

Artigos

Análise Exploratória dos Dados e Análise Espacial a partir da Geoestatística em Área Contaminada Investigada em Alta Resolução

Exploratory Data Analysis and Geospatial Analysis with Geostatistics after high-resolution site characterization

Guilherme Navarro Diogo Tavares^{1,3,4}; Raphael Martin Salaroli¹; Felipe Augusto Nascimento de Jesus²; Marcos Tanaka Riyis³✉

¹ Universidade de São Paulo (USP), São Paulo, SP, Brasil

² Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG, Brasil

³ Centro Universitário Senac, São Paulo, SP, Brasil

⁴ KOPF Ambiental Engenharia e Consultoria Ltda.

✉ guilherme.navarro.tavares@usp.br, raphael.salaroli@gmail.com, felipe.geo95@gmail.com, marcos.tanaka.r@gmail.com

Palavras-chave:

Análise Espacial;
Alta Resolução;
Investigação Ambiental;
Área Contaminada;
Geoestatística.

Keywords:

Geospatial Analysis;
High Resolution Tools;
Site Investigation;
Contaminated Sites;
Geostatistics.

Revisão por pares.

Recebido em: 12/10/2021.

Aprovado em: 03/03/2022.

Resumo

A massa de contaminantes em fase retida e sua distribuição espacial são variáveis relevantes durante a etapa de remediação ambiental. A Análise Exploratória dos Dados em conjunto com a Geoestatística, abordada através da Krigagem Ordinária (KO) e Radial Basis Function (RBF) foram comparadas às técnicas tradicionais de tratamento de dados oriundos de investigações ambientais por baixa e alta resolução com o objetivo de avaliar a diferença na estimativa de massa de contaminantes, volume de solo impactado e indicar melhorias na estratégia de investigação de áreas contaminadas. Foi verificado que os critérios tradicionais de investigação de compostos orgânicos voláteis em fase retida promovem o enviesamento dos dados e redução da confiança da estimativa de massa de contaminantes e de solo impactado. A substituição do método de interpolação tradicional (ABNT NBR 15515-3) pela geoestatística resultou na redução de massa de contaminantes de 62,73 kg para 28,16 kg (KO) e 8,45 kg (RBF), e do volume impactado de 1423,43 m³ para cerca de 525 m³. Recomenda-se a adoção de técnicas como regularização (composite) e a adição de critérios de amostragem aleatória (simples ou estratificada) ou sistemáticos em combinação aos tradicionais.

Abstract

The contaminant mass in the storage zone and their spatial distribution are relevant variables during the remediation stage. Exploratory Data Analysis in conjunction with geostatistics, approached through Ordinary Kriging (KO) and Radial Basis Function (RBF) were compared to traditional data processing techniques from low- and high-resolution environmental investigations to evaluate the difference in the estimation of contaminant mass, impacted soil volume and indicate improvements in the investigation strategy of contaminated areas. It was verified that the traditional criteria for investigating volatile organic compounds in retained phase promote the bias of the data and reduce the confidence of the estimation of the mass of contaminants and of impacted soil. The replacement of the traditional interpolation method (ABNT NBR 15515-3) by geostatistics resulted in a reduction in the mass of contaminants from 62.73 kg to 28.16 kg (KO) and 8.45 kg (RBF), and in the impacted volume of 1423.43 m³ to about 525 m³. It is recommended the adoption of techniques such as composite and the addition of random (simple or stratified) or systematic sampling criteria in combination with the traditional ones.

DOI: <http://doi.org/10.14295/ras.v36i1.30083>

1. INTRODUÇÃO E OBJETIVOS

O uso de ferramentas de alta resolução (USEPA, 2013) para investigações ambientais é crescente no Estado de São Paulo, seja pelo encorajamento da adoção dessa estratégia pela Decisão de Diretoria nº 038/2017/C da Companhia Ambiental do Estado de São Paulo (CETESB), como pelo *Return on Investigation* (ROI) evidenciado por Suthersan *et al.* (2015).

O ROI foi originalmente avaliado ao otimizar sistemas de bombeamento e tratamento por meio da identificação de zonas de

fluxo, por onde a maior parte da massa de contaminantes era transportada, descrita como a “massa que importa” (SUTHERSAN *et al.*, 2015). Essa abordagem é relevante para controle e mitigação da exposição de receptores à água subterrânea contaminada, notadamente fora da área sob avaliação, ou à adoção de técnicas de medidas de engenharia e de remediação baseadas no fluxo de massa, a exemplo de barreiras hidráulicas e bombeamento e tratamento.

Entretanto, a massa de contaminantes que apresenta relevância é determinada pela Avaliação de Risco à Saúde Humana

(ARSH) e/ou Avaliação de Risco Ecológica (ARE), e para esta devem ser privilegiadas medidas de intervenção que priorizem a remoção de massa, conforme Decreto Estadual 59.263, que regulamentou a Lei Estadual 13.577 (SÃO PAULO, 2013).

A maior parte da massa de contaminantes ocorre retida ou imobilizada ao solo (LERNER *et al.* 2003; CHERRY *et al.*, 2009; DYMENT, 2011). Para remoção dessa massa, as técnicas de remediação comumente requerem o contato entre agente remediador e contaminante ou remoção física e posterior disposição.

Em ambas as situações, torna-se necessário identificar com precisão a distribuição espacial dos contaminantes, considerando o volume impactado, concentração e características físico-químicas do meio tratado (solos ou sedimentos inconsolidados).

Dessa forma, o incremento na quantidade de dados gerados durante investigações ambientais em alta resolução e/ou sob o conceito de *smart characterization*, com o uso de técnicas como Amostragem de Solo de Perfil Completo (ASPC) (ITRC, 2015; RIYIS *et al.*, 2019), *Hydraulic Profiling Tool* (HPT), *Membrane Interface Probe* (MIP), *Optical Image Profiler* (OIP) e *Laser-Induced Fluorescence* (LIF) requerem uma nova abordagem de armazenamento e consumo.

O incremento na quantidade de dados gerados poderá gerar os benefícios esperados se acompanhados de melhorias nas ferramentas de interpretação. Este estudo busca preencher a lacuna do conhecimento quanto a aplicação de técnicas mais refinadas para interpretação de variáveis regionalizadas (concentração de contaminantes em solo) e comparação com estratégias tradicionais (ABNT NBR 15515-3).

Este estudo visa superar as investigações tradicionais direcionadas à água subterrânea ou amostragem de solo restrita à zona não saturada e complementar as investigações direcionadas à análise espacial no conceito de dados ambientais (*Geospatial Analysis for Optimization at Environmental Sites* - ITRC, 2016), caracterização do fluxo de massa de contaminantes e delimitação da massa de contaminantes e, dessa forma, otimizar eventuais intervenções por técnicas de remediação ou escavação e destinação – incrementando, portanto, o ROI.

Para que os dados obtidos em alta resolução sejam adequa-

mente tratados e contribuam para modelos conceituais robustos, a estatística clássica será amparada pela geoestatística, uma vez que se trata de variáveis regionalizadas. As estimativas de massa de contaminantes e distribuição espacial, serão adotados métodos de interpolação consagrados na indústria mineral para declaração de recursos e reservas: *Krigagem Ordinária* e *Radial Basis Function* (RBF).

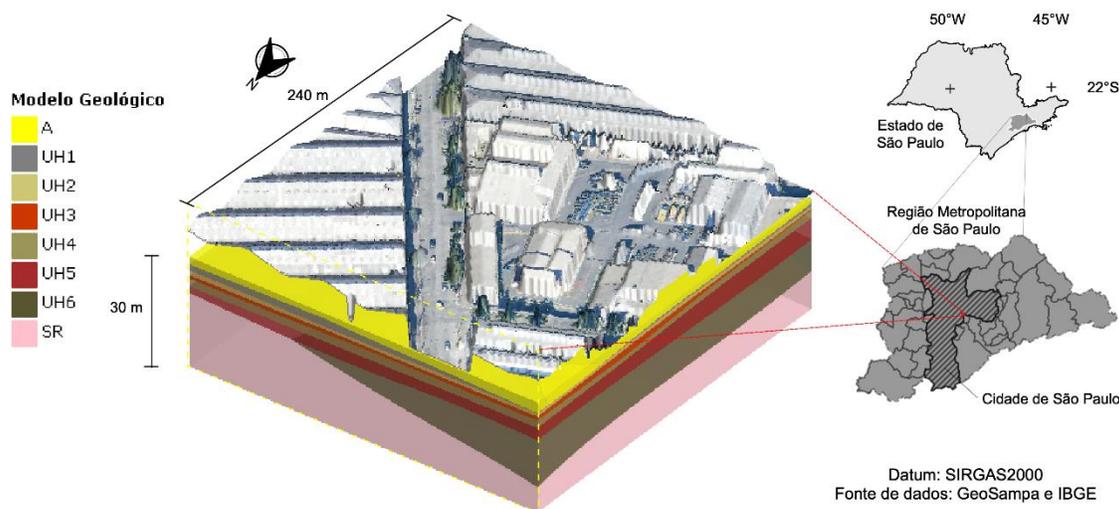
Esta estratégia foi adotada em uma área investigada sob o conceito de alta resolução, contaminada por tetracloroetano (PCE), tricloroetano (TCE), *cis-1,2-dicloroetano* (DCE) e cloreto de vinila (CV) (Solventes Clorados) e de ambiente geológico caracterizado por sedimentos inconsolidados quaternários (Bacia do Rio Tamanduateí), terciários (Bacia de São Paulo) e do solo residual do embasamento cristalino pré-cambriano (Complexo Embu).

