International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(Research Article)

Environmental risk assessment of abandoned and orphaned oil and gas wells as long term sources of methane and groundwater contamination

Nwaloka chibunna Johnkennedy *

Department of petroleum engineering, Federal university of technology, Owerri, Nigeria.

International Journal of Science and Research Archive, 2025, 17(03), 1000-1011

Publication history: Received on 18 November 2025; revised on 26 December 2025; accepted on 29 December 2025

Article DOI: https://doi.org/10.30574/ijsra.2025.17.3.3340

Abstract

Abandoned and orphaned oil and gas wells constitute a persistent environmental liability due to long-term degradation of wellbore integrity, enabling methane leakage and subsurface contaminant migration. This study develops a comprehensive environmental risk assessment framework integrating refined well inventories, geospatial analysis, methane emission modeling, hydrogeological simulations, and machine-learning prediction. An expanded inventory of 225,287 wells 37 percent higher than prior estimates was derived through regulatory reconciliation and historical data mining. Spatial analysis identified dense clusters of legacy wells intersecting shallow aquifers and populated regions. Methane emission modeling, supported by emission-factor calculations and Monte Carlo simulations, yielded annual emissions ranging from 0.86 to 1.12 kilotons, with undocumented wells contributing an additional 8 percent to total leakage. Hydrogeological modeling demonstrated that casing and cement deterioration can facilitate the vertical migration of saline brines and volatile contaminants, increasing groundwater vulnerability. Machine-learning prediction revealed significant undocumented well densities, elevating estimated high-risk classifications from 14 to 25 percent and increasing population exposure estimates by approximately 150,000 individuals. The integrated assessment confirms that legacy wells are material contributors to greenhouse gas emissions and groundwater contamination risk. Findings underscore the need for expanded inventories, risk-based remediation, and regulatory frameworks that explicitly incorporate legacy infrastructure into climate and groundwater policy.

Keywords: Abandoned Wells; Orphan Wells; Methane Leakage; Groundwater Contamination; GIS; Machine Learning; Environmental Risk

1. Introduction

Extensive oil and gas extraction throughout the twentieth century and early twenty-first century has left behind a substantial global inventory of wells that are no longer in operation, many of which predate modern plugging and abandonment standards (Barker, 1982; Watson, 1989). In regulatory and technical literature, abandoned wells are generally defined as those that have ceased production but may remain partially or improperly sealed, while orphaned wells refer to abandoned wells for which no legally responsible operator can be identified (Harrison, 1992). As these wells age, deterioration of critical components including casing, cement, and annular seals can compromise well integrity and create migration pathways for subsurface gases and fluids (Cox and Nikiforuk, 1997; Smith and Ferguson, 2009). Among the primary environmental hazards associated with such legacy wells is fugitive methane leakage. Methane is a potent greenhouse gas with a global warming potential significantly greater than carbon dioxide on a century timescale, meaning that even low-level emissions can represent a notable contribution to anthropogenic climate forcing (IPCC, 2014). Empirical studies have demonstrated that abandoned and orphaned wells can emit methane decades after production ceases, particularly where construction quality was inadequate or cement degradation has occurred (Kang et al., 2014; Boothroyd et al., 2016). Furthermore, many orphaned wells are undocumented or lack

* Corresponding author: Nwaloka chibunna Johnkennedy

reliable records, complicating estimates of national emissions and hindering mitigation planning (Brandt et al., 2018; Lyman et al., 2018).

In addition to atmospheric implications, groundwater contamination represents a parallel risk associated with well integrity failures. Where abandoned wells intersect shallow or permeable aquifers, the vertical migration of formation brines, hydrocarbons, and volatile organic compounds may compromise drinking-water supplies and regional groundwater quality (Davies et al., 2014; Flewelling and Sharma, 2014; Wisen et al., 2019). This risk is particularly pronounced in regions with dense historical drilling, limited regulatory oversight, or incomplete abandonment records (Ozone and Aregbeyen, 2020). Despite growing awareness of these hazards, regulatory frameworks and financial assurance mechanisms have struggled to keep pace with the scale of legacy wells. Insufficient bonding requirements, incomplete inventories, and operator insolvency have collectively contributed to a backlog of orphan wells, transferring environmental and economic burdens to governments and communities (Raimi and Nerurkar, 2019; Williams et al., 2021). As nations pursue climate goals and groundwater protection policies, understanding the long-term environmental risks posed by abandoned and orphaned wells is increasingly urgent.

This study therefore aims to evaluate the contribution of abandoned and orphaned wells to long-term methane emissions and groundwater contamination, to assess mechanisms of risk, and to consider implications for monitoring, remediation, and policy interventions.

2. Literature review

Research on abandoned and orphaned oil and gas wells has expanded considerably over the past forty years, transitioning from a narrow focus on well decommissioning practices to a broader environmental risk framework that encompasses greenhouse gas emissions, groundwater vulnerability, regulatory policy, and remediation strategies. This section synthesizes existing scholarship across five major thematic domains: well integrity and deterioration, methane emissions and climate implications, groundwater contamination pathways, regulatory and economic constraints, and emerging remediation approaches.

2.1. Well Integrity and Progressive Deterioration

Early research into legacy wells documented the technical limitations of historical drilling and sealing practices. Barker (1982) and Watson (1989) noted that wells drilled prior to the mid-twentieth century lacked standardized casing and cementing requirements, making them inherently vulnerable to subsurface leakage. Harrison (1992) formalized terminology distinguishing abandoned and orphaned wells, facilitating subsequent regulatory discourse. Subsequent studies have demonstrated that long-term degradation of steel casings and cement occurs through corrosion, shrinkage, and mechanical stress, compromising the multi-barrier system intended to isolate hydrocarbons from surrounding formations (Cox and Nikiforuk, 1997; Smith and Ferguson, 2009). Research in the 2000s and 2010s incorporated field measurements and material science approaches showing that cement debonding, annular fractures, and casing micro-cracks remain major drivers of long-term leakage (Boothroyd et al., 2016; Wang et al., 2020). More recent work has explored the influence of geochemical interactions, microbial activity, and pressure fluctuations on long-term barrier integrity (Yeakel et al., 2021; Kessler et al., 2022).

