

Evaluation of Machine Learning Regression Techniques for Analyzing Contaminated Soils

B ICIVIL
International Conference
On Machine Learning
2025

Rosa Virginia Encinas Quille, Felipe Valencia de Almeida University of Sao Paulo, São Paulo, Brazil

July 13-19, 2025

Abstract

Environmental monitoring of contaminated urban soils is essential for risk management and decision-making. This study applies supervised Machine Learning regression models to predict CH4 concentrations using co-measured gases (CO2, O2, H2S, CO) from 128 gas monitoring wells across 14 buildings at EACH-USP (São Paulo, Brazil), collected from 2014 to 2022.

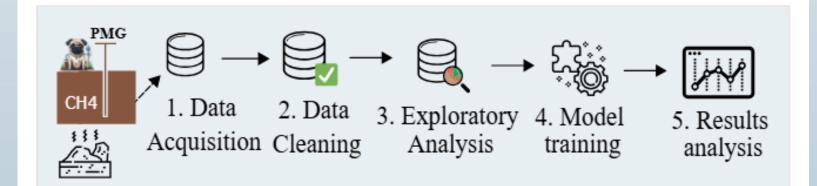
Five models were tested: Linear Regression, k-NN, Decision Tree, Random Forest, and XGBoost. Model performance was evaluated using R², MAE, and RMSE.

Random Forest showed superior results in the per-well analysis (Experiment 1), while model performance in the combined-well setting (Experiment 2) varied depending on feature composition, with DT, RF, and XGBoost performing best in different scenarios.

Introduction

- Soil contamination poses major risks to public health and urban management, particularly in densely populated areas. Polluted sites—often known as brownfields—can devalue land, pollute water resources, and threaten ecosystems.
- Environmental investigations generate large volumes of data during the diagnosis, remediation, and monitoring phases. However, conventional analysis methods are often costly, time-consuming, and dependent on laboratory procedures.
- Machine Learning (ML) regression techniques offer a promising alternative for analyzing contaminated soils. Prior studies have shown the potential of models like Random Forest, Decision Trees, and Gradient Boosting for predicting pollutant behavior.
- This study explores the use of ML regression to predict methane (CH4) concentrations in contaminated soils, using other co-measured gases (CO2, O2, H2S, CO) as input features. The goal is to support more accurate and scalable environmental diagnostics.

Methodology



1. Data Collection

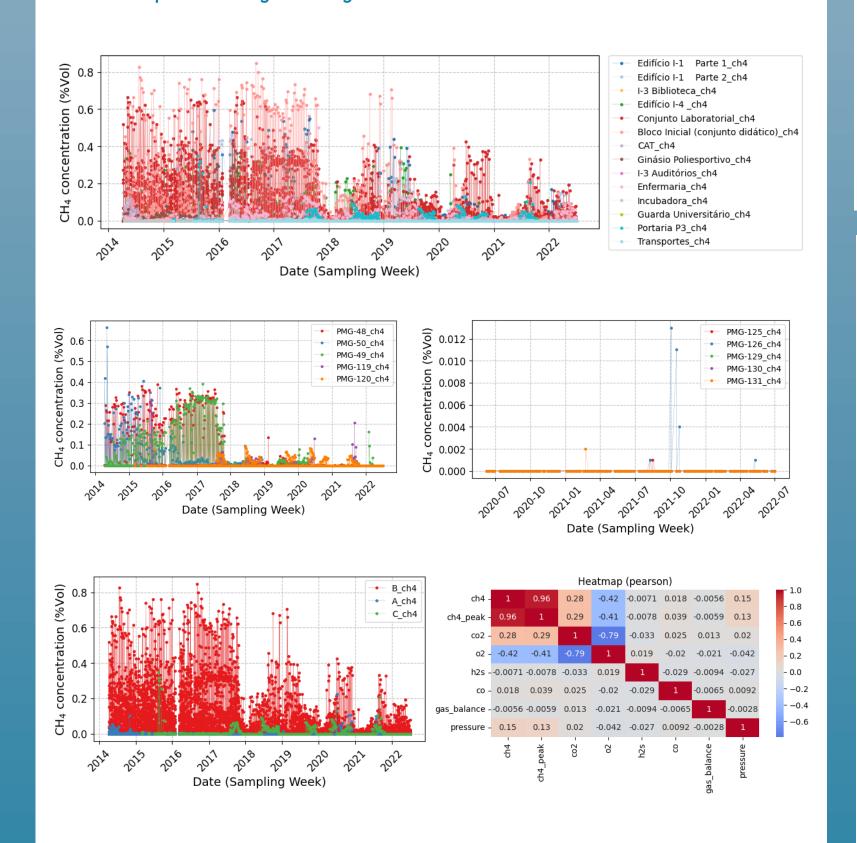
We used in situ gas concentration data collected from 128 monitoring wells across 14 buildings at EACH-USP (São Paulo) between 2014 and 2022 (totaling 100,570 records). Gases monitored include:

CH4, CO2, O2, H2S, and CO.