2. DESCRIÇÃO DA ÁREA E DA GEOLOGIA LOCAL

O estudo foi conduzido em uma área industrial localizada na cidade de São Paulo (Vila Carioca), cuja coordenada geográfica e endereço não serão disponibilizados pela existência de termo de confidencialidade e sigilo. A área sob avaliação é caracterizada por uma planície aluvial com sedimentos Quaternários (UH1 e UH2) associados ao rio Tamanduateí, sedimentos Terciários da Bacia (UH3 a UH6) de São Paulo e solo residual (pré-Cambriano) (SR) (Figura 1). Essa associação entre áreas industriais e características geológicas configuram notável complexidade, uma vez que associações de argilas, areias e estruturas primárias interferem no transporte e retenção de contaminantes. As unidades hidroestratigráficas (UH) são descritas individualmente na Tabela 6.

A área fonte considerada para o atual estudo é restrita, cerca de 15x20m (300,00 m²). O nível de água freático é raso, geralmente inferior a 1,0m e se mantém constante durante as diferentes estações do ano. A fonte primária de contaminação é antiga, provavelmente da década de 1970, e se trata de uma antiga caixa de contenção subterrânea, preenchida com aterro constituído por entulhos e material contaminado. A base da caixa não permaneceu estanque e atuou como ponto de infiltração de produto (*dense non-aqueous phase liquid* – DNAPL) diretamente sobre argilas terciárias da Bacia de São Paulo, neste estudo denominadas UH3.

Figura 1 - Mapa de Localização e Modelo Geológico. "A" faz referência ao Aterro, a sigla 'UH' às diferentes unidades hidroestratigráficas e "SR" ao solo residual do embasamento cristalino



3. METODOLOGIA

3.1 Caracterização da área fonte em alta resolução, coleta de amostras e métodos laboratoriais

O conceito de caracterização da área fonte em alta resolução compreende o uso de métodos integrados com alta densidade de amostras de solo e de água subterrânea, com interpretações hidroestratigráficas e mapeamento da permeabilidade em três dimensões. O produto é um modelo conceitual robusto, que auxilia a reduzir os custos atrelados à eventual etapa de remediação (SUTHERSAN *et al.*, 2015).

A seleção das técnicas aplicáveis se baseou em informações obtidas em estudos ambientais anteriores, que reportaram a ocorrência de aterro composto por entulhos e matriz de granulação fina, acondicionados em antiga estrutura subterrânea de 3,5 metros de profundidade, e na geologia local, onde camadas de argilas duras e rijas ocorrem entre 6,5 e 14 metros de profundidade (UH3, UH4 e UH5).

Dessa forma, foi descartado o uso ferramentas como MIP e HPT, que seriam limitados devido à presença de entulho no aterro e consistência das argilas. Foi adotada a técnica de Amostragem de Solo de Perfil Completo (ASPC) (ELLIS, 2014; RIYIS *et al.*, 2019), constituída principalmente pela alta densidade de amostras de solo em meio vadoso e saturado (principalmente em zonas de retenção).

A amostragem de solo (aterro, sedimentos inconsolidados e solo residual) para análises de compostos orgânicos voláteis (VOC)

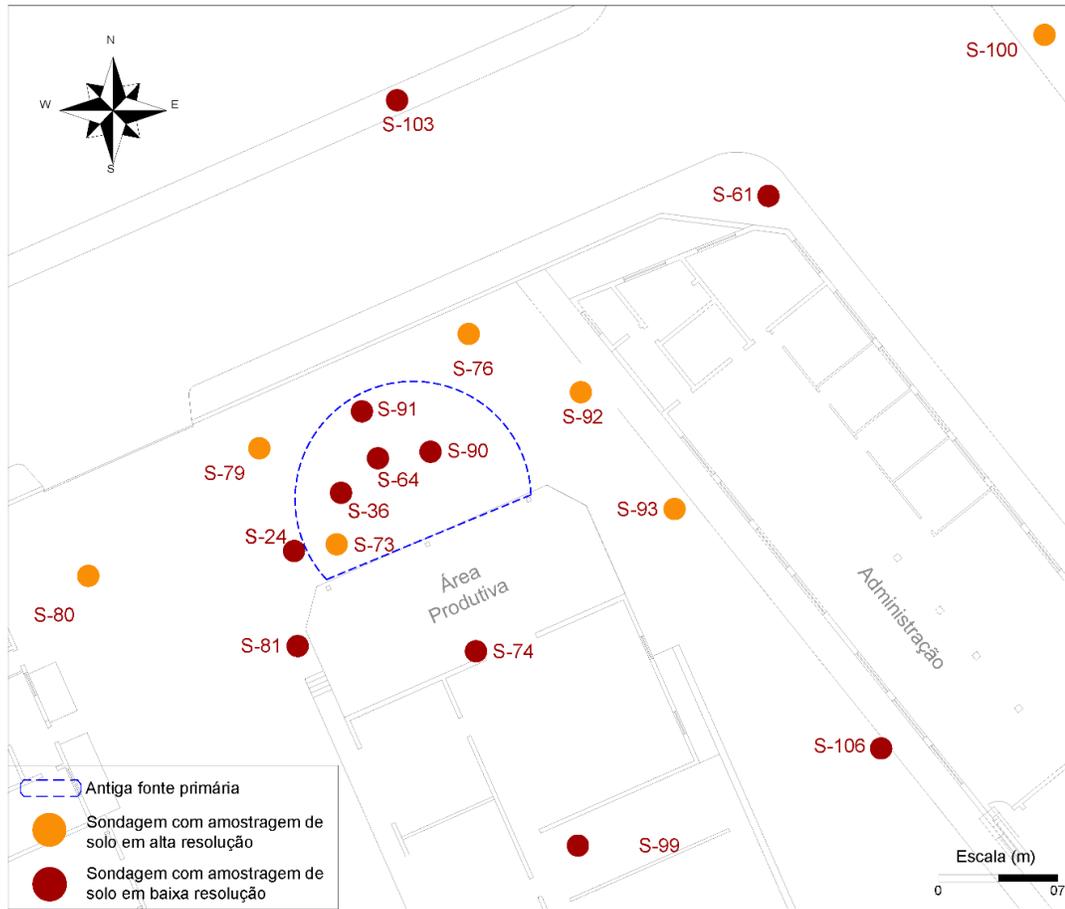
foi executada conforme a norma ABNT NBR 16.434:2015, a partir de cravação contínua (*Direct Push*) de amostradores com *liners* (onde exequível), equipados com retentores de amostra. O *Single Tube* revestido por trados ocas helicoidais ou manuais e o *Dual Tube* foram adotados como modalidade de *Direct Push*, conforme descritos por RIYIS (2019), de acordo com a profundidade alvo e espaço disponível para posicionamento de maquinário.

Nos testemunhos de sondagens foram utilizados detectores por fotoionização (Ion Science PhoCheck Tiger, calibrado) para medições de VOC a cada 10 cm, para guiar a coleta de amostras. Após a abertura do *liner*, foram descritos os litotipos, com enfoque na granulometria, estruturas deposicionais, de contato e mineralógicas.

As porções de interesse para amostragem, como zonas de contato entre granulações mais grosseiras e finas, zonas de retenção, e/ou intervalos com altas concentrações de VOC indicadas pelo detector por fotoionização, foram amostradas com auxílio de seringas descartáveis de 10 ml cortadas na ponta, de forma a transferir a amostra diretamente para um *vial* com metanol e para frasco boca larga para a quantificação de teor de umidade. Após a coleta da amostra, estas foram preservadas em baixa temperatura $-4 \pm 2^\circ\text{C}$. O método analítico empregado (EPA 8260B:1996 - SW846 - PURGE AND TRAP SYSTEM) foi conduzido em laboratórios acreditados (NBR ISO/IEC 17.025) para execução das análises químicas.

A malha de sondagens é apresentada na Figura 2.

