

SISTEMA DE DADOS AMBIENTAIS - COMPARTILHAMENTO DE DADOS PARA GERAÇÃO DE CONHECIMENTOS NO GAC: PROVA DE CONCEITO DO FATOR DE ATENUAÇÃO PARA INTRUSÃO DE VAPORES

Leandro Gomes de Freitas¹ (lfreitas@ipt.br), Calvin Stefan Iost^{2, 3}, Humberto Katsuyoshi Hagio²,
Guilherme Panichi Amaro⁴, Danilo Saunite², Atila Ferreira Pessoa², Adriel de Barros Simeão²,
Felipe Augusto Frazão Kleine¹, Rodrigo César de Araújo Cunha³

Resumo

O Sistema de Dados Ambientais (SDA) foi uma prova de conceito desenvolvida colaborativamente por profissionais do IPT, SENAC, consultorias ambientais e Earthsoft, aplicando técnicas de Gerenciamento de Dados para Áreas Contaminadas. O estudo avaliou a aplicação do fator de atenuação de 0,03, utilizado para estimar a intrusão de vapores em edificações, com base no estudo da USEPA (2012a). Foi estruturado um banco de dados com amostras de *sites* da Região Metropolitana de São Paulo, permitindo análises estatísticas e modelagem computacional, além do desenvolvimento de uma plataforma web para demonstração. Os resultados evidenciaram grande variabilidade dos fatores

de atenuação, sendo questionada a aplicação de um único fator genérico para todos os casos e reforçando a necessidade de abordagens *site-specific*. Com mais de 66.800 registros, integrados por meio de *Electronic Data Deliverables*, o SDA demonstrou a viabilidade de um sistema de dados colaborativo para promover interoperabilidade, transparência e decisões baseadas em *Big Data*, evidenciando seu potencial para otimizar a gestão de dados ambientais e o avanço das práticas de GAC no Brasil.

Palavras-chave: Intrusão de vapores; fator de atenuação; gestão de dados ambientais; *big data* ambiental; transformação digital no GAC.

1. Instituto de Pesquisas Tecnológicas do Estado de São Paulo (IPT)

2. Consultorias ambientais e demais instituições

3. Centro Universitário Senac

4. EarthSoft, Inc.

Abstract

The Environmental Data System (SDA) was a proof of concept collaboratively developed by professionals from IPT, SENAC, environmental consultancies, and Earthsoft, applying Data Management techniques for Contaminated Sites. The study evaluated the application of the attenuation factor of 0.03, used to estimate vapor intrusion into buildings, based on the USEPA (2012a) study. A database was structured with samples from sites in the São Paulo Metropolitan Region, enabling statistical analyses, computational modeling, and the development of a web-based demonstration platform. The results revealed significant variability in attenuation factors across different locations and chemical compounds, being questioned the application of a single generic factor and reinforcing the need for site-specific approaches. With over 66,800 records integrated through Electronic Data Deliverables, the SDA demonstrated the feasibility of a collaborative data system to promote interoperability, transparency, and Big Data-based decision-making, highlighting its potential to optimize environmental data management and advance contaminated site management practices in Brazil.

Keywords: *Vapor intrusion; attenuation factor; environmental data management; environmental big data; digital transformation in contaminated site management.*

1. Introdução

Com o avanço dos estudos do Gerenciamento de Áreas Contaminadas (GAC) na última década, o volume e a complexidade dos dados gerados são crescentes, demandando cada vez

mais estratégias e ferramentas adequadas para otimizar a gestão dessa massa de informações. Apesar da evolução da ciência e gestão de dados em diversas áreas do conhecimento, o mercado ambiental, principalmente no GAC, ainda vivencia uma grande defasagem na aplicação de tais tecnologias para a gestão de dados.

Um exemplo latente é a forma como são geridos os dados dos projetos pelos órgãos reguladores, na qual os relatórios ainda são arquivados em silos, em certos locais ainda contando com vias físicas ou arquivos em formatos não estruturados, tal como pdf ou docx. Mesmo com os avanços relacionados à disponibilização dos cadastros de áreas contaminadas em portais on-line, utilizando ferramentas geoespaciais interativas, como o DataGEO e o SIGAMGEO de São Paulo, ou o GEOINEA do Rio de Janeiro, ainda não há uma integração das bases de dados, o que inviabiliza a execução de análises integradas e automatizadas, bem como o processamento conjunto do valioso acervo técnico armazenado nestes estudos.

Mundo afora, são frequentes os exemplos de sistemas para submissão de resultados em formatos eletrônicos padronizados aos órgãos reguladores (USEPA 2024; SOUTH AUSTRALIA, 2019). O guia de melhores práticas para gestão de dados da Agência de Proteção Ambiental norte americana (USEPA, 2018), bem como os manuais de especificações dos escritórios regionais, destacando-se a *Region 2* (USEPA, 2019) são referências importantes acerca do uso da entrega eletrônica de dados (*Electronic data deliverable* - EDD). Desde 2010, a USEPA *Region 4* adota uma metodologia de submissão eletrônica de dados para os *sites*

inseridos no programa federal de áreas prioritárias (*Superfund*), garantindo uma padronização que possibilita otimizar a gestão e as análises dessas informações (USEPA, 2010).

Outra referência é a plataforma GeoTracker, criada pelo Conselho Estadual de Recursos Hídricos da Califórnia (EUA), que centraliza dados ambientais e laboratoriais de mais de 50.000 *sites* desde 2001, totalizando cerca de 285 milhões de registros analíticos. Segundo Beckley *et al.* (2022), esse repositório tem sido utilizado para mineração de dados e pesquisas científicas sobre a remediação ambiental, destacando sua importância na regulamentação ambiental, no monitoramento de substâncias tóxicas, como PFAS, e em investigações de intrusão de vapores.

A integração de bancos de dados do GAC é abordada na análise de Suthersan *et al.* (2016), que explora o potencial dos grandes volumes de dados (*big data*) na remediação ambiental, destacando o uso de análises estatísticas e aprendizado de máquina como métodos promissores para prever tendências, melhorar a eficiência das remediações e reduzir custos. O artigo de Horst *et al.* (2018) também analisa o impacto do *big data*, abordando os desafios e o potencial da integração de dados para aumentar a produtividade, enfatizando a necessidade de mudanças estruturais na forma como os dados são coletados, gerenciados e utilizados, para impulsionar tomadas de decisão mais eficazes e estratégicas.

Esta realidade foi a grande motivadora da articulação entre parceiros de diferentes empresas de consultoria e instituições de ensino e pesquisa do Estado de São Paulo, para o desenvolvimento de uma

Prova de Conceito (PoC) de um sistema colaborativo de gestão de dados do GAC, batizado de 'Sistema de Dados Ambientais - SDA'. A temática alvo estabelecida foi o estudo do fator de atenuação para intrusão de vapores em áreas contaminadas.

A avaliação da intrusão de vapores é uma etapa fundamental no GAC devido aos riscos que os compostos orgânicos voláteis (VOCs) representam para a saúde humana e a qualidade do ar em ambientes fechados. Compostos como os organoclorados podem migrar do solo e da água subterrânea para o interior das edificações, expondo os ocupantes a substâncias tóxicas e potencialmente carcinogênicas (Figura 1). Fatores como pressão atmosférica, características do solo e integridade das fundações influenciam as taxas de intrusão.

Conforme descrito por Johnson e Ettinger (1991), o fator de atenuação para intrusão de vapores (α) pode ser entendido como a razão entre a concentração de um determinado contaminante no ar interno do edifício e sua concentração no ar do solo, proveniente do solo ou da água subterrânea. Apesar da possibilidade de aplicar modelos matemáticos para estimar a migração de vapores do solo para ambientes internos, a variabilidade espacial e temporal das concentrações dos VOCs e as dificuldades de se diferenciar contribuições de fontes externas e subterrâneas, levou a USEPA a recomendar o uso das concentrações obtidas em amostras de *subslab* para estimar a sua concentração no ar ambiente (USEPA, 1997; USEPA, 2015). O racional por trás dessa recomendação é que as concentrações de VOCs na zona do contrapiso representam uma fonte imediata de vapores que podem penetrar na edificação, sendo,

portanto, um dado fundamental para estimar o risco de exposição e para orientar decisões sobre necessidade de intervenção e remediação. Dessa forma, é possível calcular os fatores de atenuação por meio da seguinte expressão:

$$FA = \frac{C_{in}}{C_{ss}}$$

FA = Fator de atenuação (α) para intrusão de vapores
 C_{in} = Concentração de uma determinada SQI presente na amostra de ar ambiente
 C_{ss} = Concentração de uma determinada SQI presente na amostra de ar em *subslab*

(Equação I)

O documento "EPA's Vapor Intrusion Database: Evaluation and Characterization of Attenuation Factors for Chlorinated Volatile Organic Compounds and Residential Buildings" (USE PA, 2012) apresenta os resultados de um estudo que teve como objetivo a avaliação empírica dos fatores de atenuação de diversas localidades nos EUA.

