*Georgian Geographical Journal*

# Integrating GIS and Kernel Density Estimation for Multi-Hazard Risk Assessment of Potential Onshore Pipeline Failures

*Georgian Geographical Journal*, 2025, 5(3) 14-29

DOI:
https://journals.4science.ge/index.php/GGJ

Aslan E. Babakhanov[1], Zaur T. Imrani[1]*

[1] Department of Tourism and Recreational Geography, Ministry of Science and Education, Institute of Geography named after academician H.A. Aliyev, Baku, Azerbaijan
* Corresponding author: zaur_imrani@mail.ru

**Citation**: Babakhanov, A.E.; Imrani, Z.T. Integrating GIS and Kernel Density Estimation for Multi-Hazard Risk Assessment of Potential Onshore Pipeline Failures. *Georgian Geographical Journal* 2025, 5(3), 14-29. https://doi.org/10.52340/ggj.2025.05.03.02

## Abstract

This paper introduces a newly developed risk-assessment framework that combines Kernel Density Estimation (KDE) with Geographic Information System (GIS) technology to analyze multiple hazards, including earthquakes, floods, landslides, mud volcanoes, and soil erosion, for the Baku-Tbilisi-Ceyhan pipeline in Azerbaijan. By integrating hazard-specific parameters into a unified risk matrix, each hazard's contribution is weighted, refined, and aggregated to produce a spatially explicit, combined risk map. KDE smooths hazard intensities and reveals overlaps among different risk factors. The resulting high-resolution maps enable more targeted prevention and response measures, guiding planners and stakeholders toward effective pipeline protection strategies. Although the model can demand computational power, it remains scalable and flexible, allowing for adaptation to additional hazards or expanded geographical areas. Furthermore, the methodology underscores the importance of cross-validation in setting KDE bandwidth and in calibrating hazard weights to ensure reliable outputs. Preliminary testing indicates that this integrated model improves the clarity of risk data, highlights areas needing immediate attention, and supports resilience planning across the pipeline corridor. This work can be applied more broadly to critical infrastructure projects in regions where multiple hazards coincide, thereby aiding decision-making processes for disaster risk reduction and sustainable development. Future research will focus on refining statistical models for inter-hazard correlations and incorporating machine learning for predictive analytics. The framework stands as a tool to maintain pipeline integrity in the face of evolving environmental threats.

**Keywords:** GIS, Kernel Density Estimation, Risk Matrix, Pipeline Integrity, Multi-Hazard Risk.

## Introduction

Strengthening the durability of Azerbaijan's oil and gas infrastructure against natural disasters is a critical priority for researchers and industry stakeholders. Pipelines that transport hydrocarbons, while integral to national and transnational energy supply, are especially vulnerable due to the volatile nature of their contents. A failure can lead to irreparable environmental damage, human casualties, and costly socioeconomic disruptions (Lerche et al., 2014). The growing reliance on pipelines for interregional energy exchange, coupled with the apparent increase in natural disaster frequency, has underscored the need for comprehensive risk assessments (Krausmann et al, 2011). These assessments must extend beyond purely technical criteria to include geopolitical and environmental considerations, reflecting the international implications of pipeline failures and the cross-border nature of catastrophic events.

Climate change adds further complexity by intensifying extreme weather events and altering hazard patterns, placing critical infrastructure under unprecedented levels of stress (Zio, 2016). As natural disasters become more frequent and severe, safeguarding pipelines requires a flexible and forward-looking framework capable of integrating diverse data sources, advanced modeling, and predictive analytics. This forward-thinking perspective enables risk assessments that anticipate future scenarios rather than merely responding to current conditions.

Developing robust methodologies to quantify and mitigate these risks is thus an imperative, particularly in regions like Azerbaijan where seismic activity, floods, landslides, and mud volcanoes pose substantial threats (Othman et al., 2023). With pipelines traversing extensive distances across varied terrains, even localized disruptions can have extensive consequences. Consequently, the scientific community must invest in innovative tools and interdisciplinary strategies that address physical vulnerabilities, regulatory complexities, and cross-border interdependencies. Adopting a multi-hazard approach allows for a holistic understanding of pipeline resilience, thereby guiding more effective risk management, informed policy-making, and international collaboration.

Notably, specific natural disasters recognized in Azerbaijan can directly impact onshore pipelines, highlighting the urgency for targeted, data-driven solutions (Amirova-Mammadova, 2018; Bagirov et al., 2019):

1. ***Earthquakes*:** Azerbaijan experiences substantial seismic activity due to its location in an earthquake-prone zone (Alizadeh et al., 2017). Sudden ground movements or shifts can displace or fracture pipelines. Any rupture in a pipeline carrying hydrocarbons may cause leaks and, in severe instances, lead to explosions or fires.
2. ***Floods*:** Flash floods or prolonged heavy rainfall can result in rapid water accumulation and significant soil erosion[1]. High-velocity water flows may directly damage exposed pipeline segments, while erosion can undermine the structural support of buried pipelines, increasing the risk of bending or rupturing.
3. ***Landslides*:** In mountainous or hilly terrains, landslides can occur, particularly following intense precipitation or seismic events[2]. The downward movement of rock and soil can bury pipelines or exert forces beyond their design limits, leading to material failure.
4. ***Mud Volcanoes*:** Azerbaijan is well-known for its numerous mud volcanoes (Kadirov et al., 2005). These geological formations can erupt violently, expelling hot mud and gases capable of overheating, encasing, or fracturing adjacent pipelines, ultimately causing leaks or more severe accidents (Panahi, 2005).
5. ***Soil Erosion*:** Continuous erosion, often accelerated by flooding or heavy rainfall, can gradually expose pipelines designed to remain buried (Othman et al., 2023). Without adequate coverage, pipelines become vulnerable to external impacts and stress that can lead to cracks or breaks.
6. ***Corrosion (Environmental Degradation)*:** While not considered a conventional natural hazard, corrosion is a significant environmental threat to pipelines. Humid or saline conditions accelerate metal deterioration (Cheng, 2015), and if not detected, corrosion can thin pipeline walls, resulting in leaks or complete structural failure.
7. ***Extreme Weather Conditions*:** Severe weather events—such as powerful storms, lightning strikes, or extreme temperature fluctuations—can damage pipelines directly or trigger secondary hazards like landslides or floods (Katopodis et al., 2019). While pipelines are built to withstand local climate norms, unusually harsh conditions can exceed their tolerance thresholds.
8. ***Wildfires*:** High temperatures from wildfires can damage pipelines and their protective infrastructure (Novacheck et al., 2021). Moreover, the burning of nearby vegetation may contribute to soil erosion, further exposing pipelines to potential mechanical stresses.
9. ***Ground Subsidence*:** This phenomenon, caused by natural geological activity or human practices (e.g., mining, groundwater extraction), results in the gradual sinking of the ground surface (Oruji et al., 2022). Uneven subsidence can place bending stress on pipelines, increasing the likelihood of ruptures.

Typically, the practice is to represent the impact and risk values for each hazard individually on separate maps. A risk matrix serves as a visual aid, facilitating the evaluation of the cumulative risk level associated with different occurrences, including natural disasters. This matrix is instrumental in pinpointing the probability and potential impact of a given event, thereby allocating it a precise risk rating. However, the variety of natural disasters differs by location, resulting in a diverse set of risk maps. This diversity can introduce complexities in managing and interpreting the array of maps (Samany et al., 2022).