Figura 2 - Malha de Sondagens



3.2. Análise Exploratória dos Dados e Análise Espacial por Geoestatística

A análise exploratória dos dados (AED) é uma abordagem que identifica padrões gerais nos dados coletados (USEPA, 2010), e neste caso foi direcionada à investigação de acordo com a Decisão de Diretoria nº 038/2017/C da Companhia Ambiental do Estado de São Paulo (CETESB), conforme a unidade hidroestratigráfica e substância química de interesse.

A geoestatística é direcionada ao estudo da distribuição espacial de variáveis, geralmente numéricas e que respeitam a primeira lei da geografia: “coisas mais próximas se parecem mais que coisas mais distantes” (TOBLER, 1970), úteis às aplicações práticas para avaliação de depósitos minerais (MATHERON, 1963). Entretanto, seu uso é crescente para descrição de outras variáveis regionalizadas, como na avaliação espacial de contaminações em meio hidrogeológico (HERRMANN *et al.*, 2008; LEAL *et al.*, 2017).

Para o entendimento da distribuição espacial, cálculo de massa e volume impactado dos compostos solventes clorados em fase

adsorvida ao solo, os métodos de interpolação por técnica tradicional (ABNT NBR 15515-3) e com base em geoestatística, através da Krigagem Ordinária e *Radial Basis Function* (RBF), foram implementados por meio do *Software Leapfrog Works*. O procedimento adotado para interpolação geoestatística foi o estabelecido por Yamamoto e Landim (2015), cuja técnica foi previamente aplicada em estudos similares (CAMBARDELLA *et al.*, 1994; HERRMANN *et al.*, 2009; LEAL *et al.*, 2017).

A Krigagem Ordinária considera observação estatísticas e através do variograma expressa a continuidade espacial de determinada variável. Pontos próximos aos com valores conhecidos possuem maior influência durante a interpolação.

$$Z_{KO}^*(x_o) = \sum_{i=1}^n \lambda_i Z(x_i)$$

Onde: $Z(x_o)$ é o valor da variável estimada no ponto x_o ; λ_i é o ponderador (peso) a que cada amostra $Z(x_i)$ é atribuída, e apresenta condição de restrição onde a soma dos pesos é igual a 1

A *Radial Basis Function* (RBF) gera superfícies suaves e estima valores em todo o domínio avaliado, conforme a seguinte equação (STEWART et al. 2014):

$$s(X) = \sum_{i=1}^K \omega_i \varphi(|x - x_i|) + \sum_k c_k q_k(x)$$

Onde:

x_i são as localizações espaciais nas quais a interpolação será construída;

w_i são os coeficientes RBF (pesos);

$\varphi_k(x)$ é a função de base radial.

Para o uso dos métodos de Krigagem Ordinária e RBF, foi conduzida a regularização das amostras coletadas (ou composta, ou *composite*), para que estas apresentem o mesmo suporte amostral no *Software Leapfrog Works*. Na indústria mineral o denominado “teor composto” é aplicado por intervalos de trabalho (bancada, intervalo de interesse, zona mineralizada), e neste estudo será adotada de forma a indicar a “concentração composta” do analito de interesse, cujo intervalo de trabalho é o intervalo médio de amostragem na direção vertical, é expresso pela equação abaixo, extraída de KAZUO & LANDIM (2015):

$$t_c = \frac{\sum_{i=1}^n t_i e_i}{\sum_{i=1}^n e_i}$$

Onde: n é o número de intervalos de trabalho (trechos); t_i é o teor do i -ésimo intervalo de trabalho; e_i é a espessura do i -ésimo intervalo de trabalho.

Para os analitos cuja concentração não superou o limite de quantificação, foram adotados os valores deste limite para análise estatística, numa abordagem conservadora da concentração em questão (USEPA, 2000) – entretanto esta abordagem não foi adotada durante a etapa de geoestatística, conforme abordado adiante.

A partir do *composite* fora conduzida a variografia experimental, levando em consideração a tolerância adotada e, a partir disto, determinação da direção que apresenta maior continuidade espacial e variograma melhor estruturado. A visualização em cena dos eixos variográficos no *Leapfrog Works* em conjunto com a disposição espacial dos dados analíticos auxiliou na determinação de maior continuidade dos analitos considerados.

O efeito pepita reconhecido pelo variograma foi apresentado em termos de porcentagem da semivariância total, e conforme critérios originalmente estipulados por Cambardella et al. (1994), pôde-se classificar a dependência espacial como: forte dependência, se a relação for inferior a 25%; dependência espacial moderada, se entre 25 e 75%; fraca dependência, se superior a 75%.

Após a modelagem dos variogramas, a validação cruzada foi adotada para verificar o ajuste dos parâmetros determinados pelo modelo, ajustados no *Software Leapfrog Works*. Esta técnica foi aplicada com auxílio do *Stanford Geostatistical Modeling Software (SGeMS)* e consiste em aplicar a Krigagem Ordinária aos pontos amostrais, excluindo da estimativa o ponto em análise (*leave one out*) (ROSSI et al, 2014).

Os valores estimados são posteriormente comparados com os valores da base de dados com auxílio de gráfico de dispersão entre o valor amostral e o valor estimado na validação cruzada. Desta forma, obtém-se o Coeficiente de Correlação de Pearson (r) entre os valores estimados e os medidos, o qual pode variar de -1 a 1. O sinal indica o tipo de correlação (direta ou inversa) e o valor indica a força desta relação, onde um valor 0 indica ausência de correlação entre os dados e 1 indica correlação perfeita (FIGUEIREDO; SILVA, 2009).

4. RESULTADOS E DISCUSSÃO

4.1. Análise Exploratória dos Dados

A Tabela 1 contém as estatísticas descritivas dos dados originais e dos seus logaritmos. A Figura 3 apresenta a distribuição de frequência das concentrações obtidas em relação ao valor de intervenção, em escala logarítmica, para melhor visualização.

A distribuição de frequência dos contaminantes solventes clorados apresenta elevada assimetria positiva (dados originais e regularizados), e simetria (coeficiente de assimetria próximo a zero) após a transformação logarítmica.

Verifica-se elevada variabilidade dos dados, cujo padrão pode ser explicado pelo controle da distribuição de contaminantes em virtude de estruturas subterrâneas antrópicas presentes no local e controle geológico a partir de camadas com alto potencial de transporte adjacentes a camadas onde predomina a retenção destes compostos, conforme distribuição apresentada na Figura 4 em conjunto com a unidade de medida adotada.

Tabela 1 – Estatística descritiva dos dados

TETRACLOROETENO			
VARIÁVEIS ↓ / →	DADOS ORIGINAIS	DADOS REGULARIZADOS	DADOS REGULARIZADOS (LOG)
Média (mg/kg)	1.31	1.52	0.42
Desvio Padrão	7.91	8.6	2.15
Coefficiente de Variação	6.03	5.64	1.73
Variância	62.66	74.07	4.31
Mínimo (mg/kg)	0.005	0.005	-5.30
Primeiro Quartil (mg/kg)	0.234	0.25	-1.39
Mediana (mg/kg)	0.297	0.3	-1.20
Terceiro Quartil (mg/kg)	0.343	0.35	-1.05
Máximo (mg/kg)	84.05	74.45	4.31
Coefficiente de assimetria	7.22	7.79	0.09
CIS-1,2-DICLOROETENO			
VARIÁVEIS ↓ / →	DADOS ORIGINAIS	DADOS REGULARIZADOS	DADOS REGULARIZADOS (LOG)
Média (mg/kg)	1.12	1.56	0.44
Desvio Padrão	3.71	4.51	1.51
Coefficiente de Variação	3.3	2.88	1.06
Variância	13.82	20.42	3.02
Mínimo (mg/kg)	0.005	0.005	-5.30
Primeiro Quartil (mg/kg)	0.248	0.26	-1.35
Mediana (mg/kg)	0.317	0.31	-1.17
Terceiro Quartil (mg/kg)	0.393	0.39	-0.94
Máximo (mg/kg)	33.43	31.82	3.46
Coefficiente de assimetria	6.59	4.55	-0.01

Visto que a ASPC fora direcionada ao centro de massa de contaminação e subsidiou a etapa de ARSH, foi utilizada a preservação em metanol para as amostras de solo, conforme a norma ABNT NBR 16.434:2015. Dessa forma, à medida que foram detectadas concentrações mais representativas do centro de massa, houve elevação do limite de quantificação dos analitos em questão, conforme previamente observado em NEVES *et al.* (2020).

Para TCE e CV o limite de quantificação foi superior ao limite de intervenção estabelecido na Decisão de Diretoria CETESB Nº 256/2016/E, de 22 de novembro de 2016, em amostras periféricas ao centro de massa (Figura 4). Isso prejudica a verificação da extensão espacial das plumas de contaminação. Por essa razão, esses analitos (TCE e CV) foram excluídos das análises geoespaciais.

