XLVI CILAMCE
VITÓRIA - ES - BRAZIL - 2025
XLVI IBERO-LATIN AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGINEERING

Subsidence Analysis in Buried Pipelines Due to Fluid Leakage: Numerical Modeling and Machine Learning

Guilherme Avila[1], Guilherme de L. Medeiros[1], José G. Souza[1], Juliana Farias[1], Lariane I. Farias[1], Railson P. do Carmo[1], Vinicius de A. Souza[1], Karl M. Guerra[1]
[1]*School of Engineering, Veiga de Almeida University - UVA*
Rua Ibituruna, 108, Maracanã, Rio de Janeiro – RJ - Brazil
karl.guerra2@uva.br

Abstract. Buried pipelines are vital to urban infrastructure, but leaks can disturb soil stresses, causing ground subsidence and even sinkholes. These effects arise from complex interactions between soil mechanics and unsaturated porous media flow, demanding advanced modeling. This study presents a coupled numerical and data-driven framework to simulate leakage-induced subsidence. Finite volume simulations solve the elastic equilibrium and Richards' equation using Brooks-Corey model based on the Van Genuchten parameters using Two-Point Flux Approximation (TPFA) on Matlab Reservoir Toolbox (MRST). Mechanical-hydraulic coupling is captured via effective stress and pore pressure interactions. A high-fidelity database is built to analyze key factors: pipe depth, diameter, flow rate, and soil stiffness. Tree-based symbolic regression is used to extract explicit, interpretable laws from simulation data using the symbolic regression package Python Symbolic Regression (PySR). These expressions reduce computational cost and improve physical insight that may be applied for hazard predictions.

Keywords: Soil subsidence, pipe leakage, unsaturated flow, symbolic regression, genetic algorithm.

1 Introduction

Urban subsidence caused by pressurized pipeline leaks poses a serious geotechnical risk in Brazilian cities, often resulting from suffusion processes that lead to internal erosion, soil instability, and sinkhole formation. Studies by Iwanek [1] and Soares et al. [2] highlight the role of poorly maintained pipeline networks in exacerbating such risks. The work by NIMA [3] proposes an integrated framework combining Python-based physical simulations and time-series forecasting, using real data from the Gávea neighborhood. It incorporates soil parameters from Mao et al. [4], Guo et al. [5], and considers different soil types with calibrated permeability and strength parameters. Temporal evolution of settlements is modeled using Long Short-Term Memory (LSTM) neural networks and Kalman filters in the work of Zerunian and Kellner [6], yielding 3.1 cm of settlement over 365 days with anomalies detected in some small intervals of time. Comparative analysis with Mao et al. [4] underscores the need for broader geotechnical characterization and real-world boundary conditions.

Inverse modeling (Soares et al. [2]), CFD–DEM simulations (Mao et al. [4]), and multiphase models (Cui et al. [7]; Guo et al. [5]) have enhanced understanding of erosion and leak-induced deformation. Optimization strategies using genetic algorithms (Parween & Sinha [8]; Marques [9]) and hydraulic modeling tools such as EPANET (Costa [10]; Marques [9]) and MIGHA (Castro et al. [11]) are benchmarked here. Real-time detection using WSNs (Ali & Choi [12], [13]), SAR interferometry (Theron et al. [14]), and Random Forest classifiers (Park et al. [15]) further illustrate the shift toward smart infrastructure monitoring.

In this work, the authors integrate a finite volume method implemented in MATLAB to solve the two-phase flow equations in porous media using Van Genuchten–Mualem [17] parameters, enabling the computation of pressure and saturation fields. These fields are then used to determine pore pressures and changes in effective stresses, from which ground settlements are calculated based on linear elastic theory and the constrained modulus. Finally, a symbolic regression approach, supported by genetic algorithms and the Broyden–Fletcher–Goldfarb–Shanno (BFGS) optimization technique, is employed to identify simplified yet representative analytical expressions that accurately describe the settlement behavior over

PÚBLICA

time and its final magnitude.

