Geometry-Driven Characterization of Natural Gas Pipeline Leaks Using CFD Simulations and Semi-Empirical Modeling

Mauricio Carmona[a], Alma Nouar[b], Cesar Muñoz[c], Juana Cruz[d], Lesme Corredor[e], Maicol Marenco[f] and Jhorma Medina[g]

[a] *Universidad del Norte, Barranquilla, Colombia, mycarmona@uninorte.edu.co*
[b] *Universidad del Norte, Barranquilla, Colombia, anouar@uninorte.edu.co*
[c] *Universidad del Norte, Barranquilla, Colombia, ceballosac@uninorte.edu.co*
[d] *Universidad del Norte, Barranquilla, Colombia, jcruzc@uninorte.edu.co*
[e] *Universidad del Norte, Barranquilla, Colombia, lcorredo@uninorte.edu.co*
[f] *Gases del Caribe S.A., Barranquilla, Colombia, mmarenco@gascaribe.com*
[g] *Gases del Caribe S.A., Barranquilla, Colombia, rjmedina@gascaribe.com*

Abstract:
Natural gas distribution systems experience non-operational losses due to leaks, whose irregular geometries can significantly affect methane emissions, safety, and operational efficiency. Conventional analytical models generally assume ideal circular openings, which limits their applicability to real defects. This study presents an integrated computational fluid dynamics (CFD) and semi-empirical modeling framework to estimate leakage rates as a function of leak geometry and operating conditions.
A parametric CFD campaign was conducted to characterize partial leaks with circular and amorphous geometries. Leak shape was standardized through an aspect ratio based on the dimensions of the minimum bounding rectangle, with values of 1.0 and 0.48 considered in this work. Simulations were performed for pipe diameters of 3/4″, 1″, and 2″, operating pressures of 27 psig and 60 psig, and multiple leak severity levels defined through normalized geometric parameters. Representative cases were used for mesh-independence analysis, and circular leaks were compared with theoretical predictions to support the numerical approach.
The results show that amorphous leaks consistently produce higher leakage rates than equivalent circular openings, with increases between 3% and 8% under the evaluated conditions. Based on the CFD database, four predictive approaches were assessed: a physics-based orifice model, a log-linear model, a Gamma generalized linear model, and a nonlinear physics-inspired formulation. Among them, the optimized log-linear model provided the best balance between simplicity and accuracy, reaching a maximum relative error of approximately 4%.
The proposed framework provides a computationally efficient and physically interpretable tool for estimating gas losses in natural gas distribution systems, while also supporting future experimental validation and leak characterization efforts.

Keywords:
Computational Fluid Dynamics; Gas Leakage; Leak Geometry; Methane Emissions; Pipeline Integrity; Thermodynamics.

1. Introduction

Fugitive methane emissions from natural gas distribution networks represent a significant environmental and operational concern. Although these emissions were historically considered relatively small and widely dispersed, recent measurement campaigns indicate that leakage from urban distribution systems may be significantly underestimated in national greenhouse gas inventories. Mobile leak detection surveys conducted in several U.S. cities have identified hundreds of thousands of leaks along distribution mains, revealing methane emissions considerably higher than those estimated using conventional inventory methods based on generalized emission factors [1].
Methane is a potent greenhouse gas due to its strong radiative forcing. According to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, methane exhibits a 100-year global warming potential roughly 27–29 times greater than carbon dioxide depending on its origin [2]. Consequently, even modest leakage rates can have a disproportionate influence on climate forcing. Improving the accuracy of emission estimates from natural gas infrastructure is therefore essential for both environmental accounting and mitigation planning.
Computational fluid dynamics have been extensively applied to analyze leakage mechanisms and derive predictive correlations. Ebrahimi-Moghadam et al. [4], [5] simulated natural gas leakage from both buried and above-ground pipelines, identifying clear relationships between leakage rate and key parameters such as pipe pressure, leak diameter, and pipe diameter. Their results indicated that leakage rate increases approximately linearly with pressure, quadratically with leak diameter, and with the fourth power of the ratio between leak diameter and pipe diameter. In a related study, Ding et al. [6] showed through three-dimensional CFD simulations that the geometry of the leakage opening influences jet structure and discharge characteristics, proposing correction factors that improved the accuracy of classical leakage equations.

Despite these advances, existing models fail to accurately represent irregular leak geometries because most leakage estimation methods rely on simplified geometric assumptions. Classical analytical formulations usually represent leaks as circular openings characterized by a single diameter parameter and categorized into regimes such as small-hole leaks or large-hole leaks [7]. While these models provide useful approximations for safety assessments and emission inventories, real pipeline defects rarely exhibit ideal circular geometries. Corrosion damage, mechanical impacts, and material degradation often produce irregular openings, elongated cracks, or partially detached pipe segments that cannot be adequately described by a single characteristic diameter.

To address this limitation, the present work introduces a geometry-based leakage modeling framework that combines computational fluid dynamics simulations with semi-empirical modeling techniques. The proposed methodology incorporates normalized geometric descriptors based on leak severity and aspect ratio to represent both circular and irregular leakage shapes. A parametric CFD campaign was carried out to investigate leakage behavior for different pipe diameters, operating pressures, and defect configurations. The resulting dataset was then used to develop predictive models capable of estimating leakage rates while accounting for geometric variability.

2. Leak Flow Modeling Background

2.1 Fundamental models for gas leakage

The estimation of gas leakage from pipelines has traditionally relied on analytical formulations derived from compressible flow through orifices. Within this framework, the leakage opening is modeled as an equivalent nozzle or thin-plate orifice through which pressurized gas expands from the pipe interior to the surrounding environment. The leakage rate is therefore primarily governed by upstream pressure, gas thermodynamic properties, and the effective area of the opening.

