CAAP Quarterly Report

03/31/2025

Project Name*: A Framework and Integrated Solution of a Dynamic Pipeline Hazard and Risk Data Repository for All Pipelines*

Contract Number*:693JK32450004CAAP*

Prime University*: University of Dayton*

Prepared By*: Hui Wang, Ph.D., hwang12@udayton.edu, 937-229-2039*

Hao wang, Ph.D., hw261@soe.rutgers.edu, 848-445-2874

Lei Wang, Ph.D., wang4li@ucmail.uc.edu, 513-556-7649

Homero Castaneda, Ph.D. hcastaneda@tamu.edu, 979-458-9844

Reporting Period*: 01/01/2025 – 03/31/2025*

1. Project Activities for Reporting Period:

Items Completed During this Quarterly Period:

Per the contract, Task 2 and Task 3 are associated with the second quarterly report. During this quarter, the following activities have been completed as planned in the original proposal.

Item #	***Task #***	***Activity/Deliverable***
1	2,3	2nd Quarterly Report (the 8-page main text)
2	2	Identify critical risk factors (Appendix A)
3	3	Database architecture design (Appendix B)
4	3	Initial version of the interface to public data repositories for data downloading (Appendix C)
5	-	Scheduled internal meetings & activities (Appendix D)

Items in Progress During this Quarterly Period:

Our team is currently actively working on Task 2 and 3 based on the findings from the literature review outcomes generated from Task 1. More detailed, for Task 2, based on the identified critical risk factors for hydrological, geological, and corrosion hazards from the Rutgers team, Texas A&M team, and University of Cincinnati team (i.e., risk factors as reported in this quarterly report), the following items are on-going during this quarter.

Item #	***Task #***	***Activity/Deliverable/Title***
1	2	Confirm the availability from public data sources for identified risk factors
2	2	Assess the data completeness and data quality of both public and private databases.
3	3	Data standardization and harmonization

Task #2 Objective:

The main objective of task 2 is to identify, interpret, and assess critical risk factors for the development of the proposed dynamic database. On the basis of the literature review finished in last quarter, the identified natural hazards have been categorized into geological/geotechnical, hydrotechnical, and corrosion/electrochemical hazards. The datasets of risk factors related to the three types of hazards are analyzed and assessed to evaluate their impact on pipeline integrity. This involves analyzing hazard risk factors from both public and private datasets while leveraging the expertise of respective project teams to ensure comprehensive and accurate risk evaluation.

Summary of work performed:

This task's focus is on the identification of the critical risk factors, data collection of the risk factors and assessing the feasibility to integrate the risk factors into the downstream risk models for risk assessment. A high-level summary is shown below and more detailed technical information are provided in Appendix A.

1) <u>Identify critical factors:</u>

The identification of geological, hydrological, and electrochemical risk factors are categorized based on their impact to the pipeline and organized in a spreadsheet. The risk factors data includes their influence on structure including internal and external pipeline conditions, data collection methods, current available pipeline risk models and risk categories, and industry-standard measurement practices to ensure comprehensive assessment of the data. The expertise teams have identified, compiled and added risk factors.

From the compiled risk factor table, the geohazard factors are primarily associated with the seismic activities and landslides which can be obtained from publicly available data sources. Seismic risks are assessed based on risk factors such as Ground Peak Acceleration (GPA) and Ground Peak Velocity (GPV), which are critical for assessing seismic impacts on pipeline structures. Landslide risks are influenced by the landslide slope angle, ground water level, temperature variations and also Liquefaction index. The common risk factors involve ground subsidence, frost heave, and weather-induced hazards which contribute to external damage to the pipeline.

The identified corrosion/electrochemical risk factors are which affect corrosion rates in pipelines and metallic structures, emphasizing the impact of environmental and operational conditions. The main analysis involves critical factors influencing both internal and external corrosion in pipelines. Environmental factors such as soil type, resistivity, pH levels, chloride concentration, and microbial activity play key roles in both internal and external corrosion, with acidic, high-flow, and oxidizing conditions increasing susceptibility. The risk factors which have to be derived from the other risk factors are calculated using the formula equations mentioned in Appendix A.

The hydrological risk factors are organized into river and coastal zone risks affecting the significant pipeline structure. River hydrology factors pose significant risks to pipelines at watercourse crossings, primarily due to increased water depth, velocity, and sediment transport during flood events. These risks are influenced by flood timing, duration, and magnitude, with snowmelt-driven floods causing prolonged impacts and rainfall-induced floods leading to rapid erosion. Vertical channel movement, including general and local scour, can expose buried pipelines, with severity depending on sediment type, flow conditions, and historical incident data. The Coastal zones include risks from waves, tides, currents, and storm surges, which can cause scour, erosion, and land loss. Hurricanes and coastal storms are the most severe threats, with storm surge effects influenced by storm intensity, coastline shape, and shelf characteristics.

2) <u>Assess the data completeness and data quality of both public and private databases</u>

The team has identified critical hydrological, geological, and corrosion risk factors relevant to underground pipeline risk management. To establish a comprehensive database, data from both public and private databases have been assessed for completeness. Among public sources, Google Earth Engine (GEE) serves as a major repository, providing extensive datasets related to elevation, topography, climate patterns, surface and groundwater distribution, and soil characteristics. These datasets offer strong spatial coverage and high-resolution historical records, but their accuracy and completeness can vary regionally, potentially introducing uncertainty. For seismic hazards specifically, the USGS Earthquake Repository provides authoritative earthquake data. We are in the process of systematically evaluating each data source to identify issues related to spatial resolution, temporal coverage, spatial/temporal data gaps, sensor or measurement errors, data integration inconsistencies, and regional data reliability, to quantify uncertainties and determine suitability for robust pipeline risk assessment.

Private databases maintained by pipeline operators typically provide detailed, pipeline-specific data, including precise geographical information, sensor-derived measurements, and targeted hazard assessments. To access these proprietary data sources, the research team has initiated the process through the Pipeline and Hazardous Materials Safety Administration (PHMSA). A Non-Disclosure Agreement (NDA) has already been signed, and the team is currently awaiting confirmation and access details. Additionally, efforts were made to contact pipeline operators directly by distributing a targeted questionnaire focused specifically on how hydrological and geological hazard factors are integrated into their risk assessment processes. However, responses from the operators indicated that the requested information was considered business confidential and thus could not be disclosed. This outcome highlights a significant research obstacle, namely the industry's reluctance to share proprietary risk assessment methodologies for the research team to better design and build the database.

Task #3 Objective

This task is aiming at establish a database framework that will integrate the critical factors identified from Task 2 into the database, which serves as a comprehensive repository of information related to pipeline structure, operations, maintenance, and integrity management.

Summary of work performed:

1) <u>Database architecture design</u>
The dynamic database design is planned as a one-stop solution to ensure user-friendly efficient storage, access, and analysis of critical data for pipeline risk assessment, operational status and environmental conditions. The dynamic database framework design approach is structured in three stages: organizing georeferenced data with the risk factors from the public and private data sources, interface to risk models, and visualization. A high-level summary is shown below and the detailed figures and explanation of the database architecture design are given in Appendix B.

(i) Data structure and storage of georeferenced pipeline data and risk factors: The database manages risk factor data through a lightweight, structured georeferenced GEOJSON file, which is developed into four distinct levels for fast data processing. The first level is the root data frame which stores Operator ID, Pipeline ID, Commodities, Pipeline status, and Inspection authority, which are indexed to the GEOJSON file based on the operator and pipeline identifiers. This indexing of the operator and pipeline data is stored in a .CSV format type. The second level is the structure of the light Georeferenced GEOJSON file which stores attributes like latitude, longitude, pipeline section [Section ID], pipeline segment based on pipeline mileage [Segment ID], and storage paths to public data and private data files which store the detailed risk factor data. The third level is the public data xml file which stores the identified public data such as elevation, soil type, precipitation, and historical incident data based on Section ID, Segment ID and date selection. The last level is the private data xml file which accesses the identified risk factors data and pipeline data such as pipeline dimensions, inspection data, and maintenance records from private data sources.

(ii) Risk model integration: The selected risk models from the expertise teams assess pipeline risks based on environmental factors, material properties, and historical data. The output is a severity matrix or rank that prioritizes risks and guides mitigation efforts to improve pipeline safety.

(iii) Data Visualization: The data from georeferenced indexes is integrated into risk models, where it is analyzed to evaluate potential risks and vulnerabilities. The processed data is then visualized in two formats: Tabular Representation and Map-Based GIS Visualization, highlighting the risk levels associated with the selected pipeline.

2) Data standardization and harmonization

Building a dynamic database for proactive risk management of underground pipelines requires comprehensive data standardization and harmonization. This process involves converting heterogeneous datasets—collected from multiple public and private sources—into consistent formats and structures, enabling effective integration, analysis, and decision-making. Given the variety of data types, formats, temporal scales, and collection methods, standardization and harmonization are essential to achieve interoperability. The data obtained from Google Earth Engine (GEE) are mostly available in raster and vector formats (e.g., GeoTIFF, shapefiles, KML), suitable for direct use in Geographic Information Systems (GIS). The data from the USGS Earthquake Repository are typically provided as structured geospatial datasets (e.g., GeoJSON, CSV), containing coordinates, magnitudes, depths, and timestamps of seismic events. In contrast, many of the field inspection datasets from pipeline operators are available as tabular data (e.g., CSV, XLS), often lacking spatial referencing. To integrate these diverse data formats into the proposed database, each dataset undergoes preprocessing—including format conversion, georeferencing, temporal synchronization, and semantic alignment. Additionally, we are currently developing automated data conversion functions tailored to each dataset type to streamline this integration process and enhance database efficiency.