2 Theoretical background

2.1 Modelling the unsaturated flow in porous media

The analysis of transport in unsaturated porous media is conducted using the finite volume method, applied to a system of differential equations describing mass balance in a multiphase system. The governing equations are complemented by constitutive relationships expressed as equations of state, ensuring a robust and consistent description of the physical processes involved. Let Ω be an open and bounded set and the equations for the multiple phases present in the system be:

$$\frac{\partial}{\partial t}(\phi S_\alpha \rho_\alpha) - \nabla \cdot \left(\frac{\boldsymbol{K} k_{r\alpha}}{\mu_\alpha} (\nabla p_\alpha - \rho_\alpha \boldsymbol{g}) \right) = f_\alpha(\boldsymbol{x}, t), \qquad \boldsymbol{x} \in \Omega \subset \mathbb{R}^2 \, , t \in \mathbb{R}^+ \tag{1}$$

Where ϕ is the soil porosity; S_α, ρ_α, p_α and μ_α represent the saturation, the density, the pressure and the viscosity of the fluid phase α, respectively; $\boldsymbol{g}$ is the gravity force, $k_{r\alpha}$ is the relative permeability and $\boldsymbol{K}$ is the saturated permeability matrix, and ∇ is the gradient operator. The permeability field, $\boldsymbol{K}$, is modeled as a heterogeneous, stochastically Gaussian, and elliptic field, reflecting the inherent randomness often present in porous media.

We impose a hydrostatic boundary condition along the vertical sides and a zero-pressure condition at the surface along with a zero-water saturation. At the bottom boundary, a homogeneous Neumann condition is applied, indicating impermeability. In the region of the pipe, we model a circular source with a prescribed flow rate. The mass balance equation is discretized using the Two-Point Flux Approximation (TPFA) method, so that the saturation and velocity fields can be updated as:

$$S_i^{n+1} - S_i^n = \frac{1}{\phi |\Omega_i|} \sum_k \int_{t_n}^{t_{n+1}} \int_{\Gamma_{ik}} \left[\frac{\lambda_w}{\lambda_w + \lambda_n} \boldsymbol{v} + \frac{\lambda_w \lambda_n \boldsymbol{K}}{\lambda_w + \lambda_n} \left(\Delta\rho \boldsymbol{g} + \nabla P_c(S) \right) \right] \cdot \boldsymbol{n}_{i,k} \, ds dt \, , \qquad m = n, n+1 \tag{2}$$

$$v_{i,k} \approx A_{i,k} K_i \frac{(p_i - \pi_{i,k}) c_{i,k}}{|c_{i,k}|^2} \cdot n_{i,k} = T_{i,k}(p_i - \pi_{i,k}) = \left[T_{i,k}^{-1} + T_{k,i}^{-1} \right]^{-1} (p_i - p_k) \tag{3}$$

For $\lambda_\alpha = k_{r\alpha}/\mu_\alpha$ and $P_c = -p_w$. And $A_{i,k}$ being the surface area of the interface between cells i and k. For an arbitrary cell, $\cup_i \Gamma_i = \partial\Omega_k$ for the k-th cell with piecewise continuous boundary.

For numerical simplification, the Brooks–Corey model was adopted to represent relative permeability, as it avoids the singularities commonly found in the Van Genuchten–Mualem formulation (Van Genuchten [16]). Nevertheless, the parameters of both models are mathematically correlated. The following equivalence was established:

$$\lambda_{BC} = \frac{1}{n_{VG} - 1} \, , \qquad \alpha \approx \frac{1}{h_b} \tag{4}$$

For h_b being the air-entry pressure for the Brooks-Corey model.

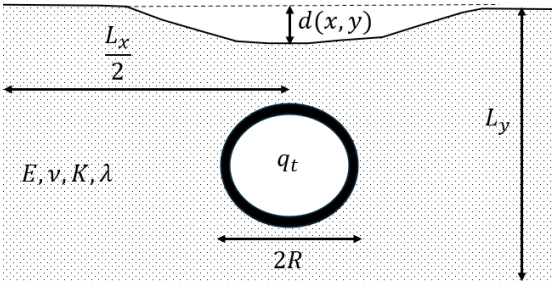


Figure 1 Scheme of the pipe leakage and infiltration problem

PÚBLICA

To generate the dataset used as input for the symbolic regression algorithm, multiple scenarios were simulated. Key properties, such as burial depth of the pipe, thickness of the soil layer, soil permeability, Brooks–Corey parameter, soil Young's modulus, Poisson's ratio, pipe flow rate, pipe diameter, and lateral extent of the soil domain, were varied within ranges found in the literature to ensure physically realistic cases (Figure 1). Once the parameter ranges were defined, Latin Hypercube Sampling was employed to generate the scenarios.

