Regional risk assessment Part I: prioritization of potentially contaminated sites

Giulia Minolfi[1,2], Fabio Rosada[2], Elisa Giubilato[2], Alex Zabeo[2], Stephane Belbèze[3], Virginie Derycke[3], Timo Tarvainen[4], Isadora Antoniano Villalobos[1], Elena Semenzin[1], Lisa Pizzol[2]

[1]*Department of Environmental Sciences, Informatics and Statistics (DAIS, Ca' Foscari University of Venice, Venice, Italy*
[2]*Greendecision Srl, Venice, Italy*
[3]*Bureau de Recherches Géologiques et Minières (BRGM), Orléans, France*
[4]*Geologian Tutkimuskeskus (GTK), Espoo, Finland*

Corresponding author

Lisa Pizzol – lisa.pizzol@greendecision.eu
Greendecision Srl, Venice, Italy

Editor's Note: This article is part of the special series "Environmental Risk Assessment of Soils: From Regulatory Frameworks to Realistic On-Site Risk Predictions of Contaminants, Mixtures and Stressors for the Environment." This joint special series (ET&C and IEAM) brings together contributions on environmental risk assessment frameworks for soil organisms exposed to chemicals, as well as research on new ecotoxicological and risk assessment advancements. The series also includes monitoring studies describing soil contamination and the associated risks for soil organisms.

Acknowledgments

The research has received fundings from the European Union through the Horizon Europe project ISLANDR (Grant agreement No. 101112889). This work was supported by the DoE 2023-2027 (MUR, AIS.DIP.ECCELLENZA2023_27.FF project).

Data availability statement

Data for case study application:
- BASIAS national database of polluted sites: https://www.data.gouv.fr/datasets/inventaire-des-sites-pollues
- Institut national de la statistique et des études économiques - INSEE (2023). Population en 2020 – Recensement de la population – Base infracommunale (IRIS). Institut national de la statistique et des études économiques. Available at: https://www.insee.fr/fr/statistiques/7704076
- Muséum national d'Histoire naturelle (MNHN). (2024). Réseau des Réserves Naturelles Nationales (RNN) [Shapefile].
- Météo-France. (2023). Climatological data for the Toulouse region [Data set]. Météo-France. https://meteofrance.com
- Observatoire Régional de l'Environnement Occitanie (OREO) / OpenIG. (2022). Espaces Naturels Sensibles (ENS) Occitanie [Shapefile].

- Région Occitanie. (2024). Réserves Naturelles Régionales (RNR) [Dataset].
Sandre. (2025). Atlas des masses d'eau souterraines [Data set]. Service d'Administration Nationale des Données et Référentiels sur l'Eau. Retrieved from https://www.sandre.eaufrance.fr/atlas
- Toulouse Métropole. (2025). Open environmental and meteorological data portal [Data set]. Toulouse Métropole. - https://data.toulouse-metropole.fr/explore/dataset/01-station-meteo-toulouse-meteopole/information/
- World Bank Group, & Technical University of Denmark (DTU). (2023). Global Wind Atlas 3.0 [Data set]. Retrieved from https://globalwindatlas.info
- National French database of polluted and potentially polluted sites (BASOL): https://www.georisques.gouv.fr/donnees/bases-de-donnees/sites-et-sols-pollues-ou-potentiellement-pollues

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Regional risk assessment Part I: prioritization of potentially contaminated sites

Abstract

Soil contamination remains a critical barrier to endurable land use and environmental health worldwide. A stepwise and risk-based approach to identify, assess and manage potentially contaminated sites across complex land uses, is presented. Unlike traditional site-specific risk assessments, which are often limited to detailed studies of individual locations, this large-scale approach (known as regional risk assessment - RRA) integrates diverse data sources - historical or current data on anthropogenic activities - with pathways relevance information and receptors vulnerability to support prioritisation when dealing with a very large number of sites. The methodology is implemented within a GIS-based framework, and integrate spatial data on hazards, pathways, and receptors to produce regional-scale risk maps. The output provides a relative ranking of potentially contaminated sites based on estimated risk levels, supporting the prioritisation of sites requiring further site-specific investigation or remediation. Applied to the metropolitan area of Toulouse, the framework demonstrates its applicability across urban and peri-urban landscapes. This methodology responds to emerging regulatory requirements, such as the Soil Monitoring and Resilience Directive, and offers practical support to strategic planning efforts aligned with EU policy initiatives. By enabling more efficient and risk-based site management, the framework contributes to the protection of human health and the land resources.

Keywords

Soil contamination; Large scale risk assessment; Risk-based prioritization; GIS-based spatial analysis; land use planning.