[1] Pipeline safety: Potential for damage to pipeline facilities caused by flooding, river scour, and river channel migration (2019). https://www.federalregister.gov/documents/2019/04/11/2019-07132/pipeline-safety-potential-for-damage-to-pipeline-facilities-caused-by-flooding-river-scour-and-river

[2] Guidelines for Constructing Natural Gas and Liquid Hydrocarbon Pipelines Through Areas Prone to Landslide and Subsidence Hazards (2009). https://rosap.ntl.bts.gov/view/dot/34640

Using a combined approach helps streamline this complexity by integrating multiple risk factors into a unified visual representation, simplifying analysis and decision-making (Falcone et al., 2022).

To derive a combined risk value encompassing all-natural disasters, one could consider aggregating the distinct risk values (Fotios et al., 2022) for each disaster. However, this method presupposes that all these incidents occur independently, a condition that might not always hold true. The overall risk assessment can be significantly swayed by the interrelations and mutual influences existing between various natural disasters. Recognizing these interdependencies is essential in accurately determining the combined risk value.

To maintain the integrity of all assessed risk values, the individual combined risk scores are connected to beforehand computed and proportionately weighted risk values for each type of natural disaster.

A critical initial phase involves quantifying the risk (Gemma et al., 2022) tied to each possible natural disaster. This could involve analyzing historical data on the frequency of natural disasters like earthquakes, floods, wildfires, and so on, within a targeted area to gauge their probability.

The core objective is to create a robust, multi-dimensional risk assessment model that not only identifies potential risks but also quantifies them in a meaningful and actionable manner (Rasouli & Imrani, 2023). To achieve this, we will employ a risk rating matrix grounded in rigorous mathematical formulations, enabling a detailed and nuanced understanding of risk levels. This matrix will integrate various risk factors listed above to generate a composite risk score for pipeline systems in the context of natural disasters. By combining the spatial analysis capabilities of GIS with the mathematical rigor of risk score calculations, this research endeavors to offer a novel and practical tool for policymakers, engineers, and disaster management professionals.

The ultimate goal is to enhance decision-making processes, facilitate proactive risk mitigation strategies, and contribute to the resilience and safety of pipeline infrastructures in the face of increasingly unpredictable natural events.

## Methods and Materials

In this research, we utilize adjusted risk matrices that correspond to each relevant hazard type - encompassing both classical natural disasters (e.g., earthquakes, floods) and environment-based threats (e.g., corrosion). Each hazard is weighted according to its relative impact on pipeline integrity. These weighted factors are then refined using Kernel Density Estimation (KDE) techniques, converting the collected risk scores into a 2D array (Gramacki, 2017). The data assigned to each cell within these matrices can be stored in any relational database (Nasser, 2018) or in straightforward file formats such as comma-separated values (CSV) or JSON. This flexibility enables seamless integration with popular GIS tools like ESRI ArcGIS or QGIS, where the 2D risk arrays can be visualized, analyzed, and used to inform decisions on pipeline protection and maintenance.

### *2.1 Area of observation*

This research area (Fig.1) situated within Azerbaijan, known for its susceptibility to various natural disasters such as earthquakes, landslides, floods, and mud volcanoes. This region is identified as the "valley of mud volcanoes" Moreover, satellite imagery revealed the presence of strong wind patterns and noticeable geological faults in the area.

### *2.2 Risk Matrix*

When analyzing the effects of natural disasters on pipeline systems, employing a risk matrix is an essential step for illustrating the degree of risk. This study utilizes a standard 5×5 risk matrix (Figure 2), which classifies threats according to their likelihood (ranging from rare to nearly certain) and potential consequences (spanning from negligible to disastrous) (Blokdyk, 2018). By adopting a five-tier system, each hazard can be systematically plotted on the grid, where higher probability and severity correspond to higher risk values.

Although summing the individual risk values for all-natural disasters can yield an overall, aggregated risk score, this approach assumes complete independence among events, which is not always accurate. Correlations and interdependencies between different types of natural disasters can significantly alter the true combined risk. Accordingly, such factors must be considered when determining an aggregated risk value.

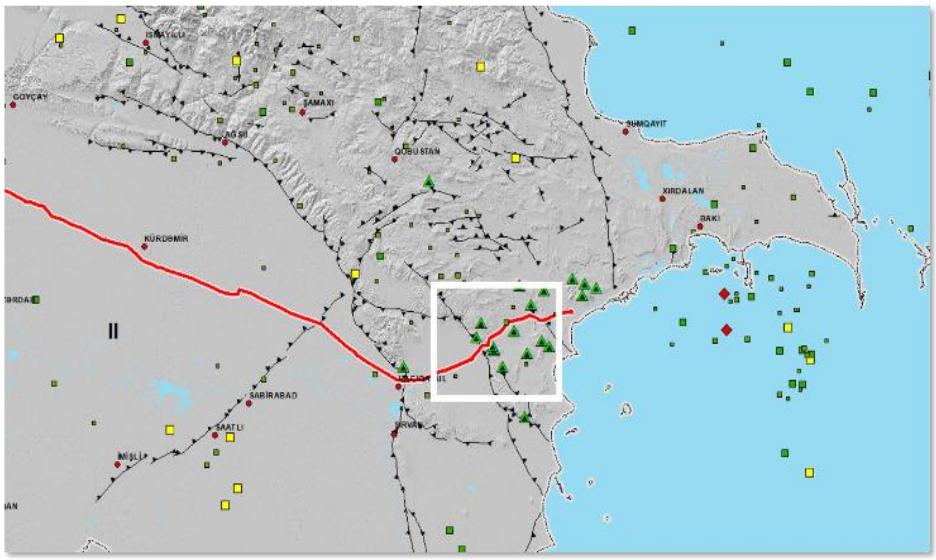


*Figure 1. Area of observation*

In spatial terms, each risk factor is mapped onto polygonal zones varying from as small as 50×50 meters to areas of several square kilometers (Han et al., 2010). These zones are categorized by risk level according to the matrix. Table 1 outlines a distribution framework for assigning integer or categorical values (e.g., "Low," "Moderate," "High") to each zone based on local hazard intensity. This classification is a cornerstone for subsequent calculations in a GIS environment, ensuring that risk values can be systematically integrated with geospatial data. As a result, risk maps visualize both individual and aggregated hazards, thereby aiding in pipeline safety assessments and informing targeted mitigation strategies.

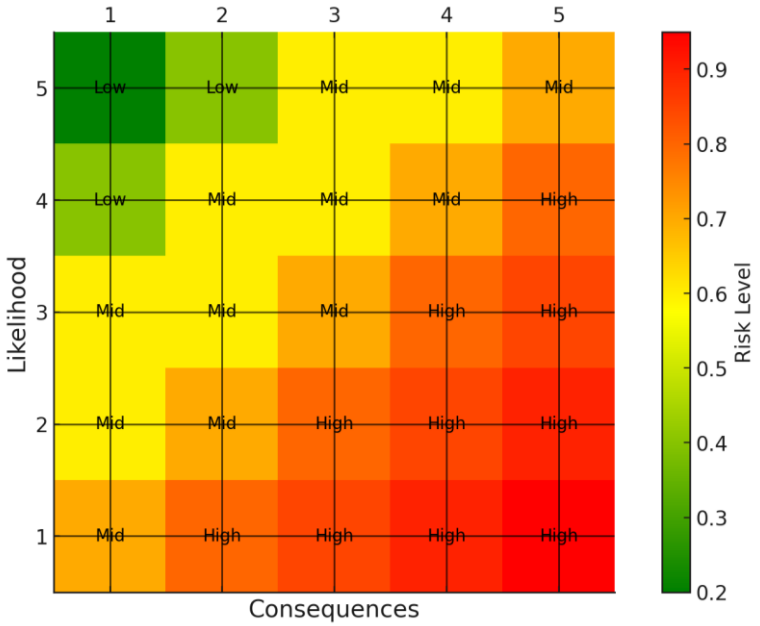


*Figure 2. Risk matrix of 5x5 elements*

*Table 1. Distribution of risk levels according to ND factor*

| Natural Disaster | Risk Level (1-5, 5 is highest) | Description | Color |
|---|---|---|---|
| Landslides | 4 | Soil shifts in hilly or mountainous regions that have the potential to disrupt or harm pipelines. | Yellow |
| Earthquakes | 5 | Seismic activity or ruptures leading to significant damages | Red |
| Flooding | 3 | Lead to erosion or the accumulation of sediment | Yellow |

| | | | |
|---|---|---|---|
| Corrosion from soil chemistry | 3 | Specific soil environments may extend the corrosion, resulting in the deterioration of pipeline materials | Yellow |