2. Data Cleaning

- Merged fragmented files into a unified dataset
- Standardized missing values (NA, #N/A, empty strings)
- Normalized formats (percent to decimals)
- Removed irrelevant columns
- Treated outliers and unified categorical labels
- Converted timestamps to standard format

3. Exploratory Analysis



A (shallow, 0.30 m, B (intermediate, 1.00 m), and C (deep, 1.30 m).

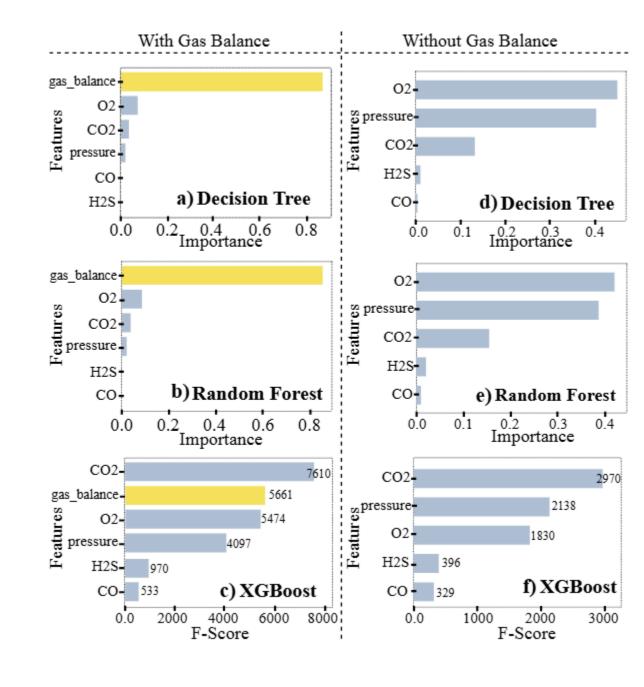
Results

Experiment 1:

- Models were trained per well to evaluate their ability to capture local gas dynamics.
- Tree-based models (Random Forest, Decision Tree, XGBoost) consistently outperformed others across most wells.
- XGBoost achieved the highest accuracy when using the gas balance feature (R² > 0.92 in high-density wells).
 Some low-data wells showed perfect or negative R², indicating potential overfitting or noise.
- Excluding gas balance led to reduced accuracy but improved interpretability.