In most analytical formulations, the gas is assumed to behave as an ideal compressible fluid, and the expansion process is treated as isentropic. Under these assumptions, classical compressible-flow relations can be applied to describe the discharge through the leakage opening. The flow regime depends on the ratio between downstream pressure $p_d$and upstream pipe pressure p_u, which determines whether the flow remains subsonic or becomes choked. For subsonic conditions, the leakage mass flow rate may be written as Eq. (1).

$$C_d A p_u \sqrt{\frac{2\gamma}{RT(\gamma-1)}\left[\left(\frac{p_d}{p_u}\right)^{2/\gamma} - \left(\frac{p_d}{p_u}\right)^{(\gamma+1)/\gamma}\right]} \qquad (1)$$

where $C_d$is the discharge coefficient, $A$the effective leakage area, T the gas temperature, R the specific gas constant, and $\gamma$the ratio of specific heats.

Under choked conditions the maximum mass flow rate becomes Eq. (2).

$$\dot{m} = C_d A p_u \sqrt{\frac{\gamma}{RT}}\left(\frac{2}{\gamma+1}\right)^{\frac{\gamma+1}{2(\gamma-1)}} \qquad (2)$$

In practical situations the discharge coefficient accounts for non-ideal effects such as viscous losses, turbulence, and geometric contraction of the jet at the leakage opening.

2.2 Classification of leakage scenarios

Engineering analyses typically classify pipeline leaks according to the size of the opening relative to the pipe diameter. Three main categories are generally considered:

• Small-hole leaks, representing minor defects such as pinholes or localized corrosion damage, typically characterized by leak diameters $d \leq 20\,mm$.

• Large-hole leaks, corresponding to intermediate failures where the leak diameter is significant but still smaller than the pipe diameter ($20\,mm < d < D$).

• Full-bore ruptures, representing catastrophic failures in which the pipe is completely severed, and the effective leak diameter approaches the pipe diameter ($d = D$).

These categories are widely used in safety analyses and risk assessments for pipeline systems. In the small-hole model the leak is approximated as a circular orifice through which compressible gas escapes according to the isentropic relations previously described. The large-hole model extends this formulation by accounting for the ratio between the leakage area and the pipe cross-section. In the case of a full-bore rupture, the discharge is controlled by the pipe cross-section itself and the resulting flow resembles the expansion from a short nozzle connected to a pressurized reservoir.

Although these simplified models provide practical estimates, they rely on idealized geometric assumptions. In most implementations the leak opening is represented by a circular orifice characterized by a single diameter or by an equivalent hydraulic diameter, which may not accurately represent real pipeline defects.

2.3 Numerical and empirical approaches for leakage estimation

To overcome the limitations of purely analytical models, several investigations have used computational fluid dynamics simulations and controlled experiments to study leakage phenomena in greater detail.

Ebrahimi-Moghadam et al. [4], [5] performed two- and three-dimensional CFD simulations of gas leakage from urban distribution pipelines considering both buried and above-ground configurations. Their results revealed systematic relationships between leakage rate and operating parameters. In particular, leakage rate increased approximately linearly with pipe pressure, quadratically with leak diameter, and with the fourth power of the ratio between leak diameter and pipe diameter. These correlations achieved prediction errors below roughly 7%.

Ding et al. [6] investigated the influence of leak geometry using three-dimensional CFD simulations. Their analysis showed that the shape of the leakage opening modifies jet development and discharge behavior. By introducing geometry-dependent correction factors, the authors reduced prediction errors to values between −1.6% and 2.6%.

Experimental studies have also contributed to improving leakage characterization. Liu et al. [3] carried out controlled experiments on buried pipelines and combined these measurements with CFD simulations to evaluate methane release and dispersion in soil. Their results indicated that leakage rate is primarily governed by pipe pressure, leak diameter, and soil properties, whereas burial depth has a smaller effect.

Additional studies have examined the migration of leaked gas through surrounding soil and underground structures. Wang et al. [9] reported that methane transport and accumulation patterns depend strongly on soil characteristics and nearby infrastructure, highlighting the need for accurate characterization of leakage sources when assessing safety risks.

2.4 Limitations of conventional leakage models

Despite the progress achieved through analytical, numerical, and experimental studies, an important limitation remains in most existing leakage models: the assumption of idealized leak geometries.

In practice, pipeline defects rarely correspond to perfect circular openings. Mechanical damage, corrosion processes, and material degradation often generate irregular cracks, elongated apertures, or partially detached pipe segments. Such defects produce leakage geometries that cannot be adequately represented by a single diameter parameter.

As a result, the use of circular-orifice assumptions can introduce significant uncertainty in leakage rate predictions, particularly when the actual geometry of the defect is unknown. Even in studies that attempt to account for different shapes, the characterization is often limited to equivalent diameters or simple geometric correction factors.

These limitations motivate the development of modeling approaches capable of incorporating geometric descriptors that better represent irregular leakage openings while remaining computationally efficient. The methodology proposed in this work addresses this issue by introducing normalized descriptors based on leak severity and aspect ratio, enabling the representation of both circular and irregular geometries within a unified modeling framework.

3. CFD methodology

3.1. Leak geometries definition

Leak geometry is a key parameter influencing the discharge characteristics of gas escaping from pressurized pipelines. Most analytical formulations used to estimate leakage rates are derived assuming ideal circular orifices, which provide a simplified representation of the leak opening. However, defects observed in natural gas distribution networks frequently exhibit irregular geometries resulting from corrosion, material degradation, or mechanical damage. Consequently, the assumption of circular openings may not adequately represent the actual flow behavior of many real leaks.