*2.3 Determining the Risk Values*

To address numerous hazard events, the contributions from each are consolidated into a unified risk score. We begin by computing the separate risk scores ($R_i$) for every hazard, following the formula:

$$R_i = L_i \times C_i$$

where $L_i$ represents the likelihood and $C_i$ denotes the impact of the $i^{th}$ hazard. After determining the individual risk scores for each hazard ($R_1, R_2, \ldots, R_n$), the average risk score ($R_{avg}$) is computed by the subsequent method:

$$R_{avg} = \frac{1}{n} \sum_{i=1}^{n} R_i$$

In this scenario, n represents the total count of hazards or events under evaluation. Averaging these values is one of several methods to consolidate the risk data, and it is particularly effective for the purposes of mapping. It's important to note that the risk scores will be visually differentiated using a color-coding system: scores ranging from 1 to 6 will be marked in Green, indicating low risk; scores between 7 and 12 will be highlighted in Yellow, signifying medium risk; and scores from 13 to 25 will be denoted in Red, representing high risk. Additionally, each specific value within these ranges could be represented by varying shades of the respective color, providing a gradient effect for more nuanced visualization Fig. 2.

Indeed, the method of calculating an aggregate risk score, as previously described, presumes that each hazard is independent and of equal importance, an assumption that might not hold true in more complex scenarios. To address this, our approach will integrate a weighting system, which allows for the differentiation in the significance of various hazards or the interdependencies between events. In this system, each hazard is assigned a weight ($\boldsymbol{w_i}$) reflecting its relative importance or impact. These weights are typically values ranging from 0 to 1, with the sum of all weights equaling 1. This ensures that the overall significance of all hazards is proportionately distributed. Following the assignment of weights, the next step involves calculating the weighted risk score ($R_{wi}$) for each hazard. This is achieved by:

$$R_{wi} = w_i \times R_i$$

To obtain the comprehensive risk score ($R_{total}$), the process involves summing all the weighted risk scores for each hazard:

$$R_{total} = \sum_{i=1}^{n} R_{\mathrm{wi}}$$

This total risk score is then categorized using the designated color-coding system based on its value:

$$Color(R_{total}) = \begin{cases} Green & if\ 1 \le R_{total} \le 6 \\ Yellow & if\ 7 \le R_{total} \le 12 \\ Red & if\ 13 \le R_{total} \le 25 \end{cases}$$

This weighted approach enhances realism by recognizing that certain hazards (e.g., high-magnitude earthquakes) may be far more consequential than others (e.g., minor flooding). However, neither the averaging nor the weighted-sum methods inherently capture potential correlations among hazards—such as how an earthquake might trigger a landslide or how floods may exacerbate soil erosion. Capturing these interdependencies typically requires more advanced statistical or probabilistic models (e.g., Copulas, Bayesian Networks, or multi-variate correlation matrices3).

Despite these limitations, weighted risk calculations represent a practical, GIS-friendly means of consolidating disparate hazards into a single metric. By overlaying the final risk values on geospatial layers, decision-makers can pinpoint pipeline segments requiring greater inspection, maintenance, or protective measures. This approach also aligns well with advanced techniques, such as Kernel Density Estimation, which can further smooth risk values spatially and highlight high-risk clusters within the study region, supporting more robust and proactive risk management.

*2. 4 Evaluation of 2D risk scores using Gauss kernel method*

The 2D risk score for natural disasters can be evaluated using geospatial parameters, such as the magnitude of previous events (for earthquakes, landslides, etc.), proximity to fault lines, soil type, slope,

[3] Multivariate correlations, https://numericalexpert.com/tutorials/statistics/multivarcorr.php

and rainfall data (for floods and soil erosion). The method also assumes that each type of disaster has its own unique contributing factors. A Gauss kernel function4 is often used in spatial data analysis because it is smooth, symmetric, and its value decreases with distance together with Kernel Density Estimation (KDE).

KDE is a non-parametric technique for estimating the probability density function of a random variable. When applied to risk levels, it allows you to smooth the risk measurements and find areas of high and low risk density. The equation (Equation 1.2) provides a smoothed estimate of the risk level distribution. From this, you could estimate probabilities of specific risk levels, find modes of risk (most common risk levels), or perform other analyses based on selection of kernel function for the KDE (Equation 1.1).

The Gaussian kernel is a common choice due to its smoothness and nice mathematical properties.

K(u) is the Gaussian kernel function, which decreases with the square of the distance from the center of the kernel (u), and f(x) is the kernel density estimate, which is an average of the kernel functions centered at each data point (Xi), with the bandwidth h which controls the amount of smoothing: a large h leads to more smoothing and a smaller h leads to less smoothing.

To apply this to measured risk levels, the data points (Xi) would be the measured risk levels, and the estimated density f(x) would give an estimate of the probability density of risk levels at each point in the risk range.

Equation 1.1: The equation for a Gauss kernel

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$$

Equation 1.2: KDE equation

$$f(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

After collecting the data on historical ND, we can then calculate the risk scores per disaster category using minimal set of crucial input parameters:

1. Earthquakes:
   Variables:
   - Magnitude (M) of the past earthquakes.
   - Distance (d) to the epicenter of past earthquakes.

   The risk score for earthquakes can be calculated as follows:

$$R(x, y) = \sum_i M_i \cdot e^{-\frac{d_i^2}{2\sigma^2}}$$

2. Landslides:
   Variables:
   - Slope of the terrain (**S**).
   - Soil moisture content (M).
   - Distance (d) to previous landslides.

   The risk score for landslides can be calculated as follows:

$$R(x, y) = \sum_i S_i \cdot M_i \cdot e^{-\frac{d_i^2}{2\sigma^2}}$$

3. Flooding:
   Variables:
   - Precipitation (P) in the area.
   - Distance (d) to water bodies such as rivers or lakes.

   The risk score for flooding can be calculated as follows:

$$R(x, y) = \sum_i P_i \cdot e^{-\frac{d_i^2}{2\sigma^2}}$$

In these equations, σ is the standard deviation, controlling the spread of the Gaussian kernel, (x, y) are the coordinates for a specific grid cell in the GIS maps, R(x, y) is the calculated risk score at point (x,

[4] The Kernel Cookbook, https://www.cs.toronto.edu/~duvenaud/cookbook/

y), and the summation i is over all the relevant events or factors within a certain radius around point (x, y).

