A closer look at underground natural gas pipeline leaks across the United States

Younki Cho1,*, Kathleen M. Smits2, Nathaniel L. Steadman3, Bridget A. Ulrich4, Clay S. Bell1, and Daniel J. Zimmerle1

Underground natural gas (NG) pipeline leakage can result in methane (CH4) buildup and migration through the soil. What is not well understood in such scenarios is how the soil conditions affect the gas migration behavior, particularly in regard to the relative contributions of specific soil properties such as soil moisture content. The objective of this study was to investigate the effects of soil properties on CH4 concentration and migration from leaking underground NG pipelines. Site characteristics such as surface cover and spatial dimensions, soil samples, and gas concentration data were collected from over 70 gas leakage sites across the United States using a standardized sampling method. Soil samples were collected from excavation sites that were 1.5–5 m in depth. The collected soil samples were analyzed in the laboratory to measure the soil texture, permeability, and moisture. Statistical analysis was performed to evaluate the effects of soil properties on CH4 migration distance and concentration. Soil texture was consistent across geographic locations due to standardized pipeline backfill protocols, allowing for the analysis of gas concentration and transport data with respect to soil conditions. Soil moisture content was the dominant influence on the gas concentration and spreading distance. High soil moisture content was associated with reduced lateral diffusion and elevated concentrations near the leak point, whereas dry conditions were associated with reduced concentrations and greater spreading distance. Increasing soil moisture content reduced the lateral diffusion of CH4 into the soil due to water-induced tortuosity, resulting in elevated concentrations close to the leak point. Lateral migration of CH4 was suspected to be by diffusion, starting at 5 m from the leaks, while transport within the immediate vicinity of the leak was controlled by advection. These findings demonstrate a pronounced effect of soil moisture content and permeability on CH4 migration distance and concentration, providing key insight into the effects of soil conditions on NG migration and how to account for such effects in leak detection surveys.

Keywords: Methane gas migration, Natural gas, Unsaturated soil, Gas transport, Underground pipeline leakage

Introduction

Demand for natural gas (NG) increased rapidly in the past decade due in part to lower costs because of more efficient drilling and production techniques (Energy Information Administration, 2020). NG is often regarded as a clean energy transition fuel, with potential climate benefits including reduced carbon dioxide (CO2) emissions relative to other fossil fuels, such as coal and oil (Lelieveld et al., 2005; Howarth, 2014; Zhang et al., 2014). However, fugitive emissions of methane (CH4) arising from NG production, processing, and distribution threaten to offset these potential climate benefits, as the global warming potential of CH4 is substantially greater than that for CO2 (Intergovernmental Panel on Climate Change, 2013). Fugitive emissions from underground pipelines are especially concerning due to the potential for gas to buildup and migrate through soil and ultimately release into the air or a substructure (e.g., a basement), which could lead to the potential of explosion as the concentration exceeds the lower explosive limits (LELs) of CH4. Although NG pipeline safety has greatly improved in recent decades (Vetter et al., 2019), incidents still occur due to aging infrastructure, excavation, and/or human error associated with construction or agricultural operations. Furthermore, failures in U.S. NG distribution pipelines over the past 10 years (significant incidents in 2011–2020) have resulted in an average of 8 fatalities, 48 injuries, and US$213 million worth of property damage annually (Pipeline and Hazardous Materials Safety
In the event of a leak, the mechanism of gas emission into the surrounding soil as well as the area of influence of the leak will vary depending upon the pressure conditions of the pipelines. NG pipeline systems encompass a wide range of both low- and high-pressure conditions, whereas low pressure conditions (1.5–2,000 kPa) prevail in distribution systems. CH₄ flow is typically dominated by advective flow near the leak point due to pipeline pressure and transitions to diffusion-dominated flow due to concentration differences at further distances away from the leak point (Forde et al., 2018; Gao et al., 2021). Furthermore, preferential flow paths from permeability variations in soil may cause increased gas migration distances, causing gas to emerge at surface locations considerable distances away from the leak location (Hendriks et al., 2010; Forde et al., 2019). Previous studies show that the NG migration length scales from a leaky pipeline may vary broadly in the range of 2–30 m (Okamoto and Gomi, 2011; Yan et al., 2015; von Fischer et al., 2017). For example, elevated gas readings were observed over a distance of 20–30 m from a leaking distribution pipeline with a relatively low leakage rate (< 6 L min⁻¹; von Fischer et al., 2017).

