TRƯỜNG ĐẠI HỌC CÔNG NGHỆ
ĐẠI HỌC QUỐC GIA HÀ NỘI

Trần Thái Dương

ACCESSING UNEXPLODED ORDNANCE (UXO) RISK AND LAND CONTAMINATION IN VIETNAM

PROJECT REPORT

Module: Visual Data Analytics - INT3137 1

Table of Content

1. Introduction

During the Vietnam War (1955–1975), the U.S. dropped an estimated 7.5 million tons of bombs—more than all sides in World War II combined. Today, around 800,000 tons of unexploded ordnance (UXO) remain scattered across 6.1 million hectares, affecting 18.7% of Vietnam's land and all 63 provinces. Some areas are over 80% contaminated. Since 1975, UXO has killed over 40,000 people and injured 60,000, many of them children or primary earners. UXO remains a major barrier to safe land use, development, and infrastructure. This project analyzes UXO contamination risk across province (ADM1), district (ADM2), and 10km² grid levels. Using bombing data, geospatial analysis, and clearance records, it identifies high-risk areas and suggests priorities for clearance to support safety and sustainable development.

Project overview:

This project is structured to systematically address the issue of UXO contamination in Vietnam. Section 2 defines the core research questions - identifying areas at risk and determining where clearance should be prioritized - while also outlining the scope of the study. Section 3 presents the data sources and abstract data requirements necessary for the analysis. Section 4 focuses on the U.S. bombing dataset, which serves as the primary input for assessing contamination risk, and details how this data is processed and cleaned. Section 5 builds the UXO risk assessment models across multiple spatial levels, including provinces (ADM1), districts (ADM2), and 10km² grids. Section 6 then integrates supplementary data to evaluate clearance priorities, exploring both geographic and budget-based strategies. Finally, Section 7 concludes the report by summarizing key findings and suggesting implications for UXO mitigation and sustainable land use planning.

2. Research questions and related problems

2.1. Research questions

The project is guided by the following key research questions designed to address the core issues of UXO contamination and clearance prioritization

2.1.1. Which areas are under the risk of UXO contamination?

This question aims to map and categorize regions across Vietnam based on their UXO risk levels, ranging from Low to Extreme, using available ADM1-level data. The analysis will highlight geographic patterns and identify specific cities or regions most affected by contamination, providing a foundation for targeted interventions.

2.1.2. Which areas should be prioritized for UXO clearance?

This question aims to figure out which contaminated areas need urgent cleanup by looking at the regions' risk level, population, terrain, and what demining work is already happening. The study will look at things like the size of the contaminated area, how easy it is to clear, and the benefits of making the land usable again for the community.

2.2. Subject and object of the study

The subject of this study is the UXO contamination risk and land clearance strategies in Vietnam, with a focus on understanding the spatial distribution and impact of explosive remnants. The object of the study comprises the ADM1-level data, including city-specific information on risk levels, land area, terrain types (e.g., residential, agricultural, mountainous), and clearance activities. This data will be analyzed to uncover insights into contamination patterns and guide policy recommendations.

3. The data

3.1. Subject structure analysis

Figure 3.1 shows how data is organized and connected in UXO management. This diagram shows the important links and relationships that help determine risk levels and how to prioritize UXO clearance work.

The subject primarily consists of location-based components, with different geographic areas (district, city, and specific locations) being the main units for analysis and planning. While time factors are present, the diagram highlights the location relationships that guide decision-making for UXO management.

The most important parts of this framework are how risk is determined and priorities are set. UXO density, which is affected by how many unexploded devices are in an area, is the main factor that determines risk levels across different geographic areas. These risk levels then help establish which areas should be cleared first, creating a step-by-step process from finding hazards to taking action.

The project looks specifically at UXO risk at two important regional levels, districts and cities, to make sure cleanup teams can take practical action and resources can be properly assigned. This multi-level approach allows for both big-picture planning at the district level and specific, targeted work at the city level.

Budget decisions follow these priority rankings, with money going to districts based on their established risk priorities. This creates a flow of resources that matches evidence-based risk assessment rather than random administrative borders.

Environmental factors, especially terrain types, add more complexity by affecting both where UXOs are found and how difficult clearance operations will be. The history of bombing missions shows where explosive materials came from, which helps predict where UXOs might be concentrated.

How people are distributed across these areas adds a human element to risk calculations, highlighting the humanitarian importance of UXO management priorities. Areas with more people living in UXO-affected zones naturally need more attention when setting priorities.

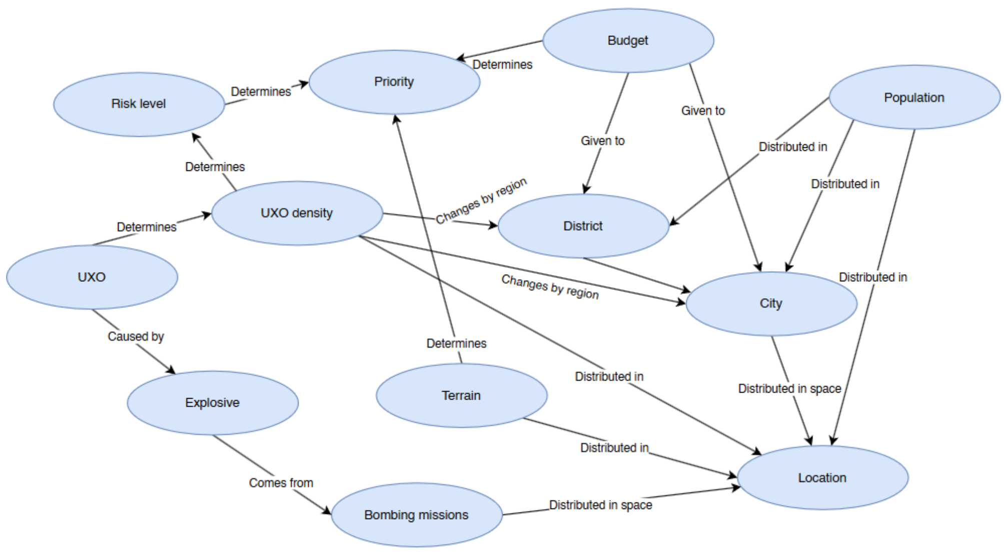


Figure 3.1. Initial subject structure analysis

3.2. Data sources

Name	Description	Source
Name Description Source THOR Vietnam War Bombing Operations	Dataset containing detailed records of U.S. bombing missions conducted over Vietnam during the Vietnam War, including coordinates, dates, and ordnance types.	U.S. Department of Defense – Theater History of Operations Reports (THOR)
Quang Tri Mine Action Center Statistics	Provincial-level data on UXO incidents, clearance activities, and contamination reports in Quảng Trị province, one of the most heavily affected areas.	Quảng Trị Mine Action Center (QTMAC)
Vietnam Population Statistics (2023)	Official demographic data by province and district, including population size and density.	General Statistics Office of Vietnam (GSO), 2023
Vietnam Administrative Boundaries	Geospatial data defining the national, provincial, and district-level boundaries of Vietnam.	GeoBoundaries – geoboundaries.org
High-Resolution Land Use and Land Cover Map	Satellite-derived land cover classifications for Vietnam, providing insights into agricultural, forest, urban, and other land use categories.	Japan Aerospace Exploration Agency (JAXA) – Aproject.jaxa.jp
Explosive dud rate	Dud rate data of specific bombs gathered from multiple sources.	Multiple sources

Table 3.1. List of datasets used

3.3. Data flow diagram

Figure 3.2. shows overview of the UXO Risk and Priority Assessment Framework The following diagram illustrates the overall structure of the UXO risk assessment and prioritization model used in this study. Raw mission data are first geolocated and aggregated into 10 km² grid cells (Level 3), where normalized metrics and dud-rate adjusted UXO estimates are computed. These are then aggregated into ADM2 (district) and ADM1 (provincial) levels to inform broader risk classification. In parallel, terrain and population data are integrated via a weighted scoring mechanism to calculate a Priority Score at the ADM1 level. This comprehensive, multi-layered approach enables informed decision-making by combining operational risk, human exposure, and clearance feasibility.

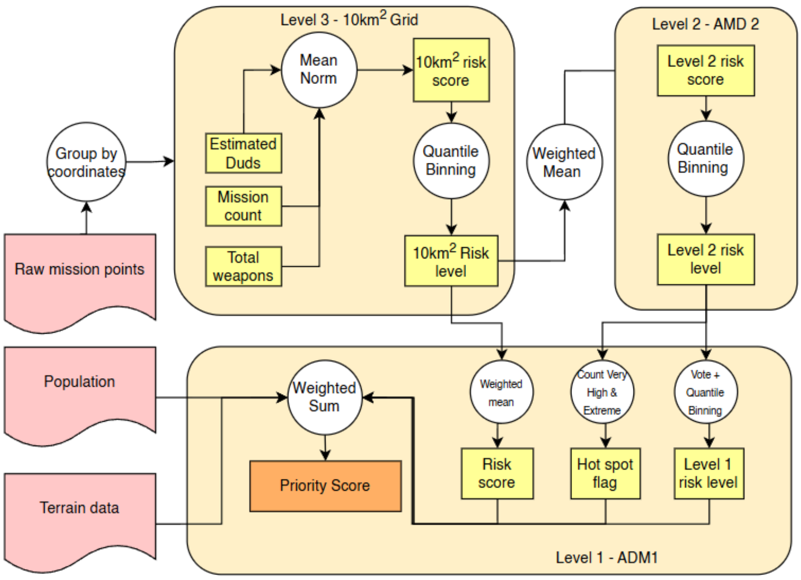


Figure 3.2. Flow of diagram for assess risk at every level and priority score for city/province level

4. The U.S. Bombing Operations Data

4.1. About the data

The THOR dataset contains detailed records of U.S. bombing missions over Vietnam, including coordinates, dates, and ordnance types. It is the primary dataset for estimating potential UXO locations due to its national coverage and historical relevance - most UXO contamination stems from the Resistance War Against America (1955–1975).

4.2. Overview of the data

The dataset contains three different files *'operations.csv'*, *'weapon.csv'*, and *'aircraft.csv'*. Aircraft data contains information on the types of aircraft used in the bombing missions included in the operation data.

	GLOSS_ID	VALIDATED_ROOT	AIRCRAFT_NAME	WEBSITE_LINK	AIRCRAFT_TYPE	AIRCRAFT_SHORTNAME	AIRCRAFT_APPLICATION	AC_MISSION_COUNT
0	1	A-1	Douglas A-1 Skyraider	http://www.navalaviationmuseum.org/attractions...	Fighter Jet	Skyraider	FIGHTER	373265
1	2	A-26	Douglas A-26 Invader	http://www.militaryfactory.com/aircraft/detail...	Light Bomber	Invader	BOMBER	36672
2	4	A-37	Cessna A-37 Dragonfly	http://www.militaryfactory.com/aircraft/detail...	Light ground-attack aircraft	Dragonfly	ATTACK	282699
3	5	A-4	McDonnell Douglas A-4 Skyhawk	http://www.fighter-planes.com/info/a4-skyhawk.htm	Fighter Jet	Skyhawk	FIGHTER	390290
4	6	A-5	North American A-5 Vigilante	http://www.militaryfactory.com/aircraft/detail...	Bomber Jet	Vigilante	BOMBER	10

Figure 4.2. Data fields in flight data - *'aircraft.csv'*

Weapons data contains information about different types of weapons used and the number of times they are used in missions.

	WEAPON_ID	WEAPONTYPE	WEAPONTYPE_COMMON_NAME	WEAPON_CLASS	WEAPONTYPE_DESC	WEAPON_COUNT
0	1	100 GP	General Purpose Bomb	BOMB	100 lb general purpose	1
1	2	1000 G	Megaboller flash powder bomb	BOMB	1000 g BKS	2
2	3	1000LB GP M-65	An-M65	BOMB	1000 lb general purpose	12776
3	4	1000LB MK-83	Mark 83 bomb	BOMB	1000 lb none guidence general purpose bomb	15522
4	5	1000LB SAP M59	AN-M59	BOMB	1000 lb semi-armor piercing bomb	454

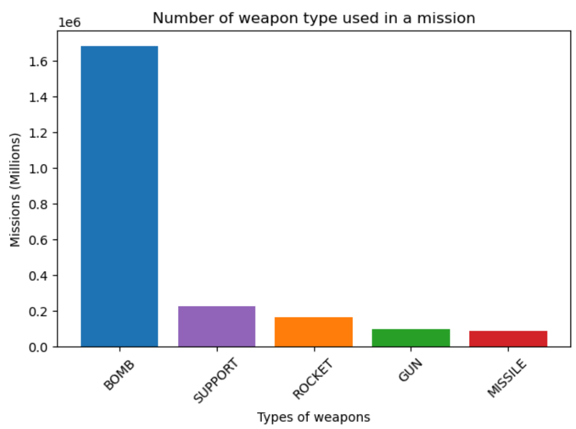
Figure 4.2. Data fields in weapon data - *weapon.csv'*

Finally, the operation dataset is the main data used for risk analysis in this project. There are 47 columns and about 4.6 million rows of data. However, only a few features are selected for the purpose of analysing UXO risk.