3) Initial version of the interface to public data repositories for data downloading

An initial version of the Python-based interface has been developed to streamline data acquisition from public repositories. This script leverages APIs and standard Python libraries, including “geopandas”, and “ee” (Google Earth Engine), enabling automated retrieval of geospatial and hazard data. The script currently supports fetching raster and vector datasets, such as elevation models, land use data, hydrological datasets from Google Earth Engine, and earthquake event data from the USGS Earthquake Repository. The interface is designed for ease of use and reproducibility, with parameters clearly defined for user customization, such as geographic boundaries, date ranges, and data resolution. Further enhancements are underway to expand data sources, optimize download speeds, and improve error handling capabilities. The detailed flowchart and explanation of the of Python-based Interface for data acquisition from public repositories is given in Appendix C.

2. Project Financial Activities Incurred during the Reporting Period:

A cost breakdown list of the expenses during this quarter in each of the categories according to the budget proposal is provided below:

Prime Contract Number: 693JK32450004CAAP
Contract Value: $774,997.00
Funded Value: $774,997.00
Cost-share amount: $116,270

	Current Period Actual	Year To Date Actual	Contract To Date Actual
Salaries & Wages FTFac-Non-Tenure	$3,491.88	$9,341.65	$9,341.65
Graduate Assistant	$2,000	$3,000	$3,000
Benefit-Faculty/Staff	$884.49	$2,366.24	$2,366.24
Student Benefits -GA	$319.2	$478.8	$478.8
Total Labor Cost	**$6,695.57**	$15,186.9	$15,186.9
Conference Registration	$0	$900	$900
G A Tuition Remission	$0	$7,105	$7,105
Total Non-Labor Cost	**$0**	**$8,005**	**$8,005**
Total Indirect Cost	**$3,361.18**	**$8075.52**	**$8075.52**
Total Expense	**$10,056.75**	**$31267.21**	**$31267.21**
Cost-share	**$0**	**$0**	**$0**

The full-time labor hour cost is for the research staff Dr. Sreelakshmi Sreeharan. Starting from 01/13/2025, the University of Dayton (UD) team has recruited a PhD student Kiranmayee Madhusudhan. Kiranmayee is currently on the Graduate Assistantship contract with 6 credits tuition remission in the Spring 2025 semester and her monthly stipend is $2,000 per month. We will recruit an undergraduate researcher Alexander Chattos during the summer. His summer research contract is expected to start from 05/15/2025 and last to 08/15/2025. As Alex will enter into his graduate study, possible contract extension will be processed into the fall semester. The PI's research time will be consolidated and charged during the summer (starting from 05/15/2025) partially using the U Dayton cost-share fund and partially using USDOT fund. During 03/02/2025 and 03/05/2025, the PI Dr. Wang attended the conference Geotechnical Frontiers 2025 at Louisville, KY. Dr. Wang attended ASCE G-I Risk Assessment Management technical committee meeting as the committee secretary and interacted with practitioners in pipeline and geotechnical engineering.

We have fished the subcontracting processes with Texas A&M University (TAMU), University of Cincinnati (UC), and Rutgers University (RU). All documents have been signed. The subcontracts start from 01/2025 or 02/2025. It is expected to have the invoices from subcontractors during the next quarter and the subcontractors have started to charge the project accordingly.

3. Project Activities with Cost Share Partners:

The meetings held with the share partners and the other project teams during the current quarter are listed below, providing an overview of key discussions and decisions accomplished.

1) Weekly progress meeting scheduled with Texas A&M team:

Data and time: 11AM - 12 PM ET; [01/09/2025], [01/16/2025], [01/23/2025], [01/30/2025], [02/06/2025], [02/13/2025], [02/20/2025], [03/06/2025], [03/13/2025]

Attendees: Hui Wang, Homero Castaneda, Myunghwan Jeong, Sasha George, Sreelakshmi Sreeharan, Kiranmayee Madhusudhan

Agenda discussed: The detailed discussion on the purpose of each column in the risk model compilation spreadsheet, particularly focusing on the data sources for electrochemical/corrosion risk factors. Incorporating the hazard risk factor data related to corrosion and identifying different risk models under each risk category. Additionally, the aim is to select one or two key risk models for each risk category and ensure all relevant risk factors are properly categorized and included.

Activities conducted (accomplishments): The electrochemical risk factors are identified through the analysis of both public and private data sources and are added in the risk data factor table alongside the corresponding risk models information.

2) Ad hoc meeting with PHMSA on NPMS and PIMMA industrial account access:

Data and time: 10:30 AM – 11:30 AM ET [2/27/2025]
Attendees: Kendrick, Ben, Jones, Stephen, Nusnin Akter, Leigha Gooding,Hui Wang, Homero Castaneda, Lei Wang, Hao Wang, Sreelakshmi Sreeharan, Jay Shah.
Agenda discussed: The UDayton team discussed the necessity of accessing private NPMS data. The PHMSA team plans to request access to NPMS Levels 1 & 2 and finalize the NDA process. Stephen Jones, Dr. Nusnin Akter, and Ben Kendrick will work together to finalize the NDA, with Stephen and Ben handling communication on the procedure.
Activities conducted (accomplishments): The NDA has been signed between PHMSA and the prime contractor. Separated NDA for data shared with subcontractors will be further discussed according to the actual needs at an ad hoc manner together with PHMSA.

3) Bi-weekly progress meeting scheduled with University of Cincinnati team:

Data and time:10 AM - 11 PM ET [02/21/2025], [03/07/2025], [03/21/2025]
Attendees: Hui Wang, Lei Wang, Yating Yang, Sreelakshmi Sreeharan, Kiranmayee Madhusudhan
Agenda discussed: The detailed discussion on the purpose of each column in the risk model spreadsheet on geological hazards, particularly focusing on the data sources for each risk factor. The team also focused on selecting risk models for geohazards related to pipelines, narrowing down to six models each from probabilistic, qualitative, and quantitative categories. They emphasized identifying detailed geohazard risk variables and creating a technical document for each model's future implementation.
Activities conducted (accomplishments): The geohazard risk factors are identified through the analysis of both public and private data sources and are added in the risk data factor table alongside the corresponding risk models information for possible implementation.

4) Bi-weekly progress meeting scheduled with Rutgers University team:

Data and time: 10AM - 11 PM ET; [02/14/2025], [02/28/2025], [03/14/2025], [03/28/2025]
Attendees: Hui Wang, Hao Wang, Jay Shah, Sreelakshmi Sreeharan, Kiranmayee Madhusudhan
Agenda discussed*:* The team focused on integrating public and private pipeline databases using remote sensing techniques and studying API Recommended Practice 1133, FEMA guidelines, and existing industry practices for hydrological risk factors. Challenges were discussed around accessing proprietary risk models, obtaining comprehensive geological and hydrological data, and approaching industry partners for guidance on workable models. They aimed to develop a list of 5-10 local pipeline operators, optimize questions for them about risk assessment factors, and map these factors to existing data sources.
Activities conducted (accomplishments)*:* The hydrological risk factors are identified through the analysis of both public and private data sources and are added in the risk data factor table. The risk model related to hydrological factors has been identified, but the team is still working on determining the process for integrating the risk factors into the potential risk models.

4. Project Activities with External Partners:

During this reporting period, two significant interactions with our industry partners took place. First, the Q4 2024 Review Meeting was conducted on 28 January 2025 during 2:00 PM to 3:00 PM, involving PHMSA and other sub university partners. This meeting provided a platform to discuss project milestones, present preliminary outcomes, gather feedback, and outline action items for subsequent tasks. Second, discussions were organized to establish a secure data transferring arrangement, focusing on obtaining proprietary pipeline data from pipeline operators and setting up secure data-sharing mechanisms. This meeting, held on 30 January 2025 during 1:00 PM to 2:00 PM, was aimed specifically to describe the project to Dominion Energy and allow them to ask questions. The attendees included Kevin Cowan (Integrity Solutions Field Services), Richard Kiser (Dominion Energy), Homero Castaneda (Texas A&M University), Hui Wang and Sreelakshmi Sreeharan from the University of Dayton. As an outcome of this meeting, we successfully received field inspection data from Integrity Solutions Field Services.

5. Potential Project Risks:

The main technical activities on risk factor identification and public database access are conducted as expected as the project is moving forward so far, no potential risks is noticed at this stage. However, during the performance of Task #2 and Task 3, we observed some delay on subcontract projects as the agreements' negotiation and corresponding paperwork process for kicking-off subcontracts costed longer time than expected. We also faced some difficulties in getting access to certain pipeline private data from industry partners and the government pipeline database such as NPMS. As the NDA has been processed, we expect optimistic progress in the coming quarter.

6. Future Project Work:

Over the next 30 days, our team will focus on finalizing access to private databases through PHMSA and initiating the retrieval of proprietary pipeline data. Additionally, we will refine and enhance the initial Python-based interface scripts to automate seamless data downloads from key public repositories such as Google Earth Engine (GEE) and the USGS Earthquake Repository. A systematic evaluation of the completeness, quality, and uncertainties of these datasets will also commence during this period. Within the following 60 days, we plan to complete comprehensive scripts for data standardization and harmonization, including developing automated functions for converting diverse datasets into a unified, compatible xml format. We will test database updating capabilities through both automatic and manual methods and integrate at least one risk assessment model each for geohazards, hydrological hazards, and corrosion. Over the subsequent 90 days, we will develop data-model compatibility checks and implement data normalization and transformation procedures to ensure consistency across different data sources and risk models. Additionally, we will select pilot pipeline sections to rigorously test the prototype database framework, validating its functionality and effectiveness in real-world scenarios.