In order to systematically characterize non-circular leak geometries, a geometric descriptor referred to as the **aspect ratio (AR)** was introduced in this study. This parameter provides a quantitative measure of the geometric irregularity of the leak opening and allows irregular shapes to be standardized for parametric analysis. The aspect ratio is defined based on the dimensions of the minimum bounding rectangle that encloses the leak geometry eq. (3):

$$AR = \frac{A}{B} \quad (3)$$

where A represents the characteristic dimension of the leak parallel to the pipe axis and Bcorresponds to the dimension perpendicular to the pipe axis.

For an ideal circular leak opening, both characteristic dimensions are equal, resulting in an aspect ratio equal to unity:

$$AR = 1 \quad (4)$$

This case represents the simplified geometry typically considered in conventional analytical leak-flow models.

In contrast, irregular leak openings may exhibit significantly different characteristic dimensions. In the present study, amorphous leak geometries were generated and characterized using the bounding rectangle approach described above.

In **Figure 1**, the amorphous geometry used for the analyses is shown. It was randomly designed to represent irregular leak shapes. From an engineering perspective, the leak has a non-uniform contour with varying curvature and no symmetry, leading to heterogeneous flow behavior and deviations from ideal orifice assumptions.

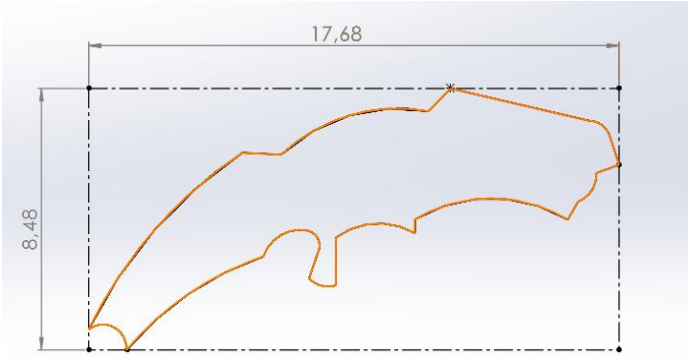

Figure. 1. *Amorphous leak geometry used for numerical analysis.*

For the representative irregular geometry considered, the bounding dimensions were **17.68 mm** in the axial direction and **8.48 mm** in the transverse direction, resulting in an aspect ratio of:

$$AR = \frac{8.48}{17.68} = 0.48 \quad (5)$$

This value represents a markedly non-circular opening and was therefore adopted as the characteristic aspect ratio for the amorphous leak cases included in the CFD parametric study.

The introduction of the aspect ratio parameter enables the systematic comparison between idealized circular leaks and irregular openings within the numerical simulations. This geometric descriptor provides a convenient means to incorporate the effect of leak morphology into the analysis, allowing the influence of shape irregularity on leakage flow rates to be explicitly evaluated.

3.2. Definition of Leak Severity Parameters

To enable consistent comparison between leak scenarios, a dimensionless severity parameter was defined by normalizing the characteristic leak dimension with respect to the pipe external diameter. Three severity levels were evaluated for both circular and irregular localized leaks: 10%, 20%, and 25%.
For localized leaks, including both circular and amorphous openings, the severity parameter was defined as the ratio between the **hydraulic diameter of the leak** and the **external pipe diameter** (Eq. 6):

$$S = \frac{D_h}{D_o} \quad (6)$$

where $D_h$is the hydraulic diameter of the leak opening and $D_o$is the external pipe diameter. This formulation allows both regular and irregular geometries to be represented using a common characteristic length scale.
For irregular openings, the hydraulic diameter was computed as $D_h = 4A_l/P_l$, where $A_l$is the leak area and $P_l$is the wetted perimeter of the opening.

3.3. Operating conditions

The simulations were performed under operating conditions representative of natural gas distribution networks. The working fluid was modeled as methane, which constitutes the primary component of natural gas and provides an adequate approximation for leakage flow analysis. A uniform inlet velocity of 20 m/s was imposed at the pipe inlet, consistent with typical gas velocities observed in distribution systems, while the inlet temperature was set to 25 °C to represent ambient conditions. Three pipe diameters 3/4″, 1″, and 2″ were considered to reflect common sizes used in distribution pipelines. For each diameter, two operating pressure levels, 27 psig and 60 psig, were evaluated, corresponding to typical medium-pressure conditions. These operating parameters were combined with the different leak configurations described in the previous sections to generate the complete set of simulation cases analyzed in this study.

3.4. CFD model setup

The computational domain was constructed to represent a straight section of pipeline containing the leak. Pipe segments of **1 m length** were modeled (as shown in **Figure 2**), with the leak located at the **midpoint of the pipe**, corresponding to slightly more than **ten times the pipe diameter**, to ensure that the internal flow was sufficiently developed at the leak location. The pipe geometry was modelled in three dimensions (3D). The internal volume of the pipe was created to represent the gas domain, and the leak opening was modelled assuming a through-wall perforation, allowing the gas to escape directly from the internal flow region to the external environment. Boundary regions were defined to represent the flow inputs and outputs. The pipe inlet was defined as Inlet Pipe, while the pipe outlet was defined as Outlet Pipe. The leak opening was defined as Outlet Leakage, and the pipe walls were treated as no-slip wall boundaries.