1. Background

Soil plays a crucial role in supporting both human life and ecosystems. A high-quality soil contributes to food security, clean water provision, sustainable agriculture, biodiversity conservation and climate change mitigation (EC, 2023). However, a wide range of human activities - including industries, waste disposal in uncontrolled landfills, mining, the use of agricultural chemicals, sewage sludge, and livestock waste, environmental accidents - has left a legacy of potentially contaminated sites all around the world (Ramón & Lull, 2019). Key contaminants include polycyclic aromatic hydrocarbons (PAHs), heavy metals, persistent organic pollutants (e.g. PCBs, dioxins), radionuclides, and pathogenic microorganisms are the main soil contaminants causing concerns. More recently, new emerging soil contaminants, such as nanomaterials, veterinary medicines, microplastics and surfactants, have raised further concerns. All these substances, both the well-known and the new emerging contaminants, may persist in soils, leading to long-term ecological degradation and spread through air, surface water and groundwater, leading to cross-borders impacts on ecosystems and human health. In some cases, contaminants are taken up by plants and may bioaccumulate in the food chain, compromising the safety of the food consumed by both humans and animals (FAO, 2018).

The systematic identification of potentially contaminated sites, combined with risk-based remediation decision-making, is widely recognized as a necessary step to public health and guide sustainable land management (Bardos et al., 2018; Nathanail & Bardos, 2004). Rather than assuming that all sites require immediate remediation, such approaches enable proportional and targeted actions where risks are likely to be unacceptable. European policy has increasingly emphasized soil protection, as evidenced by various soil-focused initiatives and strategies, including the European Green Deal, the Biodiversity Strategy, and the new EU Soil Strategy. In this context, the Soil Monitoring and Resilience Directive (SML) highlights the urgence to managing both point-source and diffuse contamination, with the goal of reducing soil pollution to non-harmful levels by 2050. The Directive (EU 2025/2360) on soil monitoring and resilience was formally adopted by the European Parliament and the Council in 2025, published in the Official Journal on 26 November 2025, and entered into force on 16 December 2025. It represents a milestone for soil protection in the EU: Member States now have a three-year period to transpose its provisions into national law and to implement risk-based approaches for identifying, investigating, and managing potentially contaminated sites (PCSs), including the definition of thresholds for unacceptable risks to human health and the environment. This process primarily concerns generic or screening-level risk assessment, which enables consistent prioritization and proportional management actions across the EU.

The scale and complexity of this task are considerable (Pérez et al., 2018). Since comprehensive soil and groundwater investigations cannot feasibly cover all EU land in a short timeframe, especially with limited resources, priority should be given to areas where the likelihood of contamination - and associated risk - is higher. In response, a new methodology for regional risk assessment (RRA) designed to identify and rank PCSs at regional scale, has been developed, especially for areas where detailed contamination data are lacking. The aim is not to replace site-specific assessments, but rather to support them by targeting locations that warrant further investigation.

Risk assessment methods for potentially contaminated sites have been developed by several national environmental agencies (e.g., APAT, 2008; DEFRA, 2011; USEPA, 1991, 2005; Health Canada, 2010a, b). At the site-specific level, risk is often expressed as the product of a hazard (the intrinsic toxicity of contaminants) and exposure (their environmental concentrations and the probability that receptors - human or ecological - come into contact with contaminants through specific pathways). These assessments typically rely on measured concentrations and detailed exposure modelling, resulting in quantitative risk indices (Chen et al., 2022). They provide absolute evaluations of risk and are tailored to individual receptors and land uses.

At a regional scale, however, applying these detailed methods is neither feasible nor appropriate, given the broader scope and different data availability. In regional applications, risk assessment often relies on simplified or generalized assumptions about exposure scenarios and land use, reflecting the limited availability of detailed site data at large scale. Consequently, such assessments typically aim to compare relative risk levels across the study area, rather than to quantify absolute risk levels. This distinction is crucial: while site-specific assessments aim to determine whether a specific location poses an unacceptable risk, RRA methods are designed to rank a large number of PCSs with a large range of level of risk, to support prioritization decisions in a stepwise approach as suggested in the SML; the evaluation of the potential risk involved in such prioritisation is much more generic than the site-specific risk assessment that is carried out on a contaminated site. This comparative approach has been widely adopted in frameworks such as the Relative Risk Model (Landis & Wiegers, 1997; Landis, 2005) and the PRA.MS methodology (EEA, 2005). Accordingly, regional assessments rely on simplified but robust frameworks that can guide further investigations or policy responses (Hunsaker et al., 1990; Landis, 2005).

Our methodology builds on those well-established methods (Pizzol et al., 2011, EEA 2005a, EEA 2005b), that has been extensively applied in recent years (Pizzol et al., 2016, Jiang et al., 2024, Hammond et al., 2021, Li et al., 2022) and that estimate the relative risk score of PCSs through three components: Source-Pathway-Receptor (S-P-R). Within this framework, the Source represents the potential source of contamination (hazard); the Pathway describes the potential connectivity and migration of contaminants through environmental media; and the Receptor identifies the potentially exposed targets. The methodology produces a relative risk-based ranking of sites to identify locations may require further site-specific investigation and to support the efficient allocation of efforts and resources, towards areas with the highest potential risk. The method includes a refinement of the PRA.MS scoring system (EEA, 2005), updated to align with the current European CLP Regulation (EC No 1272/2008), thereby ensuring consistency with hazard classification based on the Globally Harmonised System (GHS). This updated hazard evaluation captures substance-specific behaviour across environmental compartments (soil, water, air), enhancing the accuracy of risk estimations.