O primeiro passo foi a construção de uma base de dados robusta contendo informações referentes às concentrações das substâncias químicas de interesse (SQIs) em diferentes matrizes como ar ambiente, ar do solo (*subslab*) e água subterrânea. Foram consideradas informações provenientes de 41 *sites* e 913 edificações residenciais e mistas, em 15 estados diferentes do território norte americano. Para avaliar a relação entre as concentrações das SQIs quantificadas no ar ambiente e no ar em *subslab* foram calculados os fatores de atenuação para 1582 amostras pareadas (ar ambiente/*subslab*). Após a aplicação de uma série de filtros e análise da distribuição dos resultados obtidos, o percentil 95 foi adotado como um valor representativo para a população de dados analisada e assim determinou-se, como premissa, que o fator de atenuação de 0,03 fosse utilizado como um critério conservador de triagem (*screening*) para estimar as concentrações presentes no ar ambiente com base nas concentrações encontradas nos poços de *subslab*.

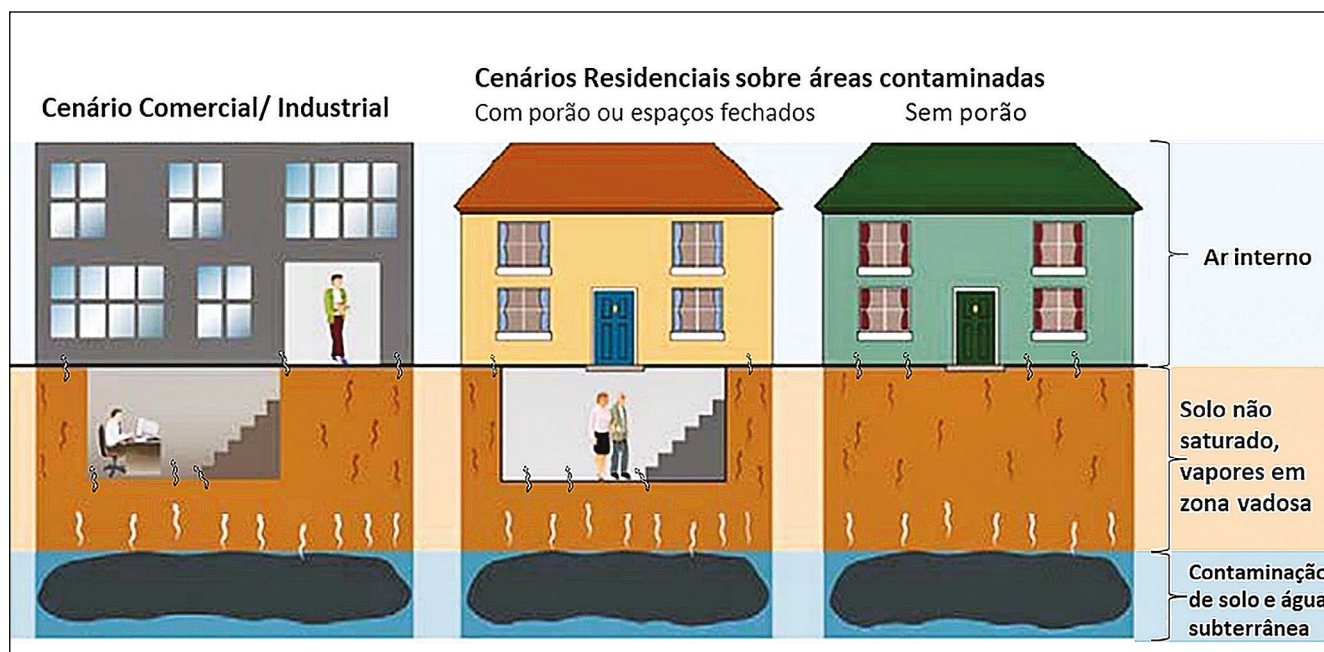


Figura 1. Representação conceitual do processo de intrusão de vapores. Fonte: Adaptada de ADEC (2023).

No Brasil, apesar de atualmente não haver uma legislação específica para valores de referência para a concentração de vapores no ar do solo, o órgão ambiental paulista (CETESB) vem recomendando a adoção do fator de atenuação de 0,03 para a definição de valores de referência para fins de varredura. Esta recomendação considera que uma concentração de referência para vapores do subsolo pode ser estimada ao dividir uma concentração máxima aceitável do ar ambiente (pré-estabelecida com base em risco toxicológico) pelo fator α padrão, o que ressalta sua importância nos estudos de intrusão de vapores.

2. Objetivo

Considerando que o fator de atenuação pode ser influenciado por uma série de variáveis ambientais, temporais e do meio físico (BREWER *et al.*, 2014), a PoC do Sistema de Dados Ambientais teve como objetivo geral avaliar a aplicabilidade do fator α sugerido pela USEPA (2012), reproduzindo o estudo com uma base de dados representativa de projetos executados na Região Metropolitana de São Paulo (RMSP). Além da possibilidade de comparação dos resultados, a PoC também buscou os seguintes objetivos específicos:

- criar uma estrutura de dados e um dicionário padrão de entrega eletrônica de dados (EDD) entre as empresas e instituições participantes;
- criar um sistema *online*, com a estratégia de *Minimum Viable Product* (MVP), incluindo um *workflow* funcional e uma plataforma web, para demonstração e publicação dos resultados e a consolidação da prova de conceito;
- subsidiar uma análise crítica da adoção do

fator de atenuação (α) padrão de 0,03 em estudos de intrusão de vapores pelos órgãos reguladores paulistas;

- demonstrar a viabilidade e o potencial de um sistema de compartilhamento colaborativo de dados entre diferentes *stakeholders* do mercado de GAC.

3. Material e Métodos

O grupo de parceiros de diferentes instituições se formou no início de 2020, com as primeiras reuniões focadas na definição de estratégias, levantamento de referências, definição das temáticas de interesse, concepção do *modus operandi*, bem como no estabelecimento da metodologia de trabalho colaborativo. Além de representantes de três consultorias ambientais, professores do curso de pós-graduação em GAC do Centro Universitário SENAC e pesquisadores do Instituto de Pesquisas Tecnológicas do Estado de São Paulo (IPT) compunham a equipe inicial do projeto.

A iniciativa buscou a estruturação de um repositório de dados unificado, colaborativo, sobre projetos de gerenciamento de áreas contaminadas. Inicialmente, o grupo se organizou em *squads* multidisciplinares nas seguintes frentes: Estratégias; Estudo de Caso; Estruturação de dados; Análise e cálculo de índices e Consolidação dos resultados.

Considerando a ausência de normativas específicas e a relevância do tema 'Intrusão de Vapores' no mercado de GAC brasileiro (GOUVÊA JÚNIOR *et al.*, 2018), a temática alvo estabelecida foi o estudo do fator de atenuação em *sites* contaminados com compostos orgânicos voláteis clorados (CVOCs).

Com o intuito de ganhar celeridade no desenvolvimento da prova de conceito, tomando

como base as diretrizes do manual de especificações para o reporte eletrônico de dados da *Region 2* da USEPA (USEPA, 2019), bem como a experiência prévia das consultorias envolvidas, o grupo optou por adotar o *software* EQuIS®, da empresa Earthsoft, como base para o tratamento de dados do projeto. A partir disto, um integrante desta empresa foi convidado a integrar o grupo de trabalho, a qual também disponibilizou uma licença piloto do *software* EQuIS para a realização da prova de conceito.

Neste modelo de colaboração, o IPT foi designado como órgão gestor dos dados e recebeu a licença piloto do EQuIS, além de centralizar o recebimento das EDDs das consultorias, sendo responsável pela curadoria, checagem e upload dos dados no *software*. Professores do Senac forneceram mentoria para o projeto. O modelo de colaboração adotado, juntamente com o *workflow* do SDA, é apresentado na Figura 2.

Definido o modelo de operação, iniciou-se a etapa de padronização da estrutura dos dados. Após um ciclo de discussões, com apoio da empresa Earthsoft, o grupo gerou uma estrutura padrão

para o tema estudado, acordada entre todos os participantes, composto por Tabelas de Referência (RTs) que integravam os *Electronic Data Deliverable* (EDD). A estrutura final se baseou em um formato desenvolvido previamente pela Earthsoft, em conjunto com a USEPA, específico para a intrusão de vapores, o qual foi posteriormente adaptado para as necessidades brasileiras. As principais colunas das RTs foram as seguintes (fora diversos outros campos de controle de qualidade):

- Localização: facility_id; facility_code; sys_loc_code; loc_type;
- Amostra: facility_id; sys_sample_code; sample_id; sample_date; sample_time; matrix_code; analysis_date;
- Resultados: facility_id; sys_sample_code; sample_id; cas_rn; chemical_name; report_result_value; report_result_unit; result_numeric.