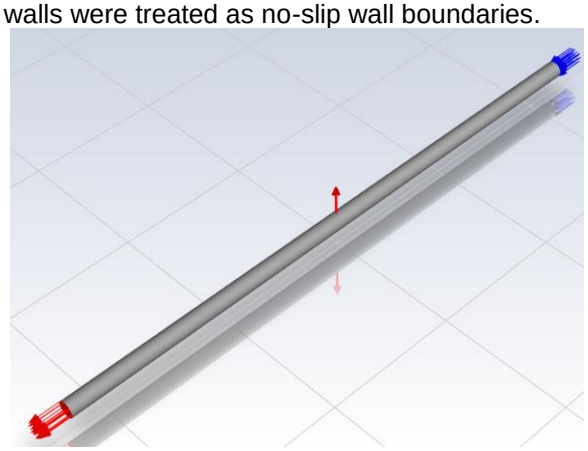

Figure. 2. *Computational domain and boundary conditions of the pipe leak model*

At the Outlet Pipe, the nominal operating pressure of the pipeline was prescribed according to the simulation case. At the leak outlet, a gauge pressure of 0 Pa was imposed to represent gas discharge to the atmosphere. In this study, all leaks were assumed to be exposed to atmospheric conditions and therefore buried leak scenarios were not considered.
The simulations were performed using a compressible flow formulation with the **energy equation enabled** to account for thermal effects during gas expansion through the leak. A **species transport model** was also activated to represent the gas composition, with methane as the primary component. The fluid properties were defined using an ideal gas assumption with constant transport properties. Turbulence effects were modeled using the standard **k–ε model**, and gravity was included in the vertical direction with a value of **−9.81 m/s²**. The numerical solution was obtained using the **SIMPLE pressure–velocity coupling scheme**, with second-order spatial discretization for momentum, pressure, and species transport to ensure adequate accuracy.

4. CFD Results

4.1. Circular leaks Results

The CFD results for circular leaks show a strong dependence of leakage flow rate on both operating pressure and pipe diameter. As expected, an increase in pressure from 27 psig to 60 psig results in a significant increase in the leakage flow rate for all pipe sizes and severity levels. Similarly, larger pipe diameters lead to higher discharge rates due to the increased mass flow capacity of the system. For a given pipe diameter, the leakage flow rate increases consistently with the severity level, defined by the normalized leak diameter. For instance, in the 1″ pipe at 27 psig, the leakage flow increases from approximately 17 Sm³/h (10%) to 103 Sm³/h (25%), demonstrating a nonlinear relationship between leak size and discharge rate.
Overall, circular leaks exhibit predictable behavior that aligns with classical orifice flow theory, where flow rate increases with both pressure differential and effective leak area.
In **Table 1**, the results obtained from the simulations are presented. The aspect ratio for these leaks is equal to 1, as they are circular.

Table 1. *Leakage flow rates for circular orifices under varying pressure and leak severity levels*

Scenario ID	Diameter [in]	Pressure [psig]	Leak Level	Leak Rate [kg/s]	Leak Rate [Sm3/h]
1	1"	27	High	0,0195	103
2	1"	27	Medium	0,0130	69
3	1"	27	Low	0,0032	17
4	1"	60	High	0,0293	155
5	1"	60	Medium	0,0193	102
6	1"	60	Low	0,0048	26
7	3/4"	27	High	0,0128	68
8	3/4"	27	Medium	0,0083	44
9	3/4"	27	Low	0,0020	11
10	3/4"	60	High	0,0191	101
11	3/4"	60	Medium	0,0124	66
12	3/4"	60	Low	0,0030	16
13	2"	27	High	0,0644	341
14	2"	27	Medium	0,0414	219
15	2"	27	Low	0,0105	56
16	2"	60	High	0,0964	510
17	2"	60	Medium	0,0617	327
18	2"	60	Low	0,0157	83

4.2. Amorphous leaks Results

The results obtained for amorphous leaks indicate consistently higher leakage flow rates compared to equivalent circular openings under the same operating conditions. This behavior is observed across all pipe diameters, pressure levels, and severity conditions.
In **Table 2**, the results obtained from the simulations are presented. The aspect ratio for these leaks is equal to 0.48, as they are amorphous.

Table 2. *Leakage flow rates for amorphous geometries under varying pressure and leak severity levels.*

Scenario ID	Diameter [in]	Pressure [psig]	Leak Level	Leak Rate [kg/s]	Amorphous Leak Rate [Sm3/h]
1	1"	27	High	0,0211	111
2	1"	27	Medium	0,0136	72
3	1"	27	Low	0,0033	18
4	1"	60	High	0,0313	166
5	1"	60	Medium	0,0203	108
6	1"	60	Low	0,0050	26
7	3/4"	27	High	0,0138	73
8	3/4"	27	Medium	0,0089	47
9	3/4"	27	Low	0,0021	11
10	3/4"	60	High	0,0205	109
11	3/4"	60	Medium	0,0133	70
12	3/4"	60	Low	0,0032	17
13	2"	27	High	0,0684	362
14	2"	27	Medium	0,0449	238
15	2"	27	Low	0,0114	60
16	2"	60	High	0,1032	546
17	2"	60	Medium	0,0670	355
18	2"	60	Low	0,0170	90

For example, in the **1″ pipe at 27 psig and 25% severity**, the leakage flow increases from approximately **103 Sm^3/h (circular)** to **111 Sm^3/h (amorphous)**. Similar trends are observed in other cases, with increases typically ranging between **3% and 8%**.
This difference can be attributed to the irregular geometry of the leak opening, which alters the local flow field and may reduce flow contraction effects compared to ideal circular orifices. As a result, amorphous leaks exhibit a higher effective discharge capacity.
These findings highlight the limitations of conventional analytical models when applied to irregular leak geometries and demonstrate the importance of incorporating geometric descriptors, such as aspect ratio, into predictive models.

4.3. Validation with theoretical models

The CFD results for circular leaks were compared against theoretical predictions based on conventional orifice flow models. The comparison shows a generally good agreement between simulated and analytical results across all evaluated conditions.
As illustrated in **Figure 3**, both CFD and theoretical results follow the same trend with increasing pressure, confirming the validity of the numerical approach. Minor deviations are observed, particularly at higher severity levels, where the theoretical model tends to slightly overpredict or underpredict the leakage rate depending on the operating condition.