In addition to the belowground transport behavior, the surface CH₄ effluxes are strongly influenced by the leak rate and subsurface heterogeneity. Previous studies found high spatial and temporal variability in fluxes (Lyman et al., 2017; Forde et al., 2018; Lyman et al., 2020). Forde et al. (2018) found that subtle heterogeneities can lead to preferential pathways for CH₄, influencing the spatial and temporal distribution at the ground surface. Lyman et al. (2017) showed that flux magnitude at NG wells is inversely related to distance from the wellhead and could not observe meaningful relationships between fluxes with temperature, atmospheric pressure, soil moisture content, or any of the other meteorological or soil properties. Although these studies assessed gas migration at the ground surface, the effects of varying subsurface soil conditions (e.g., soil texture, moisture content, and compaction) on gas migration were not well characterized. Part of the ambiguity in understanding gas migration behavior arises from difficulties in incorporating the various processes that control the gas migration behavior into a systematic evaluation approach and a lack of field data capable of testing such approaches. Consequently, the importance of soil properties is often poorly understood and hence excluded from field evaluation protocols. It is therefore of vital interest to assess the effect of soil properties on gas migration to better account for such effects in leak detection surveys.

The objective of this study was to investigate the effects of soil properties on CH₄ concentration and migration from leaking underground NG pipelines. A leak detection and repair (LDAR) survey was designed and implemented to collect comprehensive data at over 70 NG leakage sites across the United States classified as Grade 2 and 3 leaks. In general, Grade 1 leaks represent an existing or potential hazard to persons or property, requiring immediate action. Generally, Grade 2 and 3 leaks are recognized to be non-hazardous at the time of detection and are therefore prioritized for later repair (Gas Piping and Technology Administration [PHMSA], 2021), motivating stricter enforcement of safety practices. In these incidents, excavation damage was the most common cause of the leaks and then outside force damage. The environmental conditions and soil properties that affect gas migration behavior, including the migration of high concentration contours and extent, are not well understood. The effects of environmental conditions and soil properties that control the flow of NG through soils have not been statistically analyzed. Improving our understanding of the conditions that affect gas accumulation and migration behavior will improve leak response practices and ultimately aid NG operators and regulators in locating leaks.

Previous NG leakage studies mainly focus on quantifying fugitive CH₄ emissions using emission measurements from above the ground level, including well pads (Allen et al., 2013; Brantley et al., 2014; Allen et al., 2015; Omara et al., 2016; Bell et al., 2017; Yacovitch et al., 2017), gathering pipelines (Zimmerle et al., 2017), processing plants (Marchese et al., 2015; Mitchell et al., 2015; Roscioli et al., 2015; Yacovitch et al., 2015; Vaughn et al., 2017), transmission and storage facilities/pipelines (Lelièveld et al., 2005; Subramanian et al., 2015; Boothroyd et al., 2018; Thorpe et al., 2020), and distribution facilities/pipelines (Jackson et al., 2014; Lamb et al., 2015; von Fischer et al., 2017; Weller et al., 2018). Such studies used aircraft, tower, mobile, point-source, and surface samplings of CH₄ concentration in the atmosphere at various spatiotemporal scales. In parallel, the effects of atmospheric conditions including air temperature, atmospheric pressure, precipitation, and wind are taken into account, allowing for the estimation of emissions using various atmospheric transport models (Abriola and Pinder, 1985; Young, 1992; Czepiel et al., 2003; Gebert and Groengroeft, 2006). Although pipelines were occasionally included in the abovementioned studies, the influence of the complex subsurface gas migration behavior and resulting diffuse surface presentation was not considered extensively.

In subsurface NG leakage scenarios, the effect of the subsurface conditions on the resulting surface CH₄ distributions is not yet well understood even though there were several studies investigated the interconnections between subsurface CH₄ migration and emissions (Chamindu Deepagoda et al., 2016; Hendriks et al., 2016; Forde et al., 2018; Ulrich et al., 2019; Schollaert et al., 2020). Gas behavior is influenced by soil conditions, subsurface infrastructure, pipeline pressure, and gas composition (Okamoto and Gomi, 2011; Lyman et al., 2017; Ulrich et al., 2019; Lyman et al., 2020). Subsurface migration is further complicated by pressure differentials, which develop from fluctuation in barometric pressures due to wind variations and meteorologically induced long-term changes in barometric pressure, which can cause both vertical and lateral transport of gas (Forde et al., 2019; Lyman et al., 2020; Fleming et al., 2021). The effect of soil moisture contents on gas migration was not observed (Forde et al., 2018). Surface conditions such as pavement or structures create barriers to gas flow and release, increasing lateral transport or causing accumulation belowground (Ulrich et al., 2019; Schollaert et al., 2020).
Committee, 2018). Three to four soil samples at varying depths (1–3 ft), pipeline conditions, site survey data, and underground gas concentration profiles were collected for each site. The number of leak sites was selected to provide a statistically significant sample size and a variety of site conditions while not overburdening operator field teams. The collected data from LDAR surveys together with regression analysis were used to identify dominant parameters affecting underground gas migration. The soil properties analyzed in the laboratory were summarized to provide insight into the relationship between soil properties and gas migration distances. We further assessed the effects of ground cover (pavement and grassy surfaces) on gas concentrations and migration distances. Additionally, the reasons a leak occurred as cited by the operator personnel in the LDAR survey were also assessed. The results presented here provide insights into the effects of soil properties and, in part, surface conditions on CH₄ concentrations and spreading distances. The broader findings of this study can be incorporated into recommended practices for operator risk assessments to address NG leakage incidents.