Além da disponibilização de técnicos especializados e do apoio nas discussões, as consultorias tiveram um papel fundamental na disponibilização de dados, com qualidade e volume adequados ao projeto. Para isso, os dados tiveram sua localização geoespacial e procedência descaracterizadas e des-

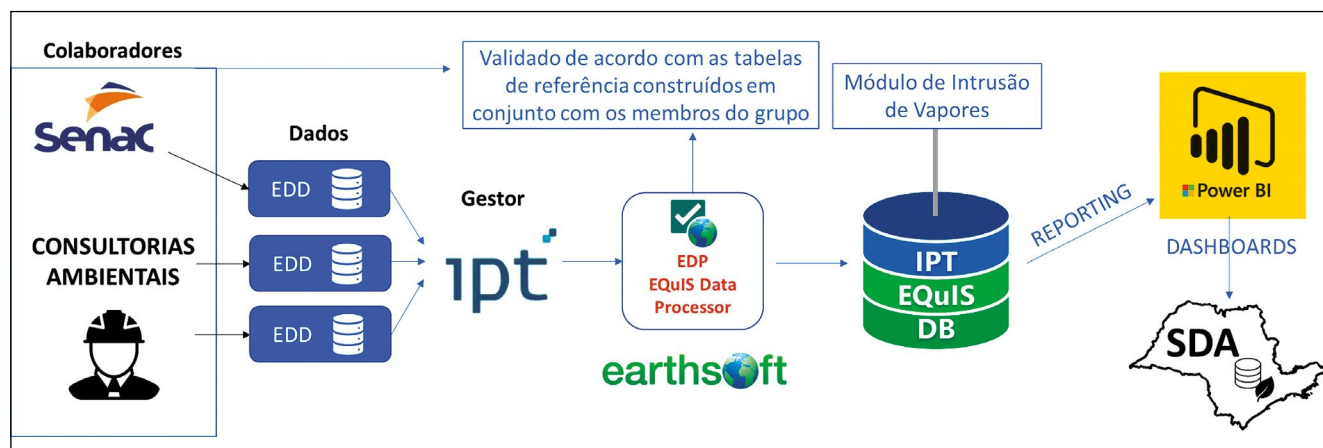
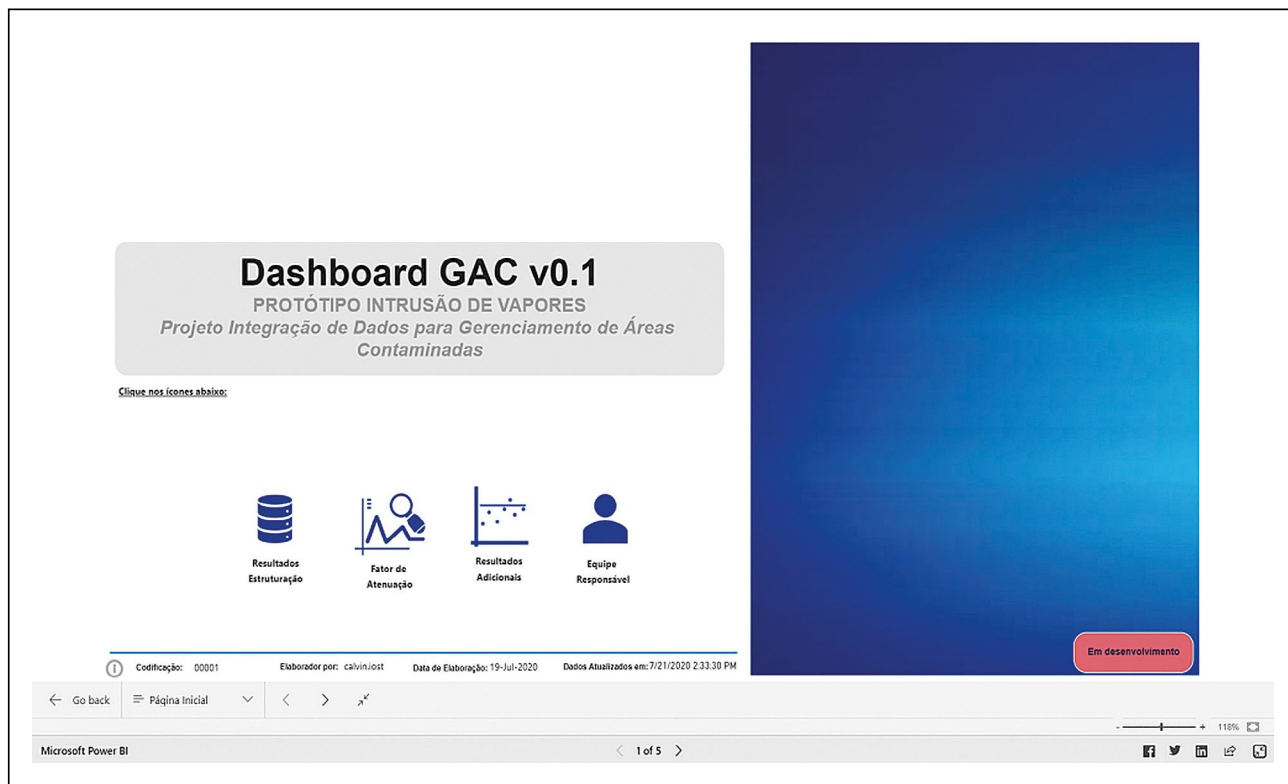


Figura 2. Modelo de colaboração entre as empresas participantes, juntamente com o *workflow* do Sistema de Dados Ambientais. Fonte: Elaborada pelos autores (2025).

a)



b)

Em desenvolvimento

Resultados Estruturação

Filtrar Unidade

Todos

Facilities

facility_id	facility_code	prp...	data_p...	facilit...	program_code	facility_name	address_1	address_2	city	county	state	country	postal_code	coord_type_code	phone_number	alt_pho
613574060...	Zanettini Bar...		ZB	29.49...		ZANETTINI...	Av. XXXXX		São Paulo	São Paulo	SP	Brasil	XXXXX-XXX	LAT LONG	(11) XXXX-XXXX	

Subfacilities

facility_id	site_code	site_type	site_name	site_task_code	site_desc1	site_desc2	contact_name	address1	address2	city	state	zipcode	phone_number	alt_phone_number
6135740...	1	Industrial	Zanettini Barossi s...					Av. XXXXX		São Paulo	São Paulo	XXXXX-XXX	(11) XXXX-XXXX	
6135740...	2	Condomínio	Condomínio XXXXX					R. XXXXX		São Paulo	São Paulo	XXXXX-XXX	(11) XXXX-XXXX	

VI Bldg Address_v4

facility_id	building_code	subfacility_code	task_code	contact_name	address1	address2	city	state	county	phone_number	alt_pho
6135740...	1	1	2020_Levantamento	TBD	Av. XXXXX	Vila Carioca, São Paulo - SP - XXXXX-XXX	São Paulo	São Paulo	São Paulo	(11) XXXX-XXXX	
6135740...	2	2	2020_Levantamento	TBD	R. XXXXX	- Ipiranga, São Paulo - SP, XXXXX-XXX	São Paulo	São Paulo	São Paulo	(11) XXXX-XXXX	

VI Locations_v4

facility_id	data_provider	sys_loc_code	buildi...	latitude	longitude	coord_sys_desc	observation...	alt_x_coord	alt_y_coord	alt_coord_type_code	alt_identifier	horz_collect_method_code	horz_accura...
6135740...	ZB	1	1	-23.60	-46.59	LAT LONG	19/07/2020 ...	338.057.52	7.388.851.13	SIRGAS2000_UTM_23S	PRIMARY		
6135740...	ZB	2	1	-23.60	-46.59	LAT LONG	19/07/2020 ...	338.043.44	7.388.868.67	SIRGAS2000_UTM_23S	PRIMARY		

VI Building Inspection_v4

facility_id	building_code	inspection_date	task_code	building_name	building_type	building_use_type	building_size	num_floors	construct_year	foundation_depth	foundation_depth_unit
61357406...	2	7/19/2020 3:47:45 PM	2020_Levantamento	Condomínio ...	Condomínio	Condomínio	TBD				
61357406...	1	7/19/2020 3:47:45 PM	2020_Levantamento	Zanettini Bar...	Industrial	Industrial	TBD				

Group member

facility_id	group_code	member_code	member_type	report_order	display_order	status_flag	remark
61357406000110	1	41	sys_loc_code	1	1	A	AA-01
61357406000110	2	42	sys_loc_code	4	4	A	AA-02
61357406000110	3	1	sys_loc_code	2	2	A	DMC-01A

VI Samples_v4

facility_id	data_provider	sys_sample_code	sample_name	sample_matrix_code	sample_type_code	sample_source	parent_sample_code	sample_delivery_gra
61357406000110	ZB	2019_09_01_AA-01_N_AA	2019_09_01_AA-01_N_AA	AA	N	TO-15		
61357406000110	ZB	2019_09_01_AA-02_N_AA	2019_09_01_AA-02_N_AA	AA	N	TO-15		

VI TestResultsQC_v4

facility_id	sys_sample_code	lab_anl_method_name	analysis_date	fraction	column_number	test_type	lab_matrix_code	analysis_location	basis	container_id	dilution_factor	prep_method
61357406000110	2019_09_01_AA-01_N_AA	TBD		N	NA	Initial	Air	FI	NA		1.00	TBD
61357406000110	2019_09_01_AA-01_N_AA	TBD		N	NA	Initial	Air	FI	NA		1.00	TBD



Figura 3a a 3d. Produtos da PoC inicial do Sistema de Dados Ambientais.
Fonte: Elaboradas pelos autores (2022).

vinculadas das áreas de onde foram gerados, mantendo, no entanto, a relação espacial fundamental entre cada conjunto de amostras. Essa descaracterização não prejudicou o estudo e impossibilitou a identificação da procedência dos dados, trazendo segurança para todos os envolvidos e protegendo a confidencialidade das informações.

4. Estruturação do banco de dados

O primeiro *dataset* foi disponibilizado pela indústria Zanettini Barossi S/A Indústria e Comércio, parceira do curso de pós-graduação em GAC do SENAC, que concordou em ceder os seus dados, de modo aberto, sem restrições de geolocalização, para a realização da prova de conceito inicial do projeto. O produto dessa PoC foram alguns painéis dinâmicos de visualização de dados (*dashboards*), gerados através do *software* Microsoft Power BI®, com o apoio de parceiros das empresas. As Figuras 3a a 3d apresentam os se-

guintes painéis: 3a) página inicial com o menu de opções; 3b) dados tabelados e estruturados nos padrões definidos; 3c) resultados dos cálculos dos fatores de atenuação por cluster de pontos e por localidade; 3d) análise geoespacial com a localização dos pontos de amostragem.