7. Potential Impacts to Pipeline Safety:

At this phase of the project, the groundwork for developing a dynamic database that incorporates geotechnical and hydrological factors has been completed. To share these preliminary outcomes, a paper titled *"Dynamic Database for Proactive and Predictive Risk Management of Underground Pipelines: A Comprehensive Review"* has been accepted to will be presented at the ASCE UESI Pipelines Conference 2025. This dissemination will help engage the broader pipeline engineering community, gather valuable feedback, and foster collaboration. A review article is currently under preparation and to be submitted to the *Journal of Pipeline Science and Engineering*.

Appendix A:

1) Categorization of Hazard Factors and Data Sources:

Task 1 literature review highlights that major risk factors to pipeline systems can be categorized into geological/geotechnical, electrochemical, and hydrotechnical threats. Geotechnical and geological hazards involve soil or rock mass displacement, imposing mechanical loads on infrastructure, while hydrotechnical hazards stem from flowing water forces, often carrying debris. Their impact on pipelines can range from gradual ground movement causing structural stress to sudden hydrodynamic forces during floods, necessitating comprehensive risk assessment and mitigation strategies. In addition to natural hazards, electrochemical threats such as soil-induced corrosion and stray current interference also pose significant long-term risks to pipeline integrity.

(i) Data Source Collection:

The identified geological/geotechnical, electrochemical, and hydrological factors are categorized based on their impact mechanisms and systematically organized with relevant data sources for downstream effective risk assessment. During the investigation, a spreadsheet has been created for the respective expertise teams to collect relevant data concerning various natural hazard factors. The data related hazard factors are categorized based on whether they impact the internal or external aspects of the pipeline, PHMSA compliance data collection methods, along with the availability of the data (private or public) and the data format in which the data is collected. Additionally, the risk model, risk category, and measurement standards followed in the industry to collect the data are included for consistency and accuracy in assessment.

The hazard factors data obtained will be used to evaluate the compatibility of the data format downloaded from the data source with the risk models [Qualitative, Relative Assessment/Index, Quantitative System, Probabilistic]. This process involves verifying that the collected geohazard and hydrological data align with the parameters and standard requirements of the established risk models to ensure accurate risk assessment. The accessibility of the data must be checked based on private/publicly available information. For private data, it is essential to implement proper management and security measures. This approach ensures the seamless integration of the data into the risk analysis while maintaining appropriate standards for data privacy and accessibility.

The data streams are essential for adapting to rapidly changing climate and geological conditions, historical data plays a vital role in informing and optimizing real-time data collection efforts. It provides valuable context for understanding long-term trends and patterns, which enhances the accuracy and effectiveness of risk models. Historical data sources include PHMSA Incident Reports, National Response Center (NRC) Reports, U.S. Federal Emergency Management Agency (FEMA) disaster declarations, and the NOAA storm database. These data sources offer insights into recurring patterns, risk factors, and mitigation strategies, as well as the frequency and distribution of hazardous substance releases.

The Public data sources, which include georeferenced climate data, geological/geotechnical data, and remote sensing data from government agencies through open-source platforms such as Google Earth Engine, NASA Earth Data, USGS Earth Explorer, NOAA Climate Data Online, and Soil Grids. Private data sources include data samples from pipeline owners/operators, which may include in-line inspection data, soil surveys, indirect inspection data (DCVG, CIPS, CP potential, etc.), and emerging technologies such as 3D scanning using LiDAR or photogrammetry. A meeting with the PHMSA industry partners was scheduled detailing access requests to private

NPMS data availability and ongoing efforts to acquire additional attributes from pipeline operators, in line with new PHMSA compliance requirements.

A spreadsheet has been created to analyze various geological/geotechnical, electrochemical, and hydrological factors related to pipeline integrity. The following columns are included and discussed with other sub university teams, providing inputs to the spreadsheet based on their respective areas of expertise:

General Factors: Specifies the hazard factor.
Internal/External: Indicates whether the factor impacts internal or external pipeline damage. If a factor affects both, create a new row to account for potential differences in data sources or other attributes.
Direct/Derived Factor: Identifies whether the factor is directly observed or derived from a formula or relationship.
Direct: The factor is directly measured or observed.
Derived: The factor is inferred or calculated based on other data or relationships.
Data Collection: Describes the method used to collect the data.
Private/Public: Specifies data availability.
Data Format: Indicates the format in which the data is collected.
Risk Model Name: The name of the risk model or system used to predict or assess geohazard risks related to pipelines.
Measurement Standards: The standards or guidelines followed for measuring and reporting data.
Risk Model Category: Classifies the risk model (e.g., Probabilistic, Qualitative, Quantitative).
Additional Information: Includes links to sources, research, detailed notes, or other relevant information. If the factor is derived, specify the equation or relationship used.

Geological/geotechnical Factors:
Based on the hazard factor data table, the majority of geological/geotechnical factors are primarily associated with seismic activities and landslides, which are the dominant risks impacting pipeline stability. For seismic activity, the key factors include Ground Peak Acceleration (GPA) and Ground Peak Velocity (GPV), which are critical for assessing seismic impacts on pipeline structures. For landslides, a primary influencing factor is the landslide slope angle. Beyond seismic and landslide-related hazards, other significant factors affecting pipeline stability in different geo-hazard categories include groundwater level, temperature variations, and liquefaction potential index. These factors are particularly relevant in areas prone to subsidence, frost heave, and weather-induced hazards. Additionally, almost all these factors contribute to external damage to the pipeline, with most being directly measurable. At the same time, fewer are calculated based on specific equations, as referenced in the Additional Information column in the spreadsheet.

Electrochemical factors:
The identified key factors mentioned in Table 1 influencing corrosion rates in pipelines and metallic structures, emphasizing the impact of environmental and operational conditions. Soil resistivity, water flow velocity, and pH levels play significant roles in corrosion risk, with acidic and high-flow conditions increasing susceptibility. Microbial activity and biofilm formation contribute to localized corrosion, while surface roughness and galvanic coupling between dissimilar metals further accelerate degradation. Moisture and UV exposure weaken protective coatings, and the presence of oxidizing agents enhances electrochemical reactions. The review

highlights the importance of selecting corrosion-resistant materials, controlling environmental conditions, and implementing regular monitoring and maintenance to mitigate corrosion risks effectively.

Table.1 Corrosion risk factors for External/Internal Corrosion

General Factor	Secondary Factor	Tertiary Factor	Corrosion risk factor	Internal/External
Environmental Conditions	Moisture level	Water condensation	Microbial activity/ H2S (SRB, IOB, etc.)	External/Internal
	Organic matter content/Soil carbon stock	Biofilm formation	Microbial activity/H2S (SRB, IOB, etc.)	External/Internal
	Soil type	Soil resistivity	Presence of corrosive ions/Electrochemical reaction kinetics	External
	Soil chloride concentration	pH	Electrochemical reaction kinetics	External
	Salinity	Sulfate concentration	Presence of corrosive ions	External
	Redox potential	Galvanic coupling		External
Operational Conditions	External pressure from buried load	Stress and mechanical load	Surface roughness	External/Internal
	Pipeline burial depth and material type	Stress and mechanical load	Surface roughness	External/Internal
	Water flow rate/Oxygen diffusion rate		Differential aeration effects	Internal
	Pressure/Pipeline geometry (bends, welds, joints)	Stress and mechanical load	Surface roughness	Internal
Chemical Composition	CO_2 concentration in liquid	pH of water	Electrochemical reaction kinetics	Internal

	O_2 concentration in liquid	Differential aeration effects	Electrochemical reaction kinetics	Internal
	Presence of inhibitors	Electrochemical reaction kinetics	Surface roughness	Internal
Material Properties	Metal composition	Galvanic coupling	Electrochemical reaction kinetics	External/Internal
	Type of coatings	Pipeline external coating degradation	Surface roughness	External/Internal

Table.1 outlines and explains the various factors influencing corrosion risks in pipelines, categorized into environmental conditions, operational conditions, chemical composition, and material properties. Understanding these categories helps in effectively identifying and mitigating potential corrosion-related risk-factors in underground pipelines.

Based on the factors presented in Table 1, critical corrosion risk factors were systematically identified for both internal and external corrosion scenarios affecting pipelines (Fig.1). For internal corrosion, key factors considered were temperature, hydrogen sulfide (H_2S) concentration, fluid movement dynamics, operating pressure, and microbial activities. Conversely, external corrosion considerations included pH, temperature, chloride ion concentration (Cl^-), soil humidity, and microbial activities. Each factor was selected due to its significant impact on corrosion mechanisms, pipeline integrity, and operational reliability. The following sections describe the significance of these factors in detail and suggest corresponding risk-based corrosion models inclusive of representative equations.

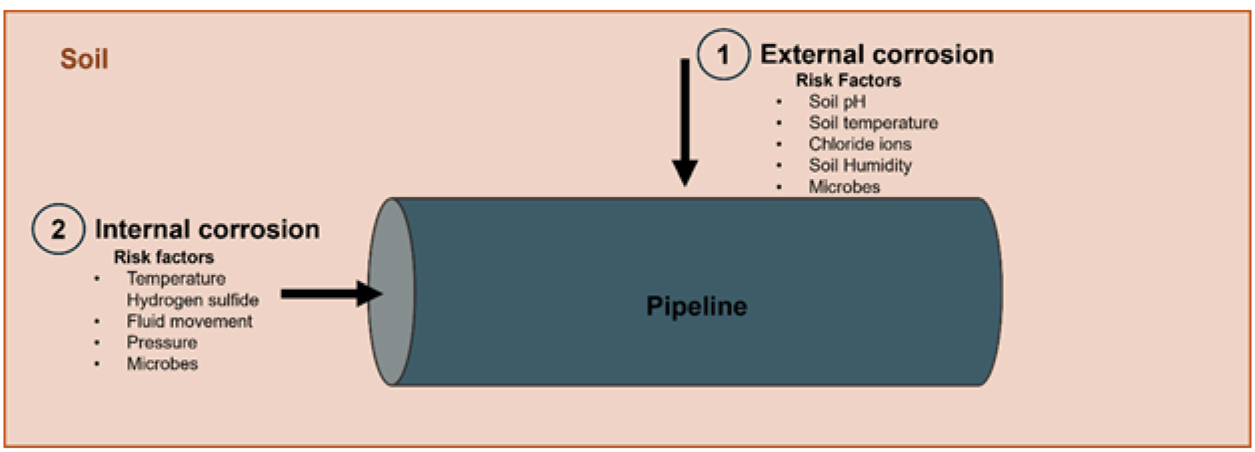

Figure 1: Schematic of risk factors affecting pipeline corrosion

Hydrological factors:

Pipelines operate in dynamic environments where natural hydrological forces continuously reshape the terrain, posing risks to their integrity. Rivers, floodplains, and coastal zones present complex challenges due to erosion, flooding, subsidence, and storm impacts. In riverine settings, channel shifts and soil erosion can expose or undermine pipelines, while floodplains introduce buoyant forces that threaten stability. Coastal zones face shoreline retreat and storm surges that can suspend pipelines, making them vulnerable to hydrodynamic forces and structural failure.