Materials and methods

Scope of study

The study team partnered with NG distribution operators to collect data from a large number of NG leakage sites using a leak response survey protocol developed specifically for this work. Distribution system operators in the United States typically grade all discovered leaks using guidelines (not standards) developed by industry regulatory agencies. For this study, we focus on leaks classified as Grade 2 and 3 leaks. This allowed time for the collection of the survey data prior to and during the leak repair. Data were collected from June 2019 to March 2020 at 77 leak locations. As our method relies on the collection of a standard set of LDAR at Grade 2 and 3 leak locations, results may not account for specific characteristics of Grade 1 leaks. However, it is important to note that the leak grade is not an indication of the leak size but rather an indication of hazard to persons or property (Railroad Commission of Texas, 2017); thus, we expect limited bias in terms of leak size. The seasonal variation was not considered in this study as repeat measurements were not available.

Survey process

Four local NG distribution companies (LDCs) participated in the data collection. Approximately, 20 LDAR surveys were conducted by each LDC, resulting in 77 site surveys over a 9-month period. LDCs were located throughout the country (Eastern, Southwestern, Midwestern, and Western), representing diverse geographic locations for assessing spatial heterogeneity. 46 were classified as Grade 2 leaks and 31 as Grade 3 leaks. The LDCs utilize slightly different leak classification criteria, and therefore, differences between Grade 2 and 3 leaks were not included in this study.

Survey design

The LDAR survey was designed in collaboration with LDCs to collect a standardized data set on underground pipeline leaks in urban and rural areas. The survey was organized to obtain data during the initial detection and final repair components of each organization’s LDAR program. Although the protocols of each LDC vary, the data collected were consistent. For the initial detection phase, the spatial dimensions of the site, pipeline location, and limited aboveground gas concentrations were recorded. This information was used to construct representative visualizations of the site and predict the leak location. For this work, the final repair component includes the time when the leak is further investigated for centering and pinpointing prior to repair as well as the actual repair operation. During leak centering and pinpointing—the processes by which the exact leak location is determined—holes were drilled (the holes are commonly known as “bar holes” and/or “pogo holes,” depending upon how they are drilled) and underground concentrations were measured and recorded throughout the site. This provided a basis for understanding the spatial extent of the leak and enabled the identification of locations with elevated concentrations that indicated gas migration away from the leak location. The number of concentration readings varied for each site based on LDC procedures as well as the specific leak location. Sometimes, the leak centering and pinpointing is performed with only a few measurements, while other times, more measurements are needed to pinpoint the leak location. On average, 12 underground concentration readings were recorded for each site. To center and pinpoint a leak location, typically, 1–3 in. diameter bar holes or < 1 in.-diameter pogo holes were drilled to the depth of the pipeline. This depth varied slightly as the pipeline depths varied between 0.5 and 1.5 m below ground surface. Gas concentrations were measured using one of two leak detection survey instruments: a Detec Pak Infrared DP-IR™ detector (accuracy: greater value of either ±0.5% or ±10% of reading from % gas on manual mode; sensitivity: 1 ppmv on auto ranging and 5 ppmv on manual ranging) with the appropriate probe and hose attachment (Heath Consultants DP-IR-1+) or a SENSIT® GOLD G2 detector (accuracy: ±10%; sensitivity: 10 ppmv) with a semiconductor sensor (SENSIT ® GOLD G2). Although the Heath instrument is highly selective to CH₄, the SENSIT instrument is sensitive to a range of volatile hydrocarbons and is selective to CH₄. In general, given the high proportion of CH₄ in NG (typically 95%), we used the measurements interchangeably and utilize the term “gas” for all measurements reported by industry partners. Although the specifics of the centering and pinpointing procedure vary by LDC, a brief guidance of the pinpointing is provided by the PHMSA (2017).