			1011	MGs		ı			Воттом			
		PMG-119			PMG-120	i		PMG-126			PMG-130	
MODEL	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
LR	0.0057	0.0199	0.2880	0.0030	0.0090	0.5350	0.0000	0.0000	0.9594	0.0000	0.0000	1.0000
KNN	0.0048	0.0229	0.0616	0.0029	0.0096	0.4641	0.0003	0.0008	-25.8308	0.0000	0.0000	1.0000
DT	0.0027	0.0092	0.8466	0.0016	0.0065	0.7514	0.0001	0.0002	-1.3077	0.0000	0.0000	1.0000
RF	0.0020	0.0066	0.9217	0.0015	0.0065	0.7549	0.0001	0.0002	-0.4892	0.0000	0.0000	1.0000
XGB	0.0019	0.0064	0.9268	0.0015	0.0061	0.7819	0.0001	0.0002	-0.0229	0.0000	0.0000	1.0000
		PMG-48			PMG-49			PMG-129			PMG-131	
MODEL	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
LR	0.0348	0.0522	0.7632	0.0302	0.0502	0.6592	0.0000	0.0000	1.0000	0.0001	0.0003	-0.0256
KNN	0.0173	0.0450	0.8243	0.0211	0.0494	0.6692	0.0000	0.0000	1.0000	0.0001	0.0003	-0.0256
DT	0.0061	0.0165	0.9763	0.0061	0.0147	0.9706	0.0000	0.0000	1.0000	0.0001	0.0003	-0.0256
RF	0.0061	0.0137	0.9838	0.0063	0.0159	0.9658	0.0000	0.0000	1.0000	0.0001	0.0003	-0.0256
XGB	0.0053	0.0137	0.9838	0.0059	0.0145	0.9715	0.0000	0.0000	1.0000	0.0001	0.0003	-0.0256
Mann	MAE	PMG-50	R^2				MAE	PMG-125	R^2			
MODEL	MAE	RMSE					MAE	RMSE				
LR	0.0258	0.0692	0.2178				0.0000	0.0001	0.4936			
KNN	0.0094	0.0450	0.6697				0.0000	0.0002	-0.1077			
DT	0.0039	0.0145	0.9656				0.0001	0.0003	-2.0769			
RF XGB	0.0039 0.0037	0.0139 0.0145	0.9684 0.9657				0.0001 0.0000	0.0002 0.0002	-0.5694 -0.0142			
			\	Witho	ut Gas	Balar	nce					
					ut Gas	Balar	nce		Вотто	M PMGs		
		PMG_110	Тор	Vitho PMGs			nce 	PMG 126		om PMGs	PMG_13	
Model	MAE	PMG-119 RMSE	Тор		PMG-120			PMG-126 RMSE		MAE	PMG-13) R ²
		RMSE	TOP R^2	PMGs MAE	PMG-120	R^2	MAE		R^2		RMSE	R^2
LR	MAE 0.0057 0.0052		R ² 0.2880	PMGs MAE 0.0030	PMG-120 RMSE 0.0090	R^2 0.5350	MAE 0.0003	RMSE	R ² -8.6053	MAE	0.0000	1.0000
LR KNN	0.0057 0.0052	0.0199 0.0216	R ² 0.2880 0.1652	MAE 0.0030 0.0030	PMG-120 RMSE 0.0090 0.0099	R ² 0.5350 0.4329	MAE 0.0003 0.0003	0.0005	-8.6053 -15.7452	MAE 0.0000	0.0000 0.0000	1.0000 1.0000
LR KNN DT	0.0057	0.0199	R ² 0.2880	PMGs MAE 0.0030	PMG-120 RMSE 0.0090	R^2 0.5350	MAE 0.0003	0.0005 0.0006	R ² -8.6053	MAE 0.0000 0.0000	0.0000 0.0000	R^2 1.0000 1.0000 1.0000
LR KNN	0.0057 0.0052 0.0037	0.0199 0.0216 0.0173	TOP R ² 0.2880 0.1652 0.4625	PMGs MAE 0.0030 0.0030 0.0030 0.0025	PMG-120 RMSE 0.0090 0.0099 0.0091	R ² 0.5350 0.4329 0.5215	MAE 0.0003 0.0003 0.0001	0.0005 0.0006 0.0002	-8.6053 -15.7452 -1.3077	MAE 0.0000 0.0000 0.0000	0.0000 0.0000 0.0000	R^2 1.0000 1.0000 1.0000 1.0000
LR KNN DT RF	0.0057 0.0052 0.0037 0.0048	0.0199 0.0216 0.0173 0.0191	R ² 0.2880 0.1652 0.4625 0.3445 0.0519	PMGs MAE 0.0030 0.0030 0.0025 0.0022	PMG-120 RMSE 0.0090 0.0099 0.0091 0.0080	R ² 0.5350 0.4329 0.5215 0.6310 0.5737	MAE 0.0003 0.0003 0.0001 0.0001	0.0005 0.0006 0.0002 0.0002	-8.6053 -15.7452 -1.3077 -0.2628 -0.0229	MAE 0.0000 0.0000 0.0000 0.0000	0.0000 0.0000 0.0000 0.0000	1.0000 1.0000 1.0000 1.0000 1.0000