Effective pipeline risk modeling requires accurate environmental data, integrating both historical failure records and hydrodynamic parameters like water velocity and erosion rates. Reliable assessments depend on comprehensive, up-to-date data sources such as satellite imagery and floodplain maps. This report explores various risk scenarios, relevant hydrological guidelines, and methods to incorporate environmental dynamics into pipeline risk modeling.

The major hydrological factors considered are river and coastal zone Hydrology and Hydro-technical Hazards.

River Hydrology factors:

Flood events pose significant risks to pipelines at watercourse crossings due to increased water depth, velocity, and sediment transport. High-energy flows can induce bending stresses, vortex-induced vibrations, and lateral forces on exposed pipelines. Scouring of riverbeds and banks may further expose pipelines to damage. The severity of these risks depends on flood timing, duration, and magnitude, which are analyzed using hydrographs. Snowmelt-driven floods lead to gradual, prolonged impacts, while rainfall-induced floods cause rapid and intense erosion. Peak flow periods increase scour, while sediment deposition during receding flows can cause uneven pipeline support. Vertical channel movement primarily affects pipelines through scour, which occur in different forms. General scour refers to the erosion of the riverbed due to increased velocity during floods, potentially exposing buried pipelines. The risk of general scour depends on sediment type, flow conditions, and historical flood data, and its depth can be estimated using hydrodynamic equations. Local scour occurs around obstructions such as pipeline supports and bridge piers, where flow acceleration leads to deeper erosion. This type of scour can be calculated using Lacey's equation or other empirical methods. Other factors, such as channel degradation, where long-term bed lowering results from sediment supply disruption, and ice scouring, where ice movement erodes the riverbed, can destabilize pipelines, increasing their exposure to external forces. Lateral channel movement presents additional threats through channel migration and bank erosion. Rivers naturally shift over time, eroding banks and depositing sediment elsewhere, which can expose pipelines or suspend unsupported sections, leading to damage. The rate of erosion can be predicted using curvature-to-width ratio models. Another major concern is channel avulsion, a sudden shift in a river's course due to extreme events like ice-melt floods or insufficient slope. This can expose pipelines to new erosional forces, significantly increasing their risk of damage. Identifying vulnerable locations relies on historical aerial imagery and hydrological data analysis. To mitigate these risks, continuous monitoring and predictive modeling of flood events, scour processes, and lateral river movements are essential for ensuring pipeline integrity and stability.

Coastal Zone Hydrology factors:

Pipelines in coastal zones face risks from waves, tides, currents, and storm surges, which can cause scoring, erosion, and land loss. Hurricanes and coastal storms create the most severe threats by pushing water inland, temporarily inundating areas and exposing pipelines to damage. Storm surge effects depend on storm intensity, approach direction, coastline shape, and continental shelf characteristics, with shallow shelves typically leading to higher surge levels. Unlike river pipelines, coastal pipelines can be affected along their entire length, not just at water crossings. Shoreline retreat is a critical coastal hazard that threatens pipelines, especially at offshore-to-land transitions. This process involves the gradual or sudden erosion of coastal material due to tidal fluctuations, wave action, and extreme weather events like hurricanes. The severity of shoreline retreat depends on shoreline composition, wave exposure, wind direction, and sea-level rise. Over time, retreat can expose buried pipelines, increasing the risk of damage. Historical and current aerial imagery comparisons are useful for assessing erosion trends and identifying high-risk areas requiring mitigation. Coastal land loss, the transformation of once-dry land into open water, further intensifies risks for pipelines, particularly around bays and estuaries. Driven by erosion, sediment depletion, land subsidence, and sea-level rise, this process leaves previously buried pipelines exposed to stronger hydrodynamic forces. The disappearance of protective features like beaches and marshes increases vulnerability to storm surge, higher water velocities, and deeper scour, heightening the likelihood of structural damage.

ii) Data Acquisition:

From the publicly available data side, the risk data factors such as temperature, precipitation, vegetation cover, seismic and soil environment will be extracted. This information will be validated, managed and stored in a georeferenced data index linked to the public data frame for risk assessment with respect to pipelines.

On the private data side, factors related to potential exposure, damage, corrosion, pipeline mechanical properties, geometry deformation, and service conditions are considered. This data, often collected in different modalities, will be georeferenced and linked with the private data frame. After interpreting the data, the information will be reorganized and grouped according to the requirements of the risk models. The inputs from subject matter experts will also be used as sources of information for interpreting portions of the data. By carefully analyzing data and employing a structured approach, epistemic uncertainty enhances the effectiveness of downstream knowledge- and data-driven analysis.

The current data acquisition sources for geological/geotechnical data heavily rely on field surveys, which directly measure the geohazard factors impacting pipeline stability. In addition to field surveys, advanced technologies are utilized, including remote sensing techniques such as InSAR (Interferometric Synthetic Aperture Radar) and LiDAR (Light Detection and Ranging), seismic exploration, and Geographic Information System (GIS) analysis. Furthermore, the most widely used public database for geo-hazard data is the United States Geological Survey (USGS).

The current data acquisition sources for hydrotechnical hazards are U.S. Geological Survey (USGS) and Environment Canada (EC) Water Survey are key agencies providing essential flood flow data, with USGS offering high-resolution streamflow data every 15 minutes, while EC provides real-time hydrometric data from over 1,800 stations and maintains an extensive historical archive. This data, often visualized through hydrographs, helps assess flood behavior, duration, and peak discharge, which are critical for evaluating flood risk and designing pipelines. Statistical flood analysis, including return periods like the 100-year or 500-year flood, further aids in establishing design thresholds for pipeline safety. In coastal environments, pipelines face risks from storm surge, coastal flooding, wave loading, and shoreline erosion. Key agencies like the USGS, National Weather Service (NWS), and FEMA provide valuable data for assessing and mitigating these risks. The NWS offers real-time flood risk forecasts, while FEMA provides mapping of flood extents and storm surge water elevations. NOAA contributes oceanographic data, including real-time hurricane forecasts, wave measurements, and tidal observations, which are crucial for understanding the hydrodynamic impacts on pipelines. Together, these agencies support comprehensive assessments of coastal zone hazards, enabling better pipeline design and operational strategies.

2) <u>Risk Assessment Models:</u>

The Risk Modeling Work Group of Pipeline and Hazardous Materials Safety Administration (PHMSA) categorizes risk models into four types: Qualitative Models, Relative Assessment/Index Models, Quantitative System Models, and Probabilistic Models. These categorizations serve as a foundation for evaluating pipeline and hazard mitigation. Based on their respective areas of expertise, each team has focused on their specialized fields for pipeline risk models which allow us to map the relevant risk factors.

The geo-hazard risk factors are predominantly applied in probabilistic models, which effectively handle uncertainties in risk assessment. For instance, In Fault Tree Analysis (FTA), Kazmi et al., 2017[1] is widely employed to systematically evaluate failure in pipeline systems, while in Probabilistic Seismic Hazard Analysis (PSHA), Hudson et al., 2022[2] plays a crucial role in quantifying seismic risks. Additionally, for the Early warning model, Ning et al., 2023[3] primarily falls within the probabilistic and quantitative system model categories, relying on real-time data, predictive analytics, and statistical simulations to assess and forecast pipeline vulnerabilities.

Furthermore, quantitative models provide a structured approach to risk assessment by integrating empirical data and simulation techniques. For Bayesian networks, Koduru, 2019[4]; Mahmood et al., 2024[5] effectively capture dependencies between multiple geo-hazard factors and update risk predictions dynamically. Similarly, for Monte Carlo simulation, Alvarado-Franco et al., 2017[6] generates probabilistic risk estimations by simulating various failure scenarios, enabling more robust decision-making.

The integration of geo-hazard risk models with qualitative risk assessment models is less common compared to their application in quantitative and probabilistic models. This is primarily because qualitative models rely on expert judgment and categorical classifications, which can be less effective in addressing the complexities of data-driven geo-hazard risks. Nonetheless, certain qualitative approaches have been applied in pipeline risk assessment. For example, Vanitha et al. (2023)[7] introduced the relative risk score technique, which assigns numerical values to qualitative risk factors, enabling a semi-quantitative assessment. Similarly, Zarei & Kalatpour (2018)[8] utilized the Hazard and Operability (HAZOP) Study, a structured method that identifies potential pipeline risks through expert analysis, structured brainstorming, and deviation analysis. While these methods provide initial risk assessments, they are typically supplemented with quantitative techniques to enhance accuracy. By integrating these diverse risk assessment models and leveraging geo-hazard data from public and private sources, researchers and industry professionals can make better decisions about future developments.

