

Dataset de Poços de Monitoramento de Gases da USP Leste

Rosa Virginia Encinas Quille^{1,2,3}, Gabriela dos Santos Luchetti Vieira⁴, Leandro Gomes de Freitas², Pedro Luiz Pizzigatti Corrêa^{1,3}, Solange Nice Alves de Souza^{1,3}, Alexandre Muselli Barbosa², Felipe Valencia de Almeida³

¹Escola de Artes, Ciências e Humanidades – Universidade de São Paulo

²Instituto de Pesquisas Tecnológicas

³Escola Politécnica - Universidade de São Paulo

⁴Instituto de Ciência e Tecnologia - Campus de Sorocaba – Universidade Estadual Paulista “Júlio de Mesquita Filho”

{encinas, pedro.correa, ssouza, fvalencia}@usp.br

{lfreitas, muselli}@ipt.br

{gabriela.luchetti}@unesp.br

Abstract. *The composition of gases in the soil can pose risks to human health and buildings. Time series analysis is an approach to understanding the impact of these gases better, as it allows us to identify patterns and trends over time. This dataset compiles historical data collected from the soils of USP Leste between 2014 and 2022, including methane, carbon dioxide, and hydrogen sulfide concentrations. In addition to providing a foundation for environmental studies, this dataset can be used to train machine learning models, allowing for the evaluation of their predictive capabilities and generalization potential.*

Resumo. *A composição de gases no solo pode oferecer riscos à saúde humana e às edificações. A análise de séries temporais é uma abordagem que permite compreender melhor o impacto desses gases, ao possibilitar a identificação de padrões e tendências ao longo do tempo. Este dataset reúne dados históricos coletados nos solos da USP Leste entre 2014 e 2022, incluindo concentrações de metano, dióxido de carbono e sulfeto de hidrogênio. Além de fornecer uma base para estudos ambientais, esse conjunto de dados pode ser utilizado no treinamento de modelos de aprendizado de máquina, possibilitando a avaliação de sua capacidade preditiva e potencial de generalização.*

1. Introdução

O Gerenciamento de Áreas Contaminadas permite avaliar os impactos ambientais em zonas urbanas, principalmente em regiões que historicamente receberam resíduos sólidos. A área da USP Leste, localizada na cidade de São Paulo e inserida na bacia do Alto Tietê, apresenta um caso, devido à presença de materiais de dragagem do Rio Tietê aterrados no terreno. Em função do elevado teor de matéria orgânica do material, o aterramento propiciou condições de decomposição anaeróbica, gerando gases como metano e dióxido de carbono, além de compostos orgânicos voláteis potencialmente tóxicos. O gás metano, em concentrações acima de 5% no subsolo, que corresponde ao seu limite inferior de

explosividade (LEL), pode representar um risco potencial aos usuários do campus, caso tenha volume e pressão suficientes para sua intrusão e acúmulo nas edificações sobrejacentes [ASTM 2016]. Desde 2014, foram implementadas soluções de engenharia e um programa contínuo de monitoramento por PMGs no campus, visando à segurança dos usuários. A rede permite a coleta sistemática de dados sobre gases em diferentes profundidades, apoiando a gestão ambiental com base em evidências técnicas e normativas [USP 2004, SERVMAR 2014, Nakao and Cagnon 2018, Weber Ambiental 2021].

Esses dados, obtidos ao longo do tempo, assumem relevância científica ao fornecerem uma base empírica para análises temporais, avaliação de riscos e desenvolvimento de modelos preditivos. A sistematização e disponibilização pública dessas medições ampliam as possibilidades de uso em estudos aplicados de classificação e regressão, com destaque para o uso de métodos estatísticos e de aprendizado de máquina voltados à compreensão da variabilidade espacial e temporal dos gases. Assim, o presente artigo tem como objetivo apresentar um conjunto de dados (*dataset*) contendo as medições realizadas nos PMGs da USP Leste entre 2014 e 2022, destacando seu potencial de aplicação em análises científicas, estratégias de remediação e ações de planejamento ambiental urbano.

Este artigo apresenta um *dataset* obtido entre 2014 e 2022 em PMGs da USP Leste, voltado ao monitoramento de gases no subsolo urbano. Os dados possuem relevância científica por subsidiar análises temporais, avaliação de riscos e modelagem preditiva. Sua sistematização favorece aplicações com métodos estatísticos e de aprendizado de máquina em estudos ambientais.

2. Coleta dos Dados

A coleta de dados foi realizada por meio de uma rede de PMGs, instalada a partir de 2014 em diferentes edificações do campus, conforme ilustrado na Figura 1. O monitoramento foi executado por empresas especializadas, como a Weber Ambiental e, posteriormente, a Engesolve Engenharia, sob supervisão técnica da USP. O monitoramento inclui gases como CH₄, CO₂, O₂, CO, H₂S, VOC e o LEL. A Tabela 1 descreve os equipamentos utilizados incluindo os parâmetros monitorados, os princípios de detecção empregados e os procedimentos operacionais adotados. A confiabilidade das medições é assegurada por protocolos de calibração regular, verificações em campo e registros padronizados. A qualidade dos dados foi assegurada por calibrações, verificações em campo e procedimentos de limpeza (exemplo, correção de valores tipo 79..%) e padronização dos dados brutos.