After data were collected, the concentration data were plotted in conjunction with the site dimensions to construct a representative visualization of each leak location. During final repair, information about the actual leak location, cause, and extent was recorded and compared to the suspected leak location based on the leak centering and pinpointing data. Three soil samples were collected at
different depths (1–3 ft) from each site using a standardized soil sampling procedure and testing kit provided by the study team. The soil sampling kit included 2 in. × 3 in. stainless steel soil sampling sleeves and 2 in. round plastic end caps to preserve intact soil cores. The study team provided written guidelines and training to operator personnel to ensure sampling consistency between LDCs.

**Survey analysis**

Collected data were postprocessed to identify trends and investigate the degree to which certain soil parameters affected CH₄ concentrations. First, knowledge of the site location was used to construct site maps that marked the location of initial detection (i.e., the suspected leak location), the location of the final repair (i.e., the actual leak location), and each measured gas concentration, placed relative to the actual leak location. Operator teams took photos of each survey and recorded additional site conditions such as pipeline depth and pressure and identified and tabulated the cause of the leak, based upon the condition of the pipeline and the type of repair performed. Due to differences in protocol among the participating companies, aboveground concentration data for the initial detection phase were only available from 34 surveys. As the main data needed for this analysis was the final repair data, the limited initial detection phase concentration data only affected a small portion of the analysis as will be described further below.

Soil samples were returned to the study team, and properties of the soil samples were determined through laboratory analysis at the University of Texas at Arlington. This included soil textural class, particle size, dry bulk density, total porosity, saturated hydraulic conductivity, intrinsic permeability, and soil gas diffusivity. In total, 199 soil samples (3–4 samples per site) were analyzed. Upon receiving the samples, the soil moisture content (g/g) of the bulk sample was measured. The samples were then oven-dried for 2 days and mechanically sieved in accordance with the ASTM E11 standard procedures to determine the particle size distribution and soil textural class. Notably, larger aggregates were not crushed to obtain smaller aggregates; rather, the natural aggregates of the different sizes were collected by sieving. The study team characterized the fractions with respect to grain-size distribution based on the U.S. Department of Agriculture soil texture classification. Soil properties were then used to calculate the soil gas diffusivity ($D_p/D_w$, the ratio of gas diffusion coefficients in soil and free air; Penman, 1940), hydraulic conductivity (cm/s; Hazen, 1892; 1911), and intrinsic permeability (m²; Hazen, 1892; 1911).

Statistical analysis was used to investigate the empirical relationship between the measured parameters. All analysis was performed using Statistical Package for the Social Sciences software. A bivariate Pearson analysis was used to screen the correlation between measured parameters (i.e., soil properties and spatial dimensions) and subsurface gas concentrations. Results of the analysis were then used in a stepwise linear regression to iteratively examine the statistical significance of each parameter in a forward selection approach. Stepwise regression was selected as an analysis procedure as it is an efficient way of choosing multiple parameters while removing insignificant parameters from the analysis. Consolidated data included gas concentration, corresponding distance from the leak location, and the average soil properties associated with leak sites. The regression model was then used to examine the relationship between the gas concentration and the soil properties. As regression models can become unstable with high standard errors, multicollinearity between measured parameters was also examined by variance inflation factors (VIFs). Analysis of variance (ANOVA) was also performed to compare significant impact of these parameters on gas concentrations.

**Results and discussion**

**General soil properties**

LDAR surveys revealed little variation in soil type regardless of geographic location. Soils were mostly classified as poorly graded (58% of coefficient of uniformity > 6 and 22% of < 1 coefficient of curvature < 3) and coarse sands (grain sizes between 0.2 and 2.0 mm). As 3–4 soil samples were analyzed per site, the presence of soil layering was also considered; while 4 surveys exhibited layered soils, most surveys exhibited a uniform soil profile. Further, since all soil samples were collected in the excavation dug to repair the pipeline, soil samples are more reflective of the pipeline backfill than of the natural soil adjacent to the pipeline. This is attributed to standard practices for underground installation and construction of pipelines (e.g., ASTM D2321 classification II SP), which recommends coarse-grained soils with poorly graded, clean gravelly sand with little to no fine material as fill.