The detailed study examines various research studies on pipeline corrosion risk assessment and modeling. Yazdi et al. (2022)[9] explored microbiologically influenced corrosion (MIC) in offshore pipelines, highlighting probabilistic and deterministic risk-based decision-making models, including Bayesian networks, Monte Carlo simulations, and mechanistic models. Wasim and Djukic (2022)[10] analyzed external corrosion due to soil conditions, emphasizing factors like soil moisture, pH, and ion content while proposing a risk matrix for corrosion likelihood and mitigation strategies. Shabarchin and Tesfamariam (2016)[11] employed Bayesian Belief Networks (BBN) to assess internal corrosion risks in oil and gas pipelines, integrating multiple risk factors to optimize mitigation strategies. Gartland et al. (2003)[12] categorized internal corrosion mechanisms into electrochemical reactions, microbiologically induced corrosion, and localized corrosion, introducing empirical, mechanistic, and data-driven models for risk assessment. These studies collectively contribute to a comprehensive understanding of corrosion mechanisms, predictive models, and proactive mitigation strategies for pipeline integrity management.

Geological/geotechnical risk models: The findings confirm that seismic activity and landslides are the most critical geo-hazard risks affecting pipeline stability, with factors such as Ground Peak Acceleration (GPA), Ground Peak Velocity (GPV), and landslide slope angle playing a significant role in risk evaluation. Additionally, factors related to subsidence, frost heave, and weather-induced hazards further contribute to risks of external damage. While field surveys remain the primary method of data acquisition, the integration of remote sensing technologies (InSAR, LiDAR), seismic exploration, and GIS-based analysis can significantly enhance data accuracy and risk prediction capabilities. Furthermore, current pipeline risk assessments rely heavily on probabilistic and quantitative models, including Fault Tree Analysis (FTA), Probabilistic Seismic Hazard Analysis (PSHA), Monte Carlo Simulation, and Bayesian Networks. These models provide a systematic, data-driven approach to pipeline risk assessment, facilitating more reliable predictions and effective risk mitigation strategies.

Electrochemical risk models: The electro-chemical risk model review emphasized preventive measures including selecting appropriate materials, controlling environmental conditions, and regular monitoring and maintenance practices based on empirical relationships and standardized guidelines. The assessment of corrosion risks in pipelines and metallic structures involves evaluating several key factors using various models, standards, and estimation methods. For soil conductivity (resistivity), an empirical resistivity measurement model is used, with the severity directly correlated to the resistivity values, guided by ASTM G57 standards. Water flow rate is assessed using a qualitative model, with corrosion severity categorized based on flow velocity ranges, following the API RP 14E standard. The pH of the liquid is evaluated through an empirical, qualitative relationship, with severity categorized based on direct pH measurement. Microbial activity and biofilm formation are assessed using a semi-quantitative microbial count (CFU/ml) and qualitative biofilm coverage model, with severity linked to microbial concentration and biofilm extent, following NACE TM0194 and ISO 16784 standards. Pipe wall roughness is evaluated using the relative roughness (ε/D) model, with severity estimation based on the Moody diagram and Colebrook-White Equation and is guided by API RP 14E and ASME B31.3 standards. For galvanic coupling, the Galvanic Series potential difference model is used, with severity assessed based on the metal pairing potential, following ASTM G82 standards. Metal composition is evaluated qualitatively, with severity estimation based on the corrosion resistance of the alloy type, as per ISO 15156 standards. Moisture level and UV exposure are assessed using qualitative, empirical models, with severity correlated to moisture percentage and UV intensity, guided by ISO 9223 (Moisture) and ISO 4892-3 (UV) standards. Finally, the presence of oxidizing agents and electrochemical reaction kinetics are assessed through a qualitative model using the Butler-Volmer equation, with severity correlated to agent concentration and reaction rates, guided by ASTM G31 and ASTM G102 standards. These models integrate qualitative, empirical, and equation-based methods to estimate corrosion severity, following industry standards to ensure consistent and reliable risk evaluation.

Risk-based decision-making models for assessing microbiologically induced corrosion (MIC) focus on understanding how microbial activity accelerates corrosion, particularly through sulfate-reducing and iron-oxidizing bacteria. These models use both probabilistic and deterministic approaches to evaluate and mitigate corrosion risks. Probabilistic models, like Bayesian networks and Monte Carlo simulations, use statistical methods to represent the relationship between risk factors (e.g., microbes, temperature) and predict pipeline failure probabilities over time. Deterministic models, such as mechanistic models (based on electrochemical and biochemical reactions) and empirical models (using field and lab data), predict corrosion rates directly. A hybrid approach combines both probabilistic and deterministic methods, incorporating fuzzy logic to assign risk degrees (e.g., low, medium, high) and machine learning AI models to detect corrosion patterns from field measurements, providing a comprehensive framework for proactive corrosion management.

We have summarized the likelihood ratings of different soil conditions affect external corrosion rates in Table 2 below.

Table 2: Likelihood ratings of different soil conditions affect external corrosion rates.

Soil conditions	Impact on corrosion rate	Likelihood rating
Soil moisture	Higher moisture levels result in an increased corrosion rate	High
Soil pH	Acidic and alkaline pH increases corrosion faster	High
Soil temperature	Higher temperature increases the corrosion rate	Medium - High
Soil Resistivity	Low soil resistivity increases the corrosion rate	High
Sulfate ion content	Higher concentrations of sulfate ions increase the corrosion rate	High
Chloride ion content	Higher concentrations of chloride ions increase the corrosion rate	High

We have summarized the likelihood ratings of different soil conditions affect internal corrosion rates in Table 3 below.

Table 3: Likelihood ratings of different soil conditions affect internal corrosion rates.

Pipeline conditions	Impact on corrosion rate	Likelihood rating
Pressure of pipeline	Higher pressure increases the corrosion rate	High
Temperature of pipeline	Higher temperature increases the corrosion rate	High
Flow regime	Turbulent flow increases the corrosion rate	Medium - High
Water presence	Water presence increases the corrosion rate	High
Carbon dioxide and Hydrogen sulfide	Both gases increase the corrosion rate	High
Sulfate-reducing bacteria	The presence of this bacteria increases the corrosion rate (MIC)	High

Probability of failure (Risk assessment): The suggested models for assessing corrosion risk include both probabilistic and deterministic approaches. Probabilistic models, such as Bayesian Networks and Monte Carlo Simulations, use probabilistic distributions and scenario generation to represent the relationships between various risk factors like microbial activity, temperature, and environmental conditions, and to estimate the likelihood of pipeline failure over time. On the other hand, deterministic models, including Mechanistic Models and Empirical Models, focus on simulating corrosion based on physical and chemical principles or historical data, respectively, to predict corrosion rates and pipeline degradation. Combining these models allows for a more comprehensive and robust approach to corrosion risk assessment and management with the given calculation.

$$PoF = e^{\alpha+\beta CR+\gamma T}$$

PoF – Probability of pipeline failure; α, β, γ – Coefficients; CR – corrosion rate; t – Service time.

Risk Model for Pipe Wall Roughness: Pipe wall roughness significantly affects fluid dynamics and corrosion behavior in pipelines. Key standards that guide this are API RP 14E (which offers fluid velocity guidelines) and ASME B31.3 (which covers process piping design). The Colebrook-White Equation is used to calculate the friction factor (f), which is influenced by the roughness of the pipe wall and affects flow and corrosion characteristics.

$$\frac{1}{\sqrt{f}} = -2log_{10}\left(\frac{\varepsilon}{3.7D_h} + \frac{2.51}{Re\sqrt{f}}\right)$$

f = Darcy friction factor; ε = Absolute roughness (m) ;D = Pipe diameter (m);Re = Reynolds number.

Risk Model for presence of Oxidizing Agents and Electrochemical Reaction Kinetics: Oxidizing agents significantly influence electrochemical corrosion by accelerating the reaction kinetics. The presence of oxidizing agents enhances electron transfer at the metal surface, leading to an increased corrosion rate. To model this process, a qualitative assessment of the impact of oxidizing agents is combined with a quantitative approach using the Butler-Volmer Equation for the corrosion current density, j.

$$j = j_0 \cdot \left\{ \exp\left[\frac{\alpha_a zF\eta}{RT}\right] - \exp\left[-\frac{\alpha_c zF\eta}{RT}\right] \right\}$$

j_0 = Exchange current density (A/m²); η = Overpotential (V); α_a, α_c = Transfer coefficients; n = Number of electrons transferred; F = Faraday constant (96485 C/mol);R = Universal gas constant (8.314 J/mol·K); T = Absolute temperature (K)

This equation helps describe the electrochemical reaction kinetics and the relationship between the corrosion current density and overpotential. The Butler-Volmer Equation accounts for both anodic and cathodic reactions, providing a detailed understanding of the corrosion process. Standards such as ASTM G31, which measures corrosion rates through weight loss, and ASTM G102, which calculates corrosion rates from electrochemical measurements, are used to assess the corrosion rate in practical applications. Together, these methods and standards offer a comprehensive approach to understanding and quantifying the effects of oxidizing agents on corrosion.

Hydrological risk models: Flood frequency analysis is an essential component of hydrologic risk evaluation for pipelines, particularly at river and floodplain crossings. In the U.S., the standard methodology for determining flood flow frequencies is outlined in USGS Bulletin 17B, published in 1981. This bulletin provides procedures for statistical analysis of annual peak discharge records from stream gauges, forming the foundation for estimating flood probabilities of various magnitudes, such as the widely used 100-year return period flood, often required by FEMA for pipeline construction projects.