THOR_DATA_VIET_ID	COUNTRYFLYINGMISSION	MILSERVICE	MSNDATE
SOURCEID	SOURCERECORD	VALID_AIRCRAFT_ROOT	TAKEOFFLOCATION
TGTLATDD_DDD_WGS84	TGTLONDDD_DDD_WGS84	TGTTYPE	NUMWEAPONSDELIVERED
TIMEONTARGET	WEAPONTYPE	WEAPONTYPECLASS	WEAPONTYPEWEIGHT
AIRCRAFT_ORIGINAL	AIRCRAFT_ROOT	AIRFORCEGROUP	AIRFORCESQDN
CALLSIGN	FLTHOURS	MFUNC	MFUNC_DESC
MISSIONID	NUMOFACFT	OPERATIONSUPPORTED	PERIODOFDAY
UNIT	TGTCLOUDCOVER	TGTCONTROL	TGTCOUNTRY
TGTID	TGTORIGCOORDS	TGTORIGCOORDSFORMAT	TGTWEATHER
ADDITIONALINFO	GEOZONE	ID	MFUNC_DESC_CLASS
NUMWEAPONSJETTISONED	NUMWEAPONSRETURNED	RELEASEALTITUDE	RELEASEFLTSPEED
RESULTSBDA	TIMEOFFTARGET	WEAPONSLOADEDWEIGHT	

Figure 4.3. Data fields in mission data - *'operations.csv'*

4.2. Data processing and insight



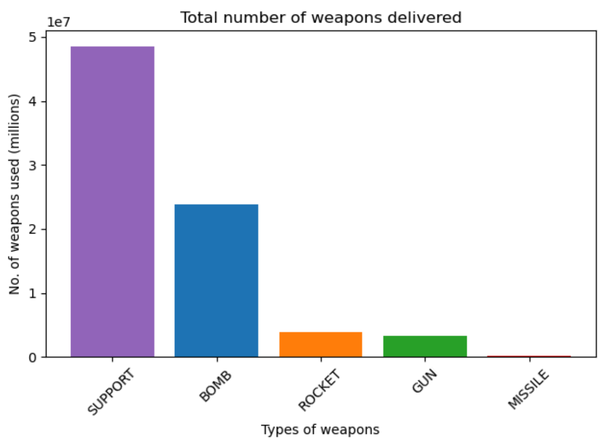


Figure 4.4. The first graph shows the number of missions by the type of weapons they used. The second graph shows the number of weapons delivered in all missions.

6,104,2533 was the total amount of bombs used in this dataset, which is close to other recorded sources of data. It was recorded that 7.5 million tons of bombs were dropped in the period of 1965-1975 [1]. About 200,000 missions are support missions but the amount of weapons they used is over 4 millions. Further inspection in *'MFUCN_DESC'* which shows the mission descriptions shows that they are not UXO related, support missions include tasks like refuelling, spreading leaflets, cargo transportation,... By looking at figure 4.4, About over 1.6 millions of missions among 4.6 millions missions are bombing missions, the number of missions that use rocket, gun, and missiles are less than 200,000 each.

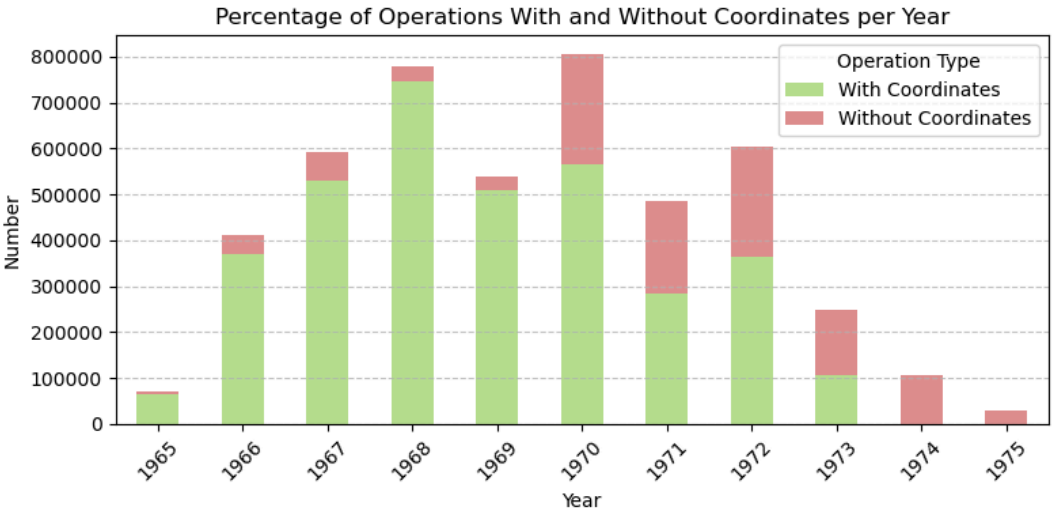


Figure 4.4. Graphs show the number of air missions recorded from 1965-1975, red color in a bar represents the number of missions without coordinates, green color shows the number of operations with coordinates.

Figure 4.5. shows an interesting insight of the number of recorded flight missions. The data reveals that 1970 was the peak year for total recorded missions, approaching 800,000, reflecting the height of U.S. aerial involvement in the conflict. Interestingly, the highest number of missions with complete coordinate data occurred in 1968, a year marked by intense combat operations such as the Tet Offensive. After 1970, while the total number of missions declined steadily, the proportion of missions lacking coordinate data increased. This trend likely correlates with the Vietnamization policy initiated in 1969, where the U.S. began transferring military responsibility to South Vietnamese forces. As American troops and administrative personnel withdrew, the completeness of mission records diminished, possibly due to reduced oversight or deprioritization of detailed reporting. By 1973, following the Paris Peace Accords and the formal end of U.S. combat operations, mission numbers dropped sharply, and the remaining records showed a high incidence of missing data. This suggests that in the later years of the war, while fewer operations were conducted, they were less systematically documented - highlighting how geopolitical shifts can directly influence the quality and completeness of wartime records.

4.3. Cleaning the data

The objective of this project is to analyze the locations that are under risk of UXO contamination, therefore, it is important that the data must contain coordinate information. Figure 5.2. shows that the proportion of data rows with missing coordinate information is 75%.

In order to decide whether to keep or drop 25% of the missing data, the impact on other fields needs to be investigated. I divided the data set into two, one with missing longitude and latitude information and one without the coordinates to analyze the possible effects of removing the missing data.

THOR_DATA_VIET_ID	100.000000	NUMWEAPONSRETURNED	100.000000	NUMOFACFT	100.000000	WEAPONTYPEWEIGHT	100.000000
ADDITIONALINFO	100.000000	ID	100.000000	MFUNC_DESC_CLASS	100.000000	NUMWEAPONSDELIVERED	100.000000
FLTHOURS	100.000000	NUMWEAPONSJETTISONED	100.000000	WEAPONSLOADEDWEIGHT	100.000000	VALID_AIRCRAFT_ROOT	100.000000
SOURCERECORD	100.000000	SOURCEID	100.000000	MSNDATE	100.000000	AIRCRAFT_ORIGINAL	99.989680
AIRCRAFT_ROOT	99.989680	UNIT	99.989444	MILSERVICE	99.930434	COUNTRYFLYINGMISSION	99.922598
TAKEOFFLOCATION	99.893564	MISSIONID	99.664484	TIMEOFFTARGET	99.434119	TIMEONTARGET	99.434119
MFUNC	97.835075	MFUNC_DESC	97.757759	PERIODOFDAY	95.722779	TGTCOUNTRY	95.358572
TGTORIGCOORDS	77.128183	TGTORIGCOORDSFORMAT	76.600157	TGTLATDD_DDD_WGS84	75.802348	TGTLONDDD_DDD_WGS84	75.802348
GEOZONE	74.990408	TGTTYPE	60.808095	OPERATIONSUPPORTED	58.889123	TGTCONTROL	53.953074
WEAPONTYPE	48.537839	WEAPON_CLASS	48.533493	TGTCLOUDCOVER	47.750950	TGTWEATHER	44.486508
CALLSIGN	29.335610	RESULTSBDA	6.110719	RELEASEALTITUDE	0.072328	AIRFORCEGROUP	0.062264
AIRFORCESQDN	0.059566	RELEASEFLTSPEED	0.036164	TGTID	0.000749	WEAPONTYPECLASS	0.000000

Figure 4.6. Proportion of missing data of every data field in operation data

First, I investigated the proportion of weapon types in each dataset to assess how much information on explosives would be lost if data with missing coordinates were excluded. By inspecting the total variance distance between distribution of each feature in the two different groups, I found out that the missing data is not missing at random (MAR) but might be missing not at random (MNAR). Figure 4.7 shows the TVD score of different features.

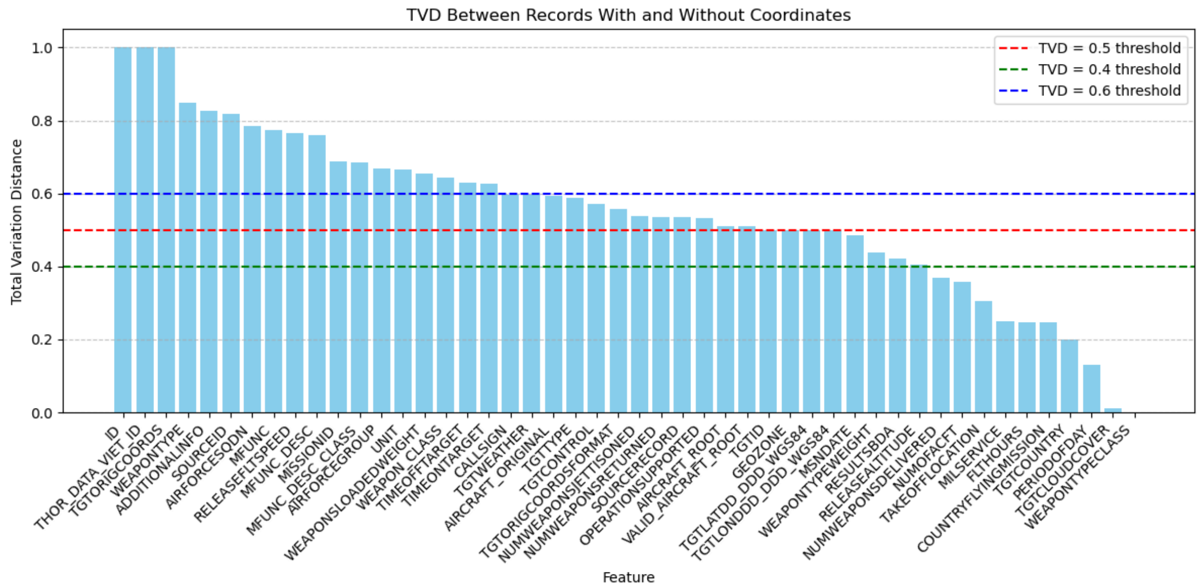


Figure 4.7. Total variance distance between features of data with coordinates and without coordinates

Based on the features with high TVD between two datasets, I have selected a few features for inspection.In figure 4.8 the first two pie charts show that the majority of weapons in the dataset without missing coordinates are for SUPPORT purposes, which do not pose a UXO threat. Excluding the data with missing coordinates would result in the loss of 20.3% of the total number of bombs. However, when considering bomb information based on tonnage, as shown in the third pie chart, only 3.3% of the total bomb tonnage would be lost.

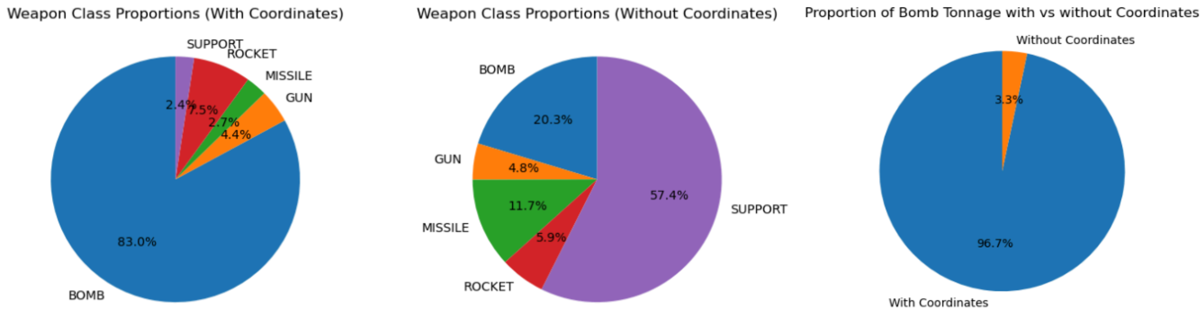


Figure 4.8. First and second pie charts showing proportion of weapons in data with coordinates and data without coordinates. Third pie chart shows the proportion of bombs of bombs in data with and without coordinates compared to the total amount of bombs in the data set.