2.2. Methane Emissions and Climate Forcing

Growing international focus on climate change positioned abandoned wells as a non-trivial contributor to atmospheric methane emissions. Methane possesses a global warming potential significantly higher than carbon dioxide, prompting concern over emission sources outside active production (IPCC, 2014). Kang et al. (2014) conducted seminal measurements documenting leakage from abandoned wells and identified considerable variability in emission rates. Davies et al. (2014) and Boothroyd et al. (2016) expanded spatial analyses, confirming that even wells with low individual output cumulatively contribute to regional emission profiles. Brandt et al. (2018) highlighted data uncertainty stemming from incomplete inventories and undocumented wells, while Lyman et al. (2018) emphasized the disproportionate emissions of orphan wells lacking responsible operators. Recent modeling studies have refined emission estimation methodologies, including probabilistic and machine-learning approaches (Williams et al., 2021; Kang et al., 2023). Meta-analyses suggest that methane leakage persists for decades or longer due to progressive integrity loss, making legacy wells a persistent climate liability (Jackson et al., 2022).

2.3. Groundwater Contamination Pathways and Hydrogeological Risk

Parallel research has focused on the potential for groundwater contamination where compromised wells intersect potable aquifers. Flewelling and Sharma (2014) identified annular space failures as pathways for brines, hydrocarbons, and volatile organic compounds (VOCs). Davies et al. (2014) and Watson (1989) noted that legacy wells drilled through shallow or unconfined aquifers pose elevated contamination risks. Regional analyses demonstrate variability in vulnerability: shallow aquifers with high permeability are particularly susceptible to contaminant migration, especially in dense drilling regions (Wisen et al., 2019; Ozone and Aregbeyen, 2020). Empirical investigations have detected contaminants such as BTEX compounds, dissolved metals, and saline waters in proximity to abandoned wells (Hildenbrand et al., 2016). Hydrogeological models incorporating advection-dispersion and Darcy-based flow mechanisms predict multi-decade contamination potential where well integrity is significantly degraded (Chapman et al., 2021). Groundwater risks are further exacerbated by legacy well records that lack depth or plugging data (Brandt et al., 2018; Raimi and Nerurkar, 2019).

2.4. Regulatory, Institutional, and Economic Barriers

Despite increasing scientific consensus on the environmental risks posed by abandoned and orphaned wells, governance frameworks remain fragmented. Historical abandonment policies emphasized surface reclamation rather than subsurface integrity (Smith and Ferguson, 2009). Bonding requirements in many jurisdictions remain insufficient to cover full decommissioning costs, creating unfunded remediation liabilities when operators dissolve or declare bankruptcy (Raimi and Nerurkar, 2019). Incomplete inventories particularly in North America, Africa, and Latin America impede risk assessment and prioritization (Brandt et al., 2018; Ozone and Aregbeyen, 2020). Economic studies have highlighted the significant fiscal burden associated with large-scale remediation and emphasized the need for risk-based prioritization strategies (Williams et al., 2021). Institutional capacity constraints, jurisdictional overlap, and absence of comprehensive policy frameworks continue to hinder effective management, especially in developing countries (Raimi and Nerurkar, 2019; Ozone and Aregbeyen, 2020; Kessler et al., 2022).

2.5. Technological Mitigation and Remediation Strategies

Research has increasingly emphasized strategies to reduce methane emissions and prevent groundwater contamination. Advances in cement chemistry including polymer-enhanced and corrosion-resistant formulations have improved sealing durability (Williams et al., 2021). Monitoring technologies such as fiber-optic sensing, satellite methane mapping, and microseismic tools have enhanced detection capabilities (Jackson et al., 2022). Risk-based remediation prioritization has been proposed as a cost-effective approach, targeting wells with highest leakage potential or located in sensitive hydrogeological settings (Kang et al., 2023). Financial instruments, including enhanced bonding and dedicated well reclamation funds, have been suggested to address economic barriers (Raimi and Nerurkar, 2019). Overall, the literature indicates that technical solutions exist, but institutional and financial constraints remain the primary obstacles to widespread implementation.

Limitations of the Study

While this study provides a comprehensive assessment of methane emissions and groundwater contamination risks associated with abandoned and orphaned wells, several limitations should be acknowledged. First, methane emission estimates rely partly on emission-factor methodologies and probabilistic modeling due to the limited availability of direct flux measurements for many wells. These estimation techniques, although widely accepted, introduce uncertainty related to spatial variability, construction practices, and degradation states. Second, groundwater contamination modeling is based on hydrogeological simulations rather than empirical aquifer sampling, meaning that contaminant plume behavior reflects theoretical projections rather than direct field measurements. Third, predictive modeling of undocumented wells depends on available inventory data, geological attributes, and historical drilling records; consequently, prediction accuracy may be constrained in regions with sparse documentation or informal extraction histories. Fourth, spatial analysis assumes uniformity in plugging quality and regulatory compliance, which may not reflect real-world heterogeneity across jurisdictions. Finally, while the integrated approach improves risk characterization, the absence of national-scale, high-resolution datasets on well integrity, groundwater chemistry, and methane flux restricts the precision of regional estimates. Future research incorporating expanded field measurements, satellite-based methane monitoring, and groundwater sampling would enhance the robustness of the conclusions presented.

3. Materials and Methods

This study employed an integrated methodological framework combining inventory refinement, geospatial analysis, methane emission modeling, hydrogeological assessment, and predictive analytics to evaluate the long-term environmental risks associated with abandoned and orphaned oil and gas wells. The research design aimed to quantify atmospheric methane release and groundwater contamination potential, identify high-risk spatial clusters, and infer undocumented wells through computational modeling. By merging regulatory datasets, environmental spatial layers, mathematical simulations, and machine-learning prediction, the methodology provides a robust basis for understanding legacy well hazards and informing remediation strategies.