Com a validação da PoC inicial, as empresas de consultoria prosseguiram com a seleção e organização de dados dos seus projetos, considerando casos na RMSP. Uma amostra do banco de dados é apresentada nas Figura 4.

5. Análise Exploratória e Análise de Dados

Com os *datasets* estruturados, foi realizado um primeiro tratamento e filtragem dos dados, impondo controles de qualidade e filtros para o estabelecimento de pares agrupados de amostras de concentração de contaminantes no contrapiso (Css) e amostras de concentração no ar ambiente adjacente (Cin), conforme a referência do estudo

ID	FACILITY_ID	FACILITY_CODE	SYS_LOC_CODE	LOC_NAME	SAMPLE_ID	SYS_SAMPLE_CODE	SAMPLE_NAME	SAMPLE_DATE	SAMPLEDATE
1	347	ZB	PMG-25A	NaN	1795	SB1906172.007	2019_09_23_PMG-25A_N_SS	2019-09-23 12:00:00	2019-09-23
2	347	ZB	PMG-25A	NaN	1795	SB1906172.007	2019_09_23_PMG-25A_N_SS	2019-09-23 12:00:00	2019-09-23
3	347	ZB	PMG-25A	NaN	1795	SB1906172.007	2019_09_23_PMG-25A_N_SS	2019-09-23 12:00:00	2019-09-23
4	347	ZB	PMG-25A	NaN	1795	SB1906172.007	2019_09_23_PMG-25A_N_SS	2019-09-23 12:00:00	2019-09-23
5	347	ZB	PMG-25A	NaN	1795	SB1906172.007	2019_09_23_PMG-25A_N_SS	2019-09-23 12:00:00	2019-09-23
...	...	...	...	...	...	...	...	...	...
66797	11326	FACILITY-6	PCP_64_43803	PCP-64	11562	64_43803_4380905760BIOAN	PCP-064	2019-12-10 13:49:00	2019-12-10
66798	11326	FACILITY-6	PCP_64_43803	PCP-64	11562	64_43803_4380905760BIOAN	PCP-064	2019-12-10 13:49:00	2019-12-10
66799	11326	FACILITY-6	PCP_64_43803	PCP-64	11562	64_43803_4380905760BIOAN	PCP-064	2019-12-10 13:49:00	2019-12-10
66800	11326	FACILITY-6	PCP_64_43803	PCP-64	11562	64_43803_4380905760BIOAN	PCP-064	2019-12-10 13:49:00	2019-12-10
66801	11326	FACILITY-6	PCP_64_43803	PCP-64	11562	64_43803_4380905760BIOAN	PCP-064	2019-12-10 13:49:00	2019-12-10

66801 rows X 68 columns

Figura 4. Amostra do banco de dados devidamente descaracterizada. Fonte: Elaborada pelos autores (2022).

da USEPA (2012). O cálculo do fator de atenuação se dá basicamente pela razão C_{in}/C_{ss} , sendo fundamental entender os mecanismos que afetam essa relação nos estudos de intrusão de vapores.

O tratamento dos dados foi realizado com programação em Python, gerando filtros e análises exploratórias preliminares. Para o estabelecimento dos pares agrupados, utilizou-se como critério um raio de busca de 10 m entre a localização dos pares de amostras (ar ambiente/*subslab*) e um período máximo de 10 dias entre suas datas de coleta. Outro exemplo de filtro aplicado foi que nenhuma amostra de ar ambiente poderia ter concentração maior que seu par no contrapiso. Segundo o estudo publicado por Yao *et al.* (2018), estes critérios se mostram bem mais rigorosos do que os adotados no estudo de referência da USEPA (2012).

Com base nestes critérios, foram estabelecidos os pares agrupados para o cálculo dos fatores de atenuação, para os compostos de interesse: tetracloroeteno (PCE), tricloroeteno (TCE), 1,2-di-

cloroeteno (DCE) e cloreto de vinila (CV). Adicionalmente, os resultados foram segregados sazonalmente, entre estação “quente” (primavera e verão) e estação “fria” (outono e inverno), com base nas datas de amostragem.

Para a apresentação dos resultados, utilizou-se o recurso estatístico do *boxplot* (ou diagrama de caixa) que é uma ferramenta gráfica que resume a distribuição de um conjunto de dados, mostrando informações importantes de forma compacta. Eis os principais elementos e como interpretá-los:

- **Box:** Representa os quartis dos dados. A borda inferior do box é o primeiro quartil (Q1, 25% dos dados) e a borda superior é o terceiro quartil (Q3, 75% dos dados).
- **Linha Central:** Dentro da caixa, uma linha indica a mediana (Q2), que divide os dados em duas metades iguais.
- **Whiskers:** São as linhas que se estendem a partir da caixa até os valores máximos e mínimos dos dados que não são considerados outliers.

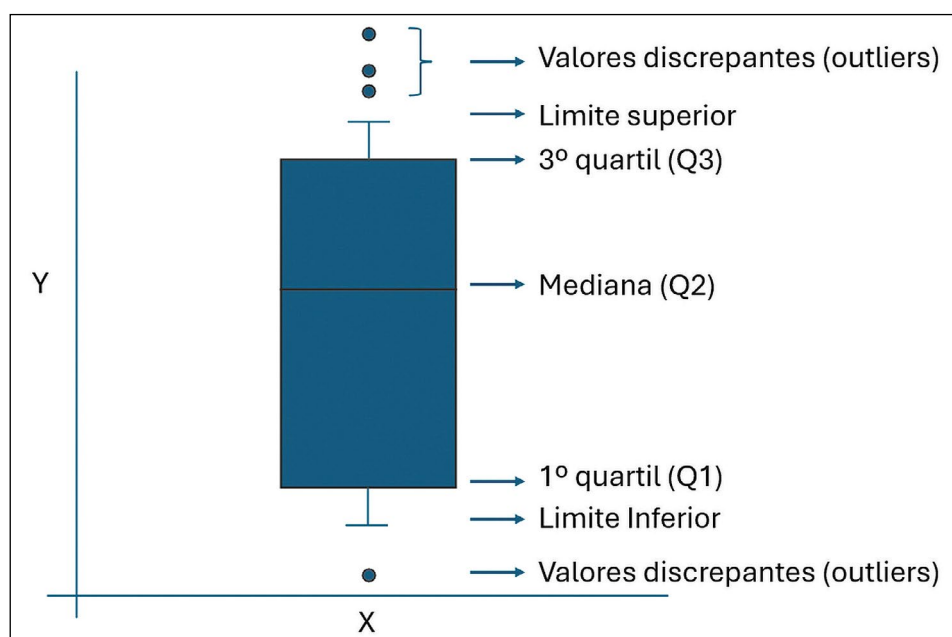


Figura 5. Representação gráfica de *boxplot*. Fonte: Elaborada pelos autores (2025).

- Outliers: Pontos que se encontram além dos *whiskers* e que podem indicar valores extremos ou erros de medição.

Em resumo, o *boxplot* permite identificar rapidamente a dispersão, a assimetria e a presença de outliers no conjunto de dados, facilitando a comparação entre diferentes grupos ou variáveis. A Figura 5 resume graficamente sua representação.

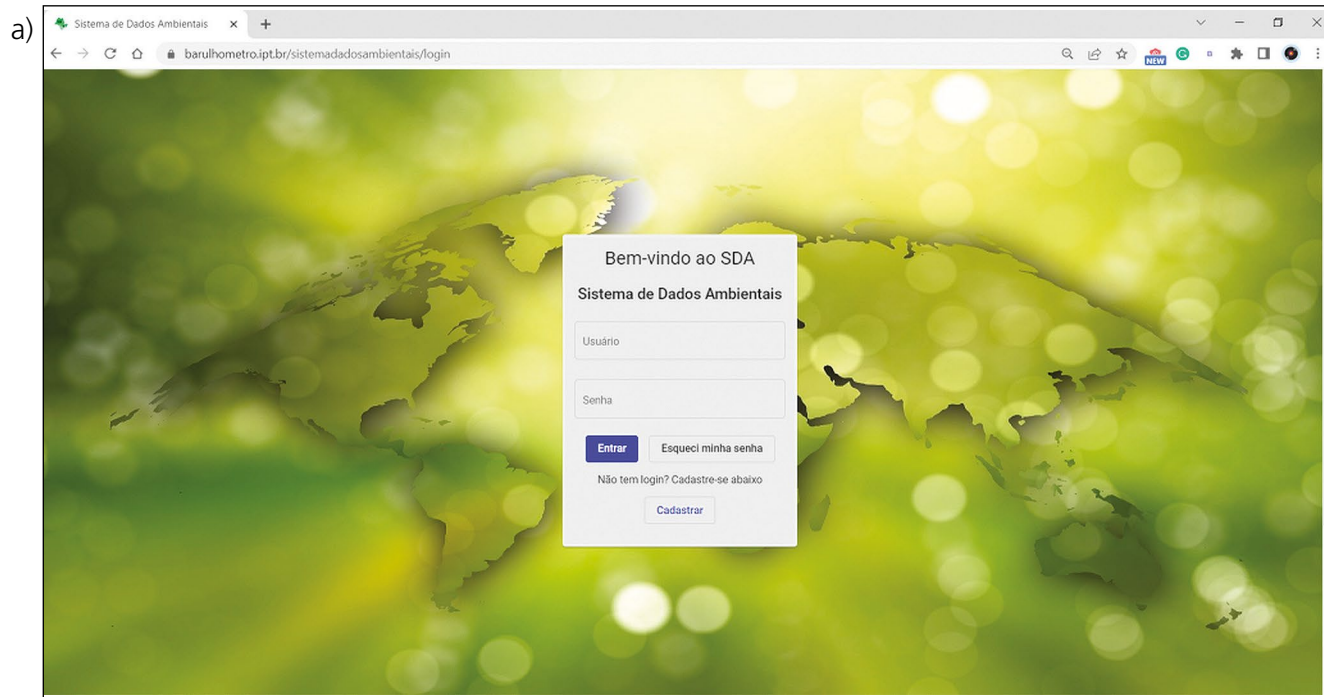
As saídas de resultados do Python foram então compiladas em gráficos analíticos, através do Microsoft Power BI, possibilitando a geração das visualizações dinâmicas em *dashboards*. O link desses painéis foi então exportado para visualização na plataforma web piloto, para a disponibilização dos resultados on-line.