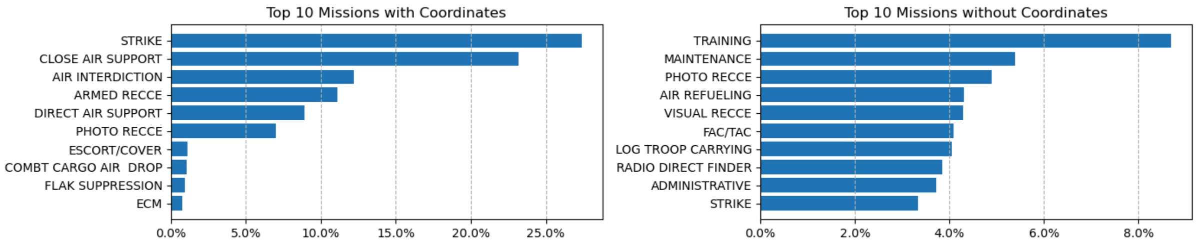


Figure 4.9. Each bar graph shows the top 10 missions of each group, with and without coordinates

Next, I examined the ‘MFUNC_DESC’ field, which provides information about the type of mission carried out. Figure 4.9 shows that the majority of missions in groups without coordinate data are non-offensive in nature - such as training, maintenance, reconnaissance, and air refueling. In contrast, the most common mission types in groups with coordinates are offensive operations, including strike, close air support, air interdiction, armed reconnaissance, and direct air support.

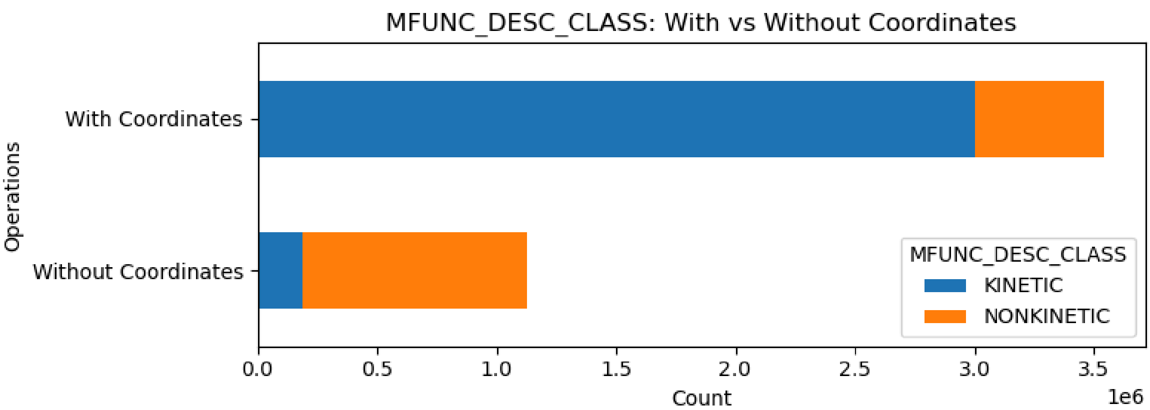


Figure 4.10. Stacked horizontal bar chart shows the number of KINETIC and NONKINETIC missions in data with and without coordinates.

I then further analyzed the ‘MFUNC_DESC_CLASS’ field, which indicates whether a mission involved the use of damage-inflicting weapons. A value of ‘KINETIC’ signifies that weapons were used, while ‘NONKINETIC’ indicates that no weapons were employed. For example, some missions may carry bombs or explosives for transportation purposes without intending to deploy them; these, along with missions like refueling, leaflet drops, and maintenance, are classified as NONKINETIC. Figure 4.10 illustrates this distinction clearly: the majority of missions without coordinates are NONKINETIC, whereas missions with coordinates are predominantly KINETIC.

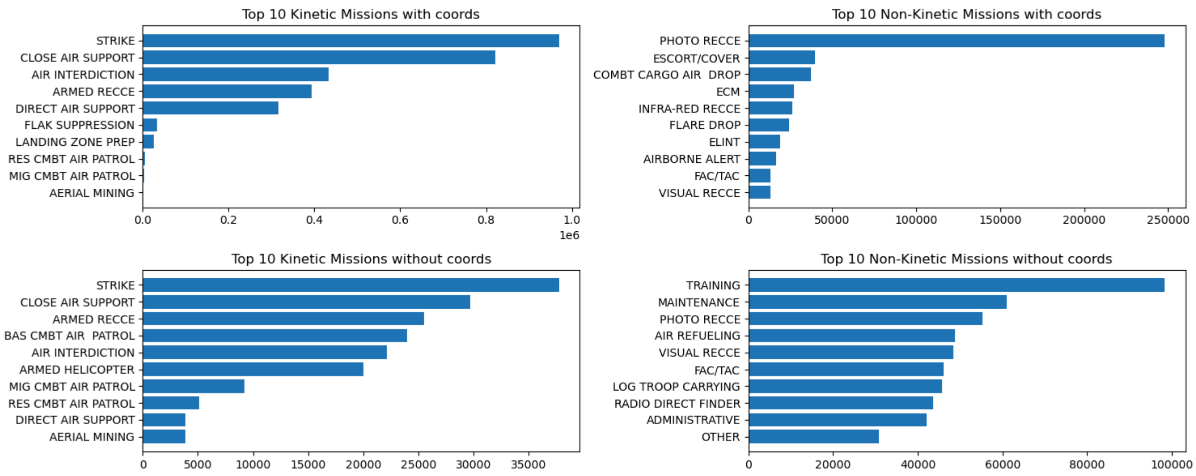


Figure 4.11. Stacked horizontal bar chart shows the number of KINETIC and NONKINETIC missions in data with and without coordinates.

With the given information, I decided to drop all the data with missing values as they not only have no valuable information but also because they have little impact on the purpose of this project.

The reason I had considered keeping the missing values is because of the *'TGTCOUNTRY'* field which has information about the targeted country that might be useful for statistical analysis. However, after plotting the locations of the mission, I concluded that the field *'TGTCOUNTRY'* does not hold geographical information and coordinate values are the only values that carry the actual mission location.

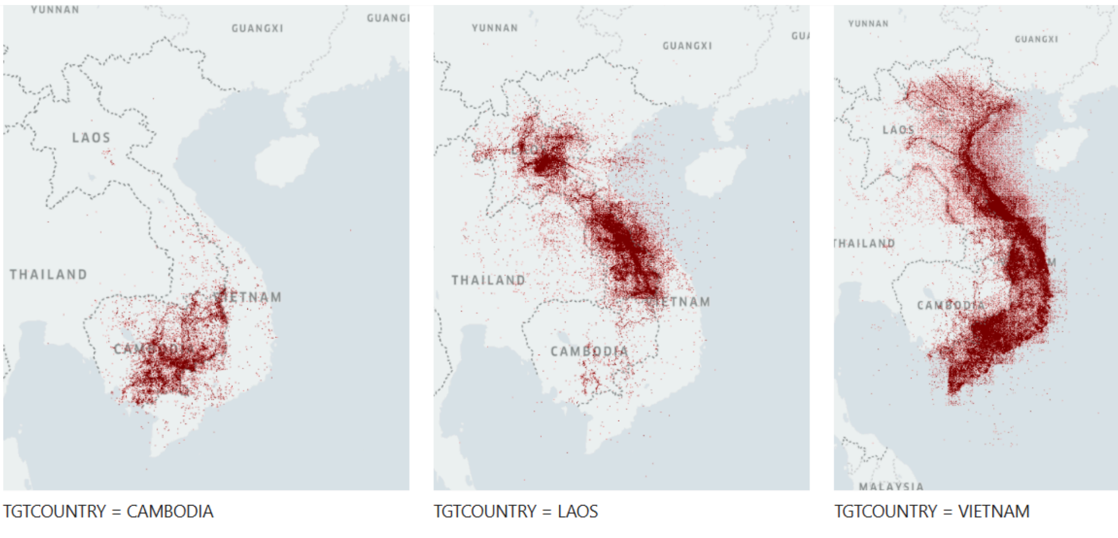


Figure 4.12. The plot shows the mission locations using target coordinate data and filtered by different TGTCOUNTRY values, from left to right, Cambodia, Laos, and Vietnam, respectively.

As a reminder, the primary goal of this project is to analyze UXO (Unexploded Ordnance) contamination within Vietnam's land borders. Therefore, only data that falls within Vietnam's national boundary will be retained for analysis. However, as shown in Figure 4.12, the spatial distribution of flight missions reveals several important patterns. Bombing activity appears to concentrate along specific routes, often following rivers or other linear features, suggesting that many targets were related to supply transportation networks. Additionally, several urban and strategic areas were repeatedly targeted, such as Phonsavan in Laos, Ho Chi Minh City, Vietnam's Central Region, and coastal areas along Vietnam's eastern shoreline.

To ensure data accuracy while accounting for possible coordinate errors, I implemented a spatial filter using administrative boundary data (ADM0). Specifically, I decided to include missions with coordinates located within Vietnam and those up to 5 kilometers beyond the national border. This buffer accounts for imprecise targeting data and the nature of certain operations - such as carpet bombing - where munitions may have crossed into Vietnam from neighboring countries like Laos and Cambodia. Additionally, missions conducted near or along the coastline are retained for the same reason. Figure 4.13 compares the dataset before and after filtering, illustrating the impact of this spatial refinement. After removing the foreign data points, 65.5% of bomb tonnage in the data remains, which means that at least 65.5% of bombs dropped in the war is on Vietnam.

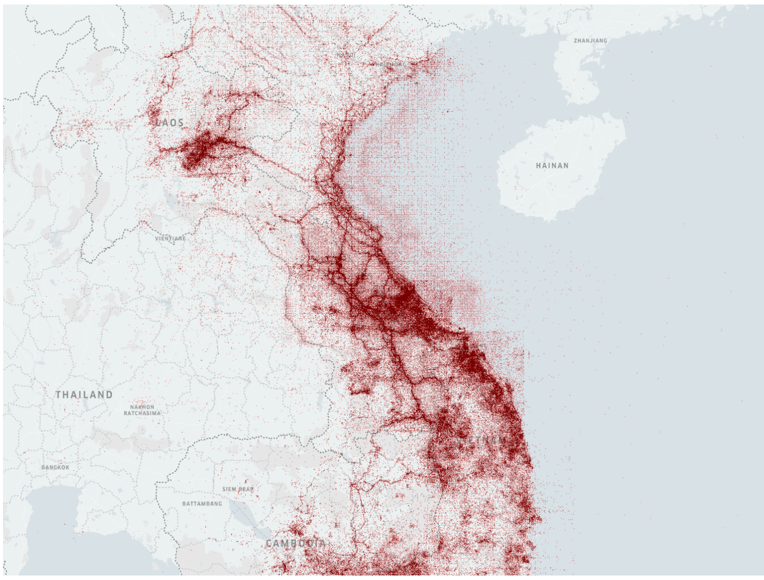


Figure 4.12. Bombing missions reveals pattern of bombs going along routes across Vietnam - Laos and Vietnam - Cambodia borders

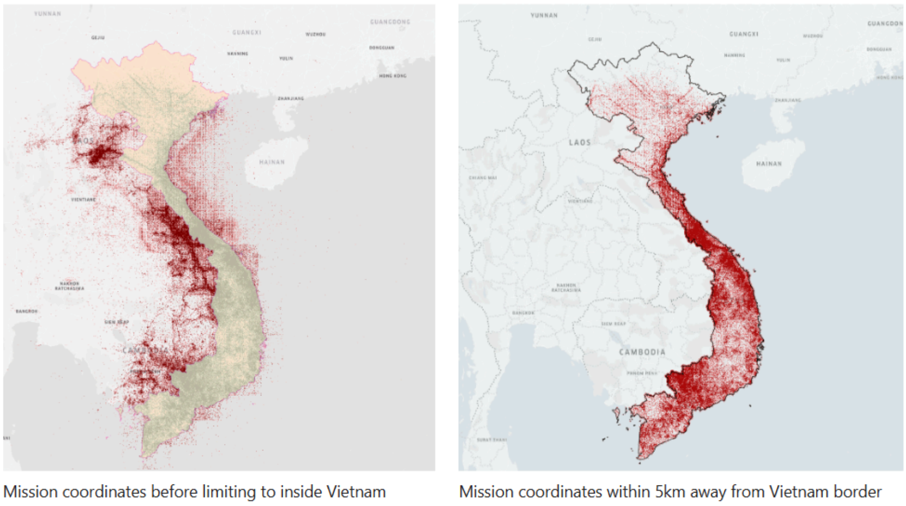


Figure 4.13. Plot of missions before removing foreign data points outside the range of 5km away from the Vietnam border and after.

5. Risk Assessment

This section aims to answer the question: Which areas are at risk of UXO contamination? by analyzing regional risk levels based on the available data.

5.1. Modelling

The risk assessment model is developed using Vietnam's administrative hierarchy, specifically at the ADM0 (national) and ADM1 (provincial) levels. These administrative levels are chosen because they align with actual decision-making structures and budget allocation mechanisms within the country. Authorities responsible for UXO response and mitigation typically operate at these levels, making them both relevant and actionable for policy and planning.