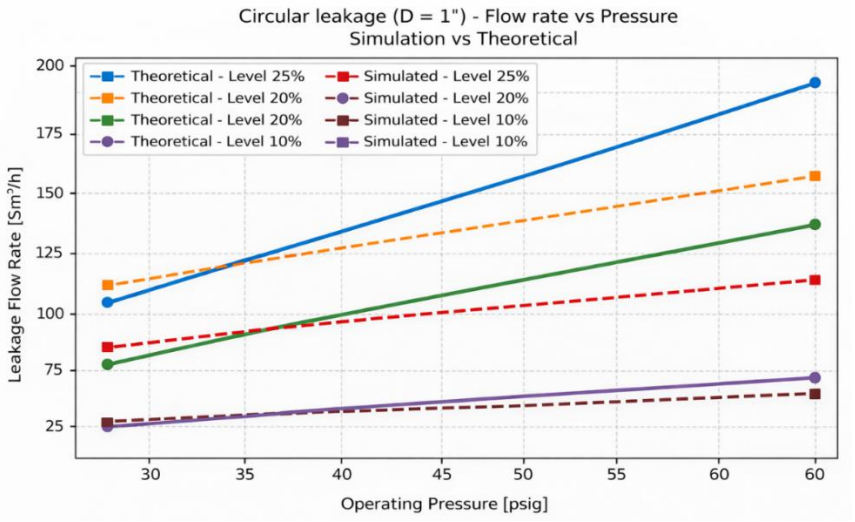

Figure. 3. *Validation of CFD results against theoretical orifice flow models*

These discrepancies can be attributed to simplifications inherent in analytical models, such as idealized flow assumptions and constant discharge coefficients. In contrast, the CFD simulations capture local flow effects, including velocity profiles and expansion behavior near the leak.
Overall, the comparison validates the CFD model as a reliable tool for predicting leakage behavior in circular openings, providing a solid baseline for the analysis of more complex geometries.

4.4. Mesh independence analysis

Figure 4 presents the mesh convergence analysis for the circular leak configurations evaluated in this study. A total of nine cases were considered, combining three pipe diameters (1 in, 1/2 in, and 3/4 in) and three damage severity levels (A1: 25%, A2: 15%, and A3: 10%). For each configuration, four mesh refinement levels were analyzed, denoted as Low, Medium, High-Medium, and High. These levels were defined based on non-overlapping ranges of mesh elements, with the Low level comprising meshes below approximately 250,000 elements, the Medium level between approximately 250,000 and 450,000 elements, the High-Medium level between approximately 450,000 and 750,000 elements, and the High level above approximately 750,000 and 1,000,000 elements. The results show that the predicted mass leakage flow rate stabilizes as the mesh is refined, indicating convergence of the numerical solution. In most cases, the difference between the High-Medium and High mesh levels is small, suggesting that further refinement has a limited effect on the predicted

leakage mass flow rate. Therefore, the selected mesh resolutions provide an adequate balance between numerical accuracy and computational cost for circular leak simulations.
This behavior was also observed for the amorphous leak geometries analyzed in this study. Therefore, the selected mesh resolutions provide an adequate balance between numerical accuracy and computational cost for leak simulations.

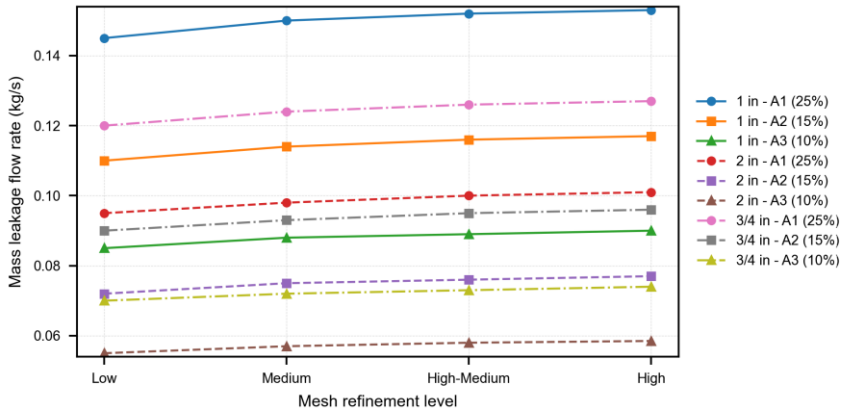

Figure. 4. *Validation of Mesh independence analysis circular leaks*

5. Development of Predictive Models

These findings highlight the limitations of conventional analytical models when applied to irregular leak geometries and demonstrate the importance of incorporating geometric descriptors, such as aspect ratio, into predictive models. The objective of this work was to develop a predictive model capable of estimating the **leakage flow rate** as a function of key **geometric and operational parameters** of the pipeline. These parameters include pipe diameter, operating pressure, leak severity, and geometric descriptors such as aspect ratio, as defined in the previous sections.
To achieve this objective, a set of **four base models** was evaluated. Each model was calibrated using the CFD simulation results, with the aim of identifying a formulation that provides both **predictive accuracy** and **physical interpretability**. The selection of these models is grounded in established approaches for modeling flow through orifices and data-driven regression techniques.

5.1. Physics-Based Orifice Model

The first model is based on classical **orifice flow theory**, where the flow rate is expressed as a function of pressure, fluid properties, and opening area. In compressible gas flow, the discharge through an orifice is commonly modeled as proportional to the **square root of the pressure difference and the characteristic area**, corrected by a discharge coefficient [10].

Based on this principle, a semi-empirical formulation was proposed Eq. (7):

$$Q = KD^2\sqrt{P}(1 + a \cdot Nivel)AR^b \quad (7)$$

This model preserves the physical structure of orifice flow, incorporating correction factors to account for leak severity and geometry. The use of such formulations is widely supported in literature, where leakage rates are commonly derived from Bernoulli-based relations corrected by empirical coefficients [11].

