a ela. O custo de cada uma das nove configurações possíveis é calculado, e a base é deslocada para a posição em que a configuração tem o menor custo. O procedimento é repetido para todas as bases, sucessivamente, até que nenhum deslocamento de nenhuma base resulte em diminuição de custo.	Resultado depende do sorteio inicial, ou seja, a "semente inicial" para os sorteios das configurações iniciais Eventualmente a posição inicial é tal que se chega a uma posição final que é um "mínimo local", não representando a melhor situação possível. O módulo de cálculo permite determinar o número de sorteios iniciais do método, que é muito rápido e permite que se calculem milhares de tentativas. Não se sabe que número de sorteios traz o melhor resultado, pois este depende da heterogeneidade da matriz de custos e das ponderações. Pode-se afirmar que, em geral, quanto maior o número, melhor o resultado.
Genético	Consiste nas seguintes etapas: 1. É sorteada uma população inicial de configurações. 2. Cada configuração tem seu custo calculado. É escolhido um número de sobreviventes, as de menor custo. 3. Cada sobrevivente gera um número de descendentes igual ao número de bases, da seguinte forma: cada base da configuração é substituída, uma por vez, por uma outra base sorteada. A configuração descendente é comparada com uma lista contendo todas as configurações testadas, para garantir a sua não repetição. 4. A nova população, junto com os sobreviventes da geração anterior, tem seu custo calculado, e os novos sobreviventes são escolhidos. 5. O processo se repete até que se atinja o número de gerações desejado.	Não está limitado aos mínimos locais. A qualidade da solução depende do número de gerações. Não tem garantia de que a solução será ótima ou a melhor de todas.
K-means	1. É sorteada uma configuração inicial. 2. Calcula-se a cobertura de cada base na configuração. 3. Em cada área de cobertura, calcula-se o melhor posicionamento da base. 4. Repete-se os dois passos anteriores (ou seja, não é feito novo sorteio) até que as posições das bases se estabilizem.	Depende do sorteio inicial. A qualidade da solução depende do número de repetições. Não há razões para acreditar que a estabilidade seja atingida no ponto de menor custo.

Figura 3 – Descrição dos Algoritmos implementados no projeto

3. Testes de implementação e desempenho

Testes foram realizados considerando de 1 a 2 bases com a simulação dos 4 algoritmos e verificação da localização das bases ótimas em comparação ao Força Bruta (solução ideal). Para 3 ou mais bases, os testes consistiram na avaliação de desempenho das soluções.

Resultados

Dentre os principais achados, destaca-se que o tráfego tem uma grande influência no posicionamento das bases, de uma maneira não trivial.

Nas condições observadas o método Hillclimbing converge mais rápido para a melhor solução; mas os resultados intermediários não se aproximam consistentemente do melhor resultado. O método Genético exige mais iterações, mas aproxima-se consistentemente do melhor resultado: quanto mais gerações, melhor.

Esses resultados influenciaram na decisão de disponibilizar um quarto método, genético com refinamento, que denominamos “GENÉTICO + HILLCLIMBING”. Esse método realiza o método genético como está previsto, e toma a melhor solução para nela aplicar o método Hillclimbing somente uma vez. A ideia é pegar a melhor geração possível e otimizar-la de um jeito que o método Genético não consegue por si. É preciso levar em conta a performance com todos os dados para avaliar os melhores usos para o algoritmo, em cada caso. O método K-means aparentemente converge para um ponto que não é o de menor custo.

A Figura 4 traz os custos de chegada e de saída de cada quadrícula da área de concessão da AES Eletropaulo (conforme descrito em 2.3). A Figura 5 apresenta um exemplo da configuração final para a simulação de 5 bases. Estima-se a manutenção geral dos indicadores de atendimento atuais (em especial, tempos médios de atendimento e deslocamento) e reduções de custo com a diminuição da área construída total das bases, que chegam a dezenas de milhões de reais.

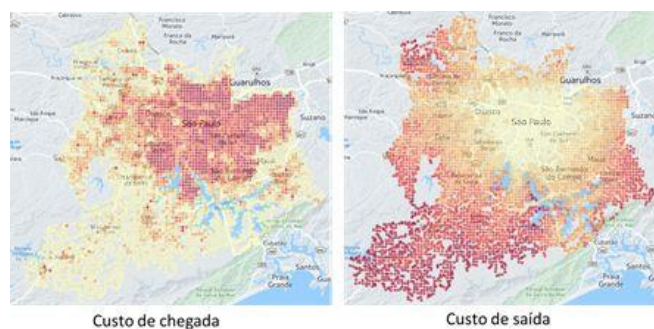


Figura 4 – Custos de Chegada e de Saída das quadrículas da área de concessão da AES Eletropaulo

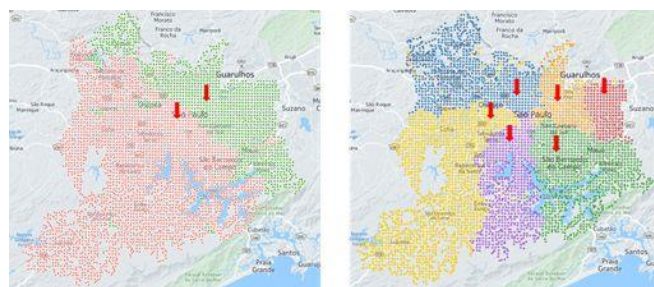


Figura 5 – Exemplo de otimização da composição de 2 (esquerda) e 5 (direita) bases operacionais