For spatial resolution, a 10 km^2 grid is used as the smallest unit of analysis. This grid size strikes a balance between being practically actionable and sufficiently detailed - it is not so large as to obscure local variations in risk, nor so small as to become impractical for intervention or resource planning. Additionally, geospatial data for ADM1 and ADM2 boundaries are readily available, further supporting the feasibility and scalability of the model.

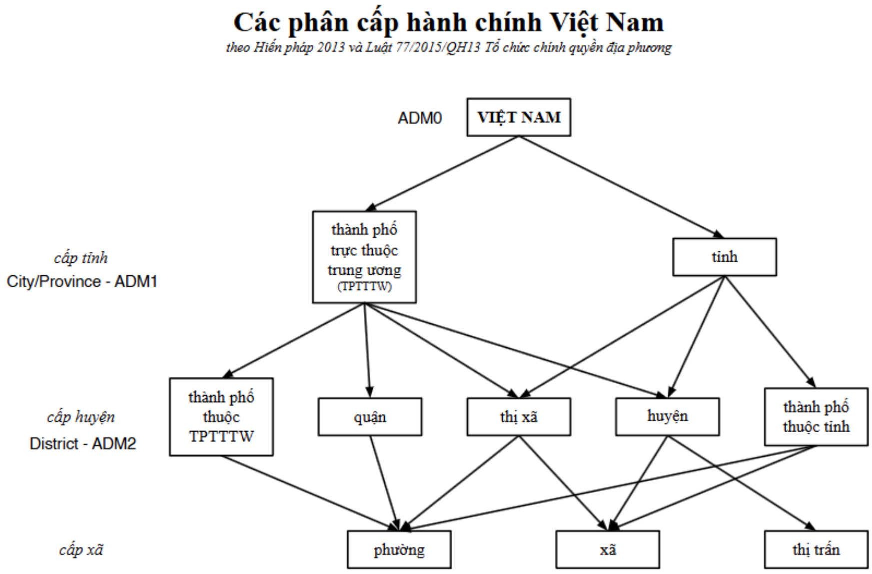


Figure 5.1. Vietnam administrative hierarchy

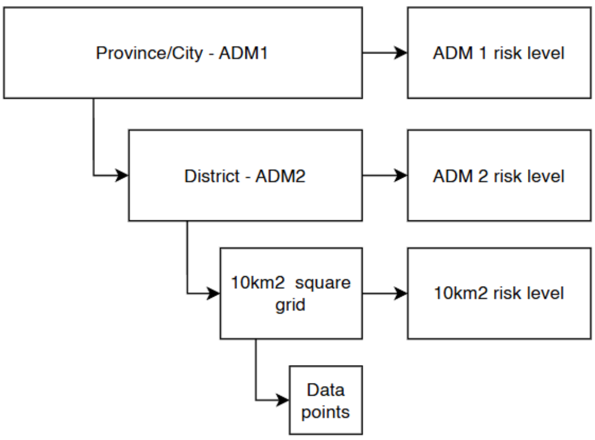


Figure 5.2. Structure of hierarchical risk assessment model

The proposed model is hierarchical, enabling UXO risk to be assessed at multiple geographic levels, i.e. national (ADM0/ADM1), district (ADM2), and 10 km² grid cells. Each level builds upon data from the levels below, ensuring that higher-level assessments reflect localized risk patterns while supporting strategic decision-making and resource allocation. At the national and ADM1 level, the model offers a broad overview of risk distribution, useful for setting national priorities and allocating funding. The ADM2 level provides a more detailed assessment suitable for provincial and district-level planning. Finally, the 10 km² grid level delivers the highest spatial resolution, supporting operational decision-making, such as identifying specific areas for UXO clearance. Each level's risk classification aggregates or summarizes the underlying data from its sub-levels. This ensures consistency, interpretability, and scalability, making the model both data-driven and actionable across administrative and operational contexts. Figure 5.2 illustrates the structure of this hierarchical risk assessment framework.

5.2. Implementation

The hierarchical UXO risk assessment model was implemented in three levels: grid-level (10 km²), ADM2 (district), and ADM1 (province/city). This structure enables both high-resolution analysis and aggregation for broader policy decisions. Each level uses geospatial and operational data, enriched with UXO dud rates, to compute composite risk indicators.

Feature	**Description**
lat, lon	Target coordinates used to map missions to specific 10 km² grid cells
WEAPONTYPE	Type of weapon used (e.g., Bomb, Rocket, Missile)
WEAPONTYPECLASS	Classifies the weapon (e.g., Explosive, Support)
WEAPONTYPEWEIGHT	Weight category of the weapon, helps assess severity
NUMWEAPONSDELIVERED	Number of weapons delivered to the location
dud_rate	Failure probability of each weapon type (from external data)
MISSIONID	Used to calculate mission density in the grid cell

Table 5.1. List of selected features for risk assessment

5.2.1. Dud rate data

Assessing the risk based on mission density alone isn't entirely accurate, since some weapons, even if dropped in large numbers, have a lower chance of exploding. That's why it's important to have an estimate of the dud rate, the probability that a weapon fails to detonate. The U.S. bombing dataset doesn't include dud rates, but it does provide the exact weapon names. Thanks to modern LLMs, finding dud rate information for hundreds of different weapons has become less painful. Without them, it would've been nearly impossible to look up dud rates for all 294 weapon types in the time available. For this study, the dud rates were gathered from demining organizations, military reports, and academic sources, and were then used to estimate the number of unexploded ordnance (UXO) likely present in each grid cell.

	WEAPONTYPE	WEAPONTYPE_COMMON_NAME	WEAPON_CLASS	DUD_RATE	SOURCE
0	100 GP	General Purpose Bomb	BOMB	0.100	US DoD UXO Technical Report, 1970
1	1000 G	Megaboller flash powder bomb	BOMB	0.035	Military Munitions Assessment Report, 1975
2	1000LB GP M-65	An-M65	BOMB	0.100	USAF Explosive Ordnance Report, 1972
3	1000LB MK-83	Mark 83 bomb	BOMB	0.057	Naval Weapons Center Study, 1974
4	1000LB SAP M59	AN-M59	BOMB	0.120	USAF EOD Historical Analysis, 1973
...	...	...	...	...	...
289	XM384 40MM AMMO	NaN	SUPPORT	NaN	Not explosive ordnance
290	XM40E5	XM40E5 Gravel Bomb	MISSILE	0.200	Cluster Munitions Assessment Report, 1975
291	XM41UNITS-CAN	NaN	MISSILE	NaN	Insufficient data
292	XM42 (XM12) GRVL BOMB	NaN	MISSILE	0.200	Cluster Munitions Assessment Report, 1975
293	XM54 MICROGRAVEL BOMB	NaN	MISSILE	0.200	Cluster Munitions Assessment Report, 1975

Figure 5.3. Dud rate dataset joint with the given weapon dataset

In the dud rate dataset, only weapons that pose an unexploded ordnance (UXO) risk have corresponding dud rate values. Some explosives lack dud rate data because no statistical information is available. Figure 5.4 shows that most weapons without dud rate data are classified as support mission weapons, which we do not require for this analysis. Notably, no weapons with dud rate data are categorized as support weapons in figure 5.5.

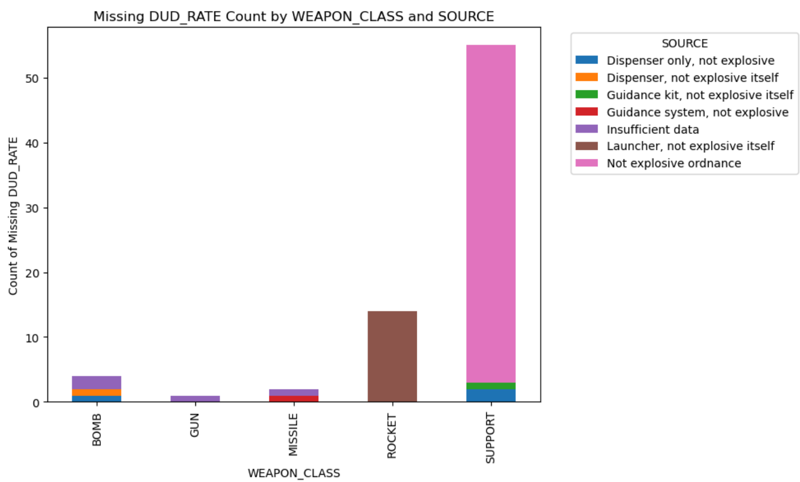


Figure 5.4. Number of weapons with unavailable dud rate

Therefore, to avoid introducing noise and irrelevant information, especially since our metric relies on mission density over an area, we will remove all data entries classified as support weapons. These weapons pose no threat and their inclusion could distort our analysis.

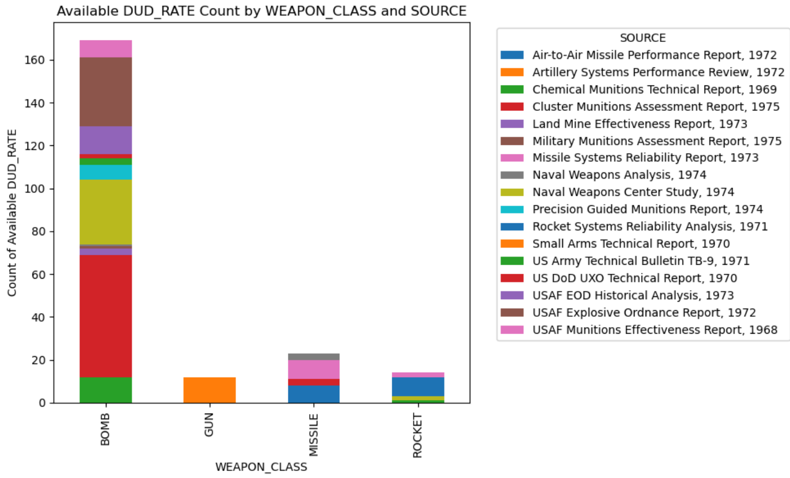


Figure 5.5. Number of weapons with available dud rate

5.2.2. Level 3 Risk Assessment - 10km2 square grids

Each mission record in the dataset containing valid geographic coordinates was assigned to one of these 10 km² cells. This spatial mapping allowed all mission-related attributes (such as the number of weapons delivered, mission type, and weapon type) to be aggregated at the cell level. Once aggregated, several metrics were computed to quantify risk.

To perform fine-grained spatial risk assessment, Vietnam's territory was divided into a uniform grid system where each cell represents an area of 10 square kilometers (10 km²). This gridding was done using a projected coordinate system, specifically UTM Zone 48N (EPSG:32648), which is suitable for Vietnam's geographic location. UTM is a common standard for regional-scale analysis as it preserves distances and areas locally, which is critical for consistent spatial measurements. Each cell in the grid is a square of 3.16 km × 3.16 km, which approximately equals 10 km² in area. The exact dimensions may vary slightly depending on the rounding, but the key principle is that the grid maintains equal-area cells to ensure comparability across the country. This resolution was chosen because it balances practical usability in the field with sufficient detail to detect local variations in risk. It is small enough to inform operational decisions such as clearance planning, yet broad enough to ensure computational feasibility across the entire country.

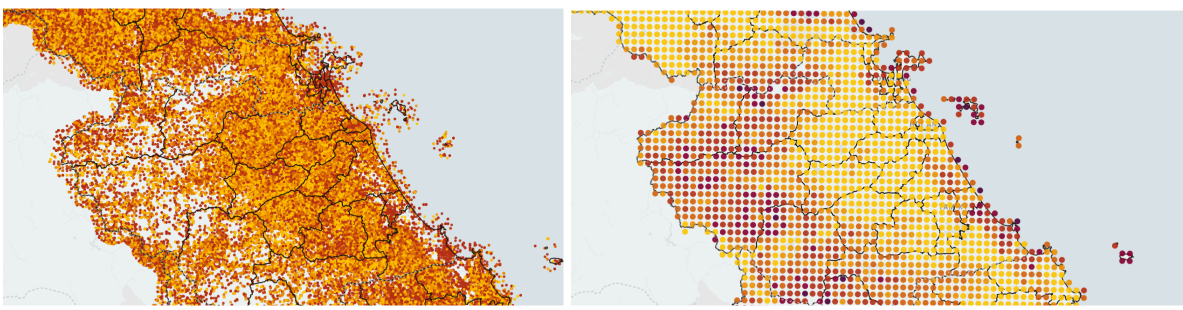


Figure 5.5. Data points binned inside a 10km^2 grid, each points in the first graph is a single mission, each points in the second image represent 10km^2 squares suspected with UXO contamination

Metric	Description
Weapon Density (per km²)	Number of weapons delivered per square kilometer.
Estimated UXO Count	Estimates how many munitions may have failed to detonate.
Composite Risk Score	Weighted and normalized score combining: – UXO density – Mission frequency – Weapon type/class (optional weighting)

Table 5.2. Features engineered for risk assessment Level 3

The weapon density per km² was calculated as the total number of weapons delivered within the grid cell divided by the cell area (10 km²). Since all cells have equal area, this simplifies to just counting the number of weapons in the cell. This metric provides a basic but essential proxy for the intensity of bombing activity.