5.2. Log-Linear Model

The second model was formulated as a **log-linear regression**, which is commonly used when the response variable is expected to follow a multiplicative or power-law-type relationship with the explanatory variables. Under this approach, nonlinear scaling laws can be transformed into a linear regression problem by taking logarithms, which allows the regression coefficients to be interpreted as exponents or sensitivities. This is especially useful in engineering applications where the dependent variable scales with pressure, size, or other dimensional parameters through approximate power-law relationships [12].
The proposed expression was Eq. (8)

$$\ln Q = \beta_0 + \beta_1 \ln P + \beta_2 \ln D + \beta_3 \ln AR + \beta_4 S \quad (8)$$

This structure provides a simple and computationally stable framework for parameter estimation and facilitates interpretation of the relative influence of each predictor. Its main limitation is that it assumes a linear effect in transformed space and may therefore underrepresent more complex nonlinear interactions. Nevertheless, log-linear models are widely used as a first approximation in the analysis of empirical scaling relationships.

5.3. Generalized Linear Model (GLM)

The third candidate model was a **Generalized Linear Model (GLM)** with a Gamma distribution and logarithmic link function Eq. (9),

$$Q = e^{X\beta} \quad (9)$$

which was selected because leakage flow rate is a **strictly positive** response variable and may exhibit **right-skewed** behavior and **non-constant variance**. GLM theory explicitly recommends Gamma models for positively skewed continuous data, while the logarithmic link ensures positive predictions and allows multiplicative effects in the original response scale. This makes the Gamma GLM a statistically consistent option for flow-rate prediction when ordinary least-squares assumptions may not hold [13].

In the present context, Gamma GLM was evaluated as a flexible alternative to conventional linearized regressions. Its main advantage is that it accommodates heteroscedasticity more naturally than standard linear regression, which is desirable when the dispersion of the leakage rate increases with its magnitude. For this reason, Gamma GLMs have been broadly adopted in applied regression problems involving asymmetric positive responses [13].

5.4. Nonlinear Physics-Inspired Model

The fourth model extended the classical orifice-based structure by introducing **nonlinear corrective terms** to better represent the interaction between leak severity and geometry Eq. (10):

$$Q = KD^2\sqrt{P}\,(1 + a\,S + b\,S^2)\,AR^c \quad (10)$$

This formulation retains the physical dependence on diameter and pressure but increases flexibility through a quadratic term in leak severity and an adjustable exponent for aspect ratio. Such nonlinear regression models are commonly used when the response cannot be adequately represented through linear combinations of coefficients and when the fitted expression must remain explicit and interpretable from an engineering standpoint.

The use of customized nonlinear or semi-empirical expressions is particularly appropriate in cases such as irregular gas leaks, where the governing behavior is physically motivated but not fully captured by first-principles equations alone. Recent engineering studies have shown the value of deriving explicit semi-empirical formulas from numerical or experimental datasets in order to preserve interpretability while improving predictive performance.

6. Model Validation and Performance

6.1. Model Selection

The model development and validation workflow was implemented in Python using the CFD database generated for circular and irregular leaks. Four baseline formulations were initially calibrated using 70% of the simulation dataset and subsequently validated against the remaining simulation results, to verify that the developed model accurately reproduced the expected values when compared with the simulation outputs.

Model performance was evaluated by comparing predicted and observed leakage flow rates in the testing dataset through relative error, parity plots, and standard goodness-of-fit indicators. To evaluate which model best represents the leakage phenomenon, several complementary statistical metrics were employed alongside the relative error with respect to the test values. The coefficient of determination (R^2)[14] was used to quantify the proportion of variance explained by each model, while the root mean square error (RMSE)[15] assessed the average prediction error in physical units. In addition, the information criteria AIC[16], AICc[17], and BIC[18] were applied to compare model performance while penalizing complexity, where lower values indicate a better balance between goodness-of-fit and parsimony. Collectively, these metrics provide a robust framework for model selection, favoring models with higher R^2 and lower RMSE, AIC, AICc, and BIC

Among the baseline formulations, the log-linear model (Model 2) provided the best compromise between predictive accuracy and interpretability, with a maximum relative error of 13%, compared to 14% for Model 1, 29% for Model 3, and 30% for Model 4. The observed-versus-predicted plots shown in **Figure 5** further indicate that Model 2 consistently reproduces the overall data trend than the other baseline alternatives, although some systematic deviations persist at higher flow rates.

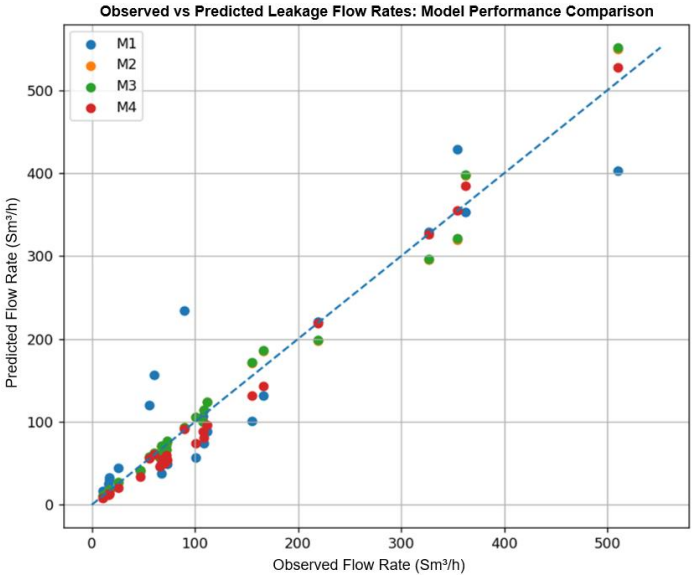

Figure. 5. *Observed vs. predicted leakage flow rates for model comparison*

6.2. Selected Model Optimization

Based on this initial comparison, Model 2 was selected as the reference formulation for refinement. Three improvement strategies were then investigated. Improvement 1 increased the flexibility of the model by incorporating nonlinear and interaction terms, leading to the following expression Eq. (11):