Flood frequency analysis is typically performed using observed flow data, often with software tools like HEC-SSP (Statistical Software Package) developed by the U.S. Army Corps of Engineers. HEC-SSP applies Bulletin 17B methods and supports various parametric statistical approaches, including the Log-Pearson Type III distribution (the default method), along with the 3-parameter lognormal and generalized extreme value distributions, each assuming different characteristics in the flood flow dataset. The reliability of these estimates depends on factors like the length of the observed hydrograph record and the quality of data at the selected gauge location.

Longer datasets allow for more statistically robust estimates. Additionally, regularly updating flood frequency analyses with current data is vital to reflect changes in climate, land use, and watershed behavior. Urbanization can significantly alter runoff patterns and flood magnitudes, making historical records insufficient in some cases.

To understand industry practices in pipeline risk assessment, we developed a targeted questionnaire to gather insights on the integration of hydrological and geological hazard factors. The objective was to create a comprehensive database schema aligned with the risk assessment strategies of pipeline practitioners. The questionnaire was designed to be concise, focusing on three key questions:

- Which type of pipeline hazard risk assessment model do you use (e.g., qualitative, quantitative, probabilistic, hybrid)?
- Do you incorporate any hydrological and geological factors into your model? If yes, please specify factors such as nearby water bodies, precipitation, temperature, snow depth, vegetation coverage, soil type along the right-of-way, pipeline structure, digital terrain model, etc.
- Do you use any specific software to perform risk assessments (e.g., in-house, open-source, commercial software)? If so, please state the software name.

The goal was to understand the diversity of risk assessment models used in the industry, the environmental factors integrated into these models, and the software tools employed. We selected pipeline operators from Ohio State using the NPMS Public Viewer and identified 40 contacts with complete details. However, we faced challenges obtaining responses due to the confidential nature of pipeline risk assessment practices. Only one response was received, which directed us to publicly available data from PHMSA DataMart, but it was inaccessible, preventing verification of the relevant data. This experience highlighted the industry's reluctance to share proprietary risk assessment methodologies and reinforced the challenges of gathering industry-specific insights. Alternative strategies may be needed to obtain the necessary data for further research.

Appendix B:

Design approach of database architecture:

The effective management and analysis of pipeline data requires a well-structured storage system that integrates geospatial information and associated risk factors. The database architecture is designed to store, organize and visualize georeferenced hazard data along with critical risk-related attributes, enabling efficient risk assessments and decision-making. The database architecture design primarily manages three key stages: georeferenced pipeline data and risk factors, interface to risk models, and visualization.

(i) Data Structure and Storage of georeferenced pipeline data and risk factors:

The database manages risk factor data through a lightweight, structured georeferenced GEOJSON file, which is developed into four distinct levels for efficient data organization and processing.

1)Root Data Frame [DatabaseID.CSV]:
The Georeferenced GEOJSON files are indexed based on the operator ID, Pipeline ID, Commodities, Pipeline Status, Inspection Authority compatible with NPMS database system. All the field attributes below are stored in the root data frame DatabaseID.CSV file as in Figure 2.

Operator ID: Accounting number assigned by the PHMSA to the company that operates the pipeline.
Pipeline ID: Unique identifier for a specific pipeline within a pipeline system.
Commodities: Specifies whether the pipeline segment carries gas or hazardous liquid commodities.
Pipeline Status: Identifies the current operational status of the pipeline segment.
Inspection Authority: The inspection authority responsible for the pipeline segment whether the pipeline is inspected by the PHMSA or state.

Based on the above attributes, each operator ID, and the respective pipeline ID is named to each Georeferenced Index file.

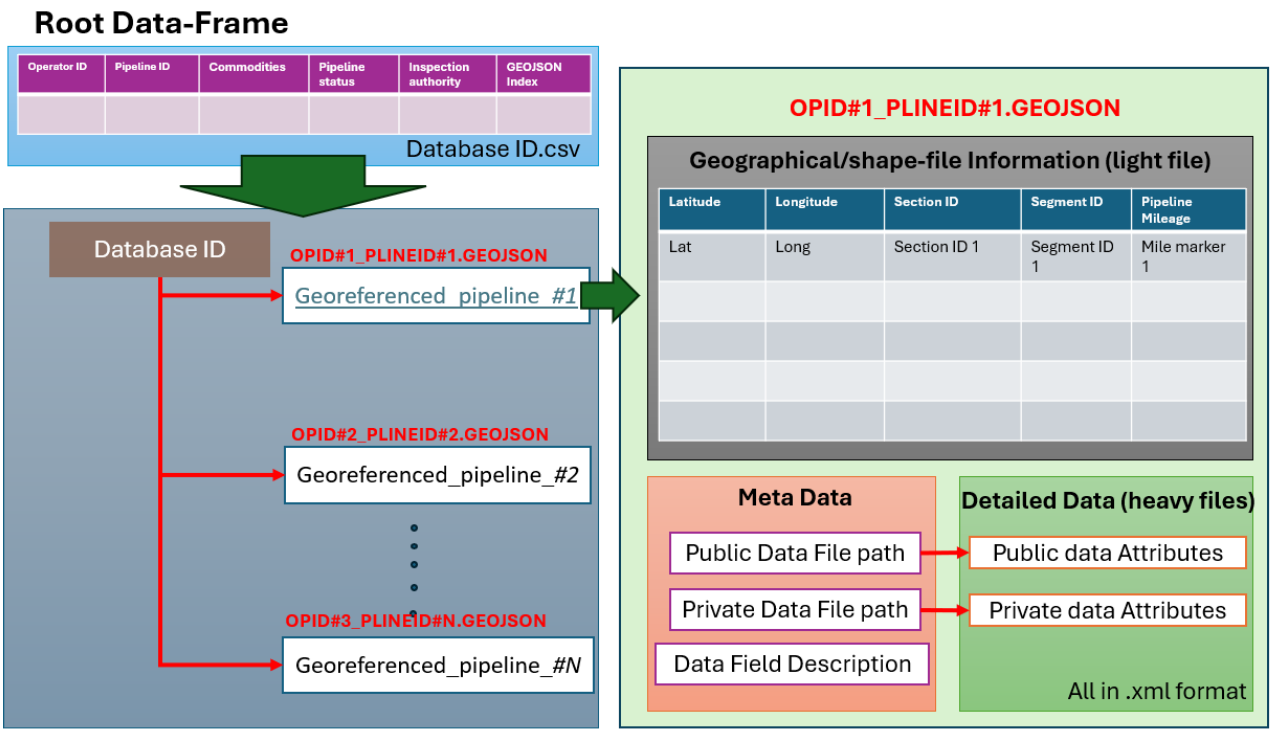

Figure 2: The root data frame and GEOJSON structure design of Database architecture.

2) Light Georeferenced GEOJSON Structure [.GEOJSON]:

The database stores pipeline data using geospatial indexes (OPID_PLINEID) in a light packet structure, specifically in the GEOJSON format. The OPID_PLINEID corresponds to each Operator ID and its respective Pipeline ID, which are mapped to their corresponding georeferenced index.

The Georeferenced Data contains the following attributes in each GEOJSON file as in Figure 2:

Latitude, longitude: The coordinates of the geographic latitude/longitude projection and displayed in decimal degrees units.

Section ID: Section for each pipeline ID

Segment ID: Pipeline divided into segments based on Pipeline Mileage(miles).

Public and Private data file path: The path to store all the information of risk factors based on public and private data sources.

Data description: Data description of all attributes.

The explanation of each GEOJSON attribute with two schematic pipeline examples can be visualized in Figure 3. The pipelines are added just for understanding purpose and do not depict any true pipeline.

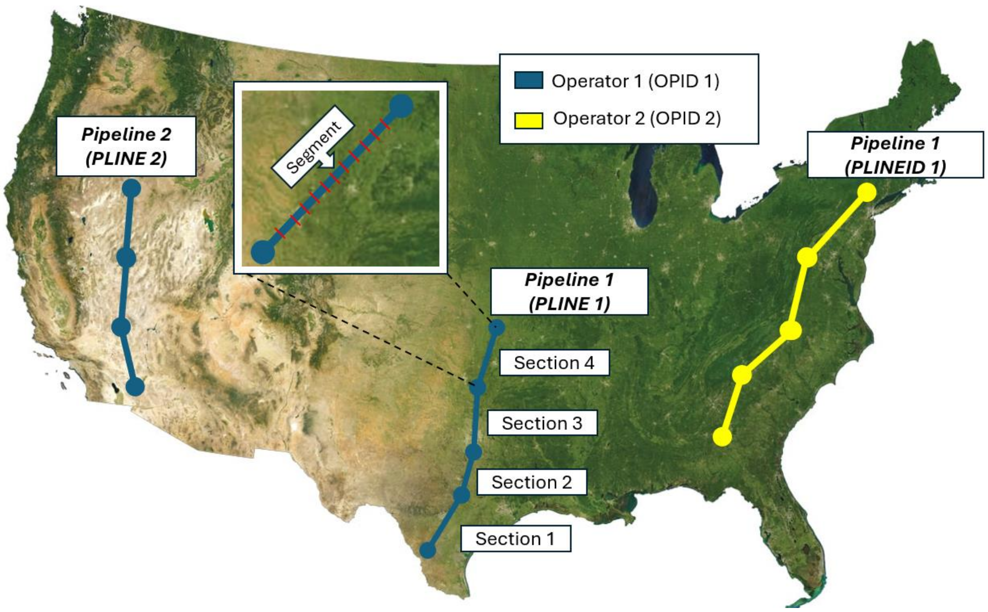

Figure 3: The schematic visualization of the GEOJSON structure attributes.

With the same approach, each georeferenced index file holds the attribute data corresponding to the respective pipeline section, maintaining the same structure specific to the Operator ID and Pipeline ID.

3)Public Attributes[.xml]:

The public data file path is linked to the corresponding GEOJSON file using the indexing keys which are Section ID, Segment ID, and date as shown in Figure 3. The public XML file contains publicly available identified risk factors essential for risk assessment, including risk factors and public data sources used to gather information based on the date of request.