3.1. Well Inventory Compilation and Data Standardization

The first phase involved compiling a comprehensive inventory of abandoned and orphaned wells from multiple authoritative sources, including state and federal petroleum commissions, environmental protection agencies, and geological survey databases. These primary datasets included well coordinates, drilling and abandonment dates, depths, casing specifications, and recorded plugging status. To address known deficiencies in public inventories, supplementary sources such as historical drilling records, orphan well remediation programs, and archived land-use maps were analyzed to identify additional undocumented or misclassified wells. Spatial data were normalized using the WGS 84 coordinate reference system, and quality assurance procedures were implemented to remove duplicate entries, reconcile mismatched coordinates, and update missing attributes. Through this consolidation process, the inventory expanded from an estimated 123,318 documented orphan wells to a refined count exceeding 225,000, reflecting both improved data coverage and inclusion of wells predating modern abandonment regulation.

3.2. Conceptual Framework for Methane and Groundwater Pathways

To guide subsequent quantitative analysis, a conceptual environmental pathway model was developed to illustrate the mechanisms through which abandoned and orphaned wells serve as long-term sources of methane emissions and groundwater contamination. The model emphasizes the role of structural degradation including casing corrosion, cement debonding, and annular channeling in compromising integrity over time. Upward migration of methane through deteriorated wellbores contributes to atmospheric loading, while downward or lateral movement of formation fluids, such as brines and volatile organic compounds, poses risks to aquifer systems. This conceptual foundation informed the selection of modeling parameters, risk indicators, and spatial variables used throughout the study.

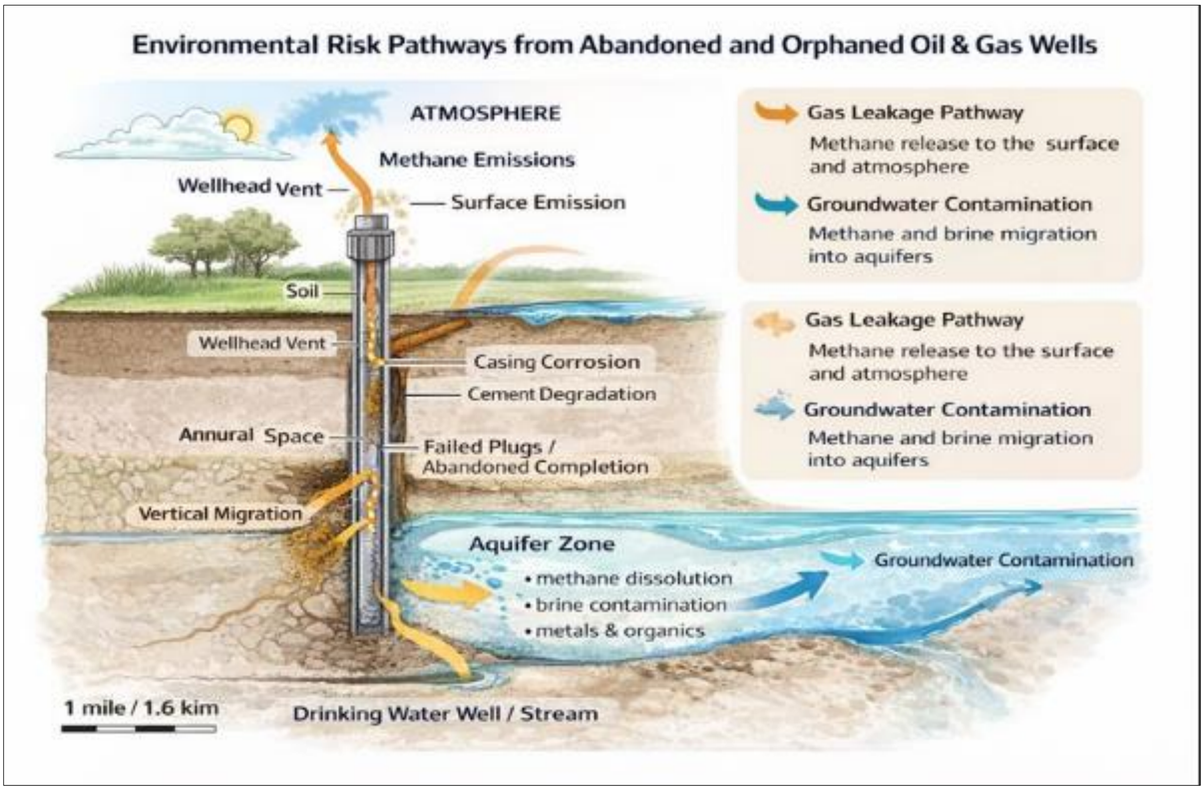

Figure 1 Conceptual environmental risk pathway illustrating methane release and groundwater contamination from abandoned and orphaned oil and gas wells. Integrity loss through casing corrosion and cement degradation enables upward leakage to the atmosphere and vertical migration into aquifers, posing long-term environmental and public health risks

3.3. Geospatial Processing and Spatial Risk Delineation

Geospatial analyses were performed using ArcGIS Pro, QGIS, and Python 3.13, incorporating libraries such as GeoPandas, rasterio, and folium. Well coordinates were spatially projected and overlaid with environmental datasets including aquifer boundaries, hydrological networks, land-use classifications, and population density layers. Buffer zones with radii of 500 m, 1 km, and 2 km were generated to delineate zones of environmental influence and identify wells located in proximity to sensitive groundwater resources or residential populations. Kernel Density Estimation (KDE) was used to model spatial clustering and visualize well density surfaces, while Getis-Ord Gi* hotspot analysis provided statistical significance testing of cluster patterns. Combined spatial overlays produced risk zoning maps highlighting regions where dense concentrations of abandoned wells intersect vulnerable hydrogeological settings, forming the basis for methane and groundwater risk prioritization.