$$\ln Q = \beta_0 + \beta_1 \ln P + \beta_2 \ln D + \beta_3 \ln AR + \beta_4 S + \beta_5 S^2 + \beta_6 (\ln P \cdot S) + \beta_7 (\ln D \cdot \ln P) \qquad (11)$$

where $Q$is the leakage flow rate, $P$is the operating pressure, $D$is the pipe diameter, $AR$is the aspect ratio, and $S$is the dimensionless leak severity parameter previously defined as $S = D_h/D_o$. This modification was introduced to capture curvature in severity and coupled effects between pressure, diameter, and leak condition. Improvement 2 adopted an alternative normalization inspired by orifice-flow behavior, reformulating the dependent variable as $Q/(D^2\sqrt{P})$ in order to isolate the dominant physical scaling and concentrate the regression on geometry-related effects. Improvement 3 retained the base log-linear structure but applied a weighted least-squares (WLS) scheme, with weights $w_i = 1/Q_i^2$, to reduce heteroscedasticity and limit the dominance of high-flow observations during calibration.

The three refined versions were subsequently re-evaluated using the testing dataset. Improvement 1 provided the best overall performance, reducing the maximum relative error from 13% to approximately 4%. In contrast, Improvement 2 yielded a maximum relative error of 27%, whereas the WLS-based formulation reduced the maximum error to 19%. These results indicate that the main limitation of the baseline log-linear model was not only heteroscedasticity, but also its inability to represent nonlinear dependence on leak severity and the interaction among pressure, diameter, and geometry. The superior performance of Improvement 1 therefore supports the inclusion of physically motivated nonlinear terms in the final predictive expression.

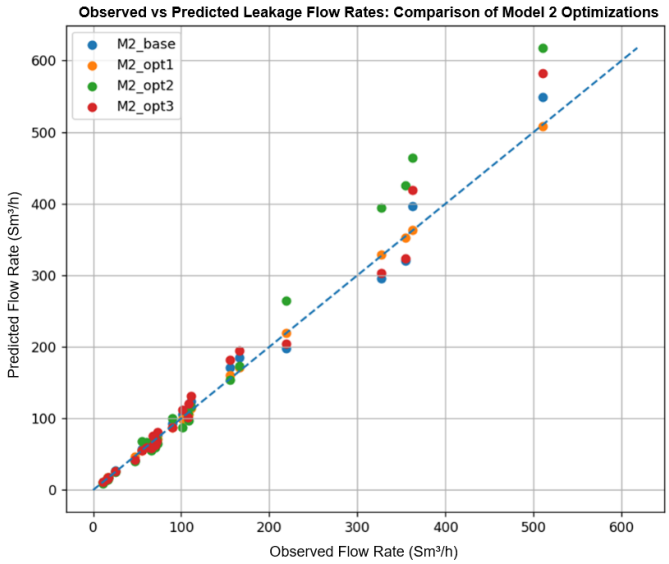

Figure. 6. *Observed vs. predicted leakage flow rates for Model 2 Optimizations*

The final correlation adopted for circular and irregular leaks is Eq. 12:

$$Q = \exp(-0.82599 + 0.48611\ln(P) + 1.64713\ln(D) - 0.09435\ln(AR) + 24.03212S - 34.68900S^2 + 0.06882\ln(P)\,S + 0.00474\ln(D)\ln(P)) \quad (12)$$

where Q is expressed in Sm^3/h, P is the operating pressure in psig, $D$is the nominal pipe diameter, AR is the aspect ratio, and $S$is the severity parameter. Within the calibrated domain of this study, this expression provided the best balance between predictive accuracy, interpretability, and computational efficiency. Its applicability is restricted to the conditions considered in the CFD campaign, namely methane discharge to atmosphere, operating pressures between 27 and 60 psig, nominal diameters of 3/4 in, 1 in, and 2 in, severity levels of 10%, 20%, and 25%, and the circular and irregular leak geometries defined in the methodology.

7. Conclusions

This study presented a comprehensive framework for the analysis and prediction of gas leakage in natural gas distribution systems, integrating computational fluid dynamics (CFD) simulations with semi-empirical modeling techniques to explicitly account for the influence of leak geometry.

The CFD results demonstrated that leakage flow rates are strongly governed by operating pressure, pipe diameter, and leak severity, showing consistent trends with classical compressible flow theory. In particular, circular leaks exhibited predictable behavior, with flow rates increasing nonlinearly with leak size and pressure. The comparison with analytical orifice models confirmed a generally good agreement, validating the numerical approach as a reliable tool for leakage prediction. Minor discrepancies observed at higher severity levels were attributed to simplifying assumptions in theoretical formulations, such as constant discharge coefficients and idealized flow conditions.

A key contribution of this work is the quantification of the impact of irregular leak geometries. The results showed that amorphous leaks systematically produce higher leakage rates than equivalent circular openings, with increases ranging between 3% and 8% under the evaluated conditions. This behavior is explained by geometric effects that modify local flow contraction and jet development, leading to an increased effective discharge capacity. These findings highlight the limitations of conventional models based on circular assumptions and demonstrate the need to incorporate geometric descriptors in leakage analysis.

To address this limitation, a set of predictive models was developed and evaluated using the CFD database, including a physics-based orifice formulation, a log-linear model, a Gamma generalized linear model, and a nonlinear physics-inspired model. Among the baseline approaches, the log-linear model provided the best balance between simplicity and predictive capability, although it exhibited limitations in capturing nonlinear interactions and heteroscedastic behavior.

Further refinement of the log-linear formulation through the inclusion of nonlinear and interaction terms significantly improved predictive performance. The optimized model reduced the maximum relative error from approximately 13% to about 4%, outperforming alternative approaches such as normalization-based transformations and weighted regression schemes. This improvement demonstrates that the accurate representation of leakage behavior requires not only physically meaningful variables, but also flexible formulations capable of capturing coupled effects between pressure, pipe diameter, leak severity, and geometry.