*2.5 Applying the minimal safe distances*

The minimal safe distance to the object (here the minimal distance to the linear part of pipeline) is a starting point to identify the way we calculate the offset for risk values. Minimal distance is a geometrical distance, a straight line. It varies from disaster factors, so let's integrate it (Dmin) into our equations:

1. Earthquakes:
   Variables:
   - Magnitude (M) of the past earthquakes.
   - Distance (d) to the epicenter of past earthquakes.
   - Minimal safe distance to the epicenter of earthquakes (Dmin).
   The risk score for earthquakes can be calculated as follows:

$$\mathrm{E}(x,y) = \sum_i M_i \cdot e^{-\frac{(d_i - D_{min})^2}{2\sigma^2}} \cdot H(d_i - D_{min})$$

Here, H (di - Dmin) is the Heaviside step function. It is equal to 0 for di < Dmin (indicating no risk inside the safe distance), and 1 for di >= Dmin.

2. Landslides:

Variables:
- Slope of the terrain (S).
- Soil moisture content (M).
- Distance (d) to previous landslides.
- Minimal safe distance to previous landslides ($D_{min}$).
  The risk score for landslides can be calculated as follows:

$$L(x,y) = \sum_i S_i \cdot M_i \cdot e^{-\frac{(d_i - D_{min})^2}{2\sigma^2}} \cdot H(d_i - D_{min})$$

3. Flooding:
   Variables:
   - Precipitation (P) in the area.
   - Distance (d) to water bodies such as rivers or lakes.
   - Minimal safe distance to water bodies ($D_{min}$).
   The risk score for flooding can be calculated as follows:

$$F(x,y) = \sum_i P_i \cdot e^{-\frac{(d_i - D_{min})^2}{2\sigma^2}} \cdot H(d_i - D_{min})$$

These equations will yield risk scores that are higher for locations closer to the dangerous object (beyond Dmin), and zero for locations within the safe distance. The more sophisticated approach is considering a gradual decrease of risk within the safe distance, instead of a sudden drop to zero (Section 2.5).

The results of evaluated risk scores (RS) of a single contributing factor for each type of disaster (e.g., a single earthquake, a single previous landslide, a single water body) using above equations can be found on Table 2:

*Table 2. Risk scores per factor, 500 meters, 5 steps*

| Distance (m) | Earthquake RS (M=5, σ=500) | Landslide RS (S=2, M=0.5, σ=500) | Flood RS (P=100, σ=500) |
|---|---|---|---|
| 200 | 0.891 | 0.577 | 19.2 |
| 700 | 0.706 | 0.353 | 14.12 |
| 1200 | 0.367 | 0.184 | 7.34 |
| 1700 | 0.135 | 0.067 | 2.7 |
| 2200 | 0.033 | 0.017 | 0.66 |

Here, M is the earthquake magnitude, S is the terrain slope, and M is the soil moisture content for landslides (not to be confused with M for earthquakes). P is the precipitation level. σ is the spread of the Gauss kernel, which is the same for all three types of disasters.

The risk scores were calculated using the updated equations from the "Section 2.4", with Dmin equal to 200 meters. Note that for the distance of 200 meters, the risk scores are 0 because of the safe distance factor. The risk scores for distances beyond Dmin were calculated by substituting the given values into the equations. The calculations assume a specific value for each parameter, and the actual risk scores may vary greatly depending on these values.

The total risk function R(x, y) at location (x, y) would be the sum of these three functions:

$$R(x, y) = E(x, y) + L(x, y) + F(x, y)$$

The function R(x, y) will give us a single risk score for each location (x, y) that we can visualize on a 2D map. We can calculate this function for each cell in a 5×5 grid, and then color each cell based on its risk score to create a 2D risk map. This is used as a model for single cell area and actual risk maps would probably use a more sophisticated model and a much larger grid.

*2.6 Splitting and scaling the grid*

In multi-hazard GIS-based risk analyses, subdividing the area of interest from larger cells (Δx, Δy) into smaller ones can greatly enhance the precision and interpretability of results. Each cell $Gi, j$ represents a discrete spatial unit for modeling hazard intensity, vulnerability, and other relevant parameters. Adopting a finer cell size $(\Delta x_{\text{small}}, \Delta y_{\text{small}})$ has several advantages:

*Increased Spatial Resolution.*

Smaller grid cells capture more localized variations in elevation, land cover, infrastructure density, and other critical attributes. In mathematical terms, the number of cells $N$ in a given region of area $A$ scales approximately as $N \propto \frac{A}{\Delta x \times \Delta y}$. Hence, decreasing cell size yields a larger $N$, facilitating higher-resolution risk modeling.

*Improved Risk Assessment.*

By refining each cell's spatial dimensions, the risk matrix (or any comparable index) can more accurately depict local hazards. For instance, in a flood risk scenario, slight changes in elevation or land cover within a 2×2 km cell may significantly affect water flow and flood extent. A coarser 10×10 km cell $(\Delta x_{\text{large}}, \Delta y_{\text{large}})$ would mask these localized variations, potentially underestimating or overestimating the actual risk.

*Localized Analysis.*

Disasters such as landslides or urban flash floods often affect small areas with disproportionately high severity. A finer grid enables targeted analysis of these hotspots, capturing the nuances of topography and land use that can influence the severity of impacts. This capability is crucial for large-scale disaster management strategies where local conditions can markedly change risk levels.

*Detailed Mitigation Planning.*

High-resolution grids reveal specific locations most prone to damage, supporting more efficient mitigation measures. Authorities can use these refined data layers to plan structural defenses (e.g., levees or retaining walls), allocate evacuation routes, and prioritize response resources, thereby reducing both immediate and long-term risk.

*Enhanced Accuracy of Predictive Models.*

Many computational models in hydrology, seismology, and other hazard-related fields use grid-based inputs. Smaller cell sizes often improve the fidelity of simulations, though at the cost of greater computational demand. By employing a grid of fine resolution, models can represent spatial heterogeneity more accurately, thereby yielding more reliable predictions of hazard behavior.

However, adopting smaller cells increases the computational burden, as the total number of cells N and the associated data complexity grow. Additionally, finer spatial resolution demands correspondingly detailed data inputs (e.g., higher-accuracy digital elevation models, land use surveys). In practical applications, the grid cell size is often a compromise between accuracy and available computational or data resources.

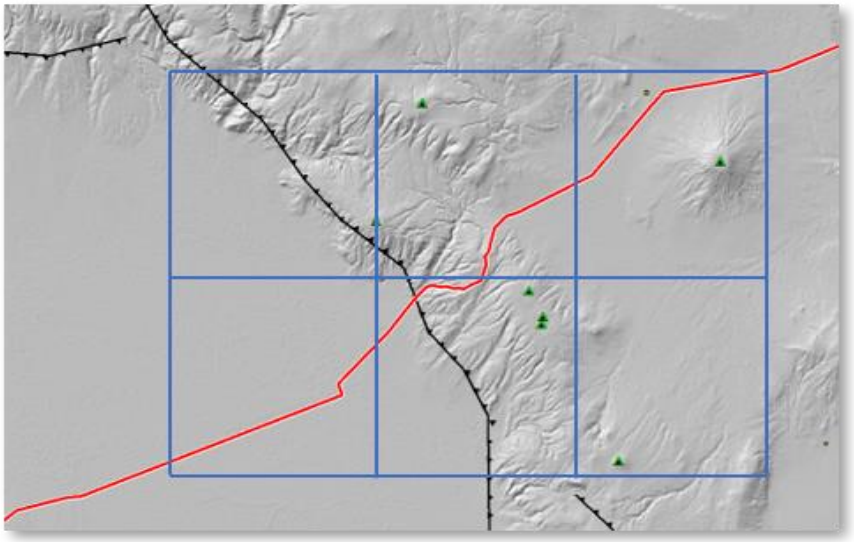

*Figure 3. The area of interest with 3x2 cells, each 10x10 km*

For example, Figure 3 may represent a coarser grid where each cell is approximately 10×10 km, capturing large-scale trends but providing only a coarse view of localized risk. Conversely, Figure 4 might illustrate a finer grid (2×2 km cells), capturing nuanced variations yet increasing both data density and processing time. Calibrating grid size becomes essential for balancing computational feasibility with the desired level of spatial detail. In many multi-hazard pipeline risk assessments, a moderate cell size is often chosen initially, followed by targeted refinement in critical zones where hazards overlap or infrastructure vulnerability is high. This approach ensures that risk maps reflect realistic spatial gradients without overwhelming computational or data-storage capacities.