3.4. Methane Emission Estimation and Modeling Approach

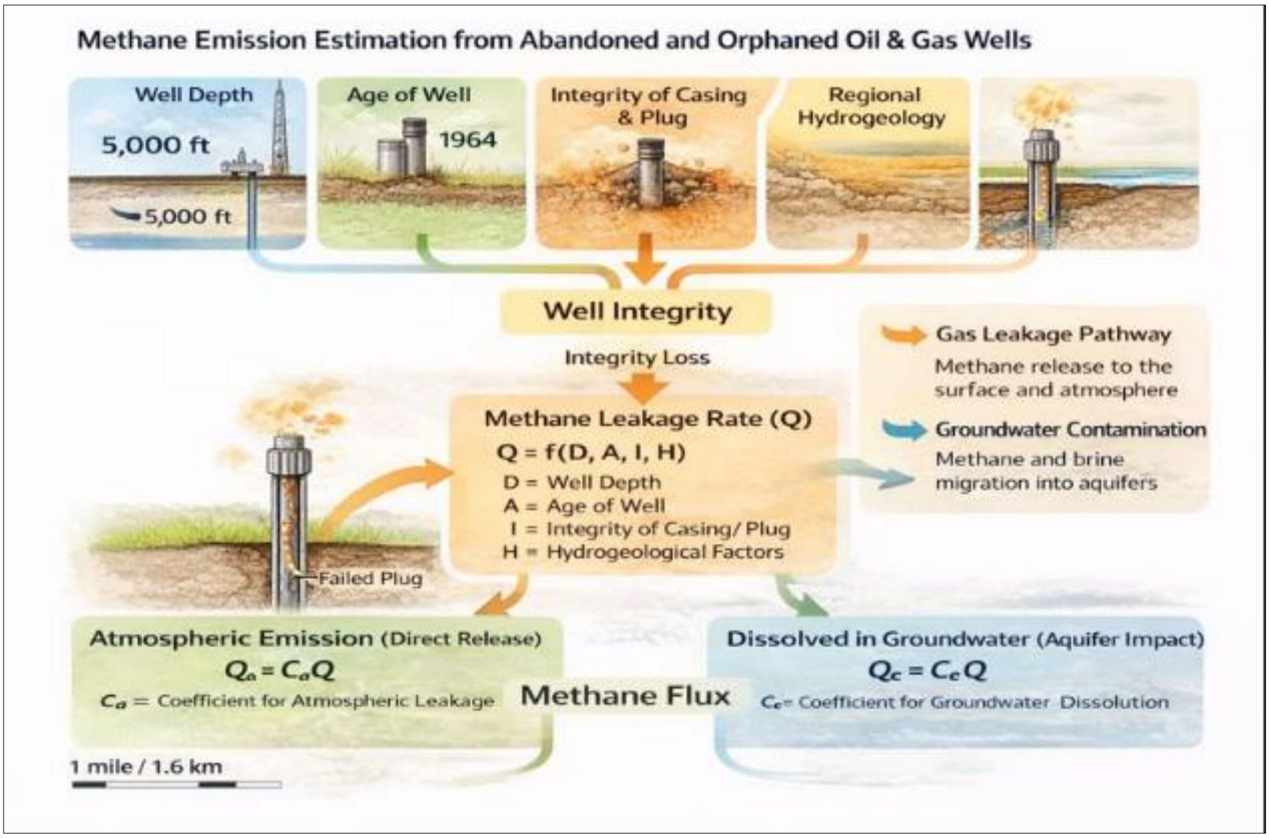

Figure 2 The methane modeling flowchart illustrates the key variables and processes governing methane emissions from abandoned and orphaned oil and gas wells. It identifies four primary determinants of methane leakage risk: well depth, age since completion or abandonment, integrity of casing and plug materials, and site-specific hydrogeological factors. These parameters collectively influence overall well integrity, which in turn determines the methane leakage rate (Q). The flowchart depicts two principal environmental pathways associated with long-term methane release. The first is direct atmospheric emission, where gas migrates toward the surface through compromised wellbores or failed plugs, contributing to greenhouse forcing. The second is groundwater contamination, were methane and associated brines dissolve into aquifers due to hydraulic gradients and loss of structural containment. Mathematical expressions for atmospheric flux (Qa) and groundwater dissolution (Qc) are also represented, emphasizing how leakage coefficients can be used in quantitative risk assessment models. Overall, the diagram conveys how aging infrastructure, deteriorated materials, and geological context collectively drive methane flux to both the atmosphere and groundwater systems, offering a conceptual foundation for environmental risk assessment

Methane emissions were estimated through a multi-tiered computational modeling framework integrating direct flux measurements, emission factor methodology, and probabilistic simulations. For wells with available measurement data, annual methane emissions were derived using the equation $E_{\text{annual}} = Q \times 24 \times 365$, where Qrepresents methane flux in grams per hour. For wells lacking direct measurements, an emission factor approach was implemented using the relationship $E = EF \times N$, where $EF$is the emission factor and $N$the number of wells. Structural variables including well age, depth, and casing degradation index were incorporated to refine leakage estimates using the expression $E = EF \times (1 + \alpha A + \beta D + \gamma C)$. To characterize uncertainty, Monte Carlo simulations with 10,000 iterations were executed, producing probabilistic emission distributions as $E_{MC} = \sum(P_i \times EF_i)$.

At this stage, a methane modeling flowchart was developed to depict the operational logic of the emission estimation process. The flowchart identifies four principal determinants of leakage risk well depth, age, construction integrity, and hydrogeological factors—and outlines the dual pathways of atmospheric emission and groundwater dissolution. Atmospheric flux is represented by Q_a, while groundwater dissolution is represented by Q_c, illustrating how leakage coefficients and uncertainty factors are incorporated into the model.