Total porosity values for soil cores ranged from 0.23 to 0.59 cm³/cm³ with an average value of 0.44 cm³/cm³ and a standard deviation of 0.07 cm³/cm³ for 199 samples, which are in typical ranges for coarse sands (Morris and Johnson, 1967; Rawls et al., 1983). The mode was observed with the bin between 0.45 and 0.50 cm³/cm³ (Figure 1A). Soil moisture content ranged from 0.0004 to 0.49 g/g with averaging 0.15 ± 0.08 g/g. Because of a unimodal distribution observed in Figure 1B, the median of the moisture content was 0.15 g/g with the mode between 0.15 and 0.20 g/g bin. These values are consistent with typical moisture content observed from loamy sand and sandy loam, which are 0.14 and 0.20 g/g, respectively (Evett, 2008). The soil moisture can vary depending on season and location, with the surface coverage type (i.e., pavement, grass, or bare soil) having a marked effect on the moisture content (Leeper et al., 2019). As expected, the soil moisture content from unpaved areas (i.e., grass-covered surfaces) was higher than that of paved areas consisting of roads or sidewalks (0.19 ± 0.08 g/g vs. 0.14 ± 0.06 g/g, respectively). As seen in Figure 1A and B, the total porosity and soil moisture content show relatively symmetrical distributions. Grain size distributions (D10 and D60 in Figure 1C and D) indicated a wide range of grain sizes among the samples. Based on the D10 grain size, the permeability estimated by the Hazen correlation (Hazen, 1892; 1911) ranges from $6 \times 10^{-11}$ to $2 \times 10^{-10}$ m² with an average of $9 \times 10^{-11}$ m² (Figure 1E), which is in...
range of typical permeability values for sand (Mahmoodlu et al., 2016).

The effect of the permeability on gas migration distance was investigated because of the relative contribution of advection and diffusion to CH₄ distribution varies with distance from the leak point based on the subsurface properties. In this work, the maximum gas migration distance is defined as the distance between the leak point and the farthest bar hole reading from the leak point with CH₄ concentrations above background. These distances were measured while centering and pinpointing the leak prior to final repair. The gas migration distance was determined by measuring the gas concentration in the bar holes until a reading of 0% gas was reached in each direction (N, E, S, and W) from the highest reading to center the leak. As seen in Figure 2A, it is observed that large maximum gas migration distances are only associated with low permeabilities (< 0.5 x 10⁻¹⁰ m²) and all high permeabilities (> 1 x 10⁻¹⁰ m²) are only associated with smaller distances (< 5 m). However, there were two cases with high permeabilities that migrated up to 6 m. The findings are limited to sands with a permeability distribution of 6 x 10⁻¹³ to 2 x 10⁻⁸ m². This behavior suggests that for the soils tested, the contribution of advection and diffusion results in varying gas behavior. As demonstrated in Gao et al. (2021), advective gas transport due to a pressure gradient caused by the gas release dominates the CH₄ transport close to a leak source, driving the migration of high concentration contours. However, in the far field (distances greater than 1.5 m), diffusion caused by the molecular concentration gradients dominates the migration. This is further complicated with differences in soil permeability.

The influence of advective transport will be greater for a high permeability soil resulting in faster gas flow at higher concentrations through the high permeability pathways close to the leak point. However, the overall spreading of the gas in high permeability soils is oftentimes less than low permeability soils due to the influence of diffusion. In the case of a low permeability soil, the advective effects are less, resulting in slow migration. However, the more predominant influence of the gas diffusion in the low permeability soils influences the spreading width of the CH₄ plume resulting in more outstretched contours.

The advection close to the source results in the gas escaping in the high permeability soils (Figure 2A; Gao et al., 2021). In the low permeability soils, diffusion was suspected to be the driving transport mechanism at distances greater than 5 m (Figure 2B). Although slow, CH₄ slowly migrated farther away from the leak point, causing spreading of CH₄ gradually in soil over time. Diffusive behavior is also seen in high permeability soils but to a lesser extent due to the competition, per say, with the vertical transport of the gas to the soil surface and escape into the atmosphere (Gao et al., 2021). As the gas leak
continues in-between LDAR, the area of influence can slowly become larger over time and reach steady state, influencing the leak classification. Therefore, frequent re-evaluation of risk assessment is recommended, especially for areas with low permeability soils in the vicinity of buildings. Further complicating the gas behavior is the influence of soil moisture, which will be discussed below.