LR KNN DT RF	0.0057 0.0052 0.0037 0.0048	0.0199 0.0216 0.0173 0.0191 0.0230	TOP R ² 0.2880 0.1652 0.4625 0.3445	PMGs MAE 0.0030 0.0030 0.0025 0.0022	PMG-120 RMSE 0.0090 0.0099 0.0091 0.0080 0.0086	R ² 0.5350 0.4329 0.5215 0.6310	MAE 0.0003 0.0003 0.0001 0.0001	RMSE 0.0005 0.0006 0.0002 0.0002 0.0002	-8.6053 -15.7452 -1.3077 -0.2628 -0.0229	MAE 0.0000 0.0000 0.0000 0.0000	RMSE 0.0000 0.0000 0.0000 0.0000 0.0000	1.0000 1.0000 1.0000 1.0000 1.0000
LR KNN DT RF XGB	0.0057 0.0052 0.0037 0.0048 0.0053 MAE	RMSE 0.0199 0.0216 0.0173 0.0191 0.0230 PMG-48 RMSE 0.0523	R ² 0.2880 0.1652 0.4625 0.3445 0.0519 R ² 0.7630	PMGs MAE 0.0030 0.0030 0.0025 0.0022 0.0024 MAE 0.0302	PMG-120 RMSE 0.0090 0.0099 0.0091 0.0080 0.0086 PMG-49 RMSE 0.0502	R ² 0.5350 0.4329 0.5215 0.6310 0.5737 R ² 0.6589	MAE 0.0003 0.0003 0.0001 0.0001 MAE 0.0000	RMSE 0.0005 0.0006 0.0002 0.0002 PMG-129 RMSE 0.0000	-8.6053 -15.7452 -1.3077 -0.2628 -0.0229	MAE 0.0000 0.0000 0.0000 0.0000 0.0000 MAE 0.0001	RMSE 0.0000 0.0000 0.0000 0.0000 PMG-13 RMSE 0.0003	R ² 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000
LR KNN DT RF XGB MODEL	0.0057 0.0052 0.0037 0.0048 0.0053 MAE 0.0348 0.0231	RMSE 0.0199 0.0216 0.0173 0.0191 0.0230 PMG-48 RMSE 0.0523 0.0535	R ² 0.2880 0.1652 0.4625 0.3445 0.0519 R ² 0.7630 0.7513	PMGS MAE 0.0030 0.0030 0.0025 0.0022 0.0024 MAE 0.0302 0.0307	PMG-120 RMSE 0.0090 0.0099 0.0091 0.0080 0.0086 PMG-49 RMSE 0.0502 0.0680	R ² 0.5350 0.4329 0.5215 0.6310 0.5737 R ² 0.6589 0.3739	MAE 0.0003 0.0003 0.0001 0.0001 0.0001 MAE 0.0000 0.0000	RMSE 0.0005 0.0006 0.0002 0.0002 0.0002 PMG-125 RMSE 0.0000 0.0000	R ² -8.6053 -15.7452 -1.3077 -0.2628 -0.0229 R ² 1.0000 1.0000	MAE 0.0000 0.0000 0.0000 0.0000 0.0000 MAE 0.0001	RMSE 0.0000 0.0000 0.0000 0.0000 0.0000 PMG-13 RMSE 0.0003 0.0003	R ² 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000
LR KNN DT RF XGB MODEL LR KNN DT	0.0057 0.0052 0.0037 0.0048 0.0053 MAE 0.0348 0.0231 0.0250	RMSE 0.0199 0.0216 0.0173 0.0191 0.0230 PMG-48 RMSE 0.0523 0.0535 0.0563	R ² 0.2880 0.1652 0.4625 0.3445 0.0519 R ² 0.7630 0.7513 0.7253	PMGs MAE 0.0030 0.0030 0.0022 0.0022 MAE 0.0302 0.0307 0.0244	PMG-120 RMSE 0.0090 0.0099 0.0091 0.0080 0.0086 PMG-49 RMSE 0.0502 0.0680 0.0611	R ² 0.5350 0.4329 0.5215 0.6310 0.5737 R ² 0.6589 0.3739 0.4942	MAE 0.0003 0.0003 0.0001 0.0001 0.0001 MAE 0.0000 0.0000	RMSE 0.0005 0.0006 0.0002 0.0002 0.0002 PMG-129 RMSE 0.0000 0.0000 0.0000	R ² -8.6053 -15.7452 -1.3077 -0.2628 -0.0229 R ² 1.0000 1.0000 1.0000	MAE 0.0000 0.0000 0.0000 0.0000 0.0000 MAE 0.0001 0.0001	RMSE 0.0000 0.0000 0.0000 0.0000 0.0000 PMG-13 RMSE 0.0003 0.0003	R ² 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000