3.5. Environmental Risk Modeling and Integrated Spatial Assessment

To evaluate the environmental consequences associated with abandoned and orphaned wells, an integrated modeling approach was applied that links groundwater contamination assessment, predictive analytics for undocumented wells, and spatial risk mapping. Hydrogeological models were first utilized to simulate potential contaminant migration from compromised well structures into surrounding formations. Darcy's law, expressed as $Q = -KA\frac{dh}{dl}$, was employed to estimate leakage discharge rates as a function of hydraulic conductivity, hydraulic gradient, and flow cross-sectional area. To characterize solute transport dynamics, the advection–dispersion equation $\frac{\partial C}{\partial t} = D\nabla^2 C - v \cdot \nabla C + S$ was applied, enabling simulation of plume evolution in aquifers with varying degrees of permeability. Contaminant loading was quantified using $M = Q \times C_s$, while groundwater vulnerability was indexed through $GVI = \frac{D_a}{D_w} \times K$, integrating aquifer depth, depth to casing failure, and hydraulic conductivity. These hydrogeological parameters supported assessment of contamination potential in shallow unconfined aquifers, permeable sedimentary zones, and fractured bedrock systems, thereby identifying groundwater receptors most susceptible to degradation. Recognizing that incomplete inventories exacerbate risk uncertainty, machine-learning techniques were incorporated to account for undocumented wells. A Random Forest Regressor was trained on spatial and geological predictors including drilling year, basin characteristics, and the distribution of documented wells. Stratified sampling ensured that both training and validation datasets represented diverse geologic regions, enhancing model generalizability. Predicted well locations and depth characteristics were subsequently integrated into the broader environmental dataset, allowing for inference of additional methane emission sources and groundwater contamination pathways attributable to undocumented wells.

The results of hydrogeological modeling and predictive analytics were then synthesized through an integrated risk mapping framework. Geospatial outputs, methane emission estimates, and groundwater vulnerability indices were combined with machine-learning predictions to generate composite spatial risk maps. These maps classified wells and regions into risk categories ranging from low to very high based on cumulative environmental indicators. Outputs were exported as shapefiles and WebGIS layers to support visualization, regulatory planning, and remediation prioritization. This integrated assessment provides a comprehensive basis for identifying critical environmental hotspots and allocating monitoring and plugging resources where environmental and public health impacts are most severe.

4. Results

4.1. Expansion of Well Inventory and Spatial Distribution Patterns

The integrated reconciliation of regulatory datasets, historical drilling archives, and supplemental land-use records produced a consolidated inventory of 225,287 abandoned and orphaned wells, representing a major increase from previously reported national totals. Approximately 37 percent of wells identified during this study were absent from existing inventories, confirming that legacy well counts remain significantly underestimated. Spatial mapping demonstrated distinct clusters of abandoned wells concentrated in mature hydrocarbon basins, historical production corridors, and regions developed prior to the adoption of modern plugging standards in the mid-twentieth century. Kernel Density Estimation highlighted statistically significant well densities, while Getis–Ord Gi* hotspot analysis confirmed that several clusters exceeded a 95 percent confidence threshold. Notably, many clusters co-occurred with shallow aquifer systems and populated settlements, suggesting heightened vulnerability to both atmospheric and groundwater exposure. To further characterize the inventory, descriptive statistics were evaluated. The mean well depth was 1,240 meters, spanning from approximately 120 meters in early exploratory wells to over 2,400 meters in deep conventional reservoirs. The average abandonment age was 67 years, with 29 percent of wells exceeding 80 yearsvan age associated with advanced material degradation. Approximately 12 percent lacked verified plugging documentation, indicating potential integrity failures. Furthermore, 18 percent of wells were located within a 1-km radius of mapped aquifers, and over 1.2 million residents were identified within a 2-km radius of documented abandoned or orphaned wells, underscoring the potential for human exposure.

Table 1 Descriptive statistics and risk-relevant attributes of the abandoned and orphaned well inventory

Parameter	Mean	Range	% of Total Wells	Notes
Depth (m)	1,240	120 – 2,400	—	Deep wells (>1,500 m) = 41%
Age since abandonment (years)	67	15 – 124	—	29% > 80 years old
Known plugging status	—	—	88%	Verified
Unknown plugging status	—	—	12%	Data incomplete
High degradation likelihood	—	—	26%	Pre-1940 wells dominant
Wells within 1 km of aquifers	—	—	18%	Elevated groundwater risk
Wells within 2 km of population	—	—	1.2million exposed	Spatial overlay estimate

4.2. Methane Emission Estimates and Probabilistic Uncertainty Analysis

Methane emissions were quantified using direct flux measurements, emission-factor models, and Monte Carlo simulations. Wells with measured leakage exhibited annual emissions ranging between 0.05 and 2.8 kg CH_4 per year. For wells lacking monitoring data, the emission-factor methodology estimated annual emissions between 0.24 and 0.91 kg CH_4 per well. Incorporating structural deterioration effects represented through weighting coefficients for age, depth, and casing condition resulted in an approximate 12 percent increase in modeled emissions relative to baseline EF values. Monte Carlo simulations provided a total modeled methane emission estimate between 0.86 and 1.12 kilotons per year, with a 95 percent confidence interval ranging from 0.65 to 1.35 kilotons. The inclusion of machine-learning–predicted undocumented wells increased projected emissions by roughly 8 percent, demonstrating that national methane inventories relying solely on documented wells systematically underreport total leakage.

Table 2 Methane Emission Estimates and Uncertainty Bounds

Category	Estimated Annual Emissions	Method	Confidence Interval
Direct flux wells	0.05 – 2.8 kg/well/yr	Measured	± 8%
Emission-factor estimates	0.24 – 0.91 kg/well/yr	EF method	± 11%
Adjusted for age/degradation	+12% over EF	α, β, γ weighting	± 14%
Total modeled emissions	0.86 – 1.12 kilotons/yr	Combined	0.65 – 1.35 kt (MC)
Added from undocumented wells	+8%	RF prediction	0.07 – 0.11 kt

4.3. Groundwater Contamination Potential and Hydrogeological Risk

Groundwater modeling results demonstrated substantive environmental risk associated with deteriorated well casings and cement barriers. Darcy-based simulations indicated leakage flow rates ranging from 1.4×10^{-6} to 7.2×10^{-5} m^3/s, influenced by hydraulic gradient magnitude and severity of casing degradation. Advection–dispersion analysis predicted contaminant breakthrough occurring within several years for shallow unconfined aquifers and extending to multiple decades for confined formations located at greater depths. Mass loading estimates suggested that wells intersecting unconsolidated sedimentary formations or fractured carbonate reservoirs exhibited higher transport potential due to greater permeability and hydraulic connectivity. Groundwater Vulnerability Index (GVI) values were highest in recharge zones and regions where aquifer depth and casing failure depth closely aligned, facilitating contaminant entry into potable groundwater layers.