Coroando a implementação, a plataforma sistêmica (*software*) foi desenvolvida completamente web-based, para permitir o acesso via browser aos analistas da AES Eletropaulo para a definição das variáveis de simulação (trânsito, dinâmica imobiliária, dados de clima e árvores), definição dos pesos dessas variáveis, seleção do número de bases, método de otimização e parâmetros específicos, simulação dos cenários e aferição dos KPIs.

Considerações finais

A introdução dos conceitos e possibilidades das geotecnologias e da geointeligência em atividades operacionais das empresas concessionárias de serviços públicos pode ser considerada como uma tendência inexorável de avanços metodológicos e da prática empresarial, notadamente da companhia que se dedicam a atividades distribuídas pelo território, seja em áreas urbanas ou rurais.

O projeto desenvolvido no âmbito do P&D demonstra que essa introdução de conceitos, considerações e metodologias de cálculos geoestatísticos ao cotidiano empresarial poderá ser realizada sem que os profissionais especialistas em planejamento e gestão operacional tenham que aprender ou incorporar toda a complexidade e nuances dos *software* de GIS (Geographic Information Systems) disponíveis no mercado.

O sistema desenvolvido (*software*) visa oferecer um instrumento de implementação automatizada da metodologia de cálculo geográfico para a realocação das bases operacionais na AES Eletropaulo, e é o principal fruto do trabalho realizado. O sistema contempla os objetivos de simplicidade para a introdução das variáveis a serem consideradas nas análises e tempos computacionais razoáveis com a possibilidade de estudos em múltiplos cenários de modificação das condições urbanas e climáticas da região de concessão.

O mesmo sistema poderá ser utilizado em outras regiões, desde que sejam disponibilizados e introduzidas variáveis locais, em combinação com os dados operacionais de qualquer companhia de distribuição de energia no país ou mesmo em outros países. Os recursos dos mapas digitais estão integrados à ferramenta desenvolvida e, por isso, os resultados deste P&D não se limitam à área geográfica da concessão da AES Eletropaulo, podendo servir também a outros tipos de concessionárias de serviços públicos.

Para as equipes de planejamento e gestão operacional das concessionárias, os resultados deste P&D oferecem a possibilidade da imediata incorporação das geotecnologias à atividade profissional, sem exigir conhecimentos aprofundados da estatística espacial. Essa possibilidade poderá induzir um novo patamar de desempenho operacional, bem como a comprovação de melhorias sensíveis na prestação de serviços públicos de distribuição de energia, constituindo, por isso mesmo, um grande sucesso metodológico para quem adotar o sistema computacional de realocação de bases operacionais desenvolvido. Trata-se, pois, de um projeto de Big Data com Otimização Geográfica sem precedentes no setor elétrico brasileiro.

Referências

- FRANCISCO, E. R: **Indicadores de renda baseados em consumo de energia elétrica**: abordagens domiciliar e regional na perspectiva da estatística espacial. (Tese de Doutorado), Fundação Getúlio Vargas - EAESP. São Paulo: 2010.
- KIRSCHNER, P: **A study of vision as a measurement of decision-making behavior**: 1971. Thesis (Ph.D.), Dept. of Industrial Engineering, Stanford University.
- LACERDA, L. **Armazenagem estratégica: analisando novos conceitos**. Mar. 2000. On-line. Disponível em <<http://www.ilos.com.br>> Acessado em: 27.nov.2017
- BALLOU, R. H. **Gerenciamento da cadeia de suprimentos/logística empresarial**. 5a Edição. Porto Alegre: Bookman, 2006.
- COELIS, E. L. **Logística empresarial**. Belo Horizonte: 2006. 10p. Instituto de Educação Tecnológica. On-line. Disponível em <http://www.techoje.com.br/site/techoje/categoria/detalhe_artigo/507>
- BARROS, M. C. **Warehouse Management System (WMS): conceitos teóricos e implementação em um centro de distribuição**. Rio de Janeiro: 2005. 132p. Dissertação de Mestrado - Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.
- WANGENHEIM, A. **Análise de agrupamentos**, 2006. On-line. Disponível em <<http://www.inf.ufsc.br/~patrec/agrupamentos.html>>
- LLOYD, C. **Spatial data analysis**. Oxford University Press, 2010.
- WANKE, P. **Aspectos fundamentais do problema de localização de instalações em redes logísticas**. 2001. On-line. Disponível em <http://www.centrodelogistica.org/new/fr-aspec.htm>

ⁱ Trabalho apresentado ao Grupo de Trabalho *Criatividade e Inovação* do **Seminário Tecnologia, Educação e Sociedade**, realizado pela Faculdade Tecnológica [Fatec] de Itaquaquecetuba, SP, nos dias 15 a 17 de março de 2017.