Estimated UXO is the important factor of this study, it can be computed by using the number of weapons delivered in the dataset and combined with the dud rate found earlier. This computation assumes that dud rates are independent and additive across weapon types. The result is an approximation of the number of live munitions still present in each cell:

Estimated UXO count *= NUM_WEAPON+DELIVERED * DUD_RATE*

To capture multiple aspects of risk in a single value, a composite risk score was developed. This score integrates the estimated UXO density, mission frequency (i.e., number of unique missions recorded in the cell), and optionally, a risk factor based on weapon type or class (e.g., bombs vs. rockets). Each of these components was scaled using min-max normalization and then combined using weighted summation:

$$\textbf{Composite Risk Score} = w_1 \times \textit{Normalized UXO Density} + w_2 \times \textit{Normalized Mission Frequency} + w_3 \times \textit{Normalized Weapon Class Risk}$$

The weights w_1 ,w_2, and w_3 can be tuned depending on policy preference or expert input, but in this study, equal weighting was initially used to balance the contribution of each component. External data is required here as well—specifically, the relative danger or persistence associated with different weapon classes (for example, fragmentation bombs versus fuel-air explosives), which can be drawn from UXO clearance reports or military ordnance databases.

The resulting composite risk score was used to classify each cell into discrete risk levels (e.g., Low, Moderate, Elevated, High, Very High, Extreme) for practical interpretation and prioritization. This scoring methodology is summarized in Table 5.2.1 and forms the basis for the broader, aggregated risk assessments at the ADM2 and ADM1 levels in the hierarchical model.

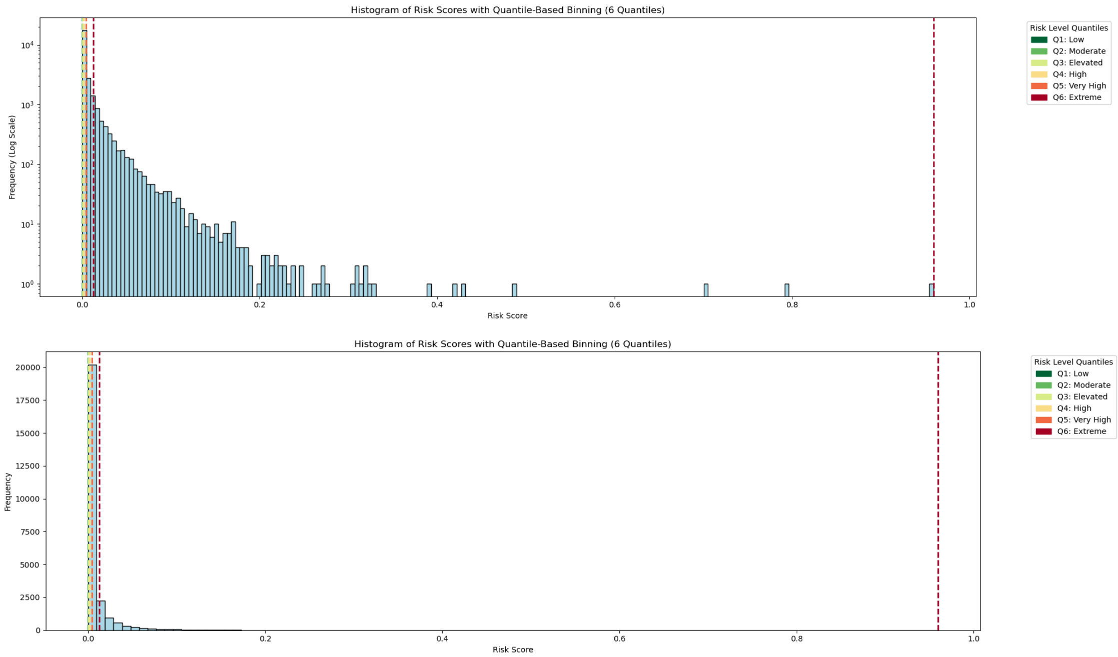


Figure 5.6. Visualization of quantile binning applied to composite risk scores. The top graph uses a logarithmic scale to highlight distribution skewness, while the bottom graph shows the same data on a normal scale.

Using 6 quantile bins means we split all grid cells into 6 groups, each with about the same number of cells, based on their composite risk scores. This way, each risk level (like Low, Moderate, High, etc.) contains a similar number of areas. It helps make the categories more balanced and easier to compare.

Other methods like fixed thresholds or clustering (e.g., natural breaks/Jenks) were also tested, but they often resulted in unbalanced group sizes. For example, with natural breaks, the highest risk class can end up containing very few cells, making it hard to detect hotspots. In contrast, quantile binning ensures that each risk level contains a similar number of grid cells, which improves visual clarity and supports more balanced, data-driven decisions in mapping and prioritization.

5.2.3. Level 2 Risk Assessment - ADM2

At Level 2 of the hierarchical UXO risk framework, the analysis aggregates risk indicators from 10 km² grid cells (Level 3) into ADM2 units, which correspond to districts or equivalent administrative areas in Vietnam. This level provides a balance between localized risk information and the administrative structure relevant for planning and resource allocation.

Feature	**Description**
Weighted Mean Risk Level	Mean of Level 3 risk with arbitrary weight assignment

Table 5.3. Features engineered for risk assessment level 2

The weighted average risk scores use weights of each risk level of every $10km^2$ grid.

$$\textbf{\textit{Weighted Mean Risk Level}} = Sum(Count[Risk\ Level_i] * w_i) / Grid\ Count$$

Where $w_i = \{w_1 : Low, w_2 : Moderate, w_3 : Elevated, w_4 : High, w_5 : Very\ High, w_6 : Extreme\}$, $sum(w) = 1$

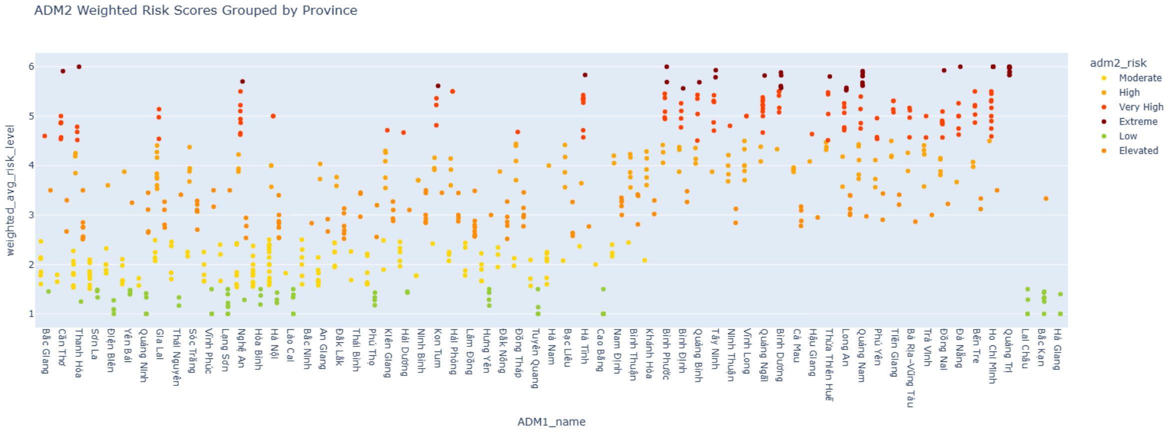


Figure 5.8. ADM mean weighted risk score grouped by province

This approach is preferred over raw Level 3 risk scores because it yields a more interpretable metric and mitigates the influence of extreme outliers, making it more suitable for policy and planning decisions at the district level. The value will then be binned into 6 quantiles in the same manner as binning Level 3 risk score into 6 risk levels.

5.2.4. Level 1 Risk Assessment - ADM1

At the ADM1 level, the model provides a strategic overview for national planning. To balance detail and interpretability, a hybrid approach was used: numerical metrics such as estimated UXO count, risk scores, and mission density were aggregated directly from 10 km² grid data for accuracy using the same weighted mean average formula for calculating Level 2 risk. Meanwhile, the overall risk level was assigned using a majority vote of ADM2 district classifications to maintain administrative relevance. Additionally, a hotspot flag was introduced to highlight provinces where more than 25% of grid cells fall into High or Extreme risk categories. This ensures that localized dangers are not lost in broader summaries.

Feature	Description
ADM1 Risk Classification	Risk level assigned to each province based on the majority of ADM2 district-level risks.
Weighted Risk Score	Average composite risk score across the province, optionally weighted by ADM2 population or area.
Hotspot Flag	Indicator set to True if ≥25% of the province's grid cells are classified as High or Extreme risk.

Table 5.4. Features engineered for risk assessment Level 1

5.3. Analysis

Which areas are under risk of UXO land contamination?

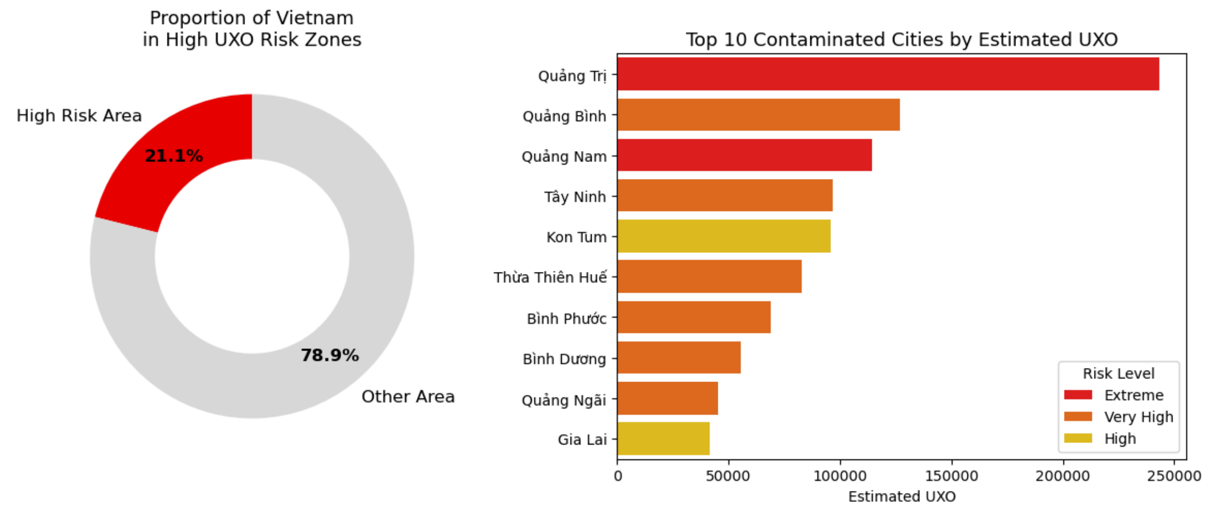


Figure 5.9. Pie chart: Proportion of land under Very High and Extreme risk of land contamination. Horizontal bar chart: Top 10 Province/City (ADM1) with highest estimated number of UXO

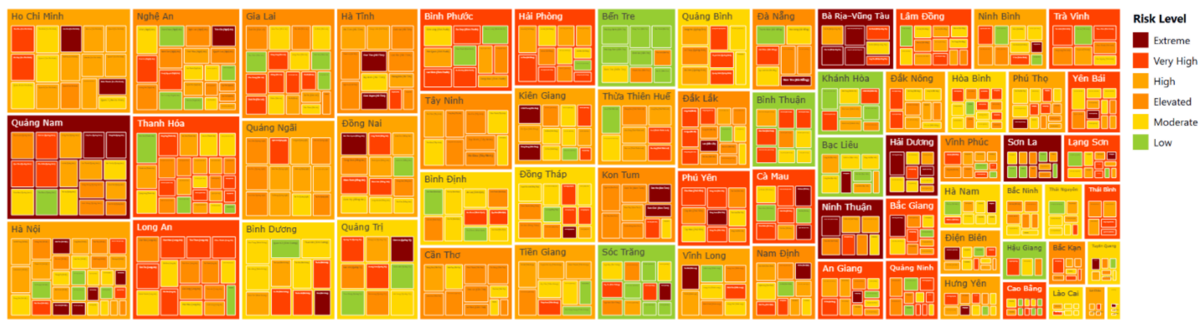


Figure 5.10. Treemap graph graph of ADM2 levels in ADM1 levels. The color indicates risk level whereas the size of each child square is the average weighted score of ADM2 districts, the size of the parent node is just the sum of all children.

By examining Figure 5.9, it is evident that Quang Tri province faces the highest risk of unexploded ordnance (UXO) contamination. This assessment aligns closely with historical facts. As a former demilitarized zone and the site of major battles such as the Battle of Quang Tri and the Easter Offensive, Quang Tri was subjected to some of the most intense aerial and ground warfare during the Vietnam War. The province endured heavy bombing campaigns, particularly during Operation Rolling Thunder and Operation Linebacker I and II, leading to one of the highest concentrations of ordnance dropped per square kilometer in history. This historical context explains the severe contamination levels observed today and reinforces the credibility of the risk analysis presented in the figure.

Figure 5.9 presents a comprehensive overview of UXO contamination in Vietnam using two visualizations: a donut chart illustrating the proportion of high-risk areas (21.1%) compared to other regions (78.9%), and a horizontal bar chart ranking the top 10 most contaminated provinces, color-coded by risk level. Quang Tri stands out clearly at the top, with significantly higher estimated UXO levels than any other province.