For each iteration, that is, for each scenario, after simulating the fluid flow, the resulting settlements were computed in a decoupled manner based on the variation in pore pressure and, consequently, in effective stress within the soil.

2.2 Elastic settlements

The settlements induced by the progressive saturation of the soil were computed using linear elasticity theory. From the vertical strains in laterally confined domains, we have:

$$d_y(x,y) = \int_\Omega \varepsilon_y(x,y)d\Omega = \int_\Omega \frac{\sigma'_y(x,y)}{M(x,y)}d\Omega = \int_\Omega \frac{\sigma_y(x,y) - p_w(x,y)}{M(x,y)}d\Omega \tag{5}$$

where $d_y(x,y)$ is the vertical displacement of the soil, $\sigma_y(x,y)$ is the total vertical stress field, $\sigma'_y(x,y)$ is the effective vertical stress field, and $p_w(x,y)$ is the pore pressure field. The term $M(x,y)$ is defined as the constrained (oedometric) modulus, given by $M = M(E,\nu)$, where E is Young's modulus and ν is Poisson's ratio.

For each time increment, saturation and pressure are computed for every cell in the domain. These values are subsequently used to calculate the effective stresses, strains, and settlements. Under the assumption of small deformations, the Jacobian of the domain transformation can be approximated as unity, eliminating the need to update the geometry at each step. Furthermore, the maximum settlements were calculated at $L_x/2$ and $\max(L_y)$,i.e., centered in the horizontal direction and at the top of the ground.

2.3 The PySR algorithm

Symbolic regression (SR) offers an interpretable alternative to black-box models by deriving concise equations directly from data, with PySR standing out as a state-of-the-art implementation that combines evolutionary algorithms and symbolic optimization (Cranmer [17]; Chawla & Makke [18]). At its core, PySR employs a multi-population asynchronous evolutionary strategy, where independent "islands" evolve using tournament selection, crossover, mutation, symbolic simplification, and constant optimization (Cranmer [18]).

Let the dataset consist of input-output pairs $(\boldsymbol{x}_i, y_i) \in \mathbb{R}^n \times \mathbb{R}$,where the goal is to discover an explicit function $f: \mathbb{R}^n \to \mathbb{R}$ such that $y_i \approx f(\boldsymbol{x}_i)$. The optimization objective in PySR is to minimize a loss function $\mathcal{L}(f)$, typically of the form:

$$\mathcal{L}(f) = \sum_{i=1}^{N} (f(\boldsymbol{x}_i) - y_i)^2 \tag{6}$$

This corresponds to the classical L2DistLoss, although PySR supports other loss functions such as LogitDistLoss and HuberLoss, allowing flexibility in how fitting quality is assessed (Cranmer [17]).

Each symbolic expression in PySR is encoded as a tree, where internal nodes represent mathematical operators and leaves represent variables or constants. To evolve these expressions, PySR applies genetic programming operations: mutation (either point mutation on a single node or macro mutation replacing subtrees) and crossover (combining subtrees from two parent trees). At each generation t, the population is represented as $P(t) = \{\alpha_1^t, \alpha_2^t, \dots, \alpha_\mu^t\}$, where each individual $\alpha_i^t \in \mathcal{I}$ (the genotype space) is mapped to its phenotypic expression by $x_i^t = \Gamma(\alpha_i^t)$. The fitness of each individual is evaluated via a transformed loss function $\phi(\alpha_i^t) = \delta(f(x_i^t))$.

The evolutionary loop consists of selection (via tournament), recombination $\theta_r: \mathcal{I}^\mu \to \mathcal{I}^\lambda$, mutation $\theta_m: \mathcal{I}^\lambda \to \mathcal{I}^\lambda$, evaluation, and survivor selection $\theta_s: (\mathcal{I}^\lambda \cup \mathcal{I}^{\lambda+\mu}) \to \mathcal{I}^\mu$, which defines the next generation. PySR employs a multi-population (or island) architecture, where each subpopulation evolves independently and asynchronously, enhancing diversity and reducing the risk of premature convergence.

PÚBLICA

Constant optimization is also applied during evolution to fine-tune numeric parameters within expressions. The process continues until a stopping criterion is reached, ultimately returning a Pareto front of candidate equations balancing accuracy and complexity.