Figura 1 - Distribuição de frequência para os analitos considerados

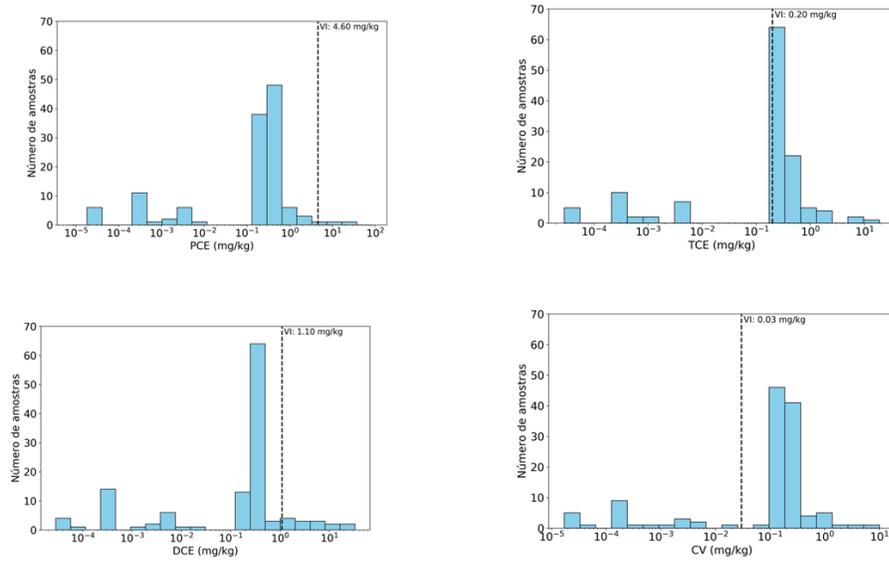
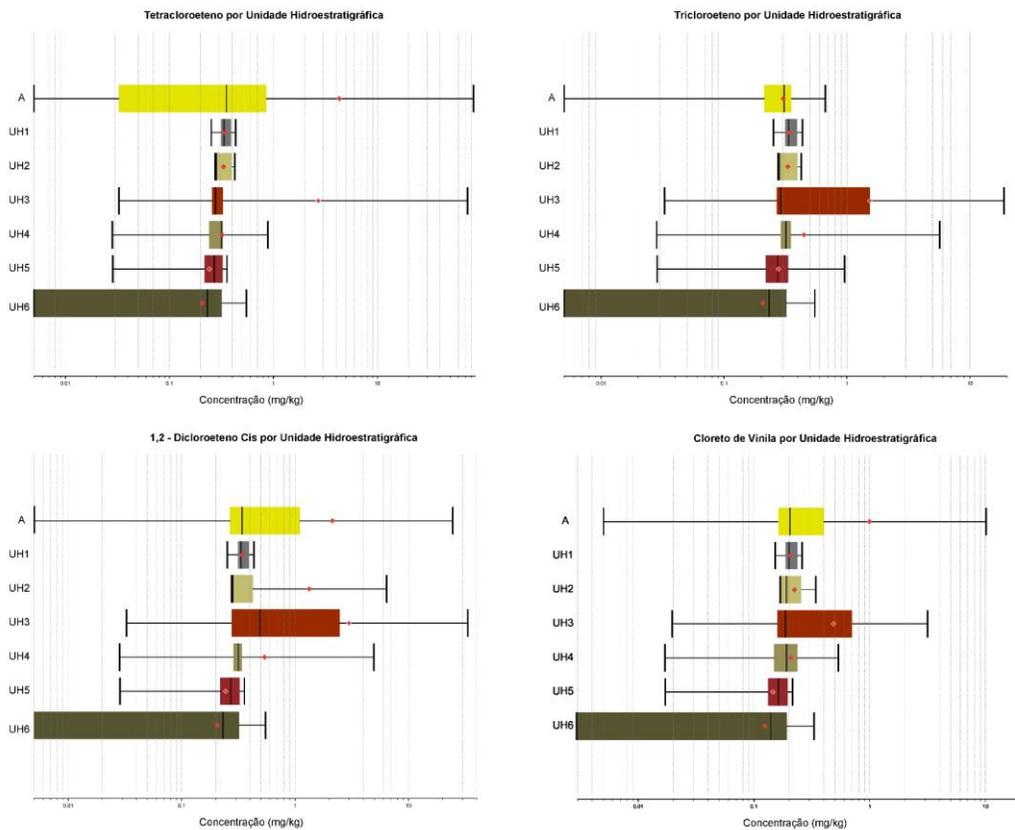


Figura 2 - Boxplot das concentrações dos analitos considerados por unidade hidroestratigráfica



4.2. Preparação das Amostras e Geoestatística

O espaçamento médio vertical entre as amostras coletadas é de 1,12 metro, considerando 07 sondagens com amostragem em alta resolução e 11 em baixa resolução. O desvio padrão refe-

rente ao espaçamento médio entre amostras, prévio a regularização, é de 0,95. O valor máximo é de 5 metros e mínimo de 0,2 metro.

Dessa forma, o valor de 1,12 metro foi adotado para a regularização. Trechos residuais de comprimento inferior a 1,12 metro

foram preenchidos gradualmente conforme as amostras adjacentes – em detrimento do descarte deste trecho ou da adoção dos valores mais próximos.

De acordo com a Figura 5 a composição de amostras aumentou a metragem com resultados de 177 para 183 metros. Observa-se que os valores do histograma de concentração dos contaminantes analisados apresentam uma forte assimetria positiva.

Desta forma, para permitir que os variogramas sejam ajustados, as concentrações foram transformadas com logaritmo natural e, a partir destes, foram realizadas as análises variográficas. Para permitir que os variogramas calculados possuam estruturas para posteriormente serem modelados, foi extraído o logaritmo natural das concentrações.

A variografia experimental para PCE e DCE considerou azimutes espaçados a cada 20°, com tolerâncias angulares de 10°. Neste processo foi determinado que a direção de maior continuidade

espacial é a de 20°/0° para PCE e 150°/0° para DCE, conforme apresentado na Figura 6, Figura 7 e Tabela 2.

Os parâmetros analisados apresentaram dependência espacial distintas, conforme propriedades de seus variogramas (Figura 6 e Figura 7). Para PCE, o efeito pepita expresso como porcentagem da semivariância total foi de 1,9%, para DCE foi de 0%. Dessa forma, ambos os analitos apresentaram forte dependência espacial (Cambardella et al., 1994), considerando a escala de investigação por ASPC.

Nesta modelagem, verifica-se a existência de anisotropia mista para ambos os analitos, ou seja, comportamentos espaciais diferentes para direções diferentes. Desta forma, os eixos maiores dos variogramas corresponderam aos azimutes onde houve maior continuidade espacial, ou seja, maior amplitude, o semi-eixo corresponde ao par ortogonal desta direção e a direção vertical é o eixo menor.

Figura 3 - Espessura dos intervalos antes e após composite

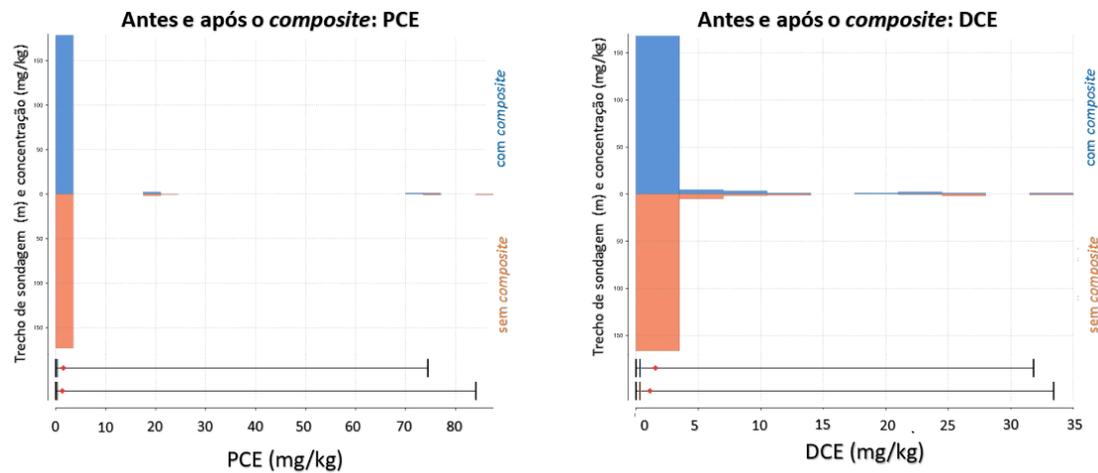


Tabela 2 – Parâmetros para os variogramas e resultado da validação cruzada

Variável	Eixo	Azimute (°)	Mergulho(°)	Pitch(°)	Passo (m)	Tolerância angular(°)	Tolerância distância (m)
PCE	<i>Major</i>	20	0	114	6	10	6
	<i>Semimajor</i>	110	0		1.12	45	1.12
	<i>Minor</i>	0	90		6	10	6
DCE	<i>Major</i>	60	0	17	6	10	6
	<i>Semimajor</i>	150	0		1.12	45	1.12
	<i>Minor</i>	0	90				