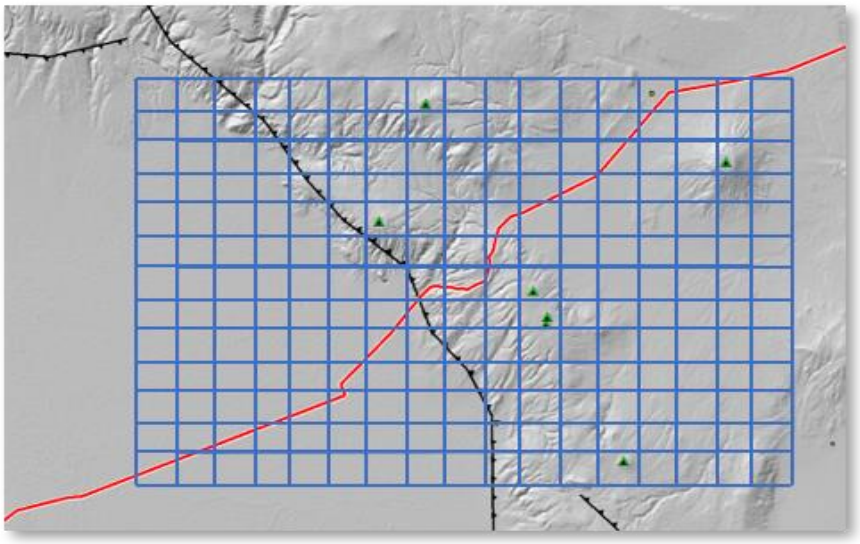

*Figure 4. The area of interest with 17x12 cells, each 2x2 km*

## 2.7 *Smoothing risk scores by categories*

When plotting discrete or granular risk values directly onto a map—especially when each value is subdivided into finer sub-scores—visual representations can appear patchy or irregular. To obtain a smoother, more continuous distribution of risk values across a two-dimensional space, Kernel Density Estimation (KDE) offers a robust solution. Under this approach, probability density functions are approximated for each designated risk category (e.g., Green, Yellow, Red), focusing attention on smaller, localized zones that require further scrutiny.

In this study, risk values lying outside the specified range of a given category are effectively excluded (or treated as zero) during computation. For instance, when constructing KDE surfaces for Green-category data, only points whose risk scores lie within the “Green” interval are included. Similarly,

Yellow and Red categories are handled in separate density estimations, thus preventing extraneous data points from skewing results.

Mathematically, for the c-th category, where $n_c$ data points $\left(x_i^{(c)}, y_i^{(c)}\right)$ satisfy that category's risk range, the two-dimensional Gaussian KDE can be written as:

Equation 2: 2D risk score using Gauss kernel

$$\hat{f}_c(x,y) = \frac{1}{n_c h^2} \sum_{i=1}^{n_c} \frac{1}{2\pi} e^{-\frac{1}{2}\left[\left(\frac{x - x_i^{(c)}}{h}\right)^2 + \left(\frac{y - y_i^{(c)}}{h}\right)^2\right]}$$

Here, categorized risk levels per impact, where $(n_c)$ is the number of data points in the (i)th category, (h) is the bandwidth parameter, (e) is the base of the natural logarithm and $x_i^{(c)}$ and $y_i^{(c)}$ are the nth pair of 2D data points.

This formulation assumes X and Y (the spatial coordinates) are independent variables in the Gaussian kernel, allowing for straightforward, radially symmetric smoothing. The choice of the bandwidth h is pivotal: smaller h values yield highly localized densities at the expense of potential overfitting, whereas larger h values give smoother distributions that may obscure important local variations. Bandwidth selection often leverages cross-validation or rule-of-thumb heuristics (Silverman, 2018).

By performing this KDE procedure separately for each risk category, the method generates multiple "layered" density surfaces that can then be visualized independently or composited. The resulting smoothed risk surfaces help highlight areas of greatest concern within each category and avoid the abrupt color transitions that simple, cell-by-cell risk mapping might produce. Consequently, decision-makers and analysts gain a clearer view of spatial risk concentrations, which supports more informed planning and more targeted mitigation strategies.

### *2.8 Data Processing*

To streamline the risk analysis within a specified Area of Interest (AIO), a systematic methodology was devised, as summarized in Table 3. The determination of the AIO's minimum size - which was initially set at 1 km² - drew on a preliminary review of notable natural disaster events within Azerbaijan, ensuring that smaller-scale events would be captured. However, the AIO dimension is not fixed and may be recalibrated for different regions or urban areas where hazard intensity or infrastructure density warrants finer resolution.

Adapting AIO dimensions typically involves additional analytical steps. First, detailed hazard datasets (e.g., seismic zonation maps, flood extents, or landslide susceptibility layers) are reviewed to confirm the spatial extent and granularity of the events. Second, the potential impact radius of each hazard is considered - particularly if one hazard (e.g., flooding) tends to spread over a wider area compared to another (e.g., localized ground subsidence). Finally, socioeconomic or administrative factors, such as population centers or critical infrastructure corridors, may also inform the selection of a larger or smaller AIO.

Within each AIO, risk levels are determined by referencing previously established matrices for natural hazards (earthquakes, floods, landslides, mud volcanoes, etc.), environment-based hazards (e.g., corrosion), and their associated weights. The resulting risk scores are then aggregated spatially, allowing analysts to pinpoint critical hotspots. By adhering to the structured workflow shown in Table 3, one ensures uniformity across different AIOs, thereby facilitating consistent comparisons - even when hazard profiles differ.

Through this stepwise approach, researchers and decision-makers can maintain a clear audit trail of how risk scores are derived, updated, and mapped onto each AIO. This enables rapid re-evaluation if new hazard data emerge - such as revised seismic models or recent flood records -and fosters an adaptable framework suitable for a variety of planning or operational needs.

*Table 3. Processing Steps for a Single AIO*

| Step | Description |
|---|---|
| 1 | Determine the AIO: The AIO should be centered around the pipeline to ensure accurate and proportionate GIS visualization of the event area and the pipeline's linear section |
| 2 | Identify (ND) and Adjust AIO Size: Identify NDs that could impact the AIO. The AIO's dimensions may be modified if an ND has a wider impact area. Concurrently, determine the risk parameters for each event and their relative weights concerning the pipeline. |
| 3 | Calculate Risk Matrices per Event: Develop risk matrices for each identified event, considering the specific risk parameters and weights. |

| | |
|---|---|
| 4 | Combine Risk Scores: Combine the individual risk scores from each event into a comprehensive risk score matrix. |

This methodical approach ensures a thorough and systematic assessment of risks within a given AIO, allowing for a nuanced understanding of potential impacts on pipeline infrastructure.

### *2.9 Aggregated Risks*

While the aggregated risk score provides a high-level view of the overall risk, it is still tied to the individual risk scores, and these can be used to understand the specific hazards in more detail. It's beneficial to consider both the aggregated and individual risk scores when planning interventions and communicating about risk. In terms of visualization of risk details in an aggregated form, the individual risk being stored as an entity within the relational database.