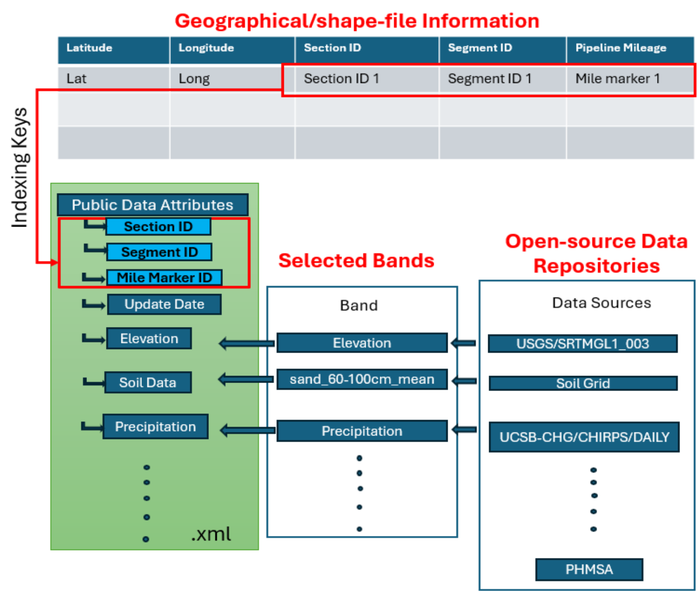

Figure 4: The structure design of Public Attributes.xml file in Database architecture

The public attributes contain information related to various risk factors, such as elevation, soil properties, precipitation, vegetation, and historical incident data from PHMSA. These factors serve as critical inputs to the risk models, which are used for assessing potential risks to pipeline integrity. The data provides valuable insights for evaluating environmental conditions, historical trends, and other key factors that may impact the safety and reliability of pipeline systems. By incorporating this information, the risk models can predict and mitigate potential threats to pipeline infrastructure more accurately.

4)Private Attributes[.xml]:

The private data file path is linked to the corresponding GEOJSON file using the indexing keys which are Section ID, Segment ID, and date as shown in Figure 4. The private XML file contains restricted access information essential for assessing pipeline dimensions, inspection data, maintenance records, and operational performance. This data is critical for in-depth analysis of the pipeline's structural integrity, ongoing monitoring efforts, and compliance with safety regulations. The private data helps support internal risk assessment models, detailed inspections, and decision-making processes to ensure the pipeline operates safely and efficiently.

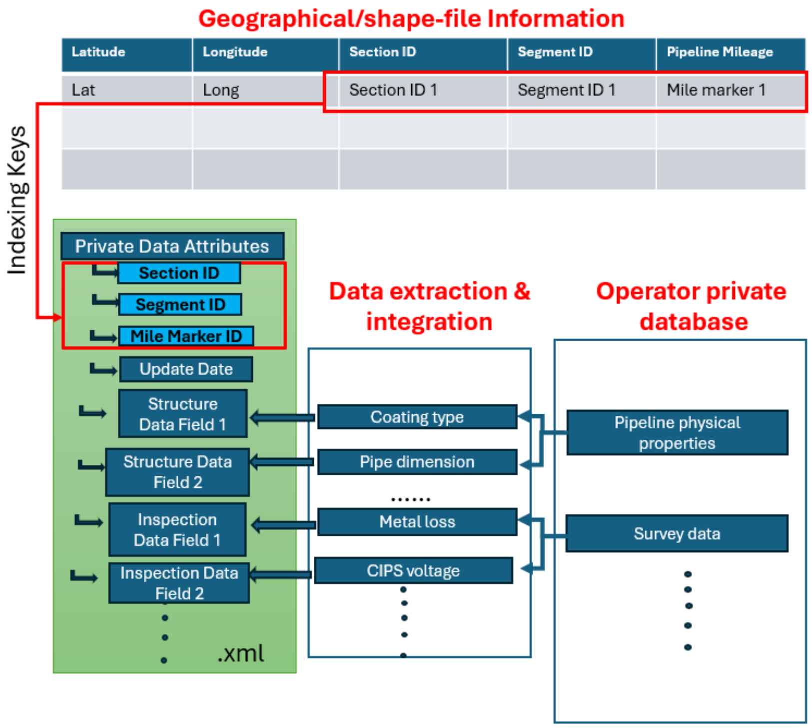

Figure 5: The structure design of Public Attributes.xml file in Database architecture

<u>(ii) Risk Model Interface:</u>

The potential risk models for assessing the risk of geo-hazard, corrosion-hazard, and hydrological-hazard are recommended based on the inputs from the three sub university teams, who provide expertise in each of these areas. These models consider various risk factors, such as environmental conditions, material properties, and historical incident data, to assess the potential risks to pipeline integrity as shown in Figure 6. Based on the risk model outputs, a severity matrix or rank will be generated, which helps prioritize risks and determine the level of intervention or mitigation required. This severity ranking plays a key role in the decision-making process, guiding actions to enhance pipeline safety and reduce the likelihood of failures or damage. We will implement this framework in the following quarters as planned in the project proposal.

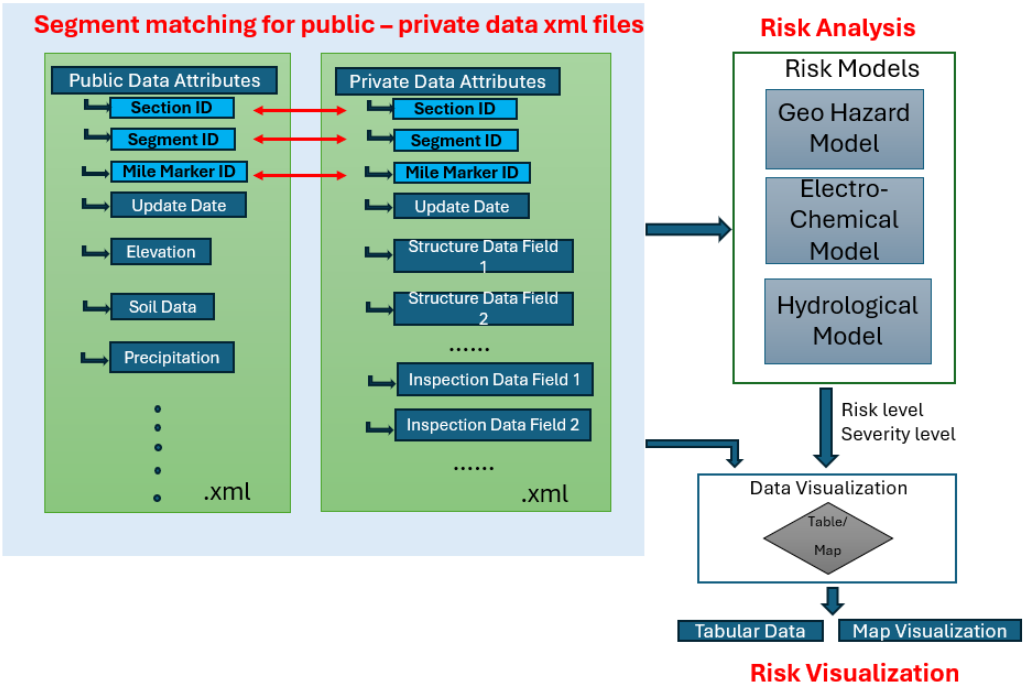

Figure 6: The Risk Model Integration with the public and private risk factors in Database architecture

<u>(iii) Data Processing & Visualization:</u>

Data from georeferenced indexes is fed into risk models, where it undergoes analysis to assess potential risks and vulnerabilities. The processed data is then visualized in two formats:

- Tabular Representation**:** Presents the data in a structured table format based on the user query, allowing for detailed value attributes of key parameters such as Section ID, Segment ID, Pipeline status, and risk level.
- Map-Based Visualization**:** Uses Geographic Information Systems (GIS) to provide spatial representation of the queried data, highlighting areas of concern, risk hotspots. This map-based visualization helps in better understanding of the geographic distribution of risks and aids in decision-making by providing a clear, interactive view of the pipeline infrastructure in relation to environmental and hazard factors.

Appendix C:

Initial version of the interface (python script) to public data repositories for data downloading

An initial version of the Python-based interface has been developed to streamline data acquisition from public data repositories. The script is designed to populate the database by retrieving relevant geospatial data corresponding to specific pipeline coordinates. The user inputs the pipeline coordinates, and the script outputs location-based data obtained from public repositories. The detailed workflow and decision logic of this implementation are illustrated in the flowchart shown in Figure 7. The data_source.csv contains information about the selected public data repositories, specifically Google Earth Engine (GEE) and the USGS Earthquake Repository. It specifies key details such as repository name, available data products, data formats, API endpoints or URLs for data retrieval, required parameters (e.g., earthquake magnitude, epicenter location), as well as spatial and temporal resolutions. Additionally, it includes instructions for preprocessing tasks, such as the reducers applied to GEE image collections (e.g., mean, first, sum), ensuring consistent data aggregation and facilitating seamless integration into the dynamic database.