Variável	Modelo	Semivariância		Range	Mínimo	Dados Condicionantes	
		Efeito Pepita	Total			Máximo	Coefficiente de Correlação
PCE (1ª estrutura)	Gaussiano	0.05	2.6	23.97	4	8	0.82
				12.97			
PCE (2ª estrutura)				5.75			
				23.97			
DCE (1ª estrutura)	Esférico	0	3.21	20	4	8	0.84
				10			
DCE (2ª estrutura)				10			
				10			

Figura 6 - Variogramas para tetracloroetano

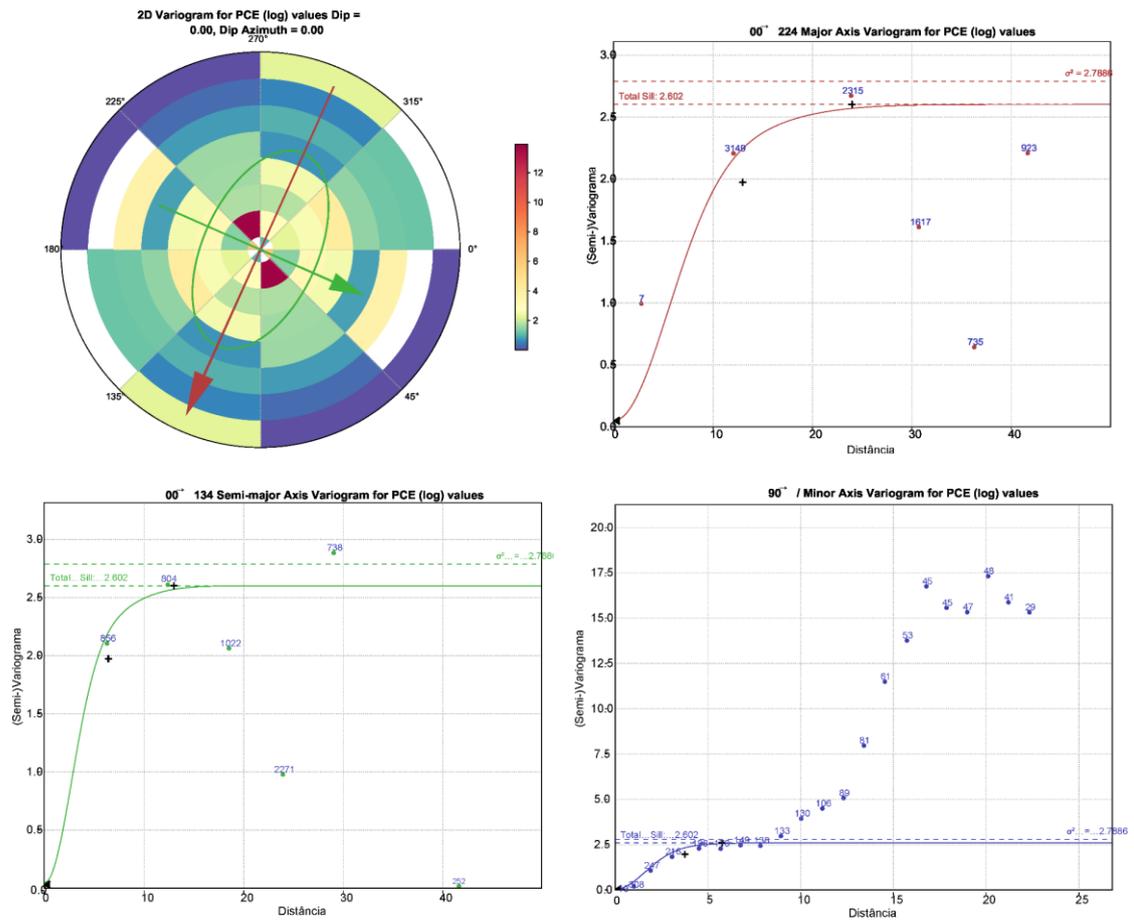
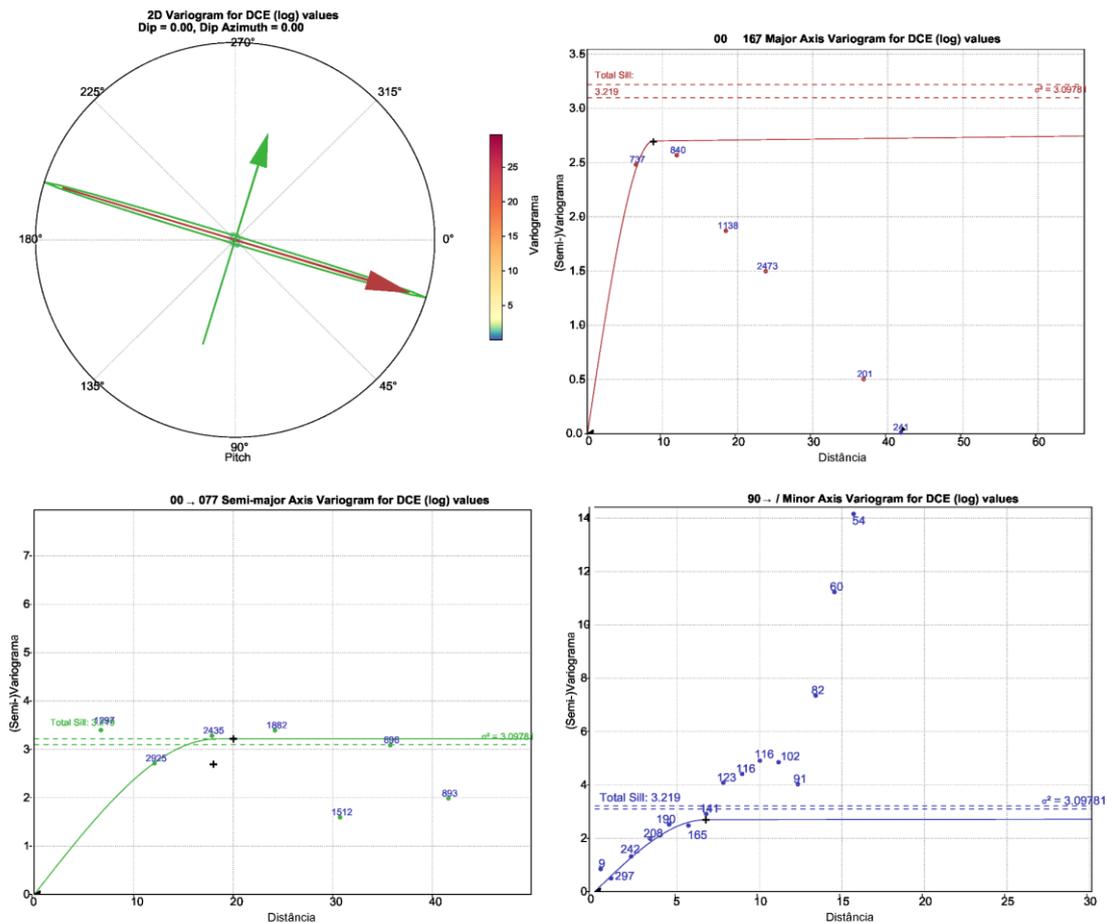


Figura 4 - Variogramas para cis-1,2-dicloroetano



4.3. Krigagem Ordinária, Radial Basis Function e Validação Cruzada

A interpolação através dos métodos de Krigagem Ordinária e *Radial Basis Function* foram conduzidas com os parâmetros de vizinhança apresentados na Tabela 2. As dimensões do modelo de bloco em relação ao domínio adotado são apresentadas na Tabela 3. As dimensões dos blocos são referentes à

metade do espaçamento médio entre sondagens com amostras de solo (eixo X e Y) e metade do espaçamento médio vertical entre as amostras (Z). Esses dados foram estimados no campo logarítmico e após as estimativas eles foram trazidos novamente ao campo do suporte amostral ao fazer o exponencial deles, conforme equação abaixo.

$Y = e^x$, onde x corresponde ao logaritmo natural dos valores estimados

Tabela 3 – Dimensões do Modelo de Blocos

Direção	Dimensões do Bloco (metro)	Número de Blocos
X	2.96	23
Y	2.96	23
Z	0.57	53

As estatísticas descritivas dos resultados de Krigagem Ordinária e *Radial Basis Function* estão apresentadas na Tabela 4 em escala real. Verificou-se a gradual suavização dos resultados em termos da estatística descritiva entre as etapas que consideram unicamente os resultados medidos, os regularizados e os estimados via Krigagem Ordinária e RBF.