Impact of surface cover on gas behavior
To investigate the effect of surface cover on the gas migration distance, we compared the observed maximum gas migration distances for paved and unpaved areas. Surface cover was binned into two categories, paved and unpaved surfaces. Paved surfaces were predominately asphalt and concrete streets, driveways, and sidewalks, and unpaved surfaces were grassy areas, such as lawns, open fields, or median areas. Surface cover could be determined by the site descriptions and plan maps provided by the LDCs. The subsurface CH$_4$ concentration was investigated regardless of surface cover; 60% of the subsurface gas concentration measurements exceeded 5 vol% (LEL of CH$_4$, which is equivalent to 50,000 ppmv); 82% of the subsurface gas concentrations that exceeded the LEL were located within 3 m of the actual leakage location. These results suggest that drilling additional bar holes or excavating around a leak to allow gas to dissipate is warranted due to the LEL levels.

For paved surfaces, the lateral transport of CH$_4$ was greater than unpaved surfaces with an average maximum distance of 12.3 m. In the case of unpaved surfaces, the average transport distance was 4.4 m. Surface concentrations, determined based on the 34 surveys collected during initial detection, were also high for the paved surfaces than unpaved. For paved surfaces, average surface concentration readings of 9,745 ± 61,571 ppmv were observed. The average surface CH$_4$ concentration of leaks with paved surfaces measured during the initial detection was 10 times higher than for unpaved surfaces. It is important to note that two outliers existed. Leaked gas traveled along water utility lines and was detected 10.5 m away from the actual leak point for both cases.

Another consideration is the difference in the initial leak detection location versus the actual leak location and how this is influenced by surface cover. On average, leaks located under unpaved surfaces were located by the initial detection survey to within 1.0 ± 0.8 m of the actual repaired leak point, and detectable surface gas concentrations were detected up to 4.4 m from the actual leak location. For paved surfaces, the difference between initially detected and final repair was greater than unpaved surfaces with an average maximum distance of 3.1 ± 3.1 m between the initially detected location the actual leak location. As mentioned in the Methods section, there were only 34 surveys that contained both initial detection locations and final repair locations. Therefore, the influence of the surface cover on the leak pinpointing is limited due to data scarcity.

Impact of distance and soil moisture on gas behavior
The effects of distance from the leak source and soil moisture on subsurface CH$_4$ concentration were evaluated. As seen in Figure 3, the CH$_4$ concentration typically dropped exponentially with an increase in distance from the leak source. On average, a 0% reading was reached 2.2 ± 2.9 m from the leak source. For one of the cases, the leaked gas migrated through the soil, along the gas pipeline, which was located under a paved surface, approximately 20.1 m from the leak source at a subsurface concentration of 210,000 ppmv. In this case, there was no water/sewer line present. In two cases, the gas was located 10.1 m from the...
leak source, adjacent to a water/sewer line with subsurface concentration readings of 1,000 ppmv.

Figure 3 provides insight into the relative change in concentration with distance from the leak location for varying soil moisture conditions (0.01–0.33 g/g). As seen previously in the literature, when the soil moisture content increased, the CH$_4$ migration distance decreased, leading to higher CH$_4$ concentrations close to the leak point. On the other hand, moving away from the leak point (e.g., 5 m from the leak source), the higher soil water saturation leaks to a lower CH$_4$ concentration. For instance, when the data after 5 m were used for comparison, when the moisture content was greater than 0.15 g/g, the average CH$_4$ concentration was 18,460 ppmv. However, when the moisture content is less than 0.15 g/g, the average concentration was much higher (225,861 ppmv). This behavior can be explained by two aspects as pointed out by Gao et al. (2021). Water saturation has reverse effect on the contribution of advection and diffusion to CH$_4$ distribution. At the higher water saturation, the tortuosity and pressure increase, resulting in increased advection and lowered lateral diffusion. Thus, CH$_4$ accumulation and high concentrations are observed near the leak point. The relative contribution of advection and diffusion to CH$_4$ distribution varies with distance from the leak point.

Regression analysis

A regression analysis was carried out to investigate the relative significance of soil parameters influencing the subsurface CH$_4$ concentration. Prior to regression analysis, a bivariate correlation was performed to screen the parameters affecting the concentration. The bivariate correlation coefficient ($r$) showed weak linear correlations between evaluated soil properties and subsurface gas concentrations. Most properties were negatively correlated with subsurface concentration (Table S1). Subsurface gas concentrations negatively correlated with distance ($r = -0.10, P < 0.01$), as expected. A weakly negative correlation was also observed for the moisture content ($r = -0.21, P < 0.001$). Regardless of grain size (D10, D30, D50, and D60), a negative correlation was consistently observed (Table S1). The total porosity showed very weak correlation with concentration ($r = -0.10, P < 0.01$). The measurement depth ($r = -0.01$) did not reach statistical significance ($P = 0.84$). In addition, all significance values from a one-way ANOVA were below 0.05, except for measurement depth and D10. Since all parameters were correlated to various degrees, it was difficult to provide a clear answer as to whether elevated gas concentrations were related to a single parameter or a combination of multiple parameters. Therefore, a regression model was then used for assessing the strength of the relationship between parameters.