LR KNN DT RF XGB MODEL LR KNN DT RF	0.0057 0.0052 0.0037 0.0048 0.0053 MAE 0.0348 0.0231 0.0250 0.0157	RMSE 0.0199 0.0216 0.0173 0.0191 0.0230 PMG-48 RMSE 0.0523 0.0535 0.0563 0.0385	TOP R ² 0.2880 0.1652 0.4625 0.3445 0.0519 R ² 0.7630 0.7513 0.7253 0.8713	PMGs MAE 0.0030 0.0030 0.0022 0.0024 MAE 0.0302 0.0307 0.0244 0.0176	PMG-120 RMSE 0.0090 0.0099 0.0091 0.0080 0.0086 PMG-49 RMSE 0.0502 0.0680 0.0611 0.0399	R ² 0.5350 0.4329 0.5215 0.6310 0.5737 R ² 0.6589 0.3739 0.4942 0.7850	MAE 0.0003 0.0003 0.0001 0.0001 0.0001 MAE 0.0000 0.0000 0.0000 0.0000	RMSE 0.0005 0.0006 0.0002 0.0002 0.0002 PMG-129 RMSE 0.0000 0.0000 0.0000 0.0000	R ² -8.6053 -15.7452 -1.3077 -0.2628 -0.0229 R ² 1.0000 1.0000 1.0000	MAE 0.0000 0.0000 0.0000 0.0000 0.0000 MAE 0.0001 0.0001	RMSE 0.0000 0.0000 0.0000 0.0000 PMG-13 RMSE 0.0003 0.0003 0.0003	R ² 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 R ² -0.0250 -0.0250 -0.0250
LR KNN DT RF XGB MODEL LR KNN DT	0.0057 0.0052 0.0037 0.0048 0.0053 MAE 0.0348 0.0231 0.0250	RMSE 0.0199 0.0216 0.0173 0.0191 0.0230 PMG-48 RMSE 0.0523 0.0535 0.0563	R ² 0.2880 0.1652 0.4625 0.3445 0.0519 R ² 0.7630 0.7513 0.7253	PMGs MAE 0.0030 0.0030 0.0022 0.0022 MAE 0.0302 0.0307 0.0244	PMG-120 RMSE 0.0090 0.0099 0.0091 0.0080 0.0086 PMG-49 RMSE 0.0502 0.0680 0.0611	R ² 0.5350 0.4329 0.5215 0.6310 0.5737 R ² 0.6589 0.3739 0.4942	MAE 0.0003 0.0003 0.0001 0.0001 0.0001 MAE 0.0000 0.0000	RMSE 0.0005 0.0006 0.0002 0.0002 0.0002 PMG-129 RMSE 0.0000 0.0000 0.0000	R ² -8.6053 -15.7452 -1.3077 -0.2628 -0.0229 R ² 1.0000 1.0000 1.0000	MAE 0.0000 0.0000 0.0000 0.0000 0.0000 MAE 0.0001 0.0001	RMSE 0.0000 0.0000 0.0000 0.0000 0.0000 PMG-13 RMSE 0.0003 0.0003	R ² 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000
LR KNN DT RF XGB MODEL LR KNN DT RF XGB	0.0057 0.0052 0.0037 0.0048 0.0053 MAE 0.0348 0.0231 0.0250 0.0166	RMSE 0.0199 0.0216 0.0173 0.0191 0.0230 PMG-48 RMSE 0.0523 0.0535 0.0563 0.0385 0.0418 PMG-50	R ² 0.2880 0.1652 0.4625 0.3445 0.0519 R ² 0.7630 0.7513 0.7253 0.8713 0.8488	PMGs MAE 0.0030 0.0030 0.0022 0.0024 MAE 0.0302 0.0307 0.0244 0.0176	PMG-120 RMSE 0.0090 0.0099 0.0091 0.0080 0.0086 PMG-49 RMSE 0.0502 0.0680 0.0611 0.0399	R ² 0.5350 0.4329 0.5215 0.6310 0.5737 R ² 0.6589 0.3739 0.4942 0.7850	MAE 0.0003 0.0003 0.0001 0.0001 0.0001 MAE 0.0000 0.0000 0.0000 0.0000	RMSE 0.0005 0.0006 0.0002 0.0002 0.0002 PMG-125 RMSE 0.0000 0.0000 0.0000 PMG-125	R ² -8.6053 -15.7452 -1.3077 -0.2628 -0.0229 R ² 1.0000 1.0000 1.0000 1.0000	MAE 0.0000 0.0000 0.0000 0.0000 0.0000 MAE 0.0001 0.0001	RMSE 0.0000 0.0000 0.0000 0.0000 PMG-13 RMSE 0.0003 0.0003 0.0003	R ² 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 R ² -0.0250 -0.0250 -0.0250