The methodology has been applied to a complex case study: the metropolitan area of Toulouse, France, which includes 37 municipalities with diverse land uses - agricultural, forested, industrial, commercial, and transportation corridors. This area presents a range of environmental pressures and contamination legacies, making it a suitable testbed for evaluating the applicability and scalability of the methodology. The objective is to prioritize PCSs, defined as those with historical or ongoing human activities that may have caused pollution (Van-Camp et al., 2004). Where official datasets are incomplete, auxiliary sources such as local knowledge, historical archives, land use records, and site inspections provide supporting evidence.

While many existing approaches are either descriptive or inventory-based, our framework is explicitly designed as a decision-support tool. It offers a pragmatic solution to the challenge of prioritizing contaminated land at regional scale and aligns with EU goals for environmental restoration and pollution prevention.

2. Methodology of large-scale risk assessment for potentially contaminated sites

The proposed methodology for large-scale risk assessment (RRA) of PCSs is structured in five sequential steps (Figure 1) to support prioritisation of risk identification, evaluation and management in a stepwise approach. The following sections describe each step in detail.

2.1 Definition of the conceptual model

The first step of the RRA involves defining the conceptual model, by identifying:

(i) the potentially contamination sources (sub-step 1.1);
(ii) the main release mechanisms and direct transport pathways (sub-step 1.2), and
(iii) the potential receptors which may be in contact with the released contaminants (sub-step 1.3).

Sub-step 1.1- Identification of potentially contamination sources

Potentially contamination sources are identified based on the type and location of anthropogenic activities with known or suspected polluting potential. These include active and historical industrial sites, mining areas, waste disposal facilities, and sites affected by environmental incidents. The identification process draws on multiple data sources, including land use records, industrial permits, accident reports, and regulatory inventories such as activities listed in Annex I of Directive 2010/75/EU, Directive 2012/18/EU, and Annex III of Directive 2004/35/CE.

The methodology considers both ongoing and past activities, recognising that legacy pollution can result in current diffuse contamination. Consistent with international guidance (ASTM, 2021; USEPA, 2020; ISO, 2018) and aligned with the risk-based principles of the EU Soil Monitoring and Resilience Directive (EU 2025/2360), all sites with potential contamination risk are included, even if direct measurements are not yet available. This sub-step supports SML Article 15 (2) that requires Member States to lay down rules concerning the timeframe, content, form and prioritisation of soil investigations.

Sub-step 1.2 – Identification of Transport Pathways

Once the sources of potential contamination have been identified, the next step involves the environmental pathways through which contaminants may travel from the source to human or ecological receptors. We adopted a simplified, yet operational modelling approach tailored to large-scale applications, incorporating the key environmental pathways. These include leaching into groundwater, runoff into surface water, volatilisation into the air, and direct contact with contaminated soil. Only the direct emission pathways are considered and not the intermediate pathways, such as deposition from air to soil/water or transfer from groundwater to surface water. This choice reflects the screening-level objective of the framework and the regional scale of analysis, where site-specific data are typically unavailable. Contaminant mobility is influenced by soil characteristics such as porosity and texture, which determine whether contaminants are retained, degraded, or mobilised. The behaviour of contaminants through these pathways is partially captured in the intrinsic hazard scores assigned to each source (paragraph 1.1.3.), which differentiate potential impacts across four transport media. Figure 2 illustrates the main transport mechanisms considered in the methodology.

Sub-step 1.3 – Identification of Receptors

Potential receptors include:

(i) humans, that can be exposed to contaminants in soil through ingestion (inadvertent eating), dermal exposure (skin contact) or inhalation (breathing);
(ii) groundwater bodies;
(iii) surface waters;
(iv) protected natural areas, where sensitive ecological receptors are expected to be present.

Groundwater and surface water are treated both as receptors and as transport media. Ecological receptors are integrated during the vulnerability analysis phase but are not tracked through complex contaminant transport chains (e.g., groundwater to lake systems), in order to maintain manageability at a large scale.

2.2 Hazard analysis

The hazard analysis quantifies the potential hazard posed by each identified site, based on the characteristics of anthropogenic activities historically or currently performed at the location. Where multiple activities are present on a single site, their hazards are estimated individually and then aggregated. Only the intrinsic characteristics of the PCSs are included in the hazard analysis.

The hazard score for each PCS (S) is composed of the following aspects:

- the intrinsic hazard of the activity, based on the substances potentially produced and released into the environment;
- a dimension A_s, expressed as the area of the PCS or estimated waste volumes;
- a duration in time T_a, defined as the number of years the activity was operational; we assumed a maximum time window of 100 years (1924 – 2024) following the reference methodologies (Pizzol et al., 2011).