5. 1. Desenvolvimento da plataforma web para visualização dos resultados

A Plataforma web para a PoC do SDA foi desenvolvida pelo time da Seção de Engenharia de Software e Transporte Inteligentes (SESTI) do IPT.

O objetivo foi produzir um protótipo, no conceito de *Minimum Viable Product* (MVP), para demonstrar as potencialidades da solução e concluir a etapa de visualização e disponibilização on-line dos resultados.

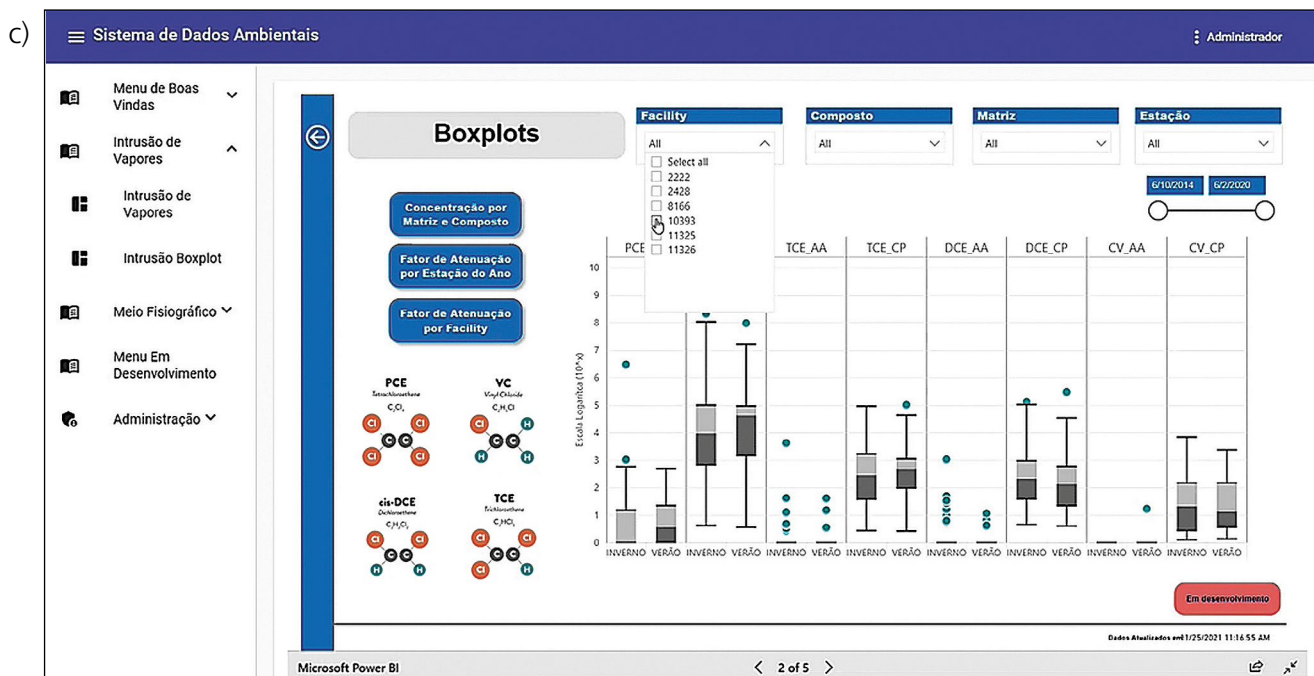
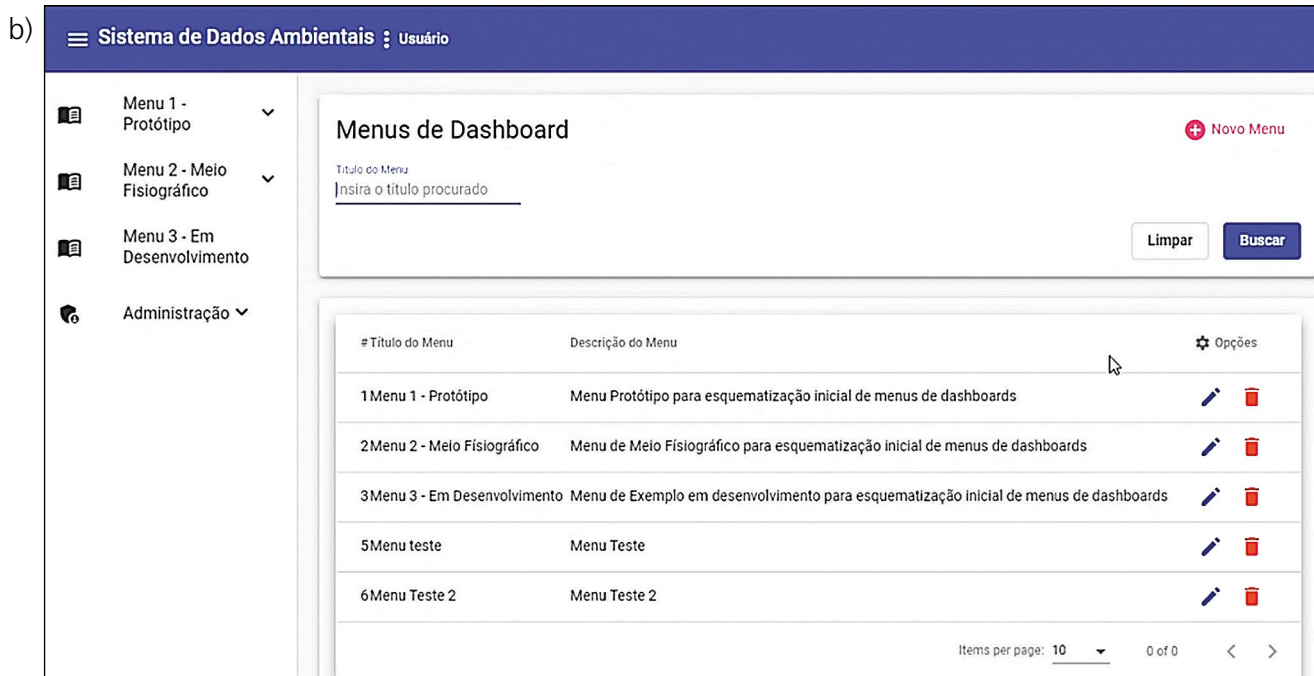
A plataforma foi desenvolvida com *framework* ANGULAR, com transmissão de dados em formato JSON por meio de uma API desenvolvida em Asp. NET Core e armazenamento de dados em uma base SQL Server. Suas funcionalidades consistem, basicamente, no controle de usuários e controle dos níveis de acessos, com opção para cadastro de novos usuários. As opções de perfis adotadas foram: administrador, parceiros ou visitantes (público). Permite ainda, para o perfil de administrador: a gestão de menus, gestão de usuários e a gestão de páginas, onde podem ser inseridos novos links para visualizações de *dashboards* externos. Possui ainda um design responsivo, que se adapta ao tamanho da tela, para uso em PC ou smartphones.