In addition to Quang Tri, several other provinces also show notable levels of contamination. Quang Binh and Quang Nam fall into the “Very High” risk category, while Tay Ninh, Kon Tum, and Thua Thien Hue are categorized as “High” risk. These findings are consistent with the locations of key military operations along the former DMZ, the Ho Chi Minh Trail, and other strategic regions. The heavy bombing of central Vietnam, particularly along the border with Laos, was aimed at disrupting supply lines and troop movements.

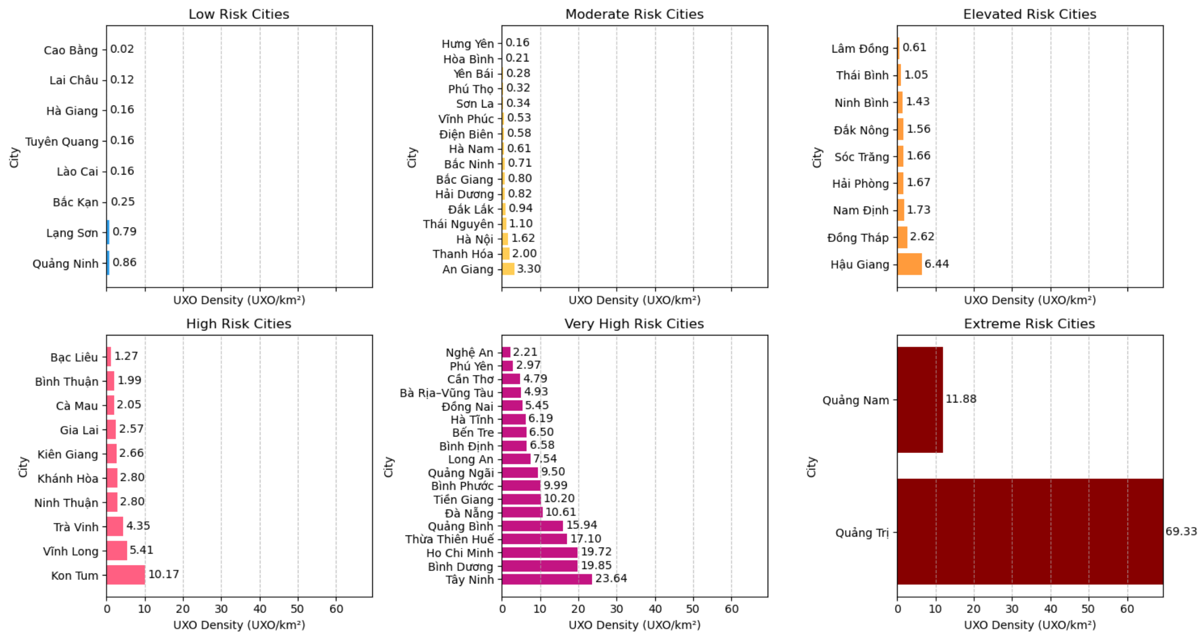


Figure 5.11. Estimated UXO density/km^2 of every City/Province, grouped by risk level

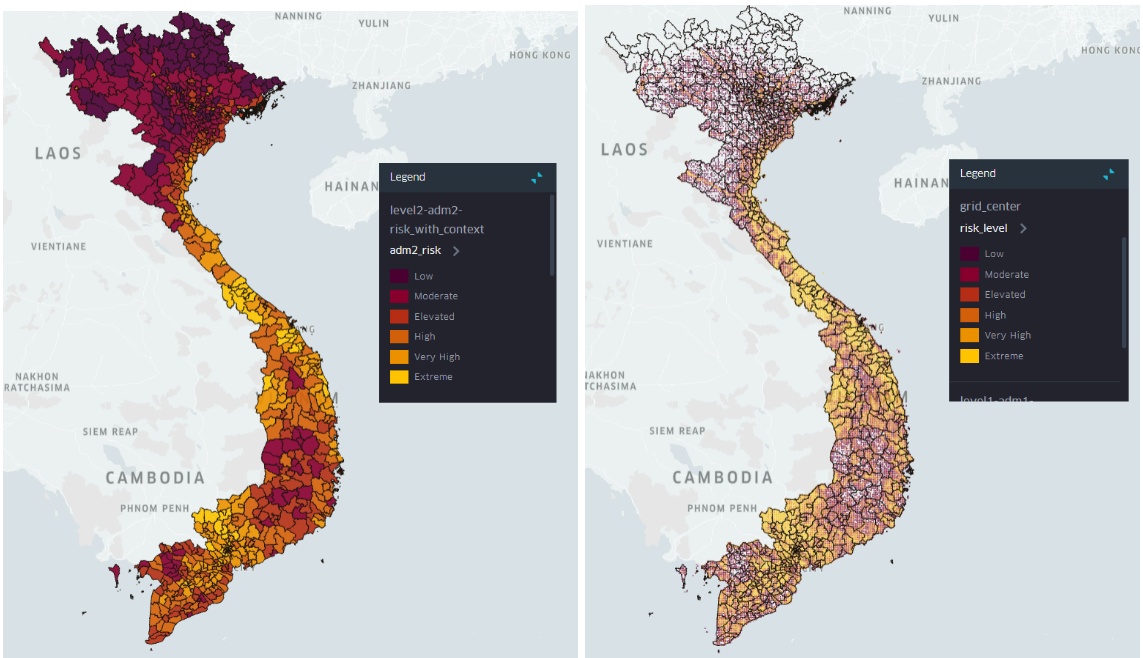


Figure 5.12. District-level (ADM2) map showing UXO risk classification by severity.

Figures 5.10, 5.11, and 5.12 provide different views for UXO risk analysis. Figure 5.11 shows histograms of estimated UXO density per km² by city/province and risk level, guiding selection for detailed study. Figure 5.12 maps district-level (ADM2) UXO risk classification by severity. Figure

5.10 presents a treemap of districts within provinces, where color indicates risk level and size reflects average weighted risk scores; parent node sizes represent the sum of district scores. Figure 5.11 offers an abstract overview, while Figures 5.10 and 5.12 provide spatial insights in hierarchical and geographic formats, respectively.

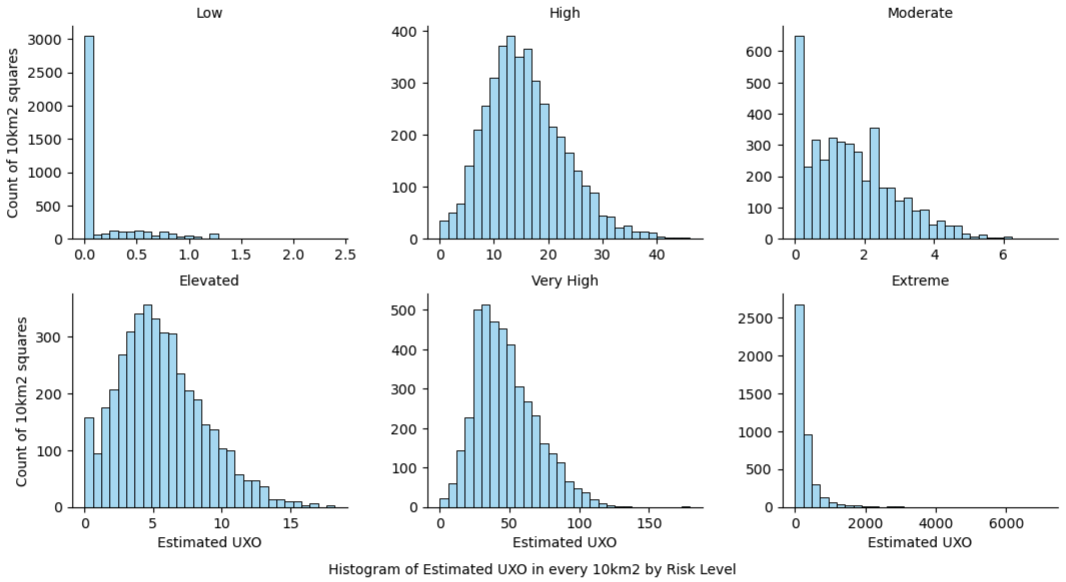


Figure 5.12. Distribution of estimated UXO among $10km^2$ squares grouped by risk level

These histograms show how many unexploded bombs (UXO) are found in different areas based on their risk level. The Extreme risk areas have by far the most UXO, with some spots containing up to 6,000 unexploded bombs in a $10km^2$ area - much higher than anywhere else. Very High risk areas have up to 150 UXO per grid cell, while High risk areas typically have between 10-30 bombs per cell. The less risky areas (Moderate, Elevated, and Low) have progressively fewer bombs, with Low risk zones mostly having less than 0.5 UXO per grid cell. A plausible approach would be to start with the worst-affected spots in Extreme risk areas, then move to Very High and High risk areas, before eventually tackling the less contaminated Low risk zones.

6. Clearance Priority Assessment

Now that we've identified areas with high UXO risk, the next step is figuring out which areas to prioritize for clearance. Risk alone isn't enough to decide this - it's just a metric. Actual clearance priorities depend on many other factors like policy, budget, feasibility, and more. Risk is important and plays a big role, but clearance decisions require a broader perspective. In the previous section, we focused on risk; here, we'll look at how to turn that into actionable clearance priorities.

6.1. Supplementary data

Like I have said, risk alone isn't enough, therefore, I have to expand the dataset. The data I have chosen to include is terrain and population, there are probably more but for the limited resource that I have access to, I find these 2 are the most important.

For the terrain data, I used the High-Resolution Land Use and Land Cover Map of Vietnam, which is derived from satellite imagery processed by the Japan Aerospace Exploration Agency (JAXA). This dataset provides comprehensive coverage of the entire country with an image size of 3264 pixels wide by 6942 pixels high. Each pixel represents an area of 250 meters by 250 meters on the ground, each pixel is labeled with a terrain type among 20 different labels.

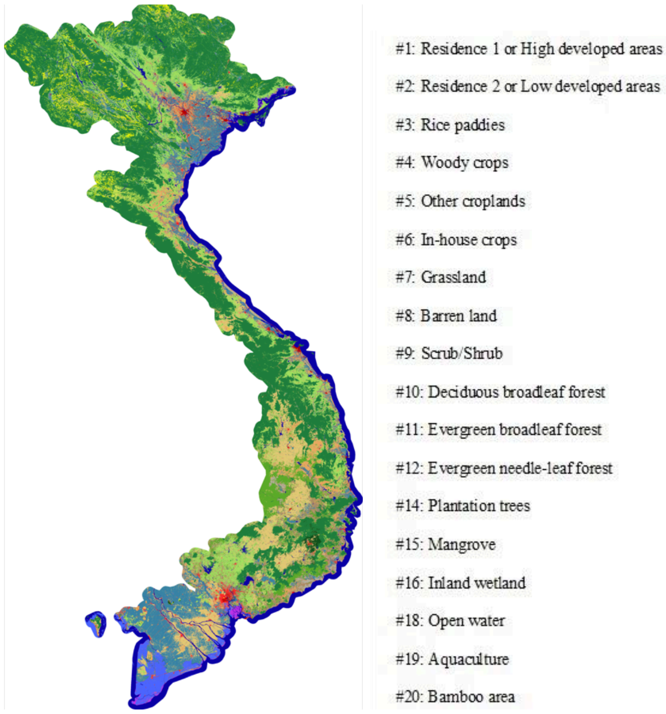


Figure 6.1. shows the Vietnam terrain image from the JAXA land use and land cover dataset alongside its 18 detailed terrain labels (2020)

To integrate this terrain dataset with the grid-level data, I calculated the number of pixels for each terrain type within every 10 km² grid, using the geographic coordinates embedded in the image's metadata. These counts were then normalized to obtain the proportion of each terrain type per grid. This approach provides both normalized terrain composition and allows for estimating absolute area coverage based on the known grid size. The dataset has 18 labels, however, I will reduce to the main types of terrain for analysis simplicity. The main groups would be

Main Terrain Group	Description / Notes
Residence Developed	Built-up areas with infrastructure
Residence	Residential areas, housing zones
Agricultural	Farmland, crops, cultivated fields
Open Terrain	Bare land, open soil, non-vegetated areas
Shrubland	Areas dominated by shrubs and small bushes
Forest	Dense tree-covered areas
Wetlands / Water	Marshes, swamps, lakes, rivers, and other water bodies

Table 6.1. Selected terrain groups

By grouping the data in this manner we can get an insight of the areas of terrains that are contaminated. Figure 6.2. Shows that the majority of contaminated land in Vietnam are forest and agricultural lands.