A key innovation of the methodology lies in the intrinsic hazard estimation of activities. We used a database developed by BRGM (2022) that correlates 345 NAF/NACE-coded activities (*Nomenclature of Economic Activities*, the European statistical classification of economic activities) with 3501 hazardous substances, exploiting for each activity and substances an emission probability. NAF codes from the French database were matched to the corresponding NACE codes. Specifically, the French NAF classification system is structured into three hierarchical levels of detail, all of which were available in the dataset provided by the French Geological Survey (BRGM). The first two levels of the NAF code correspond directly to the NACE classification system, while the third level provides a more detailed specification of the French business context. To ensure consistency and avoid mismatches between highly detailed classifications, the conversion was performed at the second level of the NAF hierarchy, corresponding to the second level of the NACE code. Although this choice introduced a simplification in the hazard assessment, it remains consistent with the screening-level objective of the framework.

The PRA.MS procedure (EEA, 2005) was adopted to assign hazard levels to each substance according to its toxicity (Risk phrases). A given substance can be associated with more hazard levels depending on the pathway/route of exposure characteristic of that chemical. The risks phrases used in the PRA.MS have been associated with the new Globally Harmonized System (GHS) hazard classes, ensuring compliance with current EU hazard classification (CLP regulation (CE 1272/2008)). The intrinsic hazard of each activity was then calculated as the sum of the products of emission probability p_u and PRA.MS hazard score h_u for all the substances associated with that activity. Different hazard scores are assigned depending on the environmental pathway, allowing the framework to partially reflect differences in contaminant behavior and exposure potential. Although substance-specific physico-chemical parameters are not explicitly modelled, their influence is partially reflected through the PRA.MS hazard scoring system, which differentiates hazard contributions according to the considered exposure pathway (groundwater, surface water, air, direct contact). As a result, this procedure yields four pathway-specific hazard values for each activity.

The adopted PRA.MS risk phrase scoring system includes mainly toxicological hazards reflecting the original scope of PRA.MS, which focused on human health and general ecological endpoints. The PRA.MS does not provide hazard scores for risk phrases related to physical hazards (e.g., flammability or explosivity) and some specific ecotoxic effects (e.g., toxicity to soil organisms, flora, fauna, or bees) and this may result in activities being assigned zero hazard despite their potential environmental relevance. To avoid this underestimation and include both environmental and human-health related effects in the hazard assessment, a conservative fallback mapping was defined (S1 in supplementary material). Physical hazards (e.g., R1–R19, R30, R44, R47) were assigned the minimum non-zero score (10), acknowledging their indirect contribution to contamination through accidental release. Aquatic toxicity classes (R50–R53) were aligned with CLP/GHS categories (EC No 1272/2008; ECHA, 2017), following a progressive scale (10–30). Specific ecotoxicity phrases (R54–R56, R57) and ozone-depleting effects (R58–R59) were assigned intermediate scores (20), reflecting their ecological significance in line with ISO (2018) and recent guidance on soil and ecosystem protection (EFSA, 2013). This precautionary approach ensures that activities linked exclusively to such risk phrases are not incorrectly classified as having zero intrinsic hazard, while remaining consistent with the PRA.MS scoring framework.

The activity-specific intrinsic hazard H_a is calculated by aggregating the probability and hazard of all the substances associated with it (Equation 1):

$$H_{a,e} = \sum_{u \in a} p_u \cdot h_{u,e}$$

Equation 1

Where:

- $H_{a,e}$ is the intrinsic hazard of a specific activity a among the 345 identified in the BRGM matrix for a specific exposure pathway e;
- p_u is the emission probability of substance u belonging to the set of substances potentially released by the anthropogenic activity a;
- $h_{u,e}$ is the hazard score of substance u potentially released by the anthropogenic activity a which is estimated from the PRA.MS procedure for each exposure pathway e (EEA, 2005).

If in a PCS, more than a single activity a has occurred over time, only the highest H_a is considered for the further calculations. The overall hazard score Hs_e for a PCS is then computed by weighing each activity's intrinsic hazard, on an exposure pathway e base, by its duration and area:

$$Hs_e = \max_{a \in s}(H_{a,e} \cdot T_a \cdot A_s) \qquad \textit{Equation 2}$$

Where:

- s is the PCS;
- a are all the activities that take place in the PCS s;
- Hs_e is the hazard score of the PCS S for the exposure pathway e;
- $H_{a,e}$ is the hazard of the activity a, relative to the exposure pathway e, as seen in equation 1;
- T_a is the time duration of activity a in PCS s;
- A_s is the area of the PCS s.

Equation 2 provide a specific source hazard score for each of the four exposure pathways: Direct contact, Air, Surface water and Groundwater. The duration of the activity is included as a parameter contributing to hazard estimation, allowing partial differentiation between sites within the same activity class. Longer operational periods may be associated with a higher likelihood of contaminant release and accumulation, and may indirectly reflect historical management practices, which were often less regulated. However, this parameter represents only a simplified proxy and does not fully account for site-specific management conditions.