The final predictive correlation derived in this study provides a computationally efficient and physically interpretable tool for estimating leakage flow rates in both circular and irregular geometries within the calibrated domain. Its applicability is limited to methane discharge under the range of conditions analyzed, including pipe diameters of 3/4 in, 1 in, and 2 in, operating pressures between 27 and 60 psig, and severity levels between 10% and 25%.

Overall, this work demonstrates that incorporating geometric descriptors such as aspect ratio significantly enhances the accuracy of leakage predictions and provides a more realistic representation of real pipeline defects. Future work should focus on extending the model to a wider range of operating conditions, including buried pipelines and transient effects, as well as performing experimental validation to further support the proposed methodology.

8. References

[1] Z. D. Weller, S. P. Hamburg, and J. C. Von Fischer, “A National Estimate of Methane Leakage from Pipeline Mains in Natural Gas Local Distribution Systems,” *Cite This: Environ. Sci. Technol*, vol. 54, p. 8967, 2020, doi: 10.1021/acs.est.0c00437.

[2] “AR6 Climate Change 2021: The Physical Science Basis — IPCC.” Accessed: Mar. 06, 2026. [Online]. Available: https://www.ipcc.ch/report/sixth-assessment-report-working-group-i/

[3] C. Liu, Y. Liao, J. Liang, Z. Cui, and Y. Li, “Quantifying methane release and dispersion estimations for buried natural gas pipeline leakages,” *Process Safety and Environmental Protection*, vol. 146, no. 44, pp. 552–563, Feb. 2021, doi: 10.1016/j.psep.2020.11.031.

[4] Y. Zeng and R. Luo, “Numerical Analysis on Pipeline Leakage Characteristics for Incompressible Flow,” *Journal of Applied Fluid Mechanics*, vol. 12, no. 2, pp. 485–494, Mar. 2019, doi: 10.29252/jafm.12.02.28612.

[5] B. Anifowose and M. Odubela, “CFD analysis of natural gas emission from damaged pipelines: Correlation development for leakage estimation,” *J. Clean. Prod.*, vol. 199, pp. 257–271, Oct. 2018, doi: 10.1016/j.jclepro.2014.12.066.

[6] Y. Ding, P. Xu, Y. Lu, M. Yang, J. Zhang, and K. Liu, “Research on Pipeline Leakage Calculation and Correction Method Based on Numerical Calculation Method,” *Energies (Basel).*, vol. 16, no. 21, pp. 1–17, 2023.

[7] DANIEL A. CROWL and JOSEPH F. LOUVAR, Eds., *Chemical Process Safety*. Accessed: Mar. 06, 2026. [Online]. Available: https://elmoukrie.com/wp-content/uploads/2022/06/daniel-crowl-joseph-louvar-chemical-process-safety_-fundamentals-with-applications-international-series-in-the-physical-and-chemical-engineering-sciences-pearson-2020.pdf

[8] L. Lu, X. Zhang, Y. Yan, J.-M. Li, and X. Zhao, “Theoretical Analysis of Natural-Gas Leakage in Urban Medium-pressure Pipelines,” *Journal of Environment and Human*, vol. 2014, no. 2, pp. 71–86, Jul. 2014, doi: 10.15764/eh.2014.02009.

[9] X. Wang, T. Hou, W. Gao, K. Yu, T. Zhang, and Y. Tan, “Experimental study on the diffusion process of natural gas from buried pipelines to underground confined spaces,” *Natural Gas Industry B*, vol. 11, no. 5, pp. 603–615, Oct. 2024, doi: 10.1016/j.ngib.2024.09.002.

[10] H. J. Bomelburg, “Estimation of Gas Leak Rates Through Very Small Orifices and Channels Prepared for the Nuclear Regulatory Commission,” 1977.

[11] W. J. Kostowski and J. Skorek, “Real gas flow simulation in damaged distribution pipelines,” *Energy*, vol. 45, no. 1, pp. 481–488, 2012, doi: 10.1016/j.energy.2012.02.076.

[12] X. Xiao, E. P. White, M. B. Hooten, and S. L. Durham, “On the use of log-transformation vs. nonlinear regression for analyzing biological power laws,” *Ecology*, vol. 92, no. 10, pp. 1887–1894, Oct. 2011, doi: 10.1890/11-0538.1.

[13] by Roger Payne, “A Guide to Regression, Nonlinear and Generalized Linear Models in GenStat ® (15 Edition) th,” 2012. [Online]. Available: http://www.genstat.co.uk/

[14] T. Gneiting and A. E. Raftery, “Strictly proper scoring rules, prediction, and estimation,” *J. Am. Stat. Assoc.*, vol. 102, no. 477, pp. 359–378, Mar. 2007, doi: 10.1198/016214506000001437.

[15] “mean_squared_error — scikit-learn 1.8.0 documentation.” Accessed: Apr. 13, 2026. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean_squared_error.html?utm_source=

[16] “statsmodels.tools.eval_measures.aic - statsmodels 0.15.0 (+976).” Accessed: Apr. 13, 2026. [Online]. Available: https://www.statsmodels.org/dev/generated/statsmodels.tools.eval_measures.aic.html?utm_source

[17] “R: Computing AIC, AICc, QAIC, and QAICc.” Accessed: Apr. 13, 2026. [Online]. Available: https://search.r-project.org/CRAN/refmans/AICcmodavg/html/AICc.html

[18] “statsmodels.tools.eval_measures.bic - statsmodels 0.15.0 (+976).” Accessed: Apr. 13, 2026. [Online]. Available: https://www.statsmodels.org/dev/generated/statsmodels.tools.eval_measures.bic.html?utm_source