Many GIS frameworks providing the automation facilities (here ArcMap, QGIS and etc.) via scripting languages like Python, JavaScript. So, by adding the referential output per aggregated risk value (here individual colored cell) we can get the detailed information on certain cell of interest in a form of table 4 or other ways, like popup tooltips.

*Table 4: Aggregated risk score for each location*

| Location | Earth-quake Risk | Land-slide Risk | Flood Risk | Mudflow Risk | Aggregated Risk |
|---|---|---|---|---|---|
| A | 5.6 | 4.2 | 7.3 | 6.4 | 23.5 |
| B | 7.1 | 2.3 | 8.5 | 5.6 | 23.5 |
| C | 3.4 | 6.7 | 6.2 | 4.5 | 20.8 |
| D | 6.9 | 4.8 | 8.1 | 5.3 | 25.1 |
| E | 4.3 | 3.5 | 7.8 | 6.7 | 22.3 |

This table clearly shows both the individual and aggregated risk scores. By comparing the aggregated scores, we can see that location D has the highest overall risk, even though it may not have the highest risk for any individual disaster category. This is a good example of how aggregated risk scores can provide a different perspective than individual risk scores.

### *2.10 KDE and risk colors*

The KDE method can be used to generate the smoothed, continuous distribution of the aggregated risk scores.

In this context, it would provide a way to understand the distribution of aggregated risk scores across the area of interest, identifying areas of high or low risk density. This could help to inform more granular risk management strategies or interventions, by focusing efforts where the density of high-risk locations is greatest.

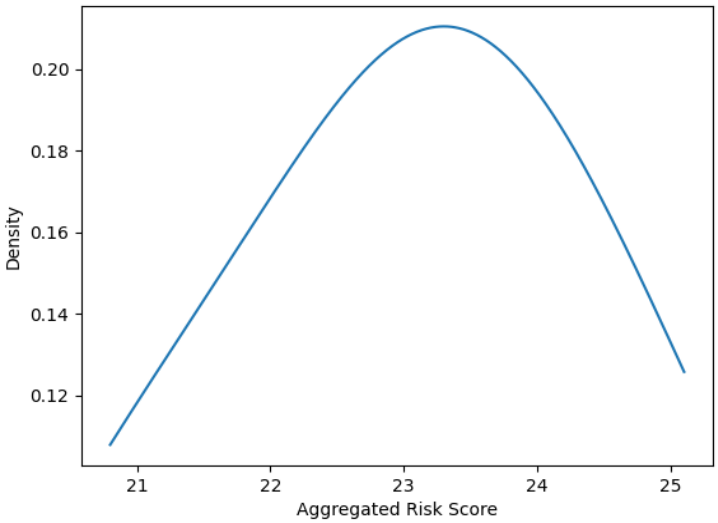


*Figure 5. Histogram of Gaussian aggregated risk score*

The Fig. 5 is a smoothed histogram (i.e., a probability density function) of the aggregated risk scores using a Gaussian kernel. The resulting plot shows the distribution of risk scores, with the y-axis representing the estimated density of each risk score.

This can help us to visualize how the risk scores are distributed and could potentially inform how we define the ranges for the 'Green', 'Yellow', and 'Red' risk rankings.

Then, the distribution of colours on the histogram be colorized based on colour ranges per risk score. On the example of given aggregated values from the table "Table 4", the risk ranking fall into 'Red' coloured risk score and be smoothed as a gradient of 'Red' colour. Following the calculation of the probability density function using the Gaussian kernel density estimation (KDE) technique, we visualize the distribution by creating a color-coded histogram plot (Fig. 6). Those gradients then be used to map smoothed aggregated risk values on final GIS maps.

This results in a spatial map where each location is coloured based on its normalized risk density, as estimated by the KDE. Darker colours indicate regions with higher risk densities.

In practical applications, this spatial risk map can be overlaid with other geographical features such as roads, buildings, and natural landmarks to provide a comprehensive view of the risk landscape. This can aid in identifying high-risk areas and prioritizing interventions.

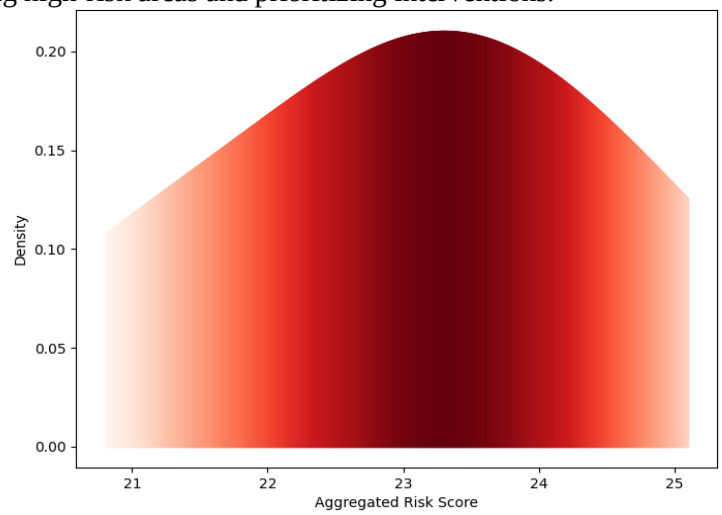


*Figure 6. Gradient histogram of Gaussian aggregated risk score*

### *2.11 Assessment challenges*

In the process of assessing risks to pipeline infrastructure due to natural disasters, several challenges arise, particularly in accurately characterizing the impact of these events. A key aspect of this challenge is the need for a comprehensive definition of risk, which encompasses a clear understanding of the probability and dimensions of potential impacts. The challenges, listed in order of increasing complexity, include:

1. Identification and Evaluation of Risk Parameters: This involves pinpointing specific risk factors and assessing their potential impact. An example is evaluating the effects of earthquakes on pipelines, including subsequent NATECH events (Krausman et al., 2016).
2. Determining Appropriate Risk Weights for Each Parameter: Assigning the correct weight to each risk parameter is crucial for accurate risk assessment.
3. Choosing the Correct Granularity for the Area of Interest: This is essential to avoid oversimplification and ensure detailed risk analysis.

### *2.12 General steps for risk visualization*

To visualize the spatial distribution of aggregated risk scores using the colorized histogram, we can employ a GIS framework. This allows us to map the risk scores onto a 2D map representation of the study area.

a. ***Define the study area***: Determine the extent and boundaries of the study area where the risk scores will be visualized.
b. ***Obtain the risk scores***: Collect or calculate the aggregated risk scores for each location within the study area.
c. ***Create a grid***: Divide the study area into a regular grid of cells. The size of each cell depends on the desired resolution for the visualization.

d. ***Interpolate the risk scores***: Assign each cell in the grid a risk score based on interpolation techniques such as inverse distance weighting or kriging. This step ensures that risk scores are assigned to locations that fall between the actual data points.
e. ***Colorize the grid***: Assign a color to each cell based on its risk score. Use a color scale or colormap that represents the risk levels of interest. For example, a colormap with shades of red can be used, where lighter shades represent lower risk and darker shades represent higher risk.
f. ***Generate the risk map***: Plot the colored grid onto the GIS map, overlaying it on relevant base layers such as roads, topography, or satellite imagery. This provides spatial context and aids in the interpretation of the risk distribution.
g. ***Include a legend***: Create a legend that clearly explains the color scale and risk levels associated with each color. This helps viewers interpret the risk map accurately.
h. ***Optional***: Add additional features: To enhance the visualization, you can incorporate other geographical features such as administrative boundaries, water bodies, or infrastructure that may be relevant to the risk assessment.

By following these steps, the colorized histogram-based visualization of aggregated risk scores can be extended to a GIS environment, allowing for a comprehensive and visually appealing representation of spatial risk distribution.