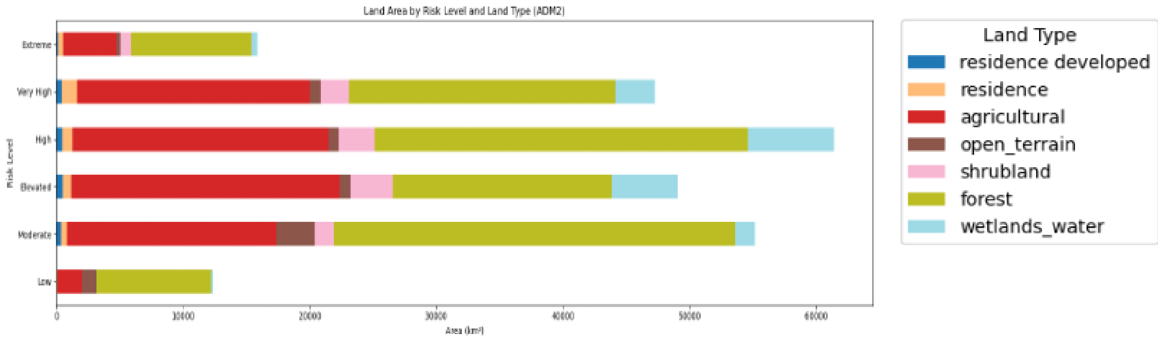


Figure 6.2. Areas of terrain (km^2) exposed to risk of UXO contamination

For population data, I used the most recent publicly available dataset provided by the General Statistics Office (GSO) of Vietnam. This dataset contains official population figures and is considered the most reliable source for national demographic information. However, the data is only available at the ADM1 level, which corresponds to the province or centrally governed city level.

	TỔNG CỤC THỐNG KÊ	
		DÂN SỐ TRUNG BÌNH
MÃ TỈNH	TÊN TỈNH	NĂM 2023
TOÀN QUỐC		100,309,209
01	Hà Nội	8,587,081
02	Hà Giang	899,900
04	Cao Bằng	547,857
6	Bắc Kạn	326,504

Figure 6.3. GSO Vietnam population data in 2023

6.2. Clean up strategy

6.2. Clean-Up Strategy

An effective clean-up strategy ideally requires access to detailed data on past and ongoing UXO clearance activities. Such information would allow us to identify already-cleared areas and assess how well our risk model aligns with real-world demining efforts, enabling further refinement. Unfortunately, much of this data remains classified or is restricted due to security and operational reasons. While some public statistics do exist, they are generally high-level figures intended for summary reporting rather than comprehensive spatial or operational analysis. Despite these constraints, this study makes use of the limited public data available to build a reasonable, data-informed clean-up model. Where exact clearance data is missing, assumptions are made based on observable patterns in risk assessments, population distribution, and terrain features.

6.2.1. Clean up strategy derived from geographical statistics

The primary source of UXO clearance data I found is from the Quang Tri Mine Action Center (QTMAC). While this data includes useful statistics like total area cleared, numbers of bombs and UXOs found by type, and common weapon categories, it has significant limitations. The data is mostly aggregated at a high level without detailed spatial or temporal resolution, making it difficult to pinpoint exactly where clearance has occurred or how effective it has been in specific locations. Additionally, the inclusion of weapons from different historical periods adds ambiguity, limiting its usefulness for current risk modeling or prioritization. For these reasons, although informative, the QTMAC data alone is not sufficient for precise analysis or to validate and improve predictive models effectively.

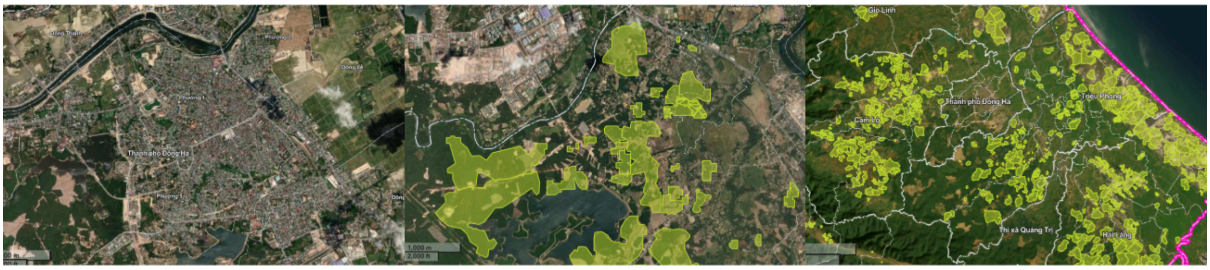


Figure 6.4. QTMAC public UXO clearance data

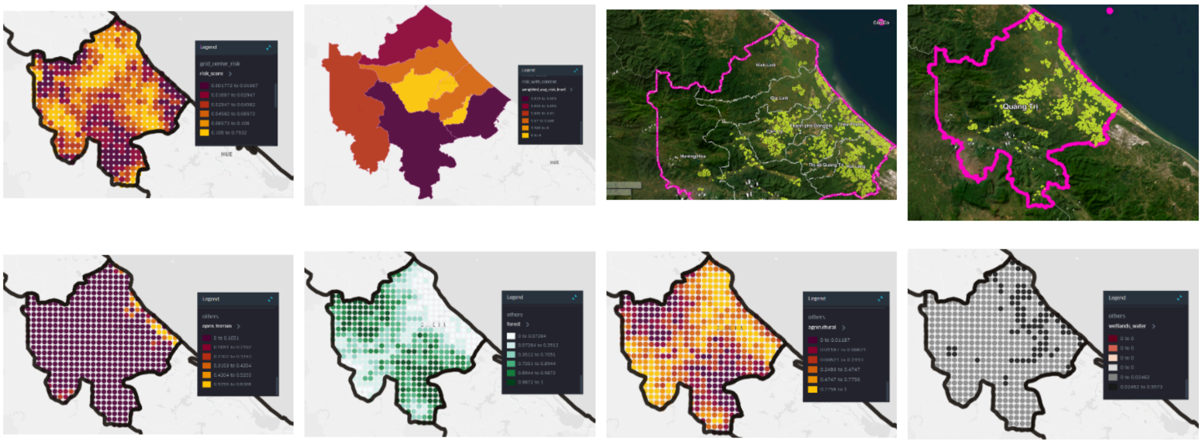
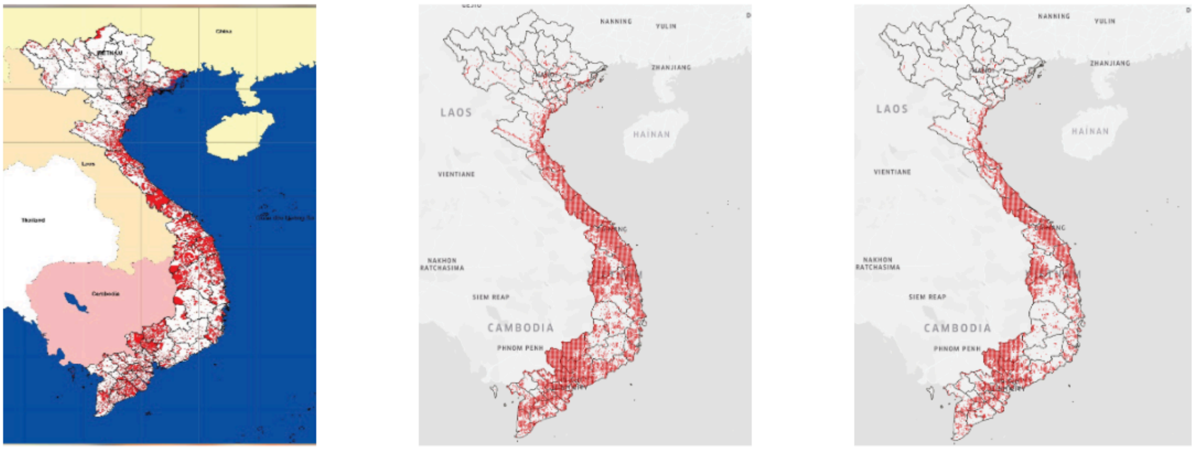


Figure 6.5. Comparison between demining QTMAC demining activities and terrain - risk level overlays in the studied dataset

Based on the map analysis in figure 6.5, it's clear that demining efforts follow a well-thought-out strategy. The teams seem to prioritize areas with high risk, visible civilian activity, and easier access which makes sense given limited resources. Most clearance is focused on residential zones in rural–urban fringe areas. These spots aren't as built-up as big cities but still have lots of people living and working there, making them high priority. The risk maps support this: places marked "Very High" or "Extreme" risk (in red and purple) often overlap with clearance zones. Terrain plays a big role too. Mountainous and remote areas are largely avoided, probably because they're harder to reach, less populated, and more expensive to clear. Instead, teams focus on flatter, more populated areas where the impact will be greatest. In short, this demining strategy balances urgency, practicality, and efficiency starting where help is needed most and where it can be delivered effectively.

Another source that has publicized statistics is Vietnam Mine Action Center (VNMAC) which shows a map of areas with recorded UXO. Figure 6.6. shows the comparison between the plotted points of the studied dataset marked with “Very High” and “Extreme” risk. The comparison shows there's a correlation between the areas marked with “Very HIgh” and “Extreme” risk level with the found UXO locations, indicating that the government has taken care of areas with high risk of UXO.

The VNMAC UXO contamination map shows UXO presence along Vietnam’s northern and western borders, notably in Ha Giang, Lao Cai, and Dien Bien Phu - areas not covered in the U.S. bombing dataset. This reflects contamination from multiple conflicts, including the 1979 Sino-Vietnamese War and the First Indochina War. Some UXOs may also stem from old depots or abandoned caches, highlighting the complex, multi-era nature of UXO risks beyond just the Vietnam War.



VNMAC data | Estimated High, Very High, Extreme | Estimated Very High, Extreme

Figure 6.6. First graph: UXO recorded in Vietnam by VNMAC. Second graph: Locations marked with risk from High to Extreme. Third graph: Locations marked with Very High to Extreme risk.

6.2.2. Clean up strategy based on budget

Another critical aspect of UXO clearance in Vietnam is the cost. Demining efforts are funded through a combination of national budget allocations and international aid. Historically, the Vietnamese government has allocated approximately $30 million per year for UXO clearance operations. In parallel, international partners—particularly the United States—have played a key role in funding. Since 1993, the U.S. has contributed over $234 million to support UXO removal, dioxin remediation, and assistance for people with disabilities (EveryCRSReport, 2020; VnExpress, 2024).

Between 2010 and 2020, Vietnam successfully cleared approximately 50,000 hectares (or 500 km²) of contaminated land. The total cost of this effort exceeded VND 12.6 trillion (roughly $540 million), with VND 10.4 trillion funded by the Vietnamese government and the remainder covered by foreign aid (VnExpress International, 2022). Recently, the Vietnamese government has set a target to clear 500,000 hectares (5,000 km²) by 2025, which translates to an annual clearance goal of 75,000 hectares (750 km²).

Based on the historical spending, the average cost to clear 5000 km² was approximately $540 million, which means the cost to clear 1 km² of UXO-contaminated land is roughly $109,000/km². Using this estimate.

This rough cost estimate offers a practical reference point for understanding the financial scale of UXO clearance and highlights the substantial investment still needed to address the legacy of war contamination across the country. Figure 6.8 groups the areas under risk and shows the corresponding estimated contaminated area and cleanup cost for each risk level. By analyzing the chart, it becomes clear that "High" and "Very High" risk areas are prioritized in clearance strategies, as they strike a critical balance between large contaminated areas and high risk to human safety. From this graph, it should be noted that areas with higher risk and lower areas are prioritized.

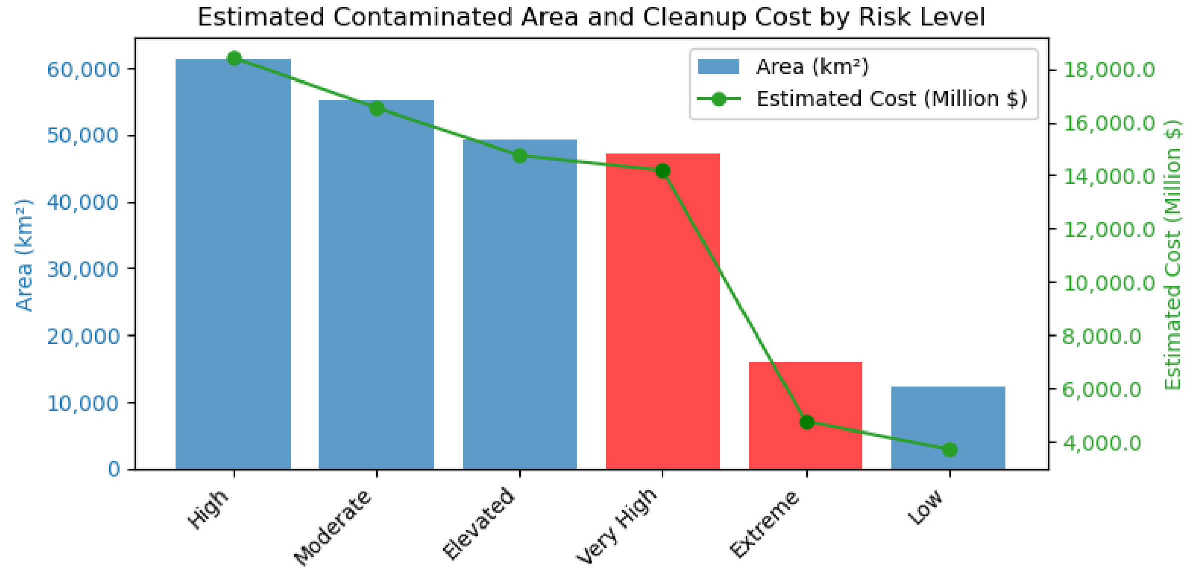


Figure 6.6. Areas (km^2) contaminated by unexploded ordnance (UXO) and the corresponding cleanup cost, grouped by risk level categories: Low, Moderate, Elevated, High, Very High, and Extreme. This chart is based on the assumption that the cleanup cost is about \$109,000/km^2.