Figura 1. Distribuição dos PMGs nos edifícios do campus

Tabela 1. Equipamentos utilizados nos PMGs e suas principais características

Equipamento	Descrição e método de medição
GEM™ 5000 (QED / Landtec) Parâmetros: Metano (CH ₄), Dióxido de carbono (CO ₂), Oxigênio (O ₂), Monóxido de carbono (CO), Sulfeto de hidrogênio (H ₂ S) e Pressão diferencial. Infravermelho: CH ₄ , CO ₂ , e O ₂ . Eletroquímico: CO e H ₂ S. Transdutor: Pressão diferencial. Operação certificada pela MCERTS; faixa de temperatura: -10 °C a 50 °C, em atmosferas com até 21 % de O ₂ ; vazão mínima de purga: 550 tsi.	Equipamento principal utilizado nos PMGs para medição direta de gases e pressão. Permite cálculo do balanço gasoso e armazenamento automatizado dos dados. Calibrado semestralmente e o método de medição segue os seguintes passos: <ol style="list-style-type: none"> 1. Verifique se o equipamento está calibrado e com bateria acima de 25 %. 2. Ligue o equipamento e aguarde o autoteste automático. 3. Confirme a configuração correta do modo GEM5000, data, hora e ID do poço. 4. Instala o filtro coletor limpo e seco, garantindo canais de gás zerados. 5. Inicia a medição acoplando a mangueira do equipamento ao poço, começando pelo nível B (profundo) e depois o nível A (raso). 6. Abre a válvula do poço, purgue e mantenha a medição por 30–40 segundos. 7. Registra os dados automaticamente no equipamento e em planilha, conferindo periodicamente por amostragem.
MX6 iBrid (Industrial Scientific) e BW GasAlert Micro 5 (Honeywell BW) Parâmetros: Compostos orgânicos voláteis (VOC) e Limite inferior de explosividade (<i>Lower Explosive Limit – LEL</i>) Sensor PID (fotoionização, 10,6 eV): VOC — intervalo de medição de 0 ppm a 1.000 ppm. O sensor não é suficientemente sensível para detectar compostos como benzeno, butadieno e cloreto de vinila antes que estes excedam seus limites de toxicidade. Sensor catalítico: LEL — intervalo de medição de 0 % a 100 % LEL, com incremento de 1 % LEL (equivalente a aproximadamente 0 % v/v a 5 % v/v de CH ₄).	Equipamentos utilizados nos PMGs para a medição de LEL e VOC, com aplicação funcional equivalente. O modelo MX6 iBrid foi empregado entre abril de 2014 e fevereiro de 2018, sendo posteriormente substituído pelo BW5. Ambos operam com sensores catalíticos (LEL) e PID (VOC), aplicados em avaliações de inflamabilidade e emissão de compostos orgânicos voláteis. Passos do método de medição: <ol style="list-style-type: none"> 1. Certificar-se de que o equipamento está devidamente calibrado. 2. No início do dia de medição da-se início ao teste da bomba, utilizando ar ambiente livre de contaminantes (O₂ ≥ 20,9 %). 3. Selecionar no equipamento o parâmetro desejado para medição: VOC ou LEL. 4. Posicionar o sensor do equipamento no ambiente ou próximo à fonte de gás a ser analisada. 5. Observar a leitura em tempo real no visor digital. A concentração detectada será exibida automaticamente, e o valor deve ser registrado manualmente. 6. Garantir que o equipamento esteja operando com o filtro de proteção contra água com porosidade de 1,0 µm, assegurando a integridade da medição.

3. Apresentação do Dataset

O dataset está disponível em um arquivo `xlsx`, contendo 16 colunas e 100.570 linhas, com um total de aproximadamente 6 kB de dados. As variáveis relacionadas aos gases medidos no solo são CH₄ (%), CO₂ (%), O₂ (%), H₂S (ppm), CO (ppm), VOC (ppm) e LEL (%) sendo que as medições podem ser representadas em porcentagem volumétrica (por exemplo, 2 %) ou em partes por milhão (por exemplo, 0,3 ppm). A Tabela 2 apresenta todas as variáveis do *dataset*, juntamente com suas unidades de medida e descrições. A resolução temporal dos dados é semanal, e o arquivo também inclui a identificação do poço onde cada medição foi realizada. O período de coleta dos dados abrange de 8 de abril de 2014 a 30 de junho de 2022. O *dataset* oficialmente disponível em: <https://doi.org/10.60748/uspgasmon1weach.00>. Para acessá-lo, é necessário cadastrar seu ORCID e aguardar a ativação pelo sistema Datamap. Alternativamente, uma cópia pública está acessível em: <https://wds.poli.usp.br/sga/discovery/>.