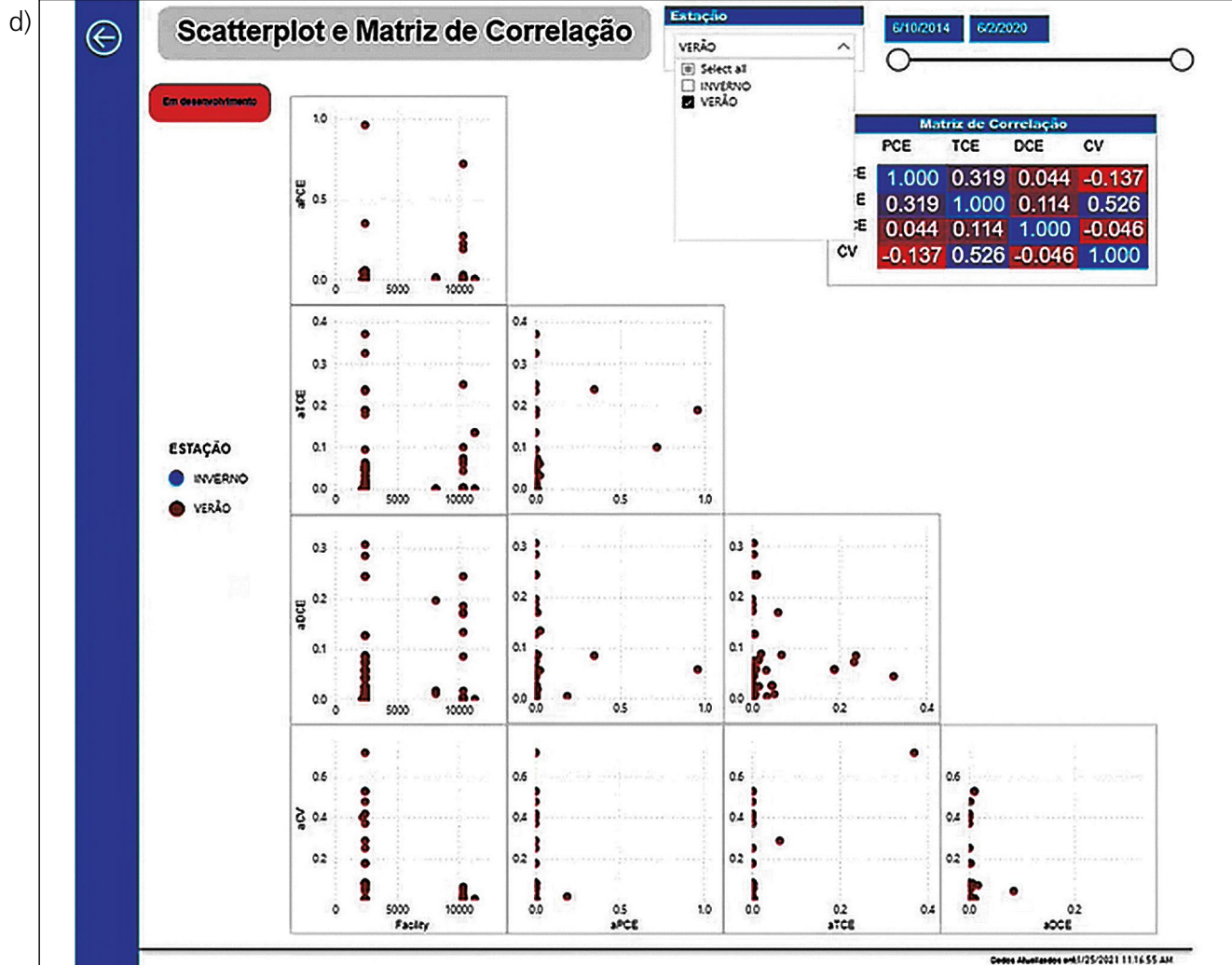


O sistema foi hospedado, temporariamente, em um servidor de projetos no IPT, visando demonstrar as funcionalidades como uma PoC. Os usuários podiam interagir com o sistema pela manipulação dos *dashboards*.

As Figuras 6a a 6d a seguir apresentam as seguintes telas ilustrativas do sistema e dos re-

sultados obtidos: 6a) página inicial para login e cadastro de usuários; 6b) home com o menu de opções e *dashboards*; 6c) painel para escolha de *softwares* por *facility*, por composto, por matriz ou por estação do ano; 6d) painel com gráficos de dispersão e matriz de correlação por composto e por estação do ano.





Figuras 6a a 6d. *Front-end* da plataforma web do SDA, com os *dashboards* de resultados da etapa piloto. Elaboradas pelos autores (2022).

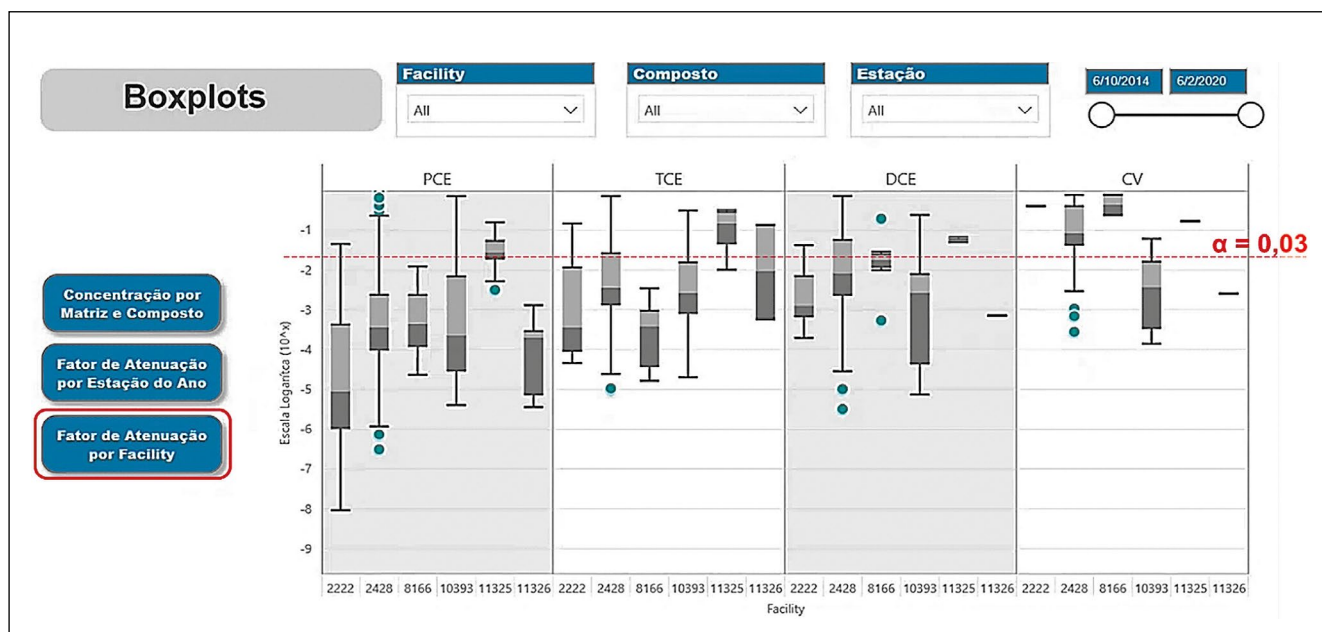


Figura 7. Resultados dos fatores de atenuação calculados por composto e por *facility*.
Fonte: Elaborada pelos autores (2022).

6. Resultados e Discussão

Ao todo foram compilados dados de 6 *facilities* (áreas ou unidades operacionais), somando 1339 amostras acreditadas analisadas pelo método TO-15, sendo 694 amostras de ar ambiente e 645 amostras de poços *subslab*, totalizando 66.801 linhas de resultados analíticos, coletadas entre junho de 2014 e novembro de 2020.

Com base no critério de estabelecimento de pares agrupados de amostras (ar ambiente/*subslab*), foram obtidos 426 pares para o cálculo dos fatores de atenuação para os compostos de interesse tetracloroetano (PCE), tricloroetano (TCE), 1,2-dicloroetano (DCE) e cloreto de vinila (CV).

A Figura 7 apresenta os resultados dos fatores de atenuação calculados para cada composto, em cada *facility*, com distribuições por *boxplots* em escala logarítmica. Para comparação com o fator de atenuação (α) padrão de 0,03, este valor é representado pela linha pontilhada vermelha.

Ao comparar os resultados de um mesmo composto químico, entre as diferentes *facilities*, observou-se variações de até 4 ordens de grandeza, sendo o PCE o composto que apresentou variações mais expressivas, com a sua mediana variando entre $9E^{-6}$ e $5E^{-2}$. Em apenas um dos *sites* a mediana deste composto se mostrou próxima ao padrão α de 0,03. Para o TCE nenhum dos *sites* apresentou fator de atenuação na mesma ordem de grandeza do padrão, enquanto para o DCE e CV, dois *sites* e um site, respectivamente, mostraram o α na mesma ordem de grandeza do padrão. Estas diferenças entre *facilities* ressaltam a natureza eminentemente *site specific* dos fatores de atenuação, sendo influenciada por características

estruturais, tais como as condições e espessura do contrapiso, condições de ventilação, presença de caminhos preferenciais, etc. (YAO *et al.*, 2013).

Ao comparar os resultados entre diferentes compostos químicos, observam-se atenuações progressivamente maiores para compostos de maior peso molecular em relação aos de menor peso molecular (FA-PCE) $\leq$ (FA-TCE) $\leq$ (FA-DCE) $\leq$ (FA-CV), lembrando que quanto maior o fator de atenuação (numericamente), menor é a atenuação da intrusão de vapores do composto e, consequentemente, maior será a concentração na amostra de ar ambiente. Para o composto PCE (MM = 165,8 g/mol), por exemplo, observam-se fatores de atenuação até 3 ordens de grandeza maiores em relação ao CV (MM = 62,5 g/mol) em um mesmo site (ex. site 8166), indicando um maior potencial de intrusão do cloreto de vinila, o que demonstra a influência da massa e do tamanho molecular, fora outras propriedades físico-químicas, na capacidade de intrusão e no fator de atenuação dos compostos (MCHUGH *et al.*, 2017).

A Figura 8 apresenta os resultados dos fatores de atenuação calculados para cada composto e a segregação sazonal entre inverno e no verão.

Apesar da sazonalidade ser um fator de grande influência nos processos de intrusão de vapores (HOLTON *et al.*, 2013), não se observaram diferenças expressivas nas medianas dos fatores de atenuação de cada composto entre verão e inverno. Apenas uma variabilidade ligeiramente maior de α no inverno pôde ser observada. Adicionalmente, para a base de dados analisada, não foram observadas variações estatisticamente significativas ($p < 0,05$) no quesito

da sazonalidade entre os fatores de atenuação para um mesmo composto.