LR KNN DT RF XGB MODEL LR KNN DT RF	0.0057 0.0052 0.0037 0.0048 0.0053 MAE 0.0348 0.0231 0.0250 0.0157	RMSE 0.0199 0.0216 0.0173 0.0191 0.0230 PMG-48 RMSE 0.0523 0.0535 0.0563 0.0385 0.0418 PMG-50 RMSE	TOP R ² 0.2880 0.1652 0.4625 0.3445 0.0519 R ² 0.7630 0.7513 0.7253 0.8713	PMGs MAE 0.0030 0.0030 0.0022 0.0024 MAE 0.0302 0.0307 0.0244 0.0176	PMG-120 RMSE 0.0090 0.0099 0.0091 0.0080 0.0086 PMG-49 RMSE 0.0502 0.0680 0.0611 0.0399	R ² 0.5350 0.4329 0.5215 0.6310 0.5737 R ² 0.6589 0.3739 0.4942 0.7850	MAE 0.0003 0.0003 0.0001 0.0001 0.0001 MAE 0.0000 0.0000 0.0000 0.0000	RMSE 0.0005 0.0006 0.0002 0.0002 0.0002 PMG-125 RMSE 0.0000 0.0000 0.0000 PMG-125 RMSE	R ² -8.6053 -15.7452 -1.3077 -0.2628 -0.0229 R ² 1.0000 1.0000 1.0000 1.0000	MAE 0.0000 0.0000 0.0000 0.0000 0.0000 MAE 0.0001 0.0001	RMSE 0.0000 0.0000 0.0000 0.0000 PMG-13 RMSE 0.0003 0.0003 0.0003	R ² 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 R ² -0.0255 -0.0255 -0.0255
LR KNN DT RF XGB MODEL LR KNN DT RF XGB	0.0057 0.0052 0.0037 0.0048 0.0053 MAE 0.0348 0.0231 0.0250 0.0157 0.0166 MAE	RMSE 0.0199 0.0216 0.0173 0.0191 0.0230 PMG-48 RMSE 0.0523 0.0535 0.0563 0.0385 0.0418 PMG-50 RMSE 0.0693	R ² 0.2880 0.1652 0.4625 0.3445 0.0519 R ² 0.7630 0.7513 0.7253 0.8713 0.8488 R ² 0.2160	PMGs MAE 0.0030 0.0030 0.0022 0.0024 MAE 0.0302 0.0307 0.0244 0.0176	PMG-120 RMSE 0.0090 0.0099 0.0091 0.0080 0.0086 PMG-49 RMSE 0.0502 0.0680 0.0611 0.0399	R ² 0.5350 0.4329 0.5215 0.6310 0.5737 R ² 0.6589 0.3739 0.4942 0.7850	MAE 0.0003 0.0003 0.0001 0.0001 0.0001 MAE 0.0000 0.0000 0.0000 0.0000 MAE 0.0000	RMSE 0.0005 0.0006 0.0002 0.0002 0.0002 PMG-125 RMSE 0.0000 0.0000 0.0000 0.0000 PMG-125 RMSE	R ² -8.6053 -15.7452 -1.3077 -0.2628 -0.0229 R ² 1.0000 1.0000 1.0000 1.0000 1.0000	MAE 0.0000 0.0000 0.0000 0.0000 0.0000 MAE 0.0001 0.0001	RMSE 0.0000 0.0000 0.0000 0.0000 PMG-13 RMSE 0.0003 0.0003 0.0003	R ² 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 R ² -0.025 -0.025 -0.025 -0.025
LR KNN DT RF XGB MODEL LR KNN DT RF XGB	0.0057 0.0052 0.0037 0.0048 0.0053 MAE 0.0348 0.0231 0.0250 0.0157 0.0166	RMSE 0.0199 0.0216 0.0173 0.0191 0.0230 PMG-48 RMSE 0.0523 0.0535 0.0563 0.0385 0.0418 PMG-50 RMSE	R ² 0.2880 0.1652 0.4625 0.3445 0.0519 R ² 0.7630 0.7513 0.7253 0.8713 0.8488	PMGs MAE 0.0030 0.0030 0.0022 0.0024 MAE 0.0302 0.0307 0.0244 0.0176	PMG-120 RMSE 0.0090 0.0099 0.0091 0.0080 0.0086 PMG-49 RMSE 0.0502 0.0680 0.0611 0.0399	R ² 0.5350 0.4329 0.5215 0.6310 0.5737 R ² 0.6589 0.3739 0.4942 0.7850	MAE 0.0003 0.0003 0.0001 0.0001 MAE 0.0000 0.0000 0.0000 0.0000 MAE	RMSE 0.0005 0.0006 0.0002 0.0002 0.0002 PMG-125 RMSE 0.0000 0.0000 0.0000 PMG-125 RMSE	R ² -8.6053 -15.7452 -1.3077 -0.2628 -0.0229 R ² 1.0000 1.0000 1.0000 1.0000	MAE 0.0000 0.0000 0.0000 0.0000 0.0000 MAE 0.0001 0.0001	RMSE 0.0000 0.0000 0.0000 0.0000 PMG-13 RMSE 0.0003 0.0003 0.0003	R ² 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 R ² -0.0255 -0.0255 -0.0255