3 Application of the method and some results

In this study, a total of $N = 1000$ numerical simulations were conducted to explore the parametric dependencies of a fluid-driven geomechanical system subjected to varying material, geometric, and boundary conditions. The simulations were governed by the following randomly sampled input parameters: Young's modulus $E \in [5 \times 10^6, 5 \times 10^7]\, Pa$, Poisson's ratio $\nu \in [0.2, 0.4]$, total volumetric flow rate $q_{total} \in [1 \times 10^{-4}, 1 \times 10^{-3}]\, m^3/s$, horizontal domain length $L_x \in [10, 30] m$, vertical domain height $L_y \in [5, 15] m$, pipe's radius $R \in [0.2, 1]$m , and mean intrinsic permeability $K_{mean} \in [1 \times 10^{-13}, 1 \times 10^{-11}]\, m^2$. Additionally, the Brooks-Corey parameter $\lambda_{BC} \in [1.5, 5]$ was introduced to vary the type of soil. Fixed parameters included porosity $\phi = 0.3$, soil density $\rho_{soil} = 1800\, kg/m^3$, and gravitational acceleration g = 9.81 m/s². The time horizon was fixed at 300 minutes, discretized with a time step Δt = 5 minutes, resulting in $n_t = 60$ time steps per simulation.

A sample size of $N = 1000$ was selected to balance statistical representativity with computational efficiency, especially given the high-dimensional, nonlinear nature of the problem; convergence tests showed that larger N values offered minimal gains in accuracy or model stability. Parameter ranges were chosen based on values commonly found in geotechnical and hydrogeological literature to ensure physical relevance.

To improve numerical conditioning and model interpretability, all input variables were normalized to a fixed interval, mitigating the effects of scale disparities that can distort optimization and bias regression models. This step is particularly important in symbolic regression, where operand magnitudes influence expression structure and convergence behavior.

Thus, to facilitate the symbolic regression process using PySR, all variable parameters were normalized to the interval [0,1]. This rescaling was performed using the standard min-max normalization rule: for each variable x_j, the transformation $\widehat{x_j} = \frac{x_j - min(x_j)}{max(x_j) - min(x_j)}$ was applied. In the case where the range $\max(x_j) - \min(x_j) = 0$, indicating a constant variable, the normalized column was filled with zeros to maintain numerical stability. The normalized dataset was stored in a structured table format and served as the input matrix for the PySR symbolic regression engine. All simulations and subsequent analysis were performed at the fixed final time t = 300 minutes, ensuring consistency in the temporal context of the output quantities across the dataset.

Additionally, the symbolic regression algorithm was configured with a restricted set of mathematical operations to ensure interpretability and physical relevance of the resulting expressions. The allowed binary operators were limited to $\{+, -, \times, \div, max, min\}$, while the set of unary functions included $\{sin, cos, exp, log, inv, \sqrt{}, \cdot^2, \cdot^3\}$, where $inv(x) = 1/x$. This constrained symbolic grammar balances expressiveness with the goal of producing equations that remain physically meaningful and easily interpretable.

The theoretical expression for vertical settlement $d_y(x, z, t)$ in a variably saturated porous medium under mechanical loading is based on Terzaghi's extended effective stress principle, accounting for the presence of pore air and water pressures. The effective stress is given by $\sigma' = \sigma - p_a = \sigma - \big(p_w + p_c(S_w)\big)$ where p_w is the water pressure and $p_c(S_w)$ is the capillary pressure as a function of saturation. Accordingly, the settlement, measured at the surface, at the final time is governed by:

$$d_y(x, t_f) = \int_0^{L_y} \frac{1}{M(E,\nu)} \left[\delta\sigma(z) - \delta p_w(z, t_f) - \frac{dp_c}{dS_w}(z, t_f) \cdot \delta S_w(z, t_f) \right] dz \tag{7}$$

With $M(E, \nu) = E/(1 - \nu^2)$ being the oedometric modulus. This integral formulation captures the influence of mechanical loading, saturation dynamics, and fluid pressure redistribution on vertical displacement in unsaturated soils.