The nature of relationships between the subsurface gas concentrations and the measured parameters was quantified, as shown in Figure 4. The multiple correlation coefficient ($R = 0.71, P < 0.001$) suggests that there was a moderate correlation between the observed subsurface gas concentrations and those predicted by the regression
model. The VIFs were less than 6, demonstrating that the input parameters were independent of each other. The estimated coefficients were not inflated. The standardized \( \beta \) coefficients estimated to compare the strength of the effects of individual parameters to the \( \text{CH}_4 \) concentrations from the regression model indicated that, of the parameters tested, soil moisture content has the strongest effect on the gas concentration, followed by particle size \( D_{60} \), distance from the leak source, uniformity coefficient, particle size \( D_{30} \), measurement depth, total porosity, and, lastly, intrinsic permeability (Table S2).

Regression analysis indicated that higher soil moisture content associates with lower gas concentrations. This suggests that water limited gas migration by displacing air in the pore spaces, hence increasing tortuosity and restricting gas movement. This in effect reduces gas diffusivity, resulting in elevated concentrations of gas close to the leak location. High concentrations close to the leak site, regardless of moisture conditions, are due to the buoyancy-driven upward movement of the constituent gasses in NG, predominantly \( \text{CH}_4 \), which are buoyant in these conditions. These results are consistent with those from previous experiments (Cuezva et al., 2011; Cohen et al., 2013; Xu et al., 2014; Van De Ven and Mumford, 2020), indicating the importance of soil moisture across sites. As expected, gas concentrations decreased exponentially with increasing distance from the leak location (Figure S3).

Gas transport through the soil is affected by pore size, which depends on grain size (Arthur et al., 2012; Benavente and Pla, 2018; Feng et al., 2020). The pore size distribution influences the gas movement and entrapment (Moldrup et al., 2001; Hamamoto et al., 2011). The relationships between soil properties and \( \text{CH}_4 \) concentration were investigated to understand soil gas dynamics and migration. The results showed that there was a negative correlation between particle diameter and gas concentration because variations in range of grain size could have attributed to higher gas concentration (Tables S1 and S2). Soil particles with high uniformity coefficient \( \text{Cu} \), considered as poorly sorted, would expect low permeability that could have attributed to gas buildup in pore space (Table S2). A positive correlation between porosity and \( \text{CH}_4 \) concentration was observed as well. For intrinsic permeability, a weak positive correlation was observed with \( \text{CH}_4 \) concentration. Although the correlation of measurement depth with concentration was not observed from the bivariate analysis, the measurement depth positively correlated with the \( \text{CH}_4 \) concentration based on the regression model. This model does not perform well for lower concentrations (< 104 ppmv), but the subsurface gas concentrations observed from the surveyed sites were considerably greater than 50,000 ppmv. For the purpose of understanding the effects of different parameters, it is suitable.

![Figure 4](http://example.com/image.png)

**Figure 4.** Gas concentrations predicted by a linear regression model. The independent parameters are soil moisture content \( (\theta_w \log \theta_w) \), distance \( (\log x) \), measurement depth \( (h) \), total porosity \( (\log \phi) \), intrinsic permeability \( (\log k) \), uniformity coefficient \( (\text{Cu}) \), logarithm of particle size \( D_{60} \) and \( D_{30} \), and a constant term. The dashed line is the 1:1 line.
**Leak causes**

Pipeline leaks were primarily attributed to material failure and pipeline corrosion at 94.6% of the 74 evaluated sites, 51.4% material failure, and 43.2% corrosion (Table S3). In this study, material failure refers to damage of the pipeline due to defective and loose parts. Corrosion refers to the metal deterioration of the pipeline due to oxidation. This analysis excludes 3 LDAR surveys, where no information regarding pipeline failure was available. Median subsurface CH₄ concentrations measured while pinpointing the actual leak locations for material failure and corrosion leaks were 25,000 ppmv and 50,000 ppmv, respectively (Figure S3A). Median detected migration distance for material failure leaks (average 2.5 ± 3.3 m with median of 1.3 m) was slightly smaller than the one for corrosion (average 2.6 ± 2.9 m with median of 1.8 m). An analysis of the gas concentrations relative to the actual emission rate of the leaks was not possible for this study because the emission rate is typically not measured in practice. Other causes of leaks, such as natural force damage (e.g., settling), incorrect operation, and third party, were not present in a sufficient number of samples to evaluate statistically.