To preserve the site-specific information of each PCS while obtaining a continuous spatial representation in GIS (Esri, 2025), for each PCS the Euclidean distance function was applied and the related map was developed. Accordingly, a raster was generated for each PCS in which each cell represents the distance from the PCS point. The resulting raster was then multiplied by the hazard score (Hs_e) of the PCS as a constant. The calculation was restricted to a maximum radius of 5 km, beyond which the influence of the contamination is assumed to be negligible. This threshold aligns with empirical findings showing that exposure estimates best correlate with pollutant levels when considering sources within approximately 5 km (Tenailleau et al., 2024). The subsequent steps of the RRA process were then executed in GIS through an automated loop for each PCS.

2.3 Pathway analysis

The movement of contaminants from point sources to receptors involves complex environmental processes, including leaching, runoff, volatilisation, and atmospheric dispersion. In this methodology, we adopt a simplified, yet operational modelling approach tailored to large-scale applications, incorporating the key environmental pathways relevant across diverse land uses, including industrial, agricultural, and urban areas.

The behavior of contaminants through these pathways is partially captured in the intrinsic hazard scores assigned to each source (Section 2.2), which differentiate potential impacts across four transport media: groundwater, surface water, air, and direct soil contact. In the next sections, pathways specific assessments will be presented.

2.3.1 Groundwater Pathway

This pathway models contaminant transport via vertical leaching into the saturated zone. Contaminants may percolate through the unsaturated soil layer and reach the aquifer, which can further act as a secondary transport medium. Water and solutes could also flow through the saturated zone until eventually reaching their discharge point, such as a spring (Lerner et al., 2009).

The infiltration process is influenced by soil properties, land cover, terrain slope, and precipitation patterns (USEPA, 1996). Areas with high permeability, low slope, and permeable land uses (e.g. vegetated surfaces) are more prone to infiltration than urban or compacted soils. Conversely, steep slopes and low-permeability soils favor surface runoff.

The groundwater pathway score for each raster cell is calculated as a function of three key spatial parameters:

- Slope (derived from a Digital Elevation Model – DEM);
- Soil infiltration capacity index;
- Land use.

These parameters are combined according to Equation 3:

$$Pg = S \cdot IC \cdot L \qquad \textit{Equation 3}$$

Where:

- Pg is the groundwater pathway score;
- S is the slope map calculated from the DEM;
- IC is the infiltration capacity index map;
- L is the normalised land use map.

2.3.2 Surface Water Pathway

Surface water contamination occurs through runoff and mobilization of contaminants during rainfall events, especially from impermeable surfaces or areas with low infiltration capacity (Braud et al., 2017). A flow accumulation map is derived from the terrain model (DEM) to simulate overland flow routes and identify zones where water (and associated contaminants) is likely to concentrate.

GIS tools are used according to the following steps:

1. Compute flow direction from the DEM;
2. Weight the flow accumulation with mean annual precipitation to account for hydrological intensity;
3. Initiate flow accumulation from PCS locations, to simulate contaminant transport from each source towards surface water bodies by integrating the outcomes of the previous steps.

Equation 4 defines the surface water pathway score:

$$Ps_s = Fa_s \qquad \textit{Equation 4}$$

Where:

- Ps_s is the map of surface water pathway for source s;
- Fa_s is the flow accumulation map for source s.

2.3.3 Air Pathway

The air pathway considers the potential atmospheric transport and dispersion of contaminants from the PCS through the atmosphere via two mechanisms:

- Volatilisation of gaseous or semi-volatile compounds;
- Particle-bound transport, where contaminants adhere to airborne dust or aerosols.

This pathway is particularly relevant for persistent contaminants that can remain in particulate form and undergo atmospheric resuspension (Csavina et al., 2011; Prabhakar et al., 2014). Atmospheric transport is often long-range and influenced by climatic and meteorological conditions (Dragović et al., 2014). We model this pathway using:

- Pathway-specific hazard score;
- The De Martonne Aridity Index, which integrates annual temperature and precipitation, to refine dispersion potential (Minolfi et al., 2018; Punia, 2021).

The De Martonne Aridity Index (AI) is evaluated in each point using meteorological data (Minolfi et al., 2018). Other climatic factors such as temperature and precipitation, in fact, significantly influence the paths of contaminants' dispersal (Punia, 2021). The AI is estimated by the following formula:

$$AI = \frac{M}{T + 10}$$ *Equation 5*

Where:

- AI is the De Martonne Aridity Index;
- M is the mean annual precipitation map (mm/yr);
- T is the mean annual temperature map (°C).

Deposition processes from air to soil are not explicitly modelled in the current framework.

2.3.4 Direct Contact

This pathway represents the potential for human exposure through direct interaction with contaminated soil, particularly relevant in urban or agricultural areas where soil access is common. No intermediate transport process is considered; risk is determined solely by the spatial overlap between potential sources and human receptors.

Hazard propagation is modelled using Inverse Distance Weighting (IDW) from the contamination source, with no additional environmental modifiers.