A dispersão absoluta dos dados foi reduzida, conforme decrés-

cimo do desvio padrão e coeficiente de variação de ambos os analitos considerados. Este decréscimo é incrementado para o RBF em virtude de o comportamento de extrapolação ser direcionado à média global do domínio. Os valores máximos foram reduzidos de 84,05 mg/kg (medido) para 78,34 mg/kg (Krigagem Ordinária) para tetracloroetano e de 33,43 mg/kg (medido) para 19,23 mg/kg (Krigagem Ordinária) para cis-1,2-dicloroetano.

A Figura 8 apresenta no modelo de blocos os resultados a partir da interpolação por Krigagem Ordinária e *Radial Basis Function* em conjunto com as amostras, consideradas nesta interpolação (após *composite*).

A fim de validar o ajuste do modelo de variograma foi realizado a validação cruzada com auxílio do software *SGeMs*, cujos detalhes estão apresentados na Tabela 2. A Figura 9 mostra a correlação entre os valores estimados com os variogramas modelados com os valores medidos, provenientes das análises químicas.

Notar que os logaritmos naturais destes valores estão sendo empregados.

Os variogramas em conjunto com os parâmetros de vizinhança adotados, foram avaliados pela validação cruzada (Figura 9). Foi verificado elevada correlação entre dados medidos e estimados, com as retas ajustadas ao gráfico de dispersão e valor de 'r' próximo a 1, respectivamente 0,81 e 0,85 para tetracloroeteno e cis-1,2-dicloroeteno.

Figura 8 - Representação no modelo de blocos da interpolação de tetracloroeteno e cis-1,2-dicloroeteno (em conjunto) por Krigagem Ordinária (imagem superior) e *Radial Basis Function* (imagem inferior)

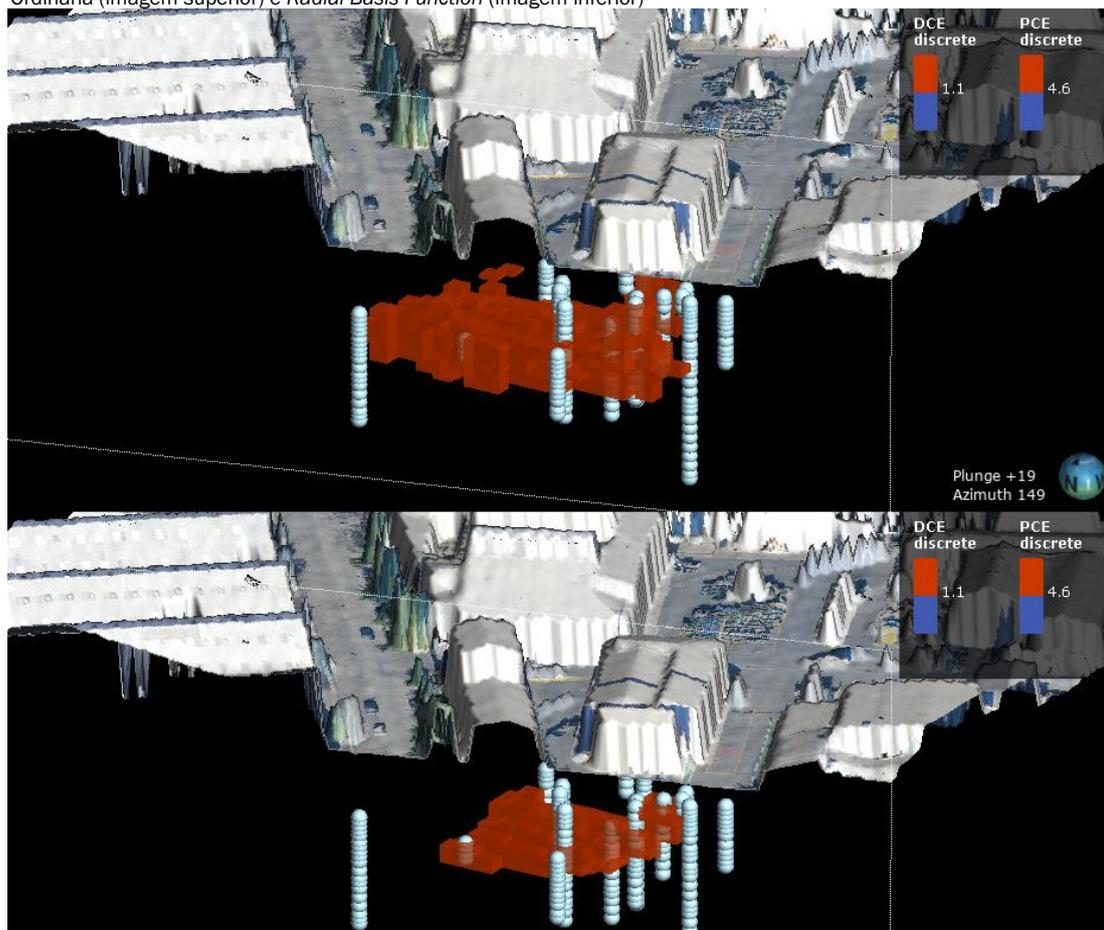
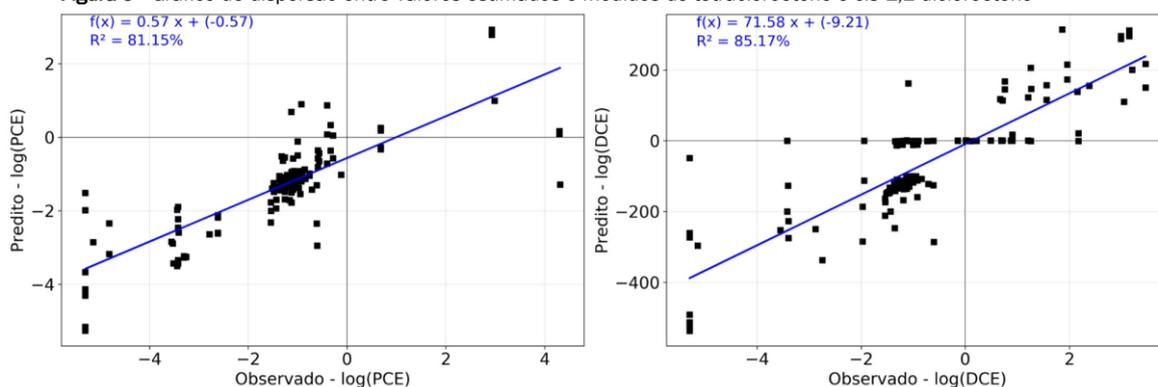


Tabela 4 - Estatística descritiva dos dados estimados

Variáveis ↓ / →	Tetracloroeteno*		Cis-1,2-Dicloroeteno*	
	Krigagem Ordinária	<i>Radial Basis Function</i>	Krigagem Ordinária	<i>Radial Basis Function</i>
Média (mg/kg)	0.26	0.081	0.44	0.32
Desvio Padrão	2.66	0.44	0.69	0.44
Coefficiente de Variação	9.88	5.48	1.56	1.36
Variância	7.08	0.19	0.48	0.19
Mínimo (mg/kg)	0.0011	0.0011	0.005	0.0051
Primeiro Quartil (mg/kg)	0.078	0.054	0.27	0.27
Mediana (mg/kg)	0.10	0.054	0.30	0.27
Terceiro Quartil (mg/kg)	0.11	0.065	0.34	0.27
Máximo (mg/kg)	78.34	32.34	19.23	14.17

*Notar que os dados estão em seu suporte amostral e não em forma logarítmica

Figura 9 - Gráfico de dispersão entre valores estimados e medidos de tetracloroeteno e cis-1,2-dicloroeteno



4.4. Análise Espacial, Massa de Contaminantes e Redução de Custo de Remediação

A massa de contaminantes pôde ser avaliada a partir de diferentes premissas de investigação de fase retida no solo – inicialmente por baixa resolução e atendo-se à obrigação legal da Decisão de Diretoria n° 038/2017/C de delimitação tridi-

mensional e em alta resolução via Amostragem de Solo de Perfil Completo (ITRC, 2015; RIYIS *et al.*, 2019) (Tabela 5). O tratamento de dados apresenta relevante consequência na avaliação espacial da distribuição de massa e na quantificação final de massa, sendo utilizado a metodologia comumente utilizada no GAC, a ABNT NBR 15515-3, e as metodologias Krigagem Ordinária e RBF.

Tabela 5 – Incremento nos pontos de observação entre a investigação em baixa resolução para alta resolução

Variáveis ↓ / →	Quantidade de Sondagens	Quantidade de Amostras
Investigação em Baixa Resolução	5	13
Investigação em Alta Resolução (ASPC)	15*	112*

*Quantitativo desconsidera os dados obtidos na investigação em baixa resolução

Como premissas para o cálculo de massa (Tabela 6) foram consideradas, por unidade hidroestratigráfica, características geotécnicas. Do topo para a base foram identificadas além do aterro, 05 unidades hidroestratigráficas impactadas – que, portanto, apresentam comportamentos hidrogeológicos particulares.