**Results**

The outcome of this study includes the successful implementation of GIS tools to create detailed visualizations of aggregated risk. A significant enhancement in our visualization technique was the incorporation of buffered zones along the pipeline route, which provided a more distinct delineation of risk areas (Petersen, 2020).

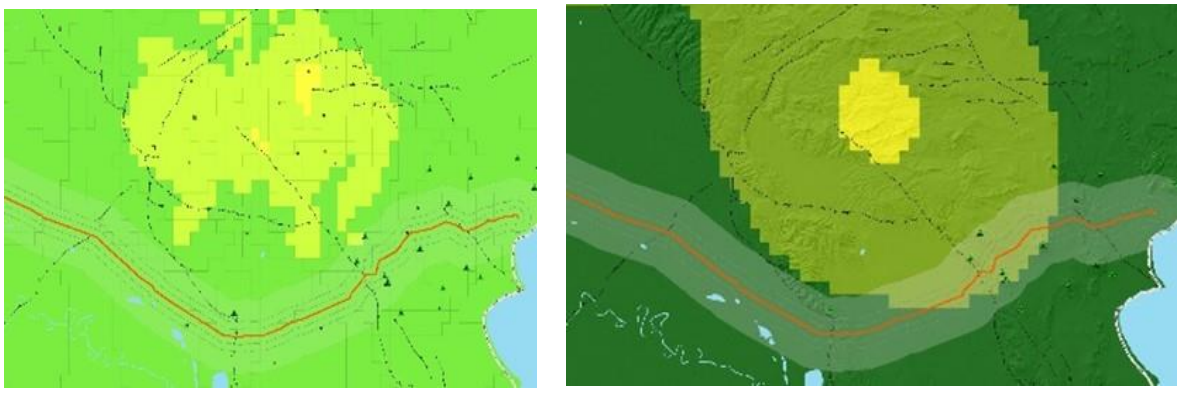

*Figure 7-8. Coarse, CRS (left), KDE based, refined CRS (right)*

The series of figures, from Fig. 7 to Fig. 10, illustrate the progression of refined combined risk scores (CRS) derived from various natural hazards within a relatively small region. The smallest unit of analysis, the AIO, covered approximately 2 km², while the entire region under assessment spanned about 100 km². The risk matrices developed were informed by a combination of data on recorded earthquakes, identified geological faults, and mud volcanoes.

The tools utilized for these visualizations were ESRI ArcMap 10.8 for mapping, output files generated (CSV files) with Python 3.8.

**Discussions**

In this study, we have navigated through the multifaceted challenges of assessing the risk to pipeline infrastructure from natural disasters. Our approach has been to meticulously identify and evaluate risk parameters, determine appropriate risk weights, and select the optimal granularity for the area of interest.

Implications of Findings: Our findings underscore the complexity inherent in assessing natural disaster risks to pipelines. The use of a three-dimensional risk assessment plot, as demonstrated in Figure 7 for earthquake impacts, provides a nuanced visualization of risk. This method allows for a more detailed understanding of how different risk parameters interact and affect the overall risk profile.

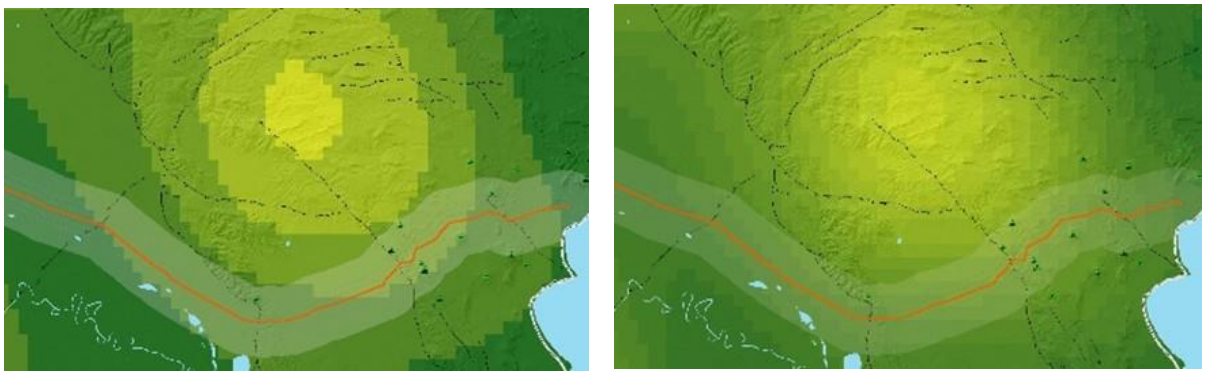

*Figure 9-10. KDE based, smaller scale CRS (left), KDE based, fine-tuned CRS (right)*

The color-coded risk categories further enhance the interpretability of these risks, making the data more accessible for decision-makers.

The application of this methodology to other natural disasters reveals its versatility. However, it also highlights the increasing complexity when more parameters are involved. Each additional parameter requires careful consideration and integration into the overall model, underscoring the need for a robust and flexible risk assessment framework.

Challenges and Limitations: One of the primary challenges encountered in this study is the sensitivity of the KDE and color mapping to the choice of parameters, especially the bandwidth h. The selection of h is critical and must be tailored to the specific characteristics of the data and the objectives of the risk assessment. This sensitivity points to a broader challenge in risk assessment: the balance between accuracy and generalizability.

Furthermore, our approach assumes the independence of risk factors, which may not always hold true in real-world scenarios.

Future Directions: Looking ahead, there are several avenues for further research. First, exploring methods to incorporate the interdependencies between different risk factors would likely yield a more comprehensive and realistic risk assessment model. This could involve the development of more sophisticated statistical models or machine learning algorithms capable of handling complex, interrelated datasets.

Second, the adaptation of our model to a wider range of natural disasters would be beneficial. Each type of disaster brings its unique set of challenges and risk factors, and a versatile, adaptable model is essential for broad applicability.

Lastly, engaging with experts in geology, meteorology, and disaster management could provide valuable insights and data, enriching the model's accuracy and relevance.

## Conclusion

This study has illuminated several key challenges in determining risk weights, which are crucial for accurate risk assessment in the context of ND and their impact on infrastructure. These challenges are summarized below:

1. ***Complexity of Risk Interactions***: ND often triggers or coincides with other events, complicating the assessment of combined risk impacts. Example: An earthquake increasing the likelihood of a landslide.
2. ***Data Scarcity***: The lack of sufficient data for rare or unprecedented risks introduces significant uncertainty in estimating probabilities and impacts.
3. ***Risk of Oversimplification***: Simplifying risk assessments can lead to underestimating the complexity and interactions of risks, potentially skewing the actual risk levels.
4. ***Model Dependence***: The accuracy of risk weights is directly linked to the efficacy of the risk assessment model. Inaccurate models can result in misleading risk weights.
5. ***Subjectivity in Risk Weighting***: Risk assessments can be influenced by individual or organizational risk tolerance, which varies widely. Balanced risk weighting often requires collective input from professionals or subject matter experts.
6. ***Evolving Risk Landscape***: Risks change over time due to environmental, technological, regulatory, and other factors, necessitating regular updates to risk weights.
7. ***Perception Bias***: Decision-makers' biases in risk perception can lead to overestimating or underestimating certain risks, influenced by factors like visibility, recency, or familiarity.

8. ***Quantifying Qualitative Risks***: Some risks are qualitative and difficult to quantify, such as reputational risks or the impact of regulatory changes, posing challenges in numerical risk weighting.