2.4 Vulnerability analysis

Receptor vulnerability to contamination from PCSs is defined as the susceptibility of a receptor to an imposed contaminant load, determined exclusively by its intrinsic characteristics (Pizzol et al., 2011) and independently from the source hazard and potentially contamination level. The methodology used to estimate the *receptor vulnerability score* follows the multi-step procedure described in Zabeo et al. (2011) and consists of five steps:

1) Identification of **regional receptors**;
2) Selection of **attributes** relevant for assessing receptor sensitivity to contamination;
3) Spatial assignment of attribute values to receptors using **GIS analysis**;
4) **Normalization** of attribute values to a common scale.

Steps 2, 3 and 4 are well described in Section 2.6.

The vulnerability score Vr for a receptor type *r* is calculated by integrating its attribute through a receptor-specific vulnerability function Fv applied to each raster cell representing the receptor (Equation 6):

$$Vr = Fv(Cr)$$ *Equation 6*

Where:

- Vr is the receptor's vulnerability score;
- Fv is the vulnerability function;
- Cr are the selected receptors' attributes.

The analysis considers four receptor categories:

- Groundwater;
- Surface water;
- Human health;
- Protected areas.

The selected attributes for each receptor are listed in Table 1.

2.4.1 Groundwater Vulnerability

Groundwater vulnerability is influenced by proximity to point sources and its role in potable water supply. The score Vg is calculated from groundwater quality and use parameters (Equation 7).

$$Vg = Q \cdot U \quad \text{Equation 7}$$

Where:

- Vg is the groundwater vulnerability score;
- Q is the groundwater quality parameter;
- U is the groundwater use parameter.

Areas where groundwater is used for drinking water supply or irrigation receive higher scores due to direct human and agricultural exposure risks.

2.4.2 Surface Water Vulnerability

Surface water vulnerability reflects its ecological and human use value and its hydrological characteristics. The score Vs is calculated as:

$$Vs = O \cdot Q \cdot U \quad \text{Equation 8}$$

Where:

- Vs is the surface water vulnerability score;
- O is the surface water order parameter;
- Q is the surface water quality parameter;
- U is the surface water use parameter.

Watercourses used for drinking water supply, irrigation, aquaculture, or recreational purposes receive higher scores due to the increased potential for human and ecological exposure. The surface water order according to Strahler classification indicates the hierarchical position of a stream within the drainage network. Lower-order streams receive higher vulnerability scores, as contaminant concentrations are generally higher and less diluted compared to those in higher-order streams.

2.4.3 Human Health Vulnerability

Human health vulnerability is determined by the land use, the population density and the percentage of vulnerable groups (e.g. children, elderly groups) according the following Equation 9:

$$Vh = L \cdot PD \cdot VG \qquad \text{Equation 9}$$

Where

- Vh is human health vulnerability;
- L is the land use parameter;
- PD is the population density parameter map;
- VG is the percentage of vulnerable group parameter map.

Residential areas with higher density and percentage of vulnerable groups receive the highest score due to human health exposure risk.

2.4.4 Protected Areas Vulnerability

Protected areas are scored according to their legal protection status, reflecting ecological sensitivity and conservation value. The score Vp is calculated as:

$$Vp = PAT \qquad \text{Equation 10}$$

Where:

- Vp is the protected areas vulnerability;
- PAT is the protected area type parameter.

Scores for protected areas are assigned based on the type of protection. Areas protected at state and regional level receive higher scores compared to ecological areas or cultural heritage monuments protected at local level.

2.5 Relative Regional Risk estimation

The spatial source risk represents the combined effect of the hazard associated with each PCS, the transport through environmental pathways, and the vulnerability of receptors. This step integrates the results from paragraphs 2.2 - 2.4 into a single spatially explicit risk score.

Each PCS is first evaluated independently. For each source and for each pathway, the spatial source risk is computed for every raster cell by combining:

- The intrinsic hazard map associated to the specific pathway (from paragraph 2.2),
- The environmental pathway score maps (from paragraph 2.3), and
- The receptor vulnerability score maps (from paragraph 2.4).

The integration follows Equation 11:

$$SRs_{e,x} = Hs_e \cdot P_e \cdot V_x \qquad \text{Equation 11}$$

Where:

- $SRs_{e,x}$ is the risk for each source associated to a specific combination of pathway and vulnerability.
- Hs_e is the hazard score for the PCS s spatially propagated through a specific pathway e ;
- P_e is the pathway score for the exposure pathway e (except for direct contact, which uses the point-source hazard map directly);
- V_x is the vulnerability score for the receptor under consideration.