LR KNN DT RF XGB MODEL LR KNN DT RF XGB	0.0057 0.0052 0.0037 0.0048 0.0053 MAE 0.0348 0.0231 0.0250 0.0157 0.0166 MAE	RMSE 0.0199 0.0216 0.0173 0.0191 0.0230 PMG-48 RMSE 0.0523 0.0535 0.0563 0.0385 0.0418 PMG-50 RMSE 0.0693	R ² 0.2880 0.1652 0.4625 0.3445 0.0519 R ² 0.7630 0.7513 0.7253 0.8713 0.8488 R ² 0.2160	PMGs MAE 0.0030 0.0030 0.0022 0.0024 MAE 0.0302 0.0307 0.0244 0.0176	PMG-120 RMSE 0.0090 0.0099 0.0091 0.0080 0.0086 PMG-49 RMSE 0.0502 0.0680 0.0611 0.0399	R ² 0.5350 0.4329 0.5215 0.6310 0.5737 R ² 0.6589 0.3739 0.4942 0.7850	MAE 0.0003 0.0003 0.0001 0.0001 0.0001 MAE 0.0000 0.0000 0.0000 0.0000 MAE 0.0000	RMSE 0.0005 0.0006 0.0002 0.0002 0.0002 PMG-125 RMSE 0.0000 0.0000 0.0000 0.0000 PMG-125 RMSE	R ² -8.6053 -15.7452 -1.3077 -0.2628 -0.0229 R ² 1.0000 1.0000 1.0000 1.0000 1.0000	MAE 0.0000 0.0000 0.0000 0.0000 0.0000 MAE 0.0001 0.0001	RMSE 0.0000 0.0000 0.0000 0.0000 PMG-13 RMSE 0.0003 0.0003 0.0003	R ² 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 R ² -0.0250 -0.0250 -0.0250
LR KNN DT RF XGB MODEL LR KNN DT RF XGB	0.0057 0.0052 0.0037 0.0048 0.0053 MAE 0.0348 0.0231 0.0250 0.0157 0.0166 MAE 0.0258 0.0197	RMSE 0.0199 0.0216 0.0173 0.0191 0.0230 PMG-48 RMSE 0.0523 0.0535 0.0563 0.0385 0.0418 PMG-50 RMSE 0.0693 0.0633	R ² 0.2880 0.1652 0.4625 0.3445 0.0519 R ² 0.7630 0.7513 0.7253 0.8713 0.8488 R ² 0.2160 0.3461	PMGs MAE 0.0030 0.0030 0.0022 0.0024 MAE 0.0302 0.0307 0.0244 0.0176	PMG-120 RMSE 0.0090 0.0099 0.0091 0.0080 0.0086 PMG-49 RMSE 0.0502 0.0680 0.0611 0.0399	R ² 0.5350 0.4329 0.5215 0.6310 0.5737 R ² 0.6589 0.3739 0.4942 0.7850	MAE 0.0003 0.0003 0.0003 0.0001 0.0001 MAE 0.0000 0.0000 0.0000 MAE 0.0000 0.0000	RMSE 0.0005 0.0006 0.0002 0.0002 0.0002 PMG-125 RMSE 0.0000 0.0000 0.0000 PMG-125 RMSE	R ² -8.6053 -15.7452 -1.3077 -0.2628 -0.0229 R ² 1.0000 1.0000 1.0000 1.0000 1.0000 -0.0000 -0.0000	MAE 0.0000 0.0000 0.0000 0.0000 0.0000 MAE 0.0001 0.0001	RMSE 0.0000 0.0000 0.0000 0.0000 PMG-13 RMSE 0.0003 0.0003 0.0003	R ² 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 R ² -0.0250 -0.0250 -0.0250
LR KNN DT RF XGB MODEL LR KNN DT RF XGB MODEL LR KNN DT LR KNN DT LR	0.0057 0.0052 0.0037 0.0048 0.0053 MAE 0.0348 0.0231 0.0250 0.0157 0.0166 MAE 0.0258 0.0197 0.0182	RMSE 0.0199 0.0216 0.0173 0.0191 0.0230 PMG-48 RMSE 0.0523 0.0535 0.0563 0.0385 0.0418 PMG-50 RMSE 0.0693 0.0653	R ² 0.2880 0.1652 0.4625 0.3445 0.0519 R ² 0.7630 0.7513 0.7253 0.8713 0.8488 R ² 0.2160 0.3461 0.3041	PMGs MAE 0.0030 0.0030 0.0022 0.0024 MAE 0.0302 0.0307 0.0244 0.0176	PMG-120 RMSE 0.0090 0.0099 0.0091 0.0080 0.0086 PMG-49 RMSE 0.0502 0.0680 0.0611 0.0399	R ² 0.5350 0.4329 0.5215 0.6310 0.5737 R ² 0.6589 0.3739 0.4942 0.7850	MAE 0.0003 0.0003 0.0001 0.0001 MAE 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.000000 0.000000 0.000000 0.00000 0.00000 0.00000 0.0000	RMSE 0.0005 0.0006 0.0002 0.0002 0.0002 PMG-129 RMSE 0.0000 0.0000 0.0000 PMG-125 RMSE	R ² -8.6053 -15.7452 -1.3077 -0.2628 -0.0229 R ² 1.0000 1.0000 1.0000 1.0000 1.0000 -0.2536 -0.1111	MAE 0.0000 0.0000 0.0000 0.0000 0.0000 MAE 0.0001 0.0001	RMSE 0.0000 0.0000 0.0000 0.0000 PMG-13 RMSE 0.0003 0.0003 0.0003	R ² 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.0025 -0.025 -0.025 -0.025