6.3. Priority score

To support strategic planning and prioritization in UXO clearance, I propose a simple method to calculate a Priority Score based on several normalized metrics. Each input variable is first rescaled to a consistent range using min-max normalization. This normalization ensures comparability across different scales. Once normalized, the Priority Score is computed as a weighted sum of the selected metrics. The Priority Score is calculated as the dot product of a weight vector (W) and a vector of normalized features (X_{norm}).

$$\textbf{\textit{Priority Score}} = W * X_{norm}$$

Where:

$W = [w_1, w_2, ..., w_{11}]$ represents the weights assigned to each feature

X_{norm} = *[estimated_uxo, population, area, is_hotspot, residence, agricultural, open_terrain, shrubland, forest, wetlands_water, residence_developed]*

Each weight reflects the importance and direction of the corresponding metric in influencing clearance priority. The table below outlines each metric, its associated weight, expected sign, and rationale. Figure shows weights are signs determined by the analysis of QTMAC cleanup strategy and strategy based on price. The importance of the weights can be tuned according to policymakers.

Metric	**Weight**	**Sign**	**Rationale**
Estimated UXO	w_1	> 0	Higher contamination indicates higher risk and urgency
Population	w_2	> 0	More people exposed increases priority
Area (km²)	w_3	< 0	Smaller areas are cheaper and faster to clear
Is Hotspot	w_4	> 0	Hotspot designation implies higher historical danger

Residence	w_5	> 0	Residential presence increases risk and aligns with patterns seen in QTMAC
Agricultural	w_6	> 0	Agricultural areas are often cleared early to support livelihoods
Open Terrain	w_7	> 0	Easier to access and aligns with common clearance patterns
Shrubland	w_8	< 0	Denser vegetation increases difficulty and cost
Forest	w_9	< 0	Similar to shrubland - challenging clearance environment
Wetlands/Water	w_{10}	< 0	Very difficult terrain for UXO operations
Residence Developed	w_{11}	< 0	Likely already cleared or prohibitively expensive to prioritize

Table 6.2. Weights for priority score model derived from public data analysis

6.3. Analysis

With the priority assessment model in place, we can now address a practical policy question: suppose you are a government decision-maker tasked with planning UXO clearance, using available data from VNMAC. The "High Risk" and "Extreme Risk" areas have already been prioritized and addressed. You are now allocated a fixed budget for continued clearance efforts, with an average cleanup cost estimated at $109,000 per square kilometer.

The next step is to determine which areas should be prioritized for clearance. Using the priority model, calibrated to reflect present-day conditions, we can make strategic decisions that optimize impact and resource use. A key adjustment in the model is the treatment of "residence developed" areas. These are assigned a negative weight, based on the rationale that highly urbanized zones are more likely to have undergone previous UXO clearance as infrastructure developed over time. For example, while Ho Chi Minh City was once a heavily bombed area, decades of rapid urbanization have likely resulted in extensive remediation efforts. Therefore, to more accurately reflect today's situation, the model penalizes areas with dense urban development, assuming a lower probability of remaining contamination. This allows the prioritization to shift toward regions where UXO presence remains a more active threat and clearance efforts are both more necessary and potentially impactful. Table 6.3. shows the choice of weights for the situation.

Feature	Weight	Rationale
Estimated UXO	5.0	Higher UXO density implies greater danger
Population	5.0	More populated areas pose higher risk to civilians
Area (km²)	-2.0	Larger areas are more costly; this discourages low-density, expansive zones
Is Hotspot	3.0	Historical hotspots are more likely to have contamination

Residence	2.5	Residential areas are prioritized due to human exposure
Agricultural	1.5	Farmland impacts livelihoods and is often accessible
Open Terrain	1.0	Easier and cheaper to clear
Shrubland	-1.0	Vegetation increases clearance difficulty
Forest	-1.5	Denser coverage complicates operations
Wetlands/Water	-3.0	Hardest terrain to clear; high cost and low accessibility
Residence Developed	-10.0	Assumed previously cleared due to infrastructure development

Table 6.3. Chosen weights for the hypothetical situation.

Which areas should be prioritized for UXO clearance?

```
Risk Level: Elevated
Đồng Tháp: 4.97      | Hậu Giang: 4.90      | Sóc Trăng: 3.62      | Thái Bình: 0.43      | Ninh Bình: -0.19      | Nam Định: -0.60
Đắk Nông: -0.85      | Lâm Đồng: -1.35      | Hải Phòng: -2.70
Risk Level: Extreme
Quảng Trị: 7.72      | Quảng Nam: 2.98
Risk Level: High
Vĩnh Long: 4.97      | Kiên Giang: 3.43     | Trà Vinh: 3.34       | Kon Tum: 3.10        | Gia Lai: 2.35         | Bình Thuận: 1.85
Khánh Hòa: 1.72      | Ninh Thuận: 1.70     | Cà Mau: 0.51         | Bạc Liêu: -0.87
Risk Level: Low
Lai Châu: -0.79      | Lạng Sơn: -0.95      | Lào Cai: -1.00       | Hà Giang: -1.10      | Tuyên Quang: -1.16    | Cao Bằng: -1.34
Bắc Kạn: -1.51       | Quảng Ninh: -3.55
Risk Level: Moderate
An Giang: 2.66       | Bắc Giang: 0.16      | Sơn La: -0.10        | Điện Biên: -0.25     | Phú Thọ: -0.31        | Vĩnh Phúc: -0.68
Thanh Hóa: -0.85     | Hà Nam: -0.90        | Đắk Lắk: -1.01       | Hưng Yên: -1.10      | Hòa Bình: -1.11       | Thái Nguyên: -1.27
Yên Bái: -1.49       | Hà Nội: -2.57        | Hải Dương: -2.63     | Bắc Ninh: -3.49
Risk Level: Very High
Long An: 5.73        | Cần Thơ: 5.57        | Tiền Giang: 5.55     | Ho Chi Minh: 4.65    | Tây Ninh: 4.30        | Bình Dương: 3.99
Quảng Bình: 3.98     | Bến Tre: 3.60        | Đồng Nai: 3.16       | Hà Tĩnh: 3.15        | Bình Phước: 3.01      | Thừa Thiên Huế: 2.34
Quảng Ngãi: 2.11     | Bình Định: 2.02      | Phú Yên: 1.76        | Bà Rịa-Vũng Tàu: 1.65 | Nghệ An: 0.46        | Đà Nẵng: -7.40
```

```
risk_level
Elevated     0.913199
Extreme      5.348366
High         2.210249
Low         -1.425295
Moderate    -0.932857
Very High    2.758319
```

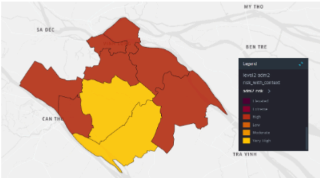
Figure 6.9. Shows the result of the priority assessment model and the mean priority score distribution of each risk level.

By examining the scores presented in Figure 6.9, and assuming that areas classified as 'Very High' and 'Extreme' risk have already been addressed, I focus on identifying priority regions within the 'High' risk category. Among these, Vĩnh Long emerges with the highest risk score (4.97). Therefore, Vĩnh Long is selected as the next target area for clearance.

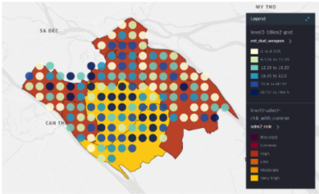
If areas under 'Extreme' and 'Very High' risk had not been cleared, they would naturally take precedence due to their greater urgency. However, given the assumption that these areas have been handled, resources can now be directed toward the highest remaining risks.

District	Risk	Est. UXO	Area (km^2)	Clearance Cost ($)
Mang Thit	High	225.08	170.00	18,666,000
Vung Liem	High	1016.47	300.00	32,940,000
Tam Binh	Very High	2509.78	290.00	31,842,000
Vinh Long	High	161.17	40.00	4,392,000
Binh Tan	High	727.78	180.00	19,764,000
Long Ho	High	528.21	190.00	20,862,000
Tra On	Very High	2013.04	270.00	29,646,000
Binh Minh	High	145.52	80.00	8,784,000
Total			1520.00	166,896,000

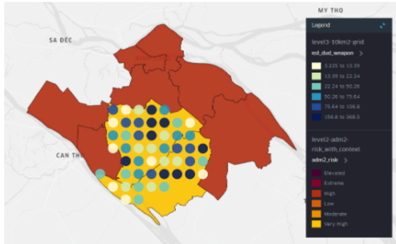
Based on previous cost estimates for similar regions, the expected cost for UXO clearance in Vĩnh Long is approximately $166,896,000 million. This figure provides a basis for policymakers to assess the feasibility and allocate necessary funding for UXO risk mitigation in the region.



Risk levels inside Vinh Long province



Risk levels of districts in Vinh Long overlaid with grid risk level



Risk levels of districts in Vinh Long overlaid with Tra On and Tam Binh 10km2 risk levels

Figure 6.10. Areas in Vinh Long's districts with high risk score by Level 2 and Level 3

In order to verify the score integrity, I have scraped data related to UXO in Vietnam from multiple sources of papers namely VNexpress, Dan Tri, Bao Hanoi, Bao Thanh Nien, and Bao Tuoi Tre, the scraped database is not very large so the verification method is performed manually by searching occurrence of a city or district name.

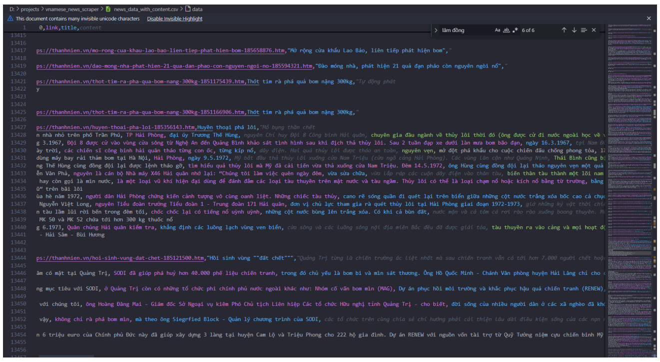
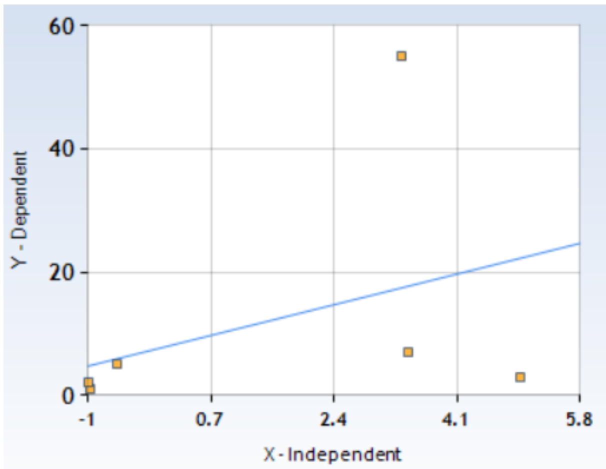


Figure 6.10. Vietnamese news data with 1300 entries

City	Risk level	Priority Score (X)	Mention count (Y)
Vĩnh Long	High	4.97	3
Kiên Giang	High	3.43	7
Trà Vinh	High	3.34	55
Lai Châu	Low	-0.97	1
Lạng Sơn	Low	-0.59	5
Lào Cai	Low	-1.0	2



The table shows that the correlation between calculated priority score and mention on UXO report can be used to improve the model to overcome the challenge of not having access to UXO demining activities.

7. Conclusion

This study has explored the complex relationship between historical U.S. bombing data and self-reported UXO (Unexploded Ordnance) risk in Vietnam, with a focus on the validity and reliability of the Vietnam Development Assessment (VDA) dataset. Despite the absence of a clear statistical correlation between bombing density and reported risk scores, our analysis indicates that local perceptions are likely influenced by other nuanced factors, such as community memory, land use patterns, and ongoing clearance efforts.

While the VDA dataset may not perfectly align with historical bombing records, it still offers value as a localized expression of risk perception. Rather than dismissing the FDA's subjective scoring, future UXO risk models might benefit from integrating both historical data and community-based assessments. This dual approach could enhance the targeting of clearance operations and resource allocation.

Further research should investigate additional explanatory variables and employ more robust geospatial models to better understand the determinants of UXO risk perception. Ground-truthing perceived risk with actual contamination remains essential to validate and refine these models. Ultimately, bridging quantitative and qualitative insights will be key to developing more effective, community-informed UXO mitigation strategies in Vietnam and similar post-conflict contexts.