For classification and prioritization purposes, these individual spatially distributed values related to specific pathways need to be aggregated to a single representative value for each source. Indeed, risk values are obtained for each pathway (groundwater, surface water, direct contact, and air), calculating both maximum and mean risk values. The final total risk score Rs for each PCS is calculated as the sum of SRs max risk values, according to Equation 12:

$$Rs = \sum_{(e,x)\in EP} SRs_{e,x} \qquad \text{Equation 12}$$

Where:
- Rs is the total risk of the PCS s;
- EP is the set of the combinations of Hazard-Pathway-Vulnerability presented in paragraph 2.1
- $SRs_{e,x}$ is the source risk (equation 11)

The regional comparison of point sources does not aim to merge or replace existing site-specific conceptual models, but rather to provide a spatially consistent overview of potential risks across the study area. This screening-level approach enables regional prioritization, identifying sites where site specific investigations would be most beneficial, without requiring data sharing from individual site or access to confidential investigation reports. As such, it complements existing national contaminated site management frameworks by offering a harmonized, regional-scale layer of analysis that can support planning, resource allocation, and policy development.

2.6 Normalisation in GIS of pathway parameters and receptor attributes

Before being able to apply any aggregation methodology, a normalization procedure should be performed in order to convert pathway parameters and receptor attributes into the same space domain, which reflect their contribution to hazard propagation or vulnerability assessment. This process ensures consistency across heterogeneous datasets by normalizing them to a common scoring framework.

Normalization values for pathways' parameters are listed in table S2 in supplementary materials, while normalization values for vulnerability attributes are reported in tables S3-S6 in supplementary materials. These normalization values are taken from Zabeo et al. (2011) and Minolfi et al. (2018), except for the flow accumulation and land use pathway parameters. Flow accumulation parameter simulates the movement of runoff water and potentially contaminant diffusion, while the land use parameter considers how permeability is conditioned by land cover (areas covered by roads, and in general residential areas, are more likely to be less permeable than agricultural and forest areas).

In the GIS environment, the pathway parameters and vulnerability attributes spatial datasets need to be normalized using the reclassification function. Categorical datasets (e.g. land use) are assigned normalized values directly based on discrete classification (e.g. agricultural or residential areas). To each point of raster maps of continuous data (e.g. slope) are assigned normalized values based on range values. The resulting normalized raster maps serve as standardized inputs for subsequent spatial analyses, including pathway modelling (paragraph 2.3) and receptor vulnerability estimation (paragraph 2.4). By harmonizing all spatial inputs in this manner, the methodology enables the integration of environmental (e.g. groundwater and surface water resources) and social factors (e.g. percentage of vulnerable groups and cultural heritage protected areas) into a unified GIS-based risk assessment framework.

2.6.1 Model Builder in GIS

The entire methodology was developed and implemented using the Model Builder tool in ArcGIS. Model Builder is a visual programming environment that allows the creation of workflows by combining geoprocessing tools, datasets, and logical operators in a graphical interface. This approach ensures transparency and reproducibility of the analysis, as each processing step is explicitly represented and can be inspected.

One of the main advantages of using Model Builder is its flexibility. The workflow designed for this study can be easily replicated in other geographical contexts: by simply replacing the input datasets (e.g., land cover, meteorological data, population density), the model automatically updates the subsequent analyses without requiring substantial modifications to the workflow.

This makes the methodology adaptable, transferable, and accessible also to users who may not have advanced programming skills, but who are familiar with the principles of GIS-based geoprocessing.

3. Case study application

3.1 Study area

The selected study area is the Toulouse Métropole ITA, situated in the Haute-Garonne department, in the Occitanie region of southwestern France. Covering an area of approximately 465 square kilometers, the metropolitan region comprises 37 municipalities, including the city of Toulouse, which serves as its administrative and economic center. With a population exceeding 800,000, Toulouse Métropole is the fourth largest region and one of the fastest-growing urban areas in France, driven by its position as a leading hub for the aerospace industry, technology sectors, and higher education institutions, such as the University of Toulouse (INSEE, 2023).

The land use within the metropolitan area is highly diverse, combining dense urban zones, industrial areas, and peri-urban agricultural lands. Since the late 1970s, in fact, Toulouse's urban fabric expanded rapidly into rural areas, creating a vast metropolitan region (Duvernoy, 2018). This variety of land use contributes to a complex contamination profile and introduces different contamination risks, from industrial emissions and waste disposal to agricultural runoff and urban sprawl. The presence of pollutants is often exacerbated by legacy contamination from past industrial practices, which can persist in urban soils long after the cessation of harmful activities (Pouyat et al., 2015).

The region's geological composition includes alluvial plains along the Garonne River, which pose specific risks for contaminant transport due to their high permeability. Toulouse experiences a temperate oceanic climate with Mediterranean influences, characterized by mild winters and warm summers. The region receives an average annual rainfall of approximately 600-700 mm, with precipitation patterns that could influence the leaching and migration of soil contaminants (Météo-France, 2023).

Given its industrial history, rapid urbanization, and agricultural practices, Toulouse Métropole presents a complex and dynamic environment for soil risk assessment. Studies have shown that urban agriculture, while beneficial for food security, can pose risks if soil contamination is not adequately managed (Moskal & Berthrong, 2018). The need for a multi-faceted approach to soil contamination risk assessment in Toulouse is underscored by the complex interplay of urban development, environmental health, and community well-being (Cheng et al., 2021). This makes it an ideal case study for evaluating contamination pathways, identifying vulnerable receptors, and developing targeted risk management strategies to support sustainable urban planning and public health protection.