According to PHMSA (2021), excavation damage is the most common cause of distribution pipeline leaks causing significant incidents from 2010 to 2019. “Significant incidents,” as reported to PHMSA (2021), include those with a fatality or injury requiring in-patient hospitalization, US$50,000 or more in total costs measured in 1984 dollars, highly volatile liquid releases of 5 barrels or more or other liquid releases of 50 barrels or more, or liquid releases resulting in an unintentional fire or explosion. However, in this study, excavation damage was not reported as a leak cause, primarily because leak damage caused by excavation damage are likely to cause major pipeline damage requiring immediate emergency response, that is, a Grade 1 leak. Typically, Grades 2 and 3 leaks do not result in a reportable incident. As a result, little consolidated, national, data exist for these leaks. However, tracking these less severe leaks would inform LDAR practices and help identify best practices for mitigating both safety risk and greenhouse gas emissions.

**Conclusions**

The findings of this study are of significant interest to ongoing efforts to prevent incidents associated with underground NG leaks. The statistical analysis of LDAR surveys on Grade 2 and 3 leaks for distribution lines revealed which factors most significantly affect subsurface CH₄ concentrations, as well as the causes most commonly associated with Grade 2 and 3 leaks. Results highlighted pronounced effects of soil moisture content and, to a lesser degree, of soil texture on subsurface CH₄ concentrations. This finding has implications for leak grading, as it is not common practice to include the influence of soil moisture as a factor when considering the leak grade. Importantly, a leak graded during high moisture conditions could have increased lateral spread of gas when the moisture decreases, potentially bringing gas in contact with soils and infrastructure that were distant from the leak at the time of grading. Thus, likely variations in soil moisture should be considered to understand how the hazard potential from a leak could evolve over time. One possible approach would be to record the moisture level when a leak was first graded and then to prioritize follow-up surveys in different moisture conditions to verify any change to the leak behavior. This is especially important in scenarios where the soil moisture may vary due to rainfall and the subsequent infiltration of water into the soil or a change in the seasonal height of the water table.

Despite the wide breadth of data collected for this study—at significant investment by the participating operators—the study did not capture all factors necessary to understand the dynamics of subsurface gas transport and nearby atmospheric transport. Interactions between the subsurface, wind, and weather remains a highly complex and variable topic that is poorly understood and difficult for LDAR personnel to characterize efficiently while in the field. An investment in simple but effective tools, associated training, and uniform reporting—perhaps of a subset of all leaks investigated and repaired—could provide substantial insight into leak dynamics, evolution over time, impact of weather and soil conditions, and other factors. These data would likely lead, over time, to improved LDAR guidance.

**Data accessibility statement**

Data sets for this research are available on the Texas Data Repository: https://doi.org/10.18738/T8/32VPN0 and in the in-text data citation reference: Cho et al. (2022).

**Supplemental files**

The supplemental files for this article can be found as follows:

- Text S1–S3.
- Figures S1–S3.
- Tables S1–S3.
- Data file. xlsx

**Acknowledgments**

Any opinion, findings, and conclusions or recommendations expressed herein are those of the authors and do not necessarily reflect the views of those providing technical input or financial support. The trade names mentioned herein are merely for identification purposes and do not constitute endorsement by any entity involved in this study. The authors would like to thank the industry representatives and site operators for their inputs and support throughout the project.

**Funding**

This material is based upon work supported by the U.S. Department of Transportation (DOT) Pipeline and Hazardous Materials Safety Administration under Grant Nos. 693JK31810013 and 693JK32010011POTA.

**Competing interests**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.
Contributed to conception and design: YC, KS, DZ.
Contributed to acquisition of data: YC, KS, CB, BU.
Contributed to analysis and interpretation of data: YC, KS, BU, NS.
Drafted and/or revised this article: YC, KS, BU, CB, DZ, NS.
Approved the submitted version for publication: YC, KS, BU, CB, DZ, NS.

References


Forde, ON, Mayer, KU, Cahill, AG, Mayer, B, Cherry, J, Parker, BL. 2018. Vadose zone gas migration and surface effluxes after a controlled natural gas release into an unconfined shallow aquifer. Vadose Zone


Morris, DA, Johnson, AL. 1967. Summary of hydrologic and physical properties of rock and soil materials, as analyzed by the hydrologic laboratory of the U.S.