*Advantages*

The utilization of aggregated risk scores in assessing various types of risks offers several advantages, as outlined below:

1. ***Simplicity in Visualization and Interpretation***: Aggregated risk scores streamline the process of understanding risks by condensing multiple risk factors into a single, comprehensive score. This simplification aids in easier interpretation and visualization.
2. ***Enhanced Comparability***: These scores enable straightforward comparisons of overall risk across different locations or regions, which is invaluable for decision-makers in resource allocation and prioritization.
3. ***Holistic Risk Understanding***: Aggregated scores offer a broad overview of risk, capturing the cumulative effect of individual risks, especially when they interact or compound.
4. ***Effective Communication***: Communicating a single aggregated risk score is often more straightforward and accessible to stakeholders and the public than explaining multiple individual risk components.

Furthermore, the aggregated risk score can be tailored to align with the risk tolerance or appetite of the concerned entity. This adaptability allows for a more relevant and targeted risk assessment.

Using a risk matrix for aggregation effectively consolidates risks from various natural disasters into a single value. This comprehensive view is crucial for effective risk management, disaster preparedness, and resource allocation. However, it's essential to periodically update the risk matrix to reflect changes in the risk landscape due to factors like climate change, urban development, and population growth.

It's also important to recognize that the aggregated risk score is an indicative tool meant to guide, but not solely dictate, decision-making processes. Factors such as societal values, financial constraints, and political considerations should also be factored into risk management strategies.

Despite the challenges, our study successfully visualized risk in an aggregated form, using a single map with finely tuned risk scores. Regular review and updating of risk weights, expert involvement, and careful consideration of risk interactions are key components of this approach.

One notable finding is the model's applicability to other risk assessments using the same matrix representation and weight calculations. The flexibility of the model is further enhanced by the potential integration of the Kernel Density Estimation (KDE) method for fine-tuning outputs and the implementation of algorithms like the Moore neighborhood (Ilachinski et al., 2001) for relative AIO region risk score determination.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contribution

All authors provided critical feedback and helped shape the research, analysis and manuscript.

## ORCID iD

*Aslan E. Babakhanov* https://orcid.org/0000-0002-8790-7945
*Zaur T. Imrani* https://orcid.org/0000-0002-0606-3753

## Reference

Alizadeh, A.A., Guliyev, I.S., Kadirov, F.A., Eppelbaum L.V. (2017). Geosciences of Azerbaijan - Volume II: Economic Geology and Applied Geophysics.

Amirova-Mammadova, S. (2018). Pipeline Politics and Natural Gas Supply from Azerbaijan to Europe: Challenges and Perspectives. Springer.

Bagirov, E., Nadirov, R., & Lerche, I. (1996). Earthquakes, mud volcano eruptions, and fracture formation hazards in the South Caspian Basin: Statistical inferences from the historical record. Energy Exploration & Exploitation, 14(6), 585-606.

Blokdyk, G. (2018). Risk Matrix a Complete Guide. 5starcooks.

Cheng, Y. F. (2015). Pipeline corrosion. Journal Title, 161-162. https://doi.org/10.1179/1478422X15Z.000000000357

Gemma C., Carmine G., John McCloskey (2022). Modelling and quantifying tomorrow's risks from natural hazards. Science of The Total Environment, Volume 817. https://doi.org/10.1016/j.scitotenv.2021.152552.

Gramacki, A. (2017). Nonparametric Kernel Density Estimation and Its Computational Aspects. Springer.

Falcone D., G. Di Bona, Forcina (2022). A new method for risk assessment in industrial processes. IFAC-PapersOnLine, 55(19) 1-6. https://doi.org/10.1016/j.ifacol.2022.09.175

Fotios P., Daniele A., Vassilios A., Mohamed Zied Babai, Devon K. Barrow. (2022). Forecasting: theory and practice, International Journal of Forecasting, 38(3), 705-871. https://doi.org/10.1016/j.ijforecast.2021.11.001

Han, Z., & Weng, W. (2010). An integrated quantitative risk analysis method for natural gas pipeline network. Journal of Loss Prevention in the Process Industries, 23(3), 428–436. https://doi.org/10.1016/j.jlp.2010.02.003

Ilachinski, A., & Zane. (2001). Cellular Automata – A Discrete Universe. Kybernetes, 32(4). https://doi.org/10.1108/k.2003.06732dae.007

Kadirov F.A, Lerche I, Guliyev I.S, Kadyrov A.G, Feyzullayev A.A, Mukhtarov A.Sh. (2005). Deep Structure Model and Dynamics of Mud Volcanoes, Southwest Absheron Peninsula (Azerbaijan). Energy Exploration & Exploitation. 23(5):307-332. doi:10.1260/014459805775992717

Katopodis, T., & Sfetsos, A. (2019). A review of climate change impacts to oil sector critical services and suggested recommendations for industry uptake. Infrastructures, 4(4), 74. https://doi.org/10.3390/infrastructures4040074

Krausmann, E., Cruz, A. M., & Salzano, E. (2016). Natech Risk Assessment and Management: Reducing the Risk of Natural-Hazard Impact on Hazardous Installations. Elsevier.

Krausmann, E., Renni, E., Campedel, M., & Cozzani, V. (2011). Industrial accidents triggered by earthquakes, floods and lightning: lessons learned from a database analysis. Natural Hazards, 59(1), 285–300. https://doi.org/10.1007/s11069-011-9754-3

Lerche, I., & Bagirov, E. (2014). Impact of Natural Hazards on Oil and Gas Extraction. Springer.

Nasser, H. (2014). Learning ArcGIS Geodatabases. Packt Publishing Ltd.

Novacheck, J., Sharp, J., Schwarz, M., Donohoo-Vallett, P., Tzavelis, Z., Buster, G., & Rossol, M. (2021). The evolving role of extreme weather events in the U.S. power system with high levels of variable renewable energy. Golden, CO: National Renewable Energy Laboratory.

Oruji, S., Ketabdar, M., & Ketabdar, M. (2022). Evaluation of land subsidence hazard on steel natural gas pipelines in California. Upstream Oil and Gas Technology.

Othman, A., El-Saoud, W. A., Habeebullah, T., Shaaban, F., & Abotalib, A. Z. (2023). Risk assessment of flash flood and soil erosion impacts on electrical infrastructures in overcrowded mountainous urban areas under climate change. Reliability Engineering & System Safety. https://doi.org/10.1016/j.ress.2023.109302

Panahi, B.M. (2005). Mud Volcanism, Geodynamics and Seismicity of Azerbaijan and the Caspian Sea Region. In: Martinelli, G., Panahi, B. (eds) Mud Volcanoes, Geodynamics and Seismicity. NATO Science Series, vol 51. Springer, Dordrecht. https://doi.org/10.1007/1-4020-3204-8_7

Petersen, K. (2020). Visualizing Risk: Drawing Together and Pushing Apart with Sociotechnical Practices.

Rasouli, A., & Imrani, Z. (2023). Object-Based Analysis for Risk and Disasters Assessment: Tutoring on Geo-Environmental Contexts of Azerbaijan. Baku: CN Poliqraf LLC. 430. https://doi.org/10.5281/zenodo.14885894

Samany, N.N., Liu, H., Aghataher, R. et al. (2022). Ten GIS-Based Solutions for Managing and Controlling COVID-19 Pandemic Outbreak. SN COMPUT. SCI. 3, 269. https://doi.org/10.1007/s42979-022-01150-9

Silverman, B.W. (2018). Density Estimation for Statistics and Data Analysis. Routledge.

Zio, E. (2016). Critical Infrastructures Vulnerability and Risk Analysis. Eur J Secur Res 1, 97–114. https://doi.org/10.1007/s41125-016-0004-2