With Gas Balance



Experiment 2:

- Random Forest delivered the best performance ($R^2 = 0.95$) when gas balance was included.
- Removing gas balance caused a drop of over 50% in R² across all models.
 Linear Regression performance remained unchanged
- Linear Regression performance remained unchanged, reflecting its limitations in capturing nonlinear relationships.

MODEL	Witi	H GAS BAL	ANCE	WITHOUT GAS BALANCE			
MODEL	MAE	RMSE	R^2	MAE	RMSE	R^2	
LR	0.0133	0.0393	0.1993	0.0133	0.0394	0.1990	
K-NN	0.0026	0.0186	0.8217	0.0086	0.0385	0.2331	
DT	0.0014	0.0127	0.9170	0.0081	0.0361	0.3249	
RF	0.0011	0.0098	0.9501	0.0078	0.0333	0.4276	
XGB	0.0015	0.0111	0.9366	0.0078	0.0336	0.4178	

4. Modeling Approach

- We trained five regression models to predict CH4 using other gases as features:
- Linear Regression, k-Nearest Neighbors (k-NN), Decision Tree, Random Forest, and XGBoost.

Model	Selected Hyperparameters				
LR	No hyperparameters (standard least squares)				
kNN	n_neighbors = 3				
DT	<pre>max_depth = 10, min_samples_split = 2, min_samples_leaf = 1, max_features = 'sqrt'</pre>				
RF	<pre>n_estimators = 300, max_depth = 20, min_samples_split = 2, min_samples_leaf = 1</pre>				
XGB	<pre>n_estimators = 100, max_depth = 3, learning_rate = 0.1</pre>				

5. Experimental Setup

- Experiment 1: Per-well models (individual time series)
- Experiment 2: All-well combined dataset
- Evaluation metrics: R2, MAE, RMSE

Conclusion and recommendation

We evaluated machine learning regressors to predict CH₄ concentrations in contaminated soils using co-measured gases. Random Forest outperformed others in well-specific models, while Decision Tree, Random Forest, and XGBoost varied in performance in aggregated data. Although the gas_balance feature improved predictions, it may cause information leakage.

Future work should explore deep learning and temporal modeling to enhance environmental gas monitoring.

Acknowledgements

R.Q. is supported by the São Paulo Research Foundation (FAPESP) under project grant 2019/21693-0 and by the Institute for Technological Research (IPT) of the State of São Paulo and F.A. is supported by National Council for Scientific and Technological Development (CNPq) under project grant 140253/2021-1.