The algorithm is designed to explore a Pareto front between high accuracy (e.g., coefficient of determination R^2) and low model complexity. Several of the discovered expressions involve nonlinear combinations of the elastic modulus E, Poisson's ratio ν, soil thickness L_y, and flow-related parameters such as the total inflow rate q. Notably, some models include q or $\sqrt{q}$ in the denominator, reflecting a physically reasonable trend wherein increased fluid flow accelerates pressure equilibration and thereby

PÚBLICA

reduces effective stress gradients, ultimately leading to reduced localized settlements. It is known that dissipative phenomena have $\sqrt{t}$ dependence and then if $q \sim \frac{1}{t}$ one must have $d \sim 1/\sqrt{q}$. For the sake of visualization, one of the selected expressions was:

$$d_y(x, t_f)_{SR} = \frac{\left(L_y + 0.462\right)^2 \cdot (1.958 - \nu)}{\sqrt{q} + \left(\frac{E + 0.1014}{0.03}\right)} \tag{8}$$

Among the suite of expressions generated, a particularly compelling one is the following simplified model:

$$d_y(x, t_f)_{SR} = L_y \cdot \frac{(1.9268 - \nu) \cdot 0.067}{E + 0.1168} \tag{9}$$

This expression was the eighth in the ranked list and represents the best trade-off between physical plausibility and mathematical simplicity. It mirrors the structure of the theoretical solution by featuring a linear dependence on L_y, an inverse dependence E, and a decreasing influence with increasing Poisson's ratio ν, consistent with the role of these parameters in the oedometric modulus. The absence of flow-related terms in this formula can be interpreted because of the relatively long simulation time ($t_f = 300s$), during which pore pressures were able to equilibrate, rendering the influence of transient flow on final settlement negligible.

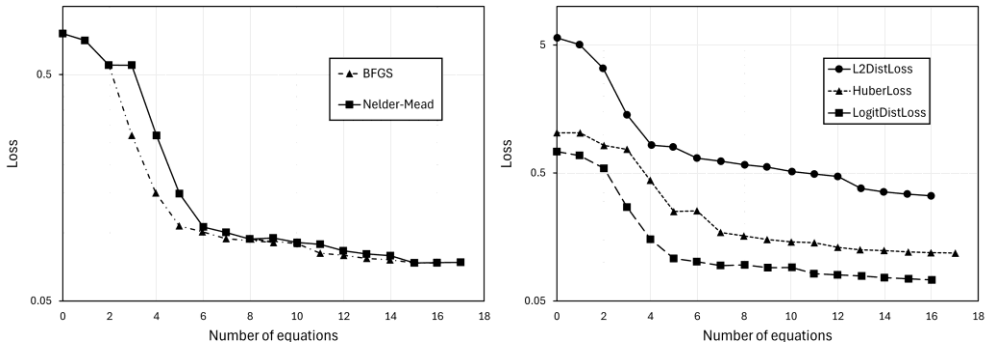


Figure 2 Loss function evolution with the number of generated equations for different optimization algorithms and loss functions