4. Considerações Finais

Este artigo apresentou um conjunto de dados sobre gases no solo, coletados ao longo de um período de oito anos na região Leste da cidade de São Paulo, especificamente no campus USP Leste. Espera-se que esses dados possam contribuir para as análises de pesquisadores e incentivar o desenvolvimento de experimentos em aprendizado de máquina. Considerando a periodicidade semanal dos dados, análises como decomposição de séries temporais, modelagem preditiva (ARIMA, SARIMA, LSTM) e correlações entre variáveis podem revelar padrões sazonais e processos geoquímicos relevantes. A inclusão de variáveis externas, como dados meteorológicos, pode auxiliar na compreensão de variações abruptas nas concentrações gasosas.

Tabela 2. Variáveis do dataset de PMGs

Variável	Unidade	Observações	Descrição
building	Não aplica	Identificador alfanumérico.	Identificador da edificação associada ao PMG. Local onde o poço está localizado.
well	Não aplica	Identificador único por poço (Identificador PMG).	Estrutura utilizada para a medição das concentrações de gases geradas no subsolo.
depth_level	Não aplica	Níveis A (0,3 m), B (1,0 m), e C (1,3 m).	Profundidade do poço utilizado para a medição dos gases.
date	Não aplica	Formato ISO 8601 (AAAA-MM-DD HH:MM:SS).	Data e horário em que foi feita a medição dos gases.
ch4	% vol	Componente do balanço gasoso e indicador de atividade microbiológica anaeróbica	Concentração de metano expressa em porcentagem volumétrica no ar solo no ponto de amostragem.
ch4_peak	% vol	Valor máximo registrado	Pico de concentração de metano em porcentagem.
co2	% vol	Componente do balanço gasoso e indicador de atividade microbiológica geral	Concentração de dióxido de carbono expressa em porcentagem volumétrica no ar do solo no ponto de amostragem.
o2	% vol	Componente do balanço gasoso e indicador da condição redox e da atividade microbiológica aeróbica	Concentração de oxigênio expressa em porcentagem volumétrica no ar do solo no ponto de amostragem.
h2s	ppm	Gás tóxico associado à atividade microbiológica anaeróbica redutora de enxofre e de odor característico	Concentração de sulfeto de hidrogênio no ar do solo no ponto de amostragem.
co	ppm	Gás inflamável e tóxico, associado a processos de combustão incompleta e degradação térmica de matéria orgânica	Concentração de monóxido de carbono no solo no ponto de amostragem.
bal	% vol	Soma estimada dos demais gases, principalmente nitrogênio, além de gases traço	Balanço atmosférico, somatória dos demais gases, principalmente nitrogênio, para compor 100% da amostra, expresso em porcentagem volumétrica.
well_pressure	mbar	Pressão diferencial interna no PMG	Pressão diferencial medida no poço de monitoramento, expressa em milibar (mbar).
field_notes	Não aplica	Observações pontuais	Anotações feitas durante a coleta em campo.
voc	ppm	Mistura de compostos orgânicos voláteis totais, potencialmente tóxicos, com alta volatilidade	Concentração de compostos orgânicos voláteis totais, medida por sensor de fotoionização (PID), expressa em partes por milhão (ppm).
lel	%	Avaliação de inflamabilidade	Percentual do Limite inferior de explosividade. Parâmetro calculado com base nas concentrações de CH ₄ e O ₂ , conforme curvas de inflamabilidade.
comments	Não aplica	Complemento técnico	Comentários técnicos registrados pelo operador.

Agradecimentos

Os autores agradecem à geóloga Fabiana Cagnon, assessora técnica da Prefeitura do Campus USP Leste, pela validação técnica das informações utilizadas neste trabalho. Agradecem também pelo apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), por meio dos processos 19/21693-0, 17/50343-2, e 17/17047-0.

Referências

ASTM (2016). Standard guide for evaluating potential hazard as a result of methane in the vadose zone. ASTM Standard E2993-16. ASTM Committee E50 on Environmental Assessment, Risk Management and Corrective Action.

Nakao, O. S. and Cagnon, F. A. (2018). *Panorama das ações ambientais na USP Leste até 2017*. Universidade de São Paulo, Superintendência do Espaço Físico, São Paulo, 2 edition. Versão digital disponível em PDF.

SERVMAR (2014). Relatório de investigação detalhada, avaliação de risco à saúde humana e plano de intervenção na ai-01 e investigação detalhada de gases – ma/12936/14/bls. Technical report, SERVMAR.

USP (2004). Relatório ambiental preliminar: Campus usp leste. Technical report, USP.

Weber Ambiental (2021). As-built completo dos sistemas de ventilação de gases instalados na USP Leste – VS.07. Relatório Técnico Projeto 311.1264.14-EGS.VS.07, Universidade de São Paulo, Superintendência do Espaço Físico (SEF), São Paulo. Contrato nº 10/2014. Processo nº 14.1.607.82.2.