A investigação da área fonte em baixa resolução (que posteriormente foi complementada pela ASPC) subestimou a massa de tetracloroeteno e superestimou a de cis-1,2-dicloroeteno, em função da reduzida quantidade de amostras, sendo 13 amostras no total – em média 2 a 3 por sondagem. A baixa densidade de amostras reduz a confiança do cálculo de massa (de forma não quantificável via ABNT NBR 15515-3).

Os poucos pontos de observação (n=13) reduzem a chance de se identificar pontos de elevada concentração e distancia os pontos para delimitar a contaminação, resultando na superestimação de massa e volume impactado.

A investigação da área fonte em alta resolução, independentemente do método de interpolação, reduziu as massas quantificadas e volumes impactados para ambos os analitos considerados em relação à abordagem tradicional em função do incremento na quantidade de amostras, sendo 125 no total. A elevada densidade amostral viabiliza a delimitação de contaminantes sem extrapolações e reduzem a massa total, além de elevar as chances de se quantificar as maiores concentrações.

Tabela 2 – Variáveis para cálculo de massa de contaminantes e dados geológicos

Variáveis → / Compartimento Impactado ↓	Idade	Comport. Hidráulico	Classificação USDA (1975)	Porosidade Total	Densid. Aparente (g/cm³)
Aterro	Recente	Variável	Variável	0.298	1.66
UH 1	Quaternário	Retenção	Franco Argilo Arenosa	0.359	1.46
UH 2	Quaternário	Transporte	Areia Franca	0.374	1.76
UH 3	Terciário	Retenção	Argila	0.359	1.52
UH 4	Terciário	Baixa Advecção	Franco Argilosa	0.363	1.76
UH 5	Terciário	Retenção	Argila	0.298	1.66

Entre a investigação por baixa resolução e por alta resolução, a massa de contaminantes estimada é reduzida à 37% dos

originais 169,91 kg, ambos as estimativas via ABNT NBR 15515-3 (Tabela 7). Resultado similar ao obtido por RIYIS *et*

al. (2015) em comparação análoga, onde a massa a partir de ASPC correspondeu a 41% da calculada via metodologias tradicionais.

No que tange à forma de interpretação dos dados obtidos em alta resolução, foi comparada a metodologia tradicional (ABNT NBR 15515-3) à Krigagem Ordinária e RBF. Desse modo, verifica-se que a massa total de contaminantes foi reduzida, respectivamente, a 44% e 13% do estimado pela norma ABNT em questão (Tabela 7). Isto ocorre em virtude do uso do variograma ajustado conforme as amostras, e cujo estimador será guiado pela média das concentrações da vizinhança – para cada bloco estimado (Krigagem Ordinária) e para a média global (RBF).

Em termos de volume de solo contaminado (Tabela 7), a diferença de cerca de 8% entre os resultados obtidos por Krigagem Ordinária e RBF apresenta pouca relevância quando comparado ao originalmente estimado via norma ABNT com dados de alta resolução (queda de cerca de 61%).

A principal justificativa é que o cálculo por meio do modelo de blocos construído, cuja dimensão de cada bloco é 2,96x2,96x0,57 metros é mais precisa que cálculos feitos conforme a ABNT, e com base em reduzidas seções horizontais e verticais são elaborados blocos de dimensões notavelmente superiores para cubagem dos contaminantes. Como motivação secundária, têm-se a redução nas massas de contaminantes estimadas em virtude do uso de interpoladores que consideram as variáveis regionalizadas.

A redução no volume estimado de solo contaminado é parâmetro de relevante influência no custo final da etapa de remediação, seja para procedimentos de escavação e disposição, como de tratamento *in situ*. É esperada, portanto, relevante redução do custo final da etapa de remediação em virtude da redução do volume de solo impactado.

A redução no volume estimado de solo contaminado é parâmetro de relevante influência no custo final da etapa de remediação, seja para procedimentos de escavação e disposição, como de tratamento *in situ*. É esperada, portanto, relevante redução do custo final da etapa de remediação em virtude da redução do volume de solo impactado.

Tabela 3 – Massa de contaminantes e volume de solo impactado

Estratégia de Investigação	Baixa Resolução	Alta Resolução (ASPC)		
		ABNT NBR 15515-3	Krigagem Ordinária	Radial Basis Function
Massa de tetracloroetano (kg)	1.56	31.21	17.06	6.05
Massa de cis-1,2-dicloroetano (kg)	168.35	31.52	11.1	2.4
Massa de tetracloroetano e cis-1,2-dicloroetano (kg)	169.91	62.73	28.16	8.45
Volume de Solo Contaminado (PCE + DCE) (m ³)	22067.68	1423.43	554.34	514.39

5. CONCLUSÃO

A análise exploratória dos dados com auxílio de ferramentas gráficas (*boxplot* e histogramas e visualização em cena no software *Leapfrog*) facilitou a observação da continuidade espacial dos contaminantes, sua distribuição entre camadas de retenção/fluxo e distribuição das concentrações em relação a valores de referência (de intervenção ou concentrações máximas aceitáveis).

A estatística descritiva indicou que para tetracloroetano e cis-1,2-dicloroetano a maior parte dos valores é inferior ao valor de intervenção e são referentes ao limite de quantificação, ainda que a estratégia de investigação tenha sido enviesada de forma direcionada às porções mais impactadas, segundo critérios estabelecidos em ABNT NBR 16.434:2015.

Em relação a tricloroetano e cloreto de vinila, a estatística descritiva indicou que a estratégia direcionada ao centro de massa pelo uso de preservação em metanol, que viabiliza a quantificação de concentrações elevadas, se mostrou insatisfatória pois reportou limites de quantificação superiores aos valores de intervenção para estes analitos na maior parte das amostras. Desta forma, houve ganho de informação no que tange à Avaliação de Risco à Saúde Humana e perda de informação referente à estimativa de massa e volume de solo impactado por estes analitos.

A análise via *boxplot* das concentrações medidas em relação às unidades hidroestratigráficas definidas no modelo geológico indicaram que as maiores médias de PCE, DCE e CV ocorrem no Aterro e UH3 – unidade onde predomina a retenção de contaminantes. Destaca-se a baixa variabilidade dos resultados oriundos da única camada de elevado fluxo advectivo (UH2), independentemente do analito considerado.

Quanto à geoestatística, os variogramas puderam ser adequadamente ajustados no software *Leapfrog* com estabelecimento da continuidade espacial para cada analito considerado. A validação cruzada da Krigagem Ordinária apresentou valores elevados de correlação entre pontos analisados e sua estimativa – superiores a 0,80.

Os variogramas referentes à direção vertical, com maior densidade de amostras, apresentaram estruturas mais bem definidas que na direção horizontal. Para que as estruturas sejam mais bem reconhecidas na direção horizontal (independente do azimute), a estratégia de amostragem não poderá ser estritamente enviesada, conforme estipulado em ABNT NBR 16.434:2015, mas deverá ter componente aleatória (simples ou estratificada) ou sistemática conforme descrito em KAZUO & LANDIM (2015), além de adequado espaçamento entre as sondagens para entendimento do fenômeno.

Entre a investigação de solo em baixa resolução e alta resolução (ASPC) houve redução da massa estimada de tetracloro-

teno e cis-1,2-dicloroeteno para cerca de 37% do valor inicial via ABNT NBR 15515-3.

A estimativa de massa de contaminantes a partir dos resultados obtidos por alta resolução com metodologia tradicional (ABNT NBR 15515-3) foi de 62.73 kg. A adoção de métodos geoestatísticos reduziu a massa estimada para 40% deste valor no caso de Krigagem Ordinária (28.16 kg) e 13,47% deste valor (8.45 kg) para *Radial Basis Function*.

Portanto, ao considerar a investigação de solo por baixa resolução, que atende aos requisitos legais mínimos de delimitação da contaminação nesta matriz, em relação a investigação por alta resolução aliada à geoestatística, houve redução de massa dos contaminantes considerados para cerca de 16% do originalmente estimado para Krigagem Ordinária e 5% considerando *Radial Basis Function*.

Outra variável relevante para a etapa de remediação – notadamente das técnicas que necessitam do contato entre agente redutor/oxidante ou da remoção física do contaminante (escavação) – é o volume de solo impactado. Ainda que haja um direcionamento pela ARSH do volume cuja ação é necessária, pôde-se verificar redução de cerca de 22 mil m³ impactados conforme delimitação em baixa resolução para 1400 m³ após a investigação em alta resolução. Após refino da análise espacial e cubagem via Krigagem Ordinária e RBF, houve redução para cerca de 500 m³ (2% do originalmente estimado).