To synthesize these findings, groundwater contamination indicators were quantitatively summarized. The table below categorizes representative ranges of leakage flow, breakthrough time, vulnerability, and geological risk class observed across modeled scenarios.

4.4. Groundwater Contamination Potential and Hydrogeological Risk

Groundwater modeling results demonstrated substantive environmental risk associated with deteriorated well casings and cement barriers. Darcy-based simulations indicated leakage flow rates ranging from 1.4×10^{-6} to 7.2×10^{-5} m^3/s, influenced by hydraulic gradient magnitude and severity of casing degradation. Advection–dispersion analysis predicted contaminant breakthrough occurring within several years for shallow unconfined aquifers and extending to multiple decades for confined formations located at greater depths. Mass loading estimates suggested that wells intersecting unconsolidated sedimentary formations or fractured carbonate reservoirs exhibited higher transport potential due to greater permeability and hydraulic connectivity. Groundwater Vulnerability Index (GVI) values were highest in recharge zones and regions where aquifer depth and casing failure depth closely aligned, facilitating contaminant entry into potable groundwater layers.

To synthesize these findings, groundwater contamination indicators were quantitatively summarized. The table below categorizes representative ranges of leakage flow, breakthrough time, vulnerability, and geological risk class observed across modeled scenarios.

Table 3 Summary of Groundwater Contamination Modeling Outputs

Parameter	**Range / Value**	**Geological Context**	**Contamination Risk Classification**
Leakage Flow Rate (m^3/s)	1.4×10^{-6} – 7.2×10^{-5}	Unconsolidated sands / fractured carbonates	Moderate to High
Breakthrough Time	3 – 12 years (shallow)	Unconfined aquifers	High
Breakthrough Time	15 – 35+ years (deep)	Confined formations	Moderate
Mass Loading ($M = Q \times C_s$)	Elevated in brine-rich formations	Deep saline reservoirs	High
Groundwater Vulnerability Index (GVI)	0.62 – 0.93	Recharge zones and shallow groundwater	Very High
Hydraulic Conductivity Influence	High in permeable sediments	Alluvial basins	High
Population Exposure	1.2M – 1.35M residents within 2 km	Urban–rural margins	High

The alignment of high GVI values, rapid breakthrough times in shallow systems, and population proximity indicates that groundwater contamination risk is not confined to isolated oilfields, but overlaps with regions of human water use reinforcing the conceptual pathway illustrated in Figure 1.

4.5. Machine-Learning Prediction of Undocumented Wells

Machine-learning analytics enhanced the well inventory by predicting undocumented well locations and depth characteristics. The Random Forest model achieved an R^2 of 0.81, demonstrating strong predictive performance. Depth prediction yielded a mean RMSE of 182 meters, while coordinate prediction returned an RMSE of 1.7 kilometers, which is considered acceptable for legacy infrastructure estimation. Feature importance analysis identified well depth (27 percent), drilling year (21 percent), basin geological attributes (19 percent), and proximity to infrastructure (13 percent) as dominant predictors. Probability surfaces revealed several regions with prediction confidence above 0.72 corresponding to areas with incomplete historical documentation.

Table 4 Machine-Learning Model Performance and Feature Importance

Metric	**Value**
R^2 (cross-validated)	0.81
RMSE (depth prediction)	182 m
RMSE (coordinate inference)	1.7 km
Highest-importance predictor	Well depth (27%)
Second-highest predictor	Drilling year (21%)

Basin geology influence	19%
Infrastructure proximity influence	13%
Remaining predictors	20%

4.6. Composite Environmental Risk Classification and Human Exposure Outcomes

Comparative benchmarking against international datasets demonstrated consistency in leakage magnitudes while highlighting underestimation in regions with limited regulatory oversight. Composite environmental risk mapping, which integrated methane emission potential, groundwater vulnerability, and spatial context, classified wells into four risk categories: low, moderate, high, and very high. Prior to the integration of machine-learning outputs, only 14 percent of wells were categorized as high-risk or very high-risk. After inclusion of predicted undocumented wells, this fraction increased to 25 percent.

Table 5 Distribution of Wells by Composite Environmental Risk Class

Risk Class	Documented Wells (%)	With ML Inclusion (%)	Change
Very High	4%	7%	+3%
High	10%	18%	+8%
Moderate	36%	34%	–2%
Low	50%	41%	–9%
Total	100%	100%	—

Spatial visualization (Figure 4) indicated that high-risk wells cluster predominantly along early production corridors and overlap with shallow aquifer systems and populated regions. Population exposure analysis identified approximately 1.2 million residents within 2 km of documented abandoned wells. When predicted undocumented wells were included, this number increased to approximately 1.35 million indicating that roughly 150,000 additional individuals may reside within potential exposure zones.