Avaliando a matriz de correlações dos fatores de atenuação entre os compostos, por época do ano, observa-se que no inverno, os fatores do PCE, TCE e DCE apresentaram maiores correlações entre si, do que no verão. Nenhum composto pai apresentou correlação significativa em relação ao CV.

A ausência de variações significativas dos fatores de atenuação entre inverno e verão pode ser associada a uma série de fatores que não pode ser comprovada no presente estudo. Entre esses fatores está a variabilidade climática da Região Metropolitana de São Paulo, pois fatores atmosféricos e de urbanização da megalópole afetam mais a variação térmica da região do que a divisão tradicional entre verão (quente) e inverno (frio) (NOBRE *et al.*, 2011).

Com relação à comparação dos resultados com o Fator α padrão de 0,03, observa-se uma

grande variabilidade entre unidades e estações do ano, ficando evidente que, para cada composto, há algumas unidades em que parte das concentrações está acima de 0,03, principalmente para CV, e outras em que os valores são majoritariamente abaixo, principalmente para PCE, TCE, DCE, que também apresentam padrão heterogêneo, dependendo da unidade. No entanto, em geral, há grupos claramente abaixo de 0,03, tanto na análise de seus *boxplots*, como na predominância dos dados abaixo do 3º quartil, seja na análise por *facility* ou por estação do ano. Nestes casos, um fator de atenuação inferior significa que houve uma maior atenuação dos compostos, implicando em menores taxas de intrusão de vapores e menores concentrações no ar ambiente.

Esse tipo de análise (em escala logarítmica) ajuda a visualizar diferenças de ordem de magnitude nas concentrações. O *boxplot*, por sua vez, facilita a comparação rápida entre me-

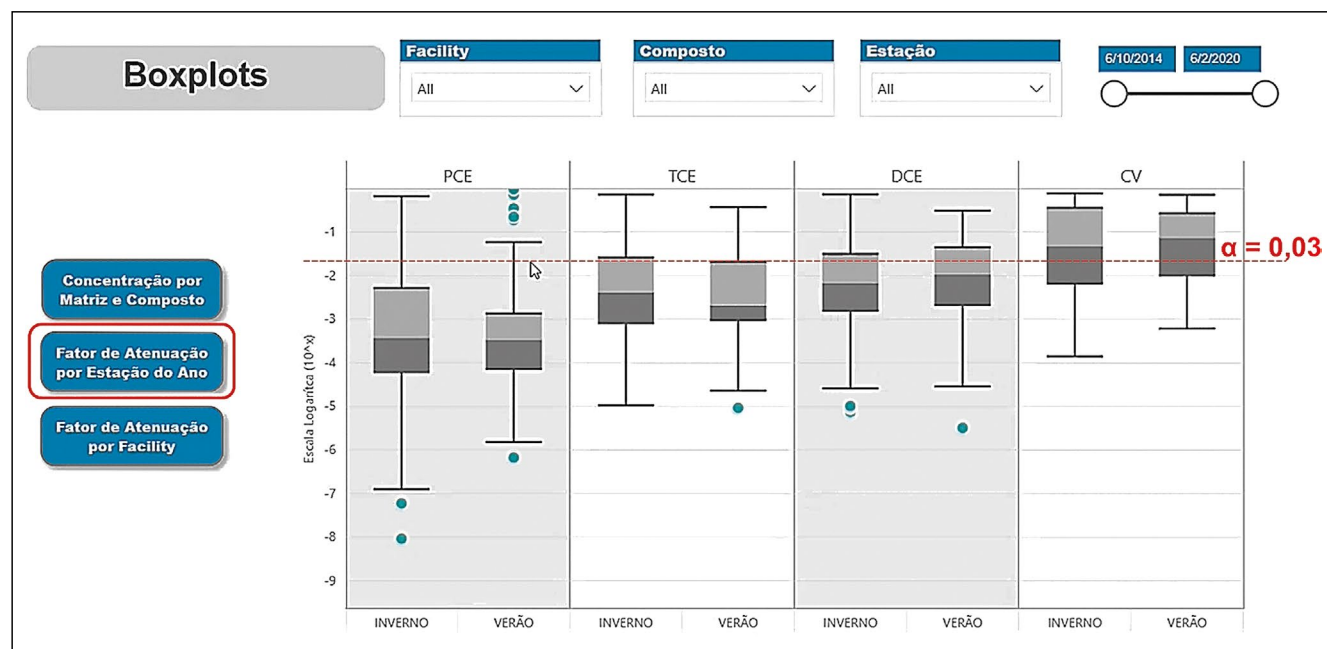


Figura 8. Resultados dos fatores de atenuação por composto e por época do ano.
Fonte: Elaborada pelos autores (2022).

dianas, dispersões e outliers, permitindo identificar onde os valores excedem ou não a linha de 0,03 de forma mais clara.

Os resultados obtidos nesta etapa piloto corroboram com as conclusões apresentadas por Yao *et al.* (2013), de que os fatores de atenuação podem variar em muitas ordens de grandeza, principalmente devido às condições das construções e pisos, em função das características do solo e em função do comportamento físico-químico de cada contaminante. Ou seja, os fatores de atenuação são inerentemente *site specific*, *building specific* e *compound specific*. Sendo assim, adotar um Fator α padrão, para todos os casos, não se mostra tecnicamente defensável.

Deve-se considerar também que o fator genérico de 0,03 sugerido pela USEPA (2012) foi derivado para cenários majoritariamente residenciais nos Estados Unidos. Apesar da natureza protetiva adotada pelo guia OSWER da USEPA (2015), a adoção deste fator para *sites* comerciais/industriais no Brasil deve ser aplicada com cautela, pois as diferenças estruturais nos pisos, fundações, circulação de ar, podem implicar em super estimativas dos riscos associados à intrusão de vapores. Além disso, seria oportuno que os órgãos ambientais brasileiros considerassem a possibilidade do estabelecimento de fatores de atenuação genéricos representativos para o meio físico dos estados, conforme as sugestões apresentadas no artigo de Brewer *et al.* (2014).

7. Discussão da Aplicação dos Resultados

Os resultados obtidos corroboram o argumento de que um fator de atenuação genérico se adequa especificamente como critério de varredura.

Em casos do GAC em que o Modelo Conceitual da área apresenta dados insuficientes e informações controversas, mostra-se plausível a adoção do $\alpha = 3\%$, como uma abordagem de triagem inicial. Entretanto, para *sites* com Modelos Conceituais mais robustos, elaborados a partir de investigações adequadas e abrangentes, mostra-se como tecnicamente mais adequada a definição de fatores de atenuação específicos, calculados com base em dados consistentes e levantamentos representativos do *site*, incluindo a coleta de amostras pareadas de *subslab* e ar ambiente, além do levantamento de informações sobre as características das edificações e a realização de vistorias no *site* para a identificação de potenciais caminhos preferenciais para a intrusão de vapores. Esta abordagem proporciona maior consistência técnica e segurança para tomadas de decisões, como na conclusão sobre a existência de risco e na consequente necessidade de implementação de medidas de intervenção, e no estabelecimento de metas de remediação mais apropriadas, conferindo justiça, economia e exequibilidade a todo o processo de remediação.

Quanto ao piloto do Sistema de Dados Ambientais, o projeto focou no uso de *Big Data* para dados laboratoriais sobre a temática de intrusão de vapores. No entanto, o grupo identificou diversas outras possibilidades de ampliação do escopo, tais como a inclusão de dados de nível da água, litologia, perfis construtivos de poços, propriedades hidráulicas de aquíferos, entre outros. Além disso, a consolidação de bases de dados, com foco nos fluxos de trabalho de órgãos ambientais, facilitaria os processos do GAC, tornando essas informações acessíveis a todos os

stakeholders. Isto possibilitaria ganhos em escala nos processos de gerenciamento, com redução de custos e geração de *insights* para decisões e uso de inteligência de dados em larga escala.

No decorrer do trabalho, o grupo também identificou algumas oportunidades de melhorias. Entre elas, pode-se mencionar a manutenção do SDA operacional além da PoC em sua fase piloto, bem como buscar linhas de financiamento para manutenção da infraestrutura digital operacional, além de incluir mais consultorias no projeto para definição em larga escala dos usos dos dados e criar procedimentos governamentais nos órgãos ambientais para estabelecer as bases e critérios padronizados, como dicionários de dados, a serem utilizados por todas as consultorias, laboratórios ou geradores de dados de forma geral.

8. Conclusões

Pode-se considerar que o protótipo desenvolvido nesta etapa piloto do Sistema de Dados Ambientais atingiu os objetivos esperados. Desde a etapa de articulação com os diversos parceiros, passando pela aplicação de técnicas estatística e computacionais de validação, tratamento, análise e visualização dos dados, até o lançamento de uma plataforma web piloto, o esforço colaborativo culminou na obtenção de resultados inéditos e com grande potencial de escalonamento.

A PoC do Sistema de Dados Ambientais (SDA) demonstrou, de forma robusta, o potencial de uma abordagem colaborativa e integrada para o gerenciamento de dados no âmbito do Gerenciamento de Áreas Contaminadas. O projeto atingiu seu objetivo geral ao avaliar a aplicabilidade do fator de

atenuação (α) sugerido pela USEPA, reproduzindo o estudo com uma base de dados de projetos na Região Metropolitana de São Paulo. Os resultados evidenciaram que os fatores de atenuação são, de fato, específicos para cada site, para cada edificação e para cada composto, com variações de até quatro ordens de grandeza, reforçando a fragilidade da aplicação de um valor genérico de 0,03 para todos os cenários e em todas as etapas do GAC.