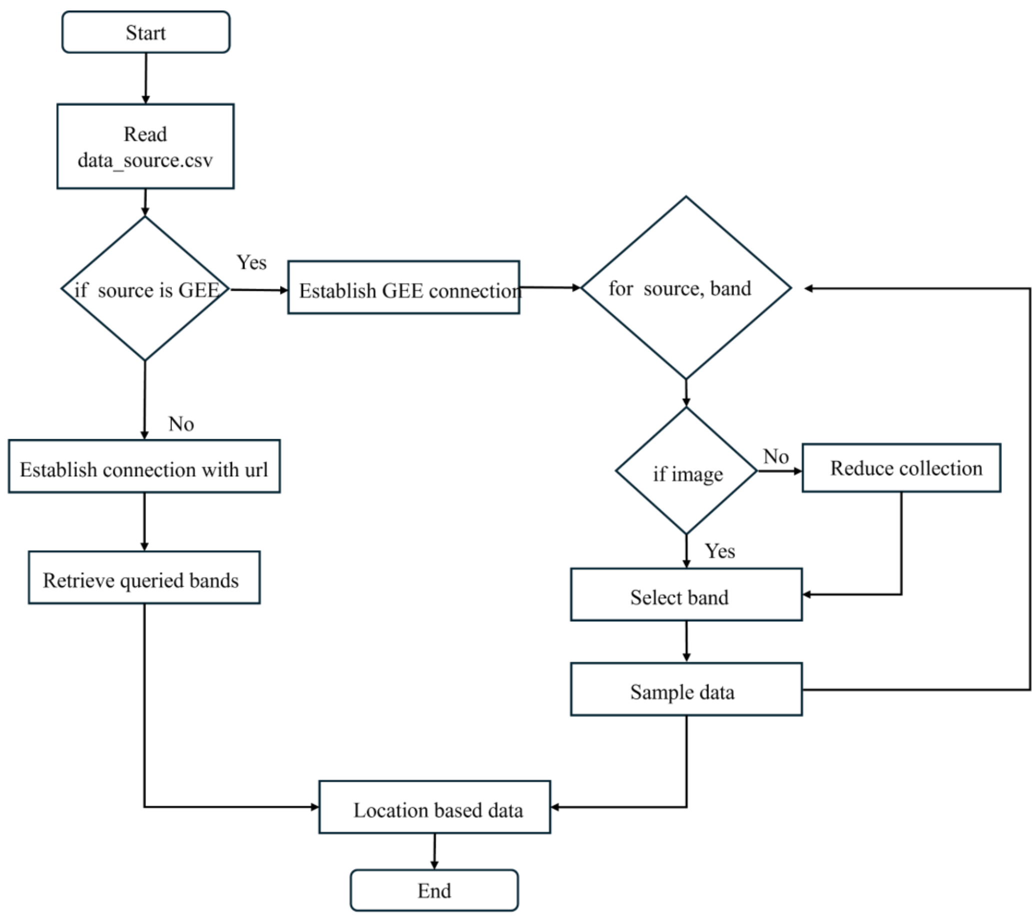

Figure 7: Flowchart of Python-based Interface for data acquisition from public repositories

Appendix D:

Scheduled University of Dayton Internal Meetings & Meeting Overview:

The detailed list of all meetings scheduled internally within the University of Dayton team, and summary of activities completed are mentioned in Table 4.

Table 4: The UDayton Internal Meetings scheduled during the current quarter period

Meeting Date	Agenda	Summary of Activities
CAAP Pipeline database progress meeting, University of Dayton **Weekly Thursday** **(10 AM – 11 AM)** [01/09/2025]	1) Discussed the plan for the next quarter and divided it into sub tasks and actions. 2) Schedule meetings with other universities for the geohazard and hydrological risk factors and model information.	The sub tasks are updated regularly and mainly divided as below: 1)Tabulate data related to hazard risk factors for the geohazard, electro-chemical and hydrological factors. 2) Downloading the data from Data sources. 3) Request access for NPMS private account 4) Check Existing Databases for Best Practices. 5) Statistical analysis of historical incident data 6) Evaluate Compatibility of Factors with Risk Models 7) Create Geographical Maps 8) Meetings with other teams
[01/16/2025]	1) Discussed and updated on the tabular format to categorize the hazard risk factors. 2) Discussed access issues to google earth engine account. 3) Discussed the literature review presentation for the quarterly meeting.	1) To check with the IT department at university of Dayton for the creation of the google earth engine project. 2) To update the "risk model name" column and share the hazard factor spreadsheet to other teams as discussed in the meeting.
[01/23/2025]	1)Discussion on Statistical analysis of historical pipeline incident data.	1) Checked and tested python scripts to run using google earth engine for the new project created. 2) Download and analyze the PHMSA incident data.

	2) Discussed the progress of the shared tabular format.	
[01/30/2025]	1)Discussion on how to get the sample US pipeline data to work on GIS visualization.	1) To organize the image data collected using python scripts and push it to github(geopipe). 2) To check the links[13][14] shared by Ben in the quarterly report meeting.
[02/06/2025]	1)Discussion on inputs received from the Texas A&M team for the hazard risk factors. 2) Discussion on the design of the database architecture.	1) Updated the shared tabular spreadsheet columns based on the discussion.
[02/13/2025]	1)Discussion on publicly available NPMS[15][16] in the links provided by Ben through mail. 2) Discussion on USGS, CDO websites for extracting data through python.	1) Check NPMS public viewer and identify at least three georisk pipeline sections. Ask for access for at least one section to test existing risk models 2)Check for missing information like historical inspection type (direct, indirect) and dates. 3) Credentials for extracting data through python for on USGS, CDO websites for
[02/20/2025]	1) Discussion on Format to display the GIS data in XML, GEOJSON files 2) To Run tests saving multiple data types (GEOJSON, XML, pandas data frame, CSV) and time test loading each data type. 3)Technical narrative of models from all the teams to each model	1) To Finalize the data type format on the GIS data. 2)To run a test saving multiple data types (GEOJSON, XML, pandas data frame, CSV) and time test loading each data type. 3)To focus on the existing 2-3 risk assessment models to implement on to display the severity level.

	mentioned in the spreadsheet document.	
[03/06/2025]	1)Discussion on the design of the database architecture. 2) Visualization structure to be made which mentions each level corresponding to the map	1) The different attributes of the georeferenced index are considered. 2) Each attribute like section ID, segment ID section is to be added to the database design as shown in Section 3.
[03/13/2025]	1)Discussion on the updates to be done to the design of the database architecture.	1)To add commodities, inspection authority in the GEOJSON index mapping layer. 2) To organize the design to the tree-based structure with the four layers.

References:

[1] Kazmi, D., Qasim, S., Harahap, I. S. H., & Vu, T. H. (2017). Analytical study of the causes of the major landslide of Bukit Antarabangsa in 2008 using fault tree analysis. *Innovative Infrastructure Solutions*, *2*(1), 55. https://doi.org/10.1007/s41062-017-0105-4.

[2] Hudson, K. S., Hudson, M. B., Hu, J., Harounian, A., & Lew, M. (2022). Quantifying Earthquake Hazards to Lifeline Systems at a Regional Scale with a Study of the Los Angeles Water System Pipeline Network. *Lifelines 2022*, 428–439. https://doi.org/10.1061/9780784484432.038.

[3]Ning, P., Jiang, Y., Tang, J., & Xie, Q. (2023). Research on the Early Warning Model for Pipelines Due to Landslide Geohazards under Multiple Influencing Factors. *Water*, *15*(4), Article 4. https://doi.org/10.3390/w15040693

[4]Koduru, S. (2019, May 26). *A Bayesian Network for Slope Geohazard Management of Buried Energy Pipelines*. https://doi.org/10.22725/ICASP13.444

[5] Mahmood, Y., Chen, J., Yodo, N., & Huang, Y. (2024). Optimizing natural gas pipeline risk assessment using hybrid fuzzy bayesian networks and expert elicitation for effective decision-making strategies. *Gas Science and Engineering*, *125*, 205283. https://doi.org/10.1016/j.jgsce.2024.205283

[6] Alvarado-Franco, J. P., Castro, D., Estrada, N., Caicedo, B., Sánchez-Silva, M., Camacho, L. A., & Muñoz, F. (2017). Quantitative-mechanistic model for assessing landslide probability and pipeline failure probability due to landslides. *Engineering Geology*, *222*, 212–224. https://doi.org/10.1016/j.enggeo.2017.04.005

[7] Vanitha, C. N., Easwaramoorthy, S. V., Krishna, S. A., & Cho, J. (2023). Efficient qualitative risk assessment of pipelines using relative risk score based on machine learning. *Scientific Reports*, *13*(1), 14918. https://doi.org/10.1038/s41598-023-38950-9

[8] Zarei, H., & Kalatpour, O. (2018). P-HAZOP: A New Extended HAZOP (Hazard and Operability) Study for Risk Analysis of Pipelines. *International Journal of Occupational Hygiene*, *10*(2), 108–113.

[9] Yazdi, S. R., Hasan, M. S., & Fadaei, R. (2022). A review of risk-based decision-making models for microbiologically influenced corrosion (MIC) in offshore pipelines. *Corrosion Science*, 192, 109854.

[10] Wasim, M., & Djukic, M. B. (2022). External corrosion of oil and gas pipelines: A review of failure mechanisms and predictive preventions. *Journal of Natural Gas Science and Engineering*, *100*, 104467. https://doi.org/10.1016/j.jngse.2022.104467.

[11] Shabarchin, M., & Tesfamariam, S. (2016). A Bayesian network-based approach for assessing internal corrosion risks in oil and gas pipelines. *Journal of Pipeline Engineering*, *15*(3), 137-149.

[12] Gartland, P. O., Johnsen, R., & Øvstetun, I. (2003). Application of internal corrosion modeling in the risk assessment of pipelines. In *NACE CORROSION 2003 Conference & Expo* (Paper No. 03359). NACE International

[13] https://www.phmsa.dot.gov/forms/gas-distribution-incident-report-form-f-7100-1

[14]https://www.phmsa.dot.gov/forms/gas-transmission-gathering-and-ungs-incident-report-form-f-71002

[15]https://www.npms.phmsa.dot.gov/Documents/NPMS_External_Pipeline_Attribute_Data_Dictionary.xlsx

[16]https://www.npms.phmsa.dot.gov/GeneralPublic.aspx