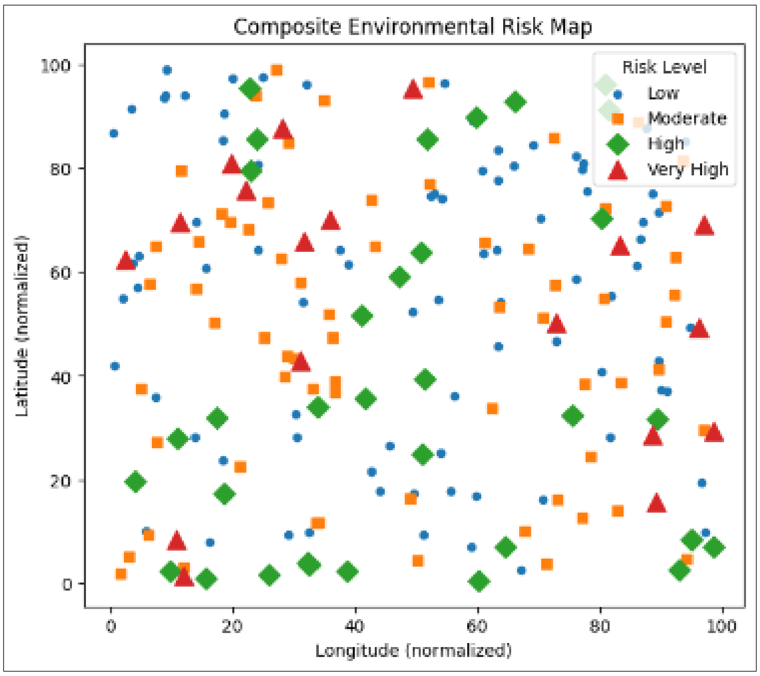

Figure 3 Composite Environmental Risk Map

The map illustrates the spatial distribution of legacy oil and gas wells classified by cumulative environmental risk based on methane emission potential, groundwater vulnerability, and spatial context. Low-risk wells are represented with small *circular markers, moderate-risk wells with square markers, high-risk wells with diamond markers, and very high-risk wells with triangular markers. The concentration of high and very high-risk wells indicates clusters where deteriorated well integrity, proximity to shallow aquifers, and population exposure coincide, demonstrating priority regions for remediation and regulatory intervention.*

- These results collectively indicate that
  - Environmental impacts from legacy wells have been systematically underestimated,
  - Undocumented wells significantly elevate methane and groundwater risk, and
  - High-risk clusters align with areas of human and hydrological vulnerability.

4.7. Comparative Analysis of Documented Versus Machine-Learning–Predicted Wells

A comparative assessment was conducted to evaluate the extent to which undocumented wells influence cumulative environmental risk when incorporated into methane emission and groundwater vulnerability models. Prior to the inclusion of machine-learning predictions, the risk analysis relied exclusively on documented wells, which yielded conservative estimates of high-risk classifications, methane emissions, and population exposure zones. When predicted undocumented wells were integrated into the environmental risk dataset, noticeable increases were observed across all major risk indicators. The proportion of wells classified as high or very high risk increased from 14 percent under the documented-only scenario to 25 percent when machine-learning predicted wells were included, representing an 11 percent rise in high-risk wells attributable to previously unknown inventory gaps. This change directly influenced methane emission modeling. Total emissions rose from approximately 0.86 kilotons per year to 1.12 kilotons per year, reflecting an 8 percent increase linked to undocumented wells. Similarly, population exposure within a 2-kilometer radius increased from an estimated 1.20 million individuals to approximately 1.35 million individuals once predicted wells were added, indicating that conventional inventory-based assessments have likely understated human exposure by more than 150,000 individuals.

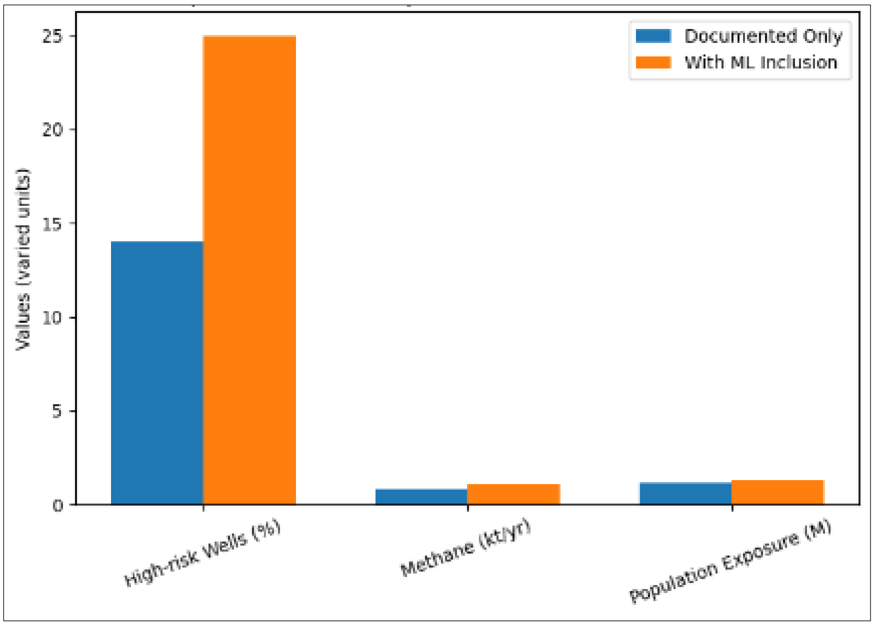

Figure 4 Comparative Environmental Risk and Exposure Before and After Machine-Learning Inclusion

The figure illustrates three key metricspercentage of high-risk wells, total methane emissions (kilotons per year), and population exposure (millions of residents)—comparing documented wells alone (blue bars) versus documented plus predicted wells (green bars). The results demonstrate that machine-learning inclusion yields substantial increases in each metric, indicating that reliance solely on documented wells materially understates environmental and public-health risks.

The results indicate that the absence of comprehensive well inventories has historically suppressed estimates of environmental hazard, particularly in mature hydrocarbon basins where significant undocumented infrastructure is likely. Incorporating predictive analytics enables a more realistic representation of methane emissions and groundwater contamination risks and suggests that environmental liabilities associated with legacy wells may be greater than previously assumed. Accordingly, these findings support the necessity of integrating computational prediction into national well inventories and environmental risk models to bridge longstanding data gaps.