No tocante aos objetivos específicos, o projeto demonstrou êxito em:

- Criar uma estrutura de dados e um dicionário padrão, por meio de *Electronic Data Deliverables* (EDD), para consolidação de dados provenientes de diferentes consultorias, garantindo qualidade, interoperabilidade e facilitando análises integradas;
- Desenvolver um sistema *online* (MVP): A implementação de uma plataforma *web* interativa evidenciou a viabilidade técnica e funcional do SDA, promovendo a visualização dinâmica dos resultados e o compartilhamento seguro dos dados entre *stakeholders*;
- Proporcionar a análise crítica do uso do fator de atenuação genérico: A comparação dos resultados específicos com o valor padrão de 0,03 evidenciou a necessidade de parâmetros mais específicos e adaptados às características de cada local e substância química, sugerindo oportunidades de melhoria dos métodos nos estudos de intrusão de vapores;
- Demonstrar a viabilidade de um sistema colaborativo: A experiência evidenciou que a cooperação entre instituições de ensino, consultorias e órgãos técnicos pode não apenas

otimizar a gestão dos dados ambientais, mas também fomentar a geração de *insights* e o avanço das práticas de GAC no Brasil.

Em síntese, o protótipo do SDA não só atingiu os objetivos propostos, mas também apontou para caminhos promissores de integração de grandes volumes de dados ambientais. A consolidação de um repositório colaborativo, aliado à padronização dos dados e à utilização de ferramentas analíticas avançadas, pode transformar a forma como as investigações de intrusão de vapores são conduzidas, subsidiando decisões mais seguras e eficientes. A continuidade e o aprimoramento desse sistema dependerão do estabelecimento de apoio institucional e financeiro, bem como da ampliação das parcerias, contribuindo significativamente para o avanço da gestão de dados ambientais no contexto do GAC brasileiro.

9. Agradecimentos

Os autores agradecem ao IPT pelo suporte técnico, institucional e material, cuja participação se respaldou no Plano de Desenvolvimento Institucional na Área de Transformação Digital (processo FAPESP 17/50343-2). Agradecimentos aos professores do SENAC pelo suporte técnico e acadêmico; à empresa Earthsoft pelo apoio institucional e técnico, e pela licença piloto do *software* Equis; bem como à indústria Zanettini Barossi, pelos dados cedidos, de modo aberto, para a prova de conceito inicial. Agradecimento especial ao apoio das consultorias ambientais, que cederam o insumo fundamental do trabalho (dados acreditados), porém optaram por não ser mencionadas nominalmente.

10. Referências Bibliográficas

ALASKA DEPARTMENT OF ENVIRONMENTAL CONSERVATION (ADEC). Vapor intrusion: fact sheet. **Spill Prevention and Response Division, Contaminated Sites Program**. 2023. Disponível em: <https://dec.alaska.gov/spar/csp/faq/vapor-intrusion/> Acesso em: 20 jan. 2025.

BECKLEY, L.; MCMASTERS, S.; COHEN, M.; CORDANO, D.; RAUCH, S.; MCHUGH, T. The California GeoTracker Database: a unique public resource for understanding contaminated sites. **Groundwater Monitoring & Remediation**, v. 42, n. 3, p. 105–115, 2022. DOI: <https://doi.org/10.1111/gwmr.12520>.

BREWER, R.; NAGASHIMA, J.; RIGBY, M.; SCHMIDT, M.; O'NEILL, H. Estimation of Generic Subslab Attenuation Factors for Vapor Intrusion Investigations. **Groundwater Monitoring & Remediation**, 2014. DOI: 10.1111/gwmr.12086.

COMPANHIA AMBIENTAL DO ESTADO DE SÃO PAULO (CETESB). **Decisão de Diretoria Nº 38/2017/C**, de 7-2-2017. São Paulo: CETESB. 2017.

GOUVÊA JÚNIOR, J. C. R.; BERTOLO, R. A.; HART, S. T. Intrusão de vapores do solo: breve histórico sobre desenvolvimento da tecnologia, cenário brasileiro e avanços recentes. **Holos Environment**, v. 18, n. 2, p. 240-270, 2018.

HOLTON, C.; LUO, H.; DAHLEN, P.; GORDER, K.; DETTENMAIER, E.; JOHNSON, P. C. Temporal variability of indoor air concentrations under natural conditions in a house overlying a dilute chlorinated solvent groundwater plume. **Environmental Science & Technology**, v. 47, n. 23, p. 13347-13354, 2013. DOI: 10.1021/es4024767.

HORST, J. et al. Digital Reinvention in the Remediation Industry. **Groundwater Monitoring & Remediation**, v. 38, n. 4, p. 15–28. 2018.

JOHNSON, P. C.; ETTINGER, R. A. Heuristic Model for Predicting the Intrusion Rate of Contaminant Vapors into Buildings. **Environmental Science & Technology**, v. 25, n. 8, p. 1445-1452, 1991.

MCHUGH, T.; LOLL, P.; EKLUND, B. Recent advances in vapor intrusion site investigations. **Journal of Environmental Management**, v. 204, p. 783–792, 2017.

NOBRE, C. et al. Vulnerability of Brazilian Megacities to Climate Change: The São Paulo Metropolitan Region (RMSP). In: MOTTA, R. S.; HAR- GRAVE, J.; LUEDEMANN, G.; GUTIERREZ, M. B. (Eds.). **Climate Change in Brazil: Economic, Social and Regulatory Aspects**. Brasília: IPEA, 2011. p. 197-219.

SOUTH AUSTRALIA. **Environment Protection Authority. Guidelines for the assessment and remediation of site contamination**. Adelaide: EPA South Australia, 2019. Disponível em: https://www.epa.sa.gov.au/files/13544_sc_groundwater_assessment.pdf. Acesso em: 05 fev. 2025.

SUTHERSAN, S.; GENTILE, M.; BELL, C.; QUINNAN, J.; HORST, J. Big Data and Environmental Remediation: Gaining Predictive Insights. **Groundwater Monitoring & Remediation**, v. 36, n. 2, p. 21–31, 2016.

U.S. ENVIRONMENTAL PROTECTION AGENCY (USEPA). User's Guide for the Johnson and Ettinger (1991) Model for Subsurface Vapor Intrusion into Buildings. **Environmental Quality Management**, Inc., 1997. Disponível em: https://www.epa.gov/sites/default/files/2015-11/documents/2004_0222_3phase_users_guide.pdf. Acesso em: 09 set. 2019.

U.S. ENVIRONMENTAL PROTECTION AGENCY (USEPA), Region 4. Environmental data submission: SESDGUID-106-R0. **Science and Ecosystem Support Division**, Athens, GA. 2010. 29 p. Disponível em: <https://www.epa.gov/sites/default/files/2015-06/documents/Environmental-Data-Submission.pdf>. Acesso em: 01 mar. 2025.

U.S. ENVIRONMENTAL PROTECTION AGENCY. **U.S. EPA's Vapor Intrusion Database: Evaluation and Characterization of Attenuation Factors for Chlorinated Volatile Organic Compounds and Re-**

sidential Buildings. EPA 530-R-10-002, 2012a. Disponível em: http://www.epa.gov/oswer/vapor-intrusion/documents/OSWER_2010_Database_Report_03-16-2012_Final.pdf.

U.S. ENVIRONMENTAL PROTECTION AGENCY. **OSWER Technical Guide for Assessing and Mitigating the Vapor Intrusion Pathway from Subsurface Vapor Sources to Indoor Air**. Washington, DC: USEPA, 2015. (OSWER Publication 9200.2-154). Disponível em: <https://www.epa.gov/sites/default/files/2015-09/documents/oswer-vapor-intrusion-technical-guide-final.pdf>. Acesso: 09 set. 2019.

U.S. ENVIRONMENTAL PROTECTION AGENCY. **Best Practices for Data Management - Technical Guide**. 2018. Disponível em: <https://semspub.epa.gov/work/HQ/100001798.pdf>. Acesso em: 09 set. 2019.

U.S. ENVIRONMENTAL PROTECTION AGENCY. Electronic Data Deliverable (EDD) - **Comprehensive Specification Manual 6. Region 2 Superfund**. EPA R2 EDD format version 4. 2019.

U.S. ENVIRONMENTAL PROTECTION AGENCY. **Electronic Data Deliverable (EDD) Comprehensive Manual Version 4.3. 2024**. Disponível em: https://www.epa.gov/system/files/documents/2023-02/R5ComprehensiveManual_20230215_0.pdf.

YAO, Y.; SHEN, R.; PENNELL, K. G.; SUUBERG, E. M. Examination of the Influence of Environmental Factors on Contaminant Vapor Concentration Attenuation Factors Using the U.S. EPA's Vapor Intrusion Database. **Environmental Science & Technology**, v. 47, n. 2, p. 906–913, 2013. DOI: 10.1021/es303441x.

YAO, Y. et al. Examining the Use of USEPA's Generic Attenuation Factor in Determining Groundwater Screening Levels for Vapor Intrusion. **Groundwater Monitoring and Remediation**, v. 38, n. 2, p. 79–89, 2018. DOI: 10.1111/gwmmr.12276.