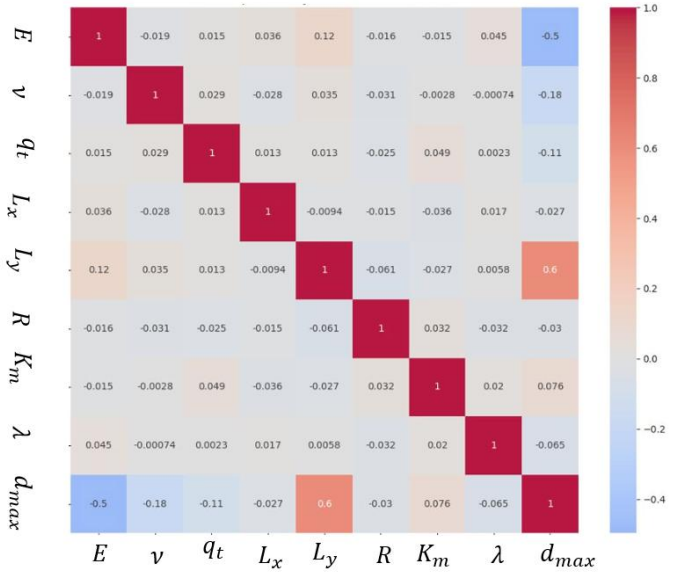


Figure 3 Map of Pearson's correlation

PÚBLICA

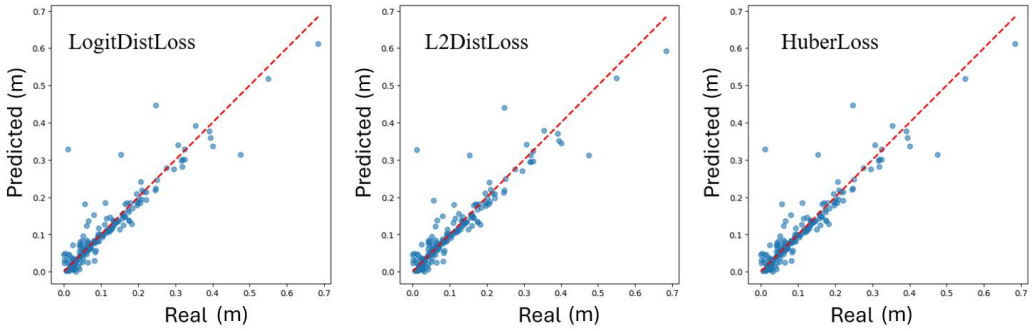


Figure 4 Comparison between real and predicted settlement for different loss functions for the test data

To preprocess the dataset, the Interquartile Range (IQR) method was applied to remove outliers. A total of 73 outliers were identified and excluded from the analysis. The dataset was partitioned using scikit-learn's `*model_selection.train_test_split*` function. Following a conventional 80/20 split, 80% of the observations were allocated to the training set and the remaining 20% to the test set.

To assess the statistical associations between variables, Pearson's correlation coefficient was used to measure the linear relationship between pairs of numerical variables. The results are presented in Figure 3 and show consistency with the physical phenomenon as settlement shows high correlation with the elastic constants and with the soil's layer size.

For model optimization, two algorithms were employed: BFGS and Nelder–Mead. After training, the model optimized with BFGS achieved a higher coefficient of determination ($R^2 = 86.01\%$) compared to the Nelder–Mead model ($R^2 = 83.21\%$). Both algorithms were configured with a maximum of 10,000 iterations and a population size of 500. During the optimization process, 18 candidate equations were generated, each with a progressively lower loss value and correspondingly increased complexity. A comparative plot of the loss values across models is provided in Figure 2.

Three loss functions were evaluated using the BFGS optimizer: LogitDistLoss, L2DistLoss, and HuberLoss. Among them, LogitDistLoss delivered the best performance, yielding an R^2 of 86.01%, followed closely by HuberLoss ($R^2 = 86.00\%$) and L2DistLoss ($R^2 = 85.79\%$), as seen in Figure 4.

4 Conclusions

Symbolic regression has proven capable of identifying physically consistent laws governing leakage-induced settlement. In the absence of analytical solutions for the underlying nonlinear and coupled partial differential equations, the numerical results gained interpretability through the equations discovered by PySR. This approach bridges the gap between data-driven modeling and physical insight, offering concise expressions that reflect the underlying mechanics. However, the process remains computationally intensive, requiring significant resources for training and optimization. Ongoing research aims to extend this framework to transient solutions that capture the temporal evolution of settlement, enabling a more complete understanding of subsidence dynamics over time.

Acknowledgements. The authors gratefully acknowledge the Directorate of Graduate Studies, Research, Extension, and Innovation at Universidade Veiga de Almeida for its cooperation and support in the development of this work.

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

PÚBLICA

References

[1] M. Iwanek, "Parameters characterizing leakages from damaged water pipes in the aspect of environmental security," Applied Water Science, vol. 12, no. 126, 2022. https://doi.org/10.1007/s13201-022-01641-3.

[2] A. K. Soares, P. B. Cheung, L. F. R. Reis, and M. P. Sandim, "Avaliação das perdas físicas de um setor da rede de abastecimento de água de Campo Grande-MS via modelo inverso," Engenharia Sanitária e Ambiental, vol. 9, no. 4, pp. 312–321, out./dez. 2004. https://www.scielo.br/j/esa/a/7qKMvMbPSM9sMzhC8JSkntL/.

[3] Núcleo Interdisciplinar de Meio Ambiente (NIMA/PUC-Rio), "Diagnóstico das presentes condições da Estação Gávea do Metrô do Rio de Janeiro e alternativas para minimização de riscos associados," Relatório técnico, Pontifícia Universidade Católica do Rio de Janeiro, set. 2019. https://www.nima.puc-rio.br/wp-content/uploads/2021/04/RELATÓRIO_PUCRIO.pdf.