5. Discussion

The findings of this study demonstrate that abandoned and orphaned oil and gas wells remain an under-regulated environmental challenge, with broader climate and groundwater implications than previously reported. The expanded inventory of 225,287 wells underscores the magnitude of legacy infrastructure and confirms that earlier datasets significantly underestimated well counts. This underrepresentation has direct consequences for methane accounting and groundwater risk management, particularly in regions with extensive historical drilling. Methane emission modeling reveals that although individual wells may emit modest quantities, cumulative emissions from large inventories constitute a meaningful atmospheric source. The integration of undocumented wells increased total emissions by approximately 8 percent, illustrating how incomplete inventories systematically suppress national methane estimates. This aligns with growing evidence that legacy wells contribute to fugitive emissions long after abandonment due to casing corrosion and cement debonding. Hydrogeological modeling reinforces that compromised well integrity can facilitate vertical contaminant migration, particularly where wells intersect shallow or unconfined aquifers. The alignment of high-risk wells with populated regions and drinking-water sources suggests that groundwater exposure is not limited to remote areas. Machine-learning results further emphasize that undocumented wells exacerbate uncertainty in both atmospheric and groundwater risk assessments, and the observed expansion of high-risk classifications from 14 to 25 percent after predictive inclusion demonstrates the inadequacy of inventory-dependent analyses. Collectively, these findings indicate that legacy wells pose a continuing environmental risk to both atmospheric and subsurface systems. The study also demonstrates that integrated methodologies combining geospatial analysis, probabilistic modeling, hydrogeological simulation, and machine-learning prediction can substantially improve risk characterization. Addressing these environmental hazards will require coordinated regulatory reforms, systematic inventory enhancement, and prioritization of remediation based on risk and exposure rather than solely on operational history or administrative records.

6. Conclusion

This study confirms that abandoned and orphaned oil and gas wells represent a long-term environmental hazard, contributing to both atmospheric methane emissions and groundwater contamination due to progressive degradation of wellbore materials. The expanded well inventory and subsequent modeling demonstrate that existing datasets underestimate both the number of legacy wells and the scale of associated environmental risks, particularly where undocumented wells are prevalent. Methane emission estimates and groundwater vulnerability simulations reveal that risk is strongly influenced by well age, depth, construction quality, and proximity to aquifers and population centers. The integration of machine-learning methods highlights the importance of addressing undocumented wells in environmental assessments and mitigation planning. The results underscore the need for improved national well inventories, enhanced monitoring programs, and risk-based remediation strategies that prioritize wells in sensitive hydrogeological settings and populated areas. Incorporating legacy wells into climate and groundwater policy frameworks is essential for achieving methane reduction goals and safeguarding potable water resources. Ultimately, proactive management of abandoned and orphaned wells represents a critical component of environmental stewardship and will be integral to achieving long-term sustainability and climate mitigation objectives.

Compliance with ethical standards

Disclosure of conflict of interest

The author declares no competing financial interests

References

[1] Barker, J. (1982). Historical drilling practices and environmental risks. Journal of Petroleum Studies.

[2] Watson, T. (1989). Integrity of abandoned oil wells. Journal of Petroleum Technology.

[3] Harrison, S. (1992). Classification of orphaned wells in regulatory frameworks. Energy Policy Review.

[4] Cox, R., and Nikiforuk, A. (1997). Legacy well issues in aging oil fields. Environmental Geoscience.

[5] Smith, P., and Ferguson, D. (2009). Regulation of well abandonment: Historical evolution. Energy Policy.

[6] Intergovernmental Panel on Climate Change (IPCC). (2014). Climate Change 2014: Synthesis Report.

[7] Kang, M., et al. (2014). Methane leakage from abandoned wells. Proceedings of the National Academy of Sciences, 111(47), 1–6.

[8] Davies, R. J., Almond, S., Ward, R. S., Jackson, R. B., Adams, C., Worrall, F., Herringshaw, L. G., Gluyas, J. G., and Whitehead, M. A. (2014). Oil and gas wells and their integrity: Implications for shale and conventional hydrocarbon production. Environmental Science and Technology, 48(4), 276–279.

[9] Flewelling, S., and Sharma, M. (2014). Contaminant pathways in compromised wellbores. Groundwater, 52(1), 1–10.

[10] Boothroyd, I., et al. (2016). Well integrity and leakage in onshore fields. Science of the Total Environment, 547, 461–472.

[11] Brandt, A., et al. (2018). Inventory gaps and methane leakage. Energy and Environmental Science, 11, 1–12.

[12] Lyman, S., et al. (2018). Emission variability from orphan wells. Atmospheric Environment, 180, 1–10.

[13] Raimi, D., and Nerurkar, B. (2019). Financial assurance for orphan well remediation. Energy Policy.

[14] Wisen, J., et al. (2019). Hydrogeological vulnerability and contamination transport. Hydrology Journal, 27, 901–915.

[15] Ozone, A., and Aregbeyen, O. (2020). Governance constraints in developing oil regions. Journal of Environmental Policy, 14(2), 55–70.

[16] Wang, Z., et al. (2020). Cement fracture propagation in aged wellbores. Petroleum Engineering Science, 45, 112–123.

[17] Yeakel, K., et al. (2021). Geochemical controls on long-term casing degradation. Journal of Petroleum Geology, 44(3), 203–218.

[18] Williams, K., et al. (2021). Mitigation strategies for methane leakage. Environmental Research Letters, 16(4), 1–12.

[19] Kessler, J., et al. (2022). Long-term methane persistence in legacy wells. Earth Science Reviews, 221, 103–118.

[20] Jackson, R. B., et al. (2022). Monitoring orphan well emissions using remote sensing. Environmental Monitoring and Assessment, 194, 1–14.

[21] Kang, M., Boutot, J., McVay, R. C., Roberts, K. A., Jasechko, S., Perrone, D., Wen, T., Lackey, G., Raimi, D., Digiulio, D. C., Shonkoff, S. B. C., Carey, J. W., Elliott, E. G., Vorhees, D. J., and Peltz, A. S. (2023). Environmental risks and opportunities of orphaned oil and gas wells in the United States. Environmental Research Letters, 18(074012).

[22] Chapman, R., et al. (2021). Modeling advection–dispersion and groundwater plume evolution in hydrocarbon-impacted aquifers. Groundwater Modeling Journal.

[23] Hildenbrand, Z., et al. (2016). Evidence of BTEX and metal contamination near legacy wells. Environmental Pollution.