Para descrição de variáveis regionalizadas no gerenciamento de áreas contaminadas a geoestatística deve ser priorizada. Para fenômenos retratados em alta resolução e que apresentam variogramas estruturados – ou seja, que não apresentam efeito pepita puro – a Krigagem pode ser utilizada. Para variáveis que não apresentam variograma estruturado ou cuja disponibilidade de dados é reduzida, a função RBF pode ser adotada, uma vez que se trata de um estimador global.

A manutenção da estratégia atualmente adotada pelo mercado de áreas contaminadas de coleta de amostras de solo para posterior análise química de compostos orgânicos voláteis – que basicamente corresponde ao enviesamento das coletas e do agrupamento de sondagens em regiões de interesse para identificação de contaminação – prejudicam o entendimento espacial do fenômeno e a estimativa de massa confiável, conforme corroborado pela variografia dos dados.

Para que a estratégia de amostragem atenda adequadamente ao objetivo de quantificação de massa de contaminantes em fase retida (ABNT NBR 15515-3 e DD n° 38/2017C) de compostos orgânicos voláteis e subsídio à Avaliação de Risco à Saúde Humana (RAGS, 1989 e DD n° 38/2017C), aos aspectos de enviesamento estabelecidos em ABNT NBR 16.434:2015 deverão ser acrescidos critérios de amostragem aleatória (simples ou estratificada) ou sistemáticos, além de técnicas de regularização (*composite*).

A Análise Exploratória dos Dados e ferramentas geoestatísticas se demonstraram efetivas na melhoria do Modelo Conceitual gerado a partir de uma investigação por estratégia de alta resolução. Foram incrementados aspectos quanto à estimativa de massa, volume impactado e distribuição espacial dos contaminantes avaliados, com relevante alteração na estratégia de uma eventual etapa de remediação – que deverá atuar em volume impactado significativamente menor.

6. AGRADECIMENTOS

Os autores agradecem à KOPF Ambiental pelo apoio fornecido para elaboração dessa pesquisa.

Os autores agradecem à Seequent por fornecer licença de estudante para o software *Leapfrog Works* com a extensão *Contaminants*, que tornou este trabalho possível.

A análise de saída/código/dados para este artigo/relatório foi gerada usando o *Leapfrog Software*. Copyright © Seequent Limited. *Leapfrog* e todos os outros nomes de produtos ou serviços da *Seequent Limited* são marcas registradas ou marcas comerciais da *Seequent Limited*.

REFERÊNCIAS

CAMBARDELLA CA, MOORMAN TB, NOVAK JM, PARKIN TB, KARLEN DL, TURCO RF, KONOPKA AE. Field-scale variability of soil properties in central Iowa soils. *Soil science society of America journal*, v. 58, n. 5, p. 1501-1511, 1994. <https://doi.org/10.2136/sssaj1994.03615995005800050033x>

CETESB – Companhia Ambiental do Estado de São Paulo. *Decisão de Diretoria 038/2017/C*. São Paulo. Disponível em: <https://cetesb.sp.gov.br/wp-content/uploads/2014/12/DD-038-2017-C.pdf>. Acesso em: 10 dez. 2021.

CHERRY, J. PARKER. B. DNAPL Contamination on Groundwater. In: CONGRESSO INTERNACIONAL DO MEIO AMBIENTE SUBTERRÂNEO (CIMAS), 1., 2009. [Anais...]. São Paulo-SP. Setembro. 2009.

DYMENT, Stephen. Using high resolution site characterization to improve remedy design and implementation. In: *Federal Remediation Technologies Roundtable Meeting*. 2011.

HERRMANN JC, BERNARDI JVE, BASTOS WR, LACERDA LD. Dispersão Espacial X Amostragem Pontual: A Geoestatística como Ferramenta de Análise do Mercúrio em Solos de Rondônia, Amazônia Ocidental. *Geochimica Brasiliensis*, v. 23, n. 1, 2009.

FIGUEIREDO FILHO, Dalson Britto; SILVA JÚNIOR, José Alexandre. Desvendando os Mistérios do Coeficiente de Correlação de Pearson (*r*). *Revista Política Hoje*, v. 18, n. 1, p. 115-146, 2009.

- ITRC (Interstate Technology & Regulatory Council). *Integrated DNAPL Site Characterization and Tools Selection (ISC-1)*. Washington, DC: Interstate Technology & Regulatory Council, DNAPL Site Characterization Team, 2015. http://www.itrcweb.org/DNAPL-ISC_tools-selection. Acesso em: 07 mar. 2021.
- ITRC (INTERSTATE TECHNOLOGY & REGULATORY COUNCIL). *Integrated DNAPL Site Characterization and Tools Selection (ISC-1)*. 2015. Disponível em: https://www.itrcweb.org/DNAPL-ISC_tools-selection/Content/3%20Distribution%20of%20DNAPLS.htm. Acesso em: 09 Mar. 2021.
- ITRC. Interstate Technology & Regulatory Council. *Geospatial Analysis For Optimization At Environmental Sites*, 2016. Disponível em: <http://gro-1.itrcweb.org/>. Acesso em: 5 set. 2020.
- LEAL PR, MOSCHINI-CARLOS V, LÓPEZ-DOVAL JC, CINTRA JP, YAMAMOTO JK, BITENCOURT MD, SANTOS RF, ABREU GC, POMPÊO MLM. Impact of copper sulfate application at an urban Brazilian reservoir: A geostatistical and ecotoxicological approach. *Sci Total Environ*, 2017. <http://dx.doi.org/10.1016/j.scitotenv.2017.07.095>
- LERNER, D. N. et al. *An illustrated handbook of DNAPL transport and fate in the subsurface*. 2003.
- MATHERON, Georges. Principles of geostatistics. *Economic geology*, v. 58, n. 8, p. 1246-1266, 1963. <https://doi.org/10.2113/gsecongeo.58.8.1246>
- RIYIS, Marcos Tanaka; ARAKAKI, Edson; RIYIS, Mauro Tanaka; GIACHETI, Heraldo Luiz. A importância da amostragem de solo de perfil completo (ASPC) para a investigação de alta resolução em áreas contaminadas. *Águas Subterrâneas*, v. 33, n. 4, 2019. <https://doi.org/10.14295/ras.v33i4.29735>
- ROSSI, Mario E.; CLAYTON V. Deutsch. *Mineral resource estimation*. Springer Science & Business Media, 2014.
- SÃO PAULO. *Decreto N° 59.263, de 05 de Junho de 2013*. Regulamenta a Lei n° 13.577, de 08 de julho de 2009, que dispõe sobre as diretrizes e procedimentos para a proteção da qualidade do solo e gerenciamento de áreas contaminadas, e dá providências correlatas. São Paulo. 2013. Legislação Estadual.
- STEWART, M., DE LACEY, J., HODKIEWICZ, P. F., & LANE, R.. Grade estimation from radial basis functions—how does it compare with conventional geostatistical estimation. *In: NINTH INTERNATIONAL MINING GEOLOGY CONFERENCE*, Adelaide, Australia. 2014. p. 139.
- SESEQUENT. *Validating geological models against drill hole data*. Disponível em: <https://www.seequent.com/validating-geological-models-against-drill-hole-data/>. Acesso em: 24 Dez. 2020.
- TOBLER, Waldo R. A computer movie simulating urban growth in the Detroit region. *Economic geography*, v. 46, n. sup1, p. 234-240, 1970. <https://doi.org/10.2307/143141>
- USDA (US. Science; Education Administration). *Soil taxonomy: a basic system of soil classification for making and interpreting soil surveys*. US Department of Agriculture, 1975.
- USEPA (US Environmental Protection Agency). *Exposure Factors Handbook*. EPA/ 600/8-89/043, Office of Health and Environmental Assessment, Washington, DC, USA, 1989.
- USEPA (US Environmental Protection Agency). *Superfund Program Implementation Manual Fiscal Year 2019*. Washington, DC, USA, 2018. Disponível em: <https://sems-pub.epa.gov/work/HQ/100001805.pdf>. Acesso em: 16 mar. 2021.
- USEPA QA/G-4HW. *Guidance for the data quality objectives process for hazardous waste sites, 2000*. (EPA/600/R-00/007, January 2000) Internet Available.
- USEPA. Environmental Protection Agency. *Causal analysis/diagnosis decision information system (CADDIS)*, 2010. Disponível em: <http://www.epa.gov/caddis>. Acesso em: 5 set. 2020.
- USEPA. *Method 5035 Closed-system Purge-and-trap and Extraction for Volatile Organics in Soil and Waste Samples*.
- YAMAMOTO, Jorge Kazuo; LANDIM, Paulo M. Barbosa. *Geostatística: conceitos e aplicações*. Oficina de textos, 2015.