[4] J. Mao, Y. Wang, H. Zhang, and X. Jing, "Study on the Influence of Urban Water Supply Pipeline Leakage on the Scouring Failure Law of Cohesive Soil Subgrade," Water, vol. 16, no. 93, 2024. https://doi.org/10.3390/w16010093.

[5] J. Guo et al., "Study on urban ground collapse induced by defective pipelines based on physical model experiments and numerical simulation," Scientific Reports, vol. 15, no. 1, p. 6085, 2025. DOI: https://doi.org/10.1038/s41598-025-90146-5.

[6] F. Zerunian and J. Kellner, "A Kalman-based online detection algorithm for water pipeline leak localization using distributed pressure sensors," Water, vol. 16, no. 1, 2024. https://www.mdpi.com/2073-4441/16/1/75.

[7] J.-Y. Cui, F.-Y. Liu, Y.-W. Geng, C. Pu, and M. Miao, "Numerical simulation of ground subsidence factors resulting from unpressurized pipeline rupture below the water table," Applied Sciences, vol. 13, no. 17, p. 9536, 2023. https://www.mdpi.com/2076-3417/13/17/9536.

[8] S. Parween and R. C. Sinha, "An integrated model for water resource allocation, distribution and management: a genetic algorithm-based optimisation model," Water Resources Management, vol. 39, no. 7, 2025. https://link.springer.com/article/10.1007/s11269-025-04217-3.

[9] S. M. Marques, "Aplicação de algoritmos genéticos em operação de redes de distribuição de água com o uso de softwares R e EPANET visando ao controle de perdas de água – estudo em uma rede teórica," MSc thesis, Universidade Federal de Itajubá, 2023. https://repositorio.unifei.edu.br/jspui/handle/123456789/3531.

[10] L. H. M. Costa, "Uso conjunto do EPANET e da metodologia de Morgan para otimização de redes de abastecimento de água," MSc thesis, Universidade Federal do Ceará, Fortaleza, 2005. http://www.repositorio.ufc.br/handle/riufc/18828.

[11] M. A. H. Castro et al., "Detecção de vazamentos em redes de distribuição de água pela aplicação do método MIGHA," Universidade Federal do Ceará, 2021. https://repositorio.ufc.br/bitstream/riufc/68778/1/2021_art_mahcastro.pdf.

[12] H. Ali and J. Choi, "Data on manmade sinkholes due to leakage in underground pipelines in different subsurface soil profiles," Data in Brief, vol. 34, p. 106740, 2021. https://www.sciencedirect.com/science/article/pii/S2352340921000263.

[13] H. Ali and J. Choi, "Risk prediction of sinkhole occurrence for different subsurface soil profiles due to leakage from underground sewer and water pipelines," Sustainability, vol. 12, no. 1, p. 310, 2020. https://doi.org/10.3390/su12010310.

[14] A. Theron, J. Engelbrecht, J. Kemp, W. Kleynhans, and T. Turnbull, "Detection of sinkhole precursors through SAR interferometry: radar and geological considerations," IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 6, pp. 871–875, jun. 2017. DOI: https://doi.org/10.1109/LGRS.2017.2684905.

[15] J. H. Park, J. Kang, J. Kang, and D. Mun, "Machine-learning-based ground sink susceptibility evaluation using underground pipeline data in Korean urban area," Scientific Reports, vol. 12, p. 20911, 2022. DOI: https://doi.org/10.1038/s41598-022-25237-8.

[16] M. Th. van Genuchten, "A closed-form equation for predicting the hydraulic conductivity of unsaturated soils," Soil Science Society of America Journal, vol. 44, no. 5, pp. 892–898, 1980. DOI: https://doi.org/10.2136/sssaj1980.03615995004400050002x.

[17] M. Cranmer, "Interpretable Machine Learning for Science with PySR and SymbolicRegression.jl," arXiv preprint, 2023. https://arxiv.org/pdf/2305.01582.

[18] S. Chawla and N. Makke, "Interpretable scientific discovery with symbolic regression: a review," Artificial Intelligence Review, vol. 52, no. 2, 2024. DOI: https://doi.org/10.1007/s10462-023-10622-0.

PÚBLICA