A key consideration in interpreting the results is how representative the Toulouse case is of other European urban areas. The selected test area encompasses a heterogeneous landscape, including a major urban center, peri-urban and industrial zones, and surrounding agricultural and natural environments. This diversity mirrors the spatial and functional complexity found in many European metropolitan regions, where legacy industrial activities coexist with expanding residential and agricultural land uses. Therefore, while local characteristics may influence specific results, the overall configuration of Toulouse provides a relevant and transferable setting for testing the large-scale applicability of the proposed risk assessment framework.

3.2 Hazard analysis

For this study, we used the French BASIAS database, which is the national repository for regional historical inventories of industrial sites and activities, managed by BRGM. It includes activities considered potentially capable of generating soil or groundwater contamination, although inclusion in the database does not imply the actual presence of contamination.

The original BASIAS dataset contained 3943 PCSs in the Toulouse metropolitan area, both active and inactive, based on historical inventories. Older archival documents are often incomplete, imprecise, or lacking exploitable address information, preventing reliable spatial positioning of some sites. Consequently, sites lacking reliable geolocation information were excluded from the GIS-based spatial analysis, since spatial positioning is required to evaluate S-P-R interactions. While this reduces the completeness of the inventory, it does not affect the relative prioritisation of the analysed sites, since the methodology is designed to rank individual PCSs rather than assess potentially contaminated areas. The absence of geolocation data for part of the BASIAS database is mainly related to the historical nature of the inventory and is not related to specific categories or activities. Therefore, no systematic underrepresentation of activity types was identified within the analysed dataset. The final working database included 2128 sites, a number deemed sufficient for robust testing of the proposed methodology.

Each PCS is classified according to the French NAF code. These were converted into the corresponding European NACE codes, which were then matched with the activities–pollutants matrix. Hazard scores were calculated exclusively on the basis of the intrinsic hazard of activities ($H_{a,e}$), which reflects the substances potentially produced and released into the environment according to their NACE classification. Information on site dimension or operational duration was not included due to data limitations in the BASIAS database.

3.3 Pathway analysis

The Corine Land Cover, the Digital Elevation Model (DEM) and the Infiltration Capacity index (ICI) maps were used for groundwater and surface water pathway analysis. The Corine Land Cover map included 20 classes that were aggregated to fit in the land use classes reported in the supplementary material (S2).

The slope raster derived from the DEM was generated using the Slope tool in GIS with the PERCENT_RISE option, in order to express the elevation change as a percentage and to assign the corresponding normalisation values. The Digital Elevation Model was also used to generate the Flow Accumulation map, required for the computation of the surface water pathway. The model accounts for the actual amount of water accumulated during precipitation events. Moreover, by integrating the geolocation of PCSs into the Flow Direction analysis, it was possible to identify the potential pathways and locations of contaminants transported by surface water runoff from these sites.

The Infiltration Capacity Index was calculated based on 277 geotechnical surveys conducted in the Toulouse area (BRGM, 2022). Four values were found in the studied area corresponding to very favourable, favourable, not very favourable and not favourable areas to infiltration. In accordance with the normalisation classes proposed in the supplementary material (S2), the normalisation values 1, 0.5, 0.2, and 0 were assigned to those areas, respectively.

Meteorological data (temperature and precipitation), obtained from the open dataset of Toulouse Métropole (Toulouse Métropole, 2025), were used both for the calculation of the Aridity Index (AI) and as input weights in the Flow Accumulation analysis. The AI was found to be less than 25 in all the area, giving the highest normalisation value throughout the whole area for the Air pathway.

3.4 Vulnerability analysis

The dataset of groundwater abstraction points for drinking, agricultural, and industrial uses, provided by BRGM, was used for **groundwater vulnerability analysis**. The shapefile from the SANDRE Atlas Catalogue, which contains information on groundwater bodies was also used for this purpose. We focused on the shallowest aquifers, reclassifying the different areas according to the abstraction points located within them. Where multiple points were present, the highest value was retained. Areas lacking information on aquifers as well as those where no abstraction points were identified within an aquifer, were conservatively assigned the normalisation value corresponding to potable water use.

The **surface water vulnerability analysis** includes two parameters. As the surface water order parameter was not available for the case study area, we used a shapefile from the SANDRE Atlas Catalogue containing information on the width of different sections of the river network. According to Horton–Strahler laws, the stream order is directly related to discharge, depth, and width. We therefore assigned the following normalisation values based on river width:

- 0–15 m: value 1 (as a precautionary measure)
- 15–50 m: value 0.5
- >50m: value 0.3
- NC: value 0.8

For the surface water use parameter, shapefiles containing abstraction points for drinking water and agricultural use were downloaded from the Système d'Information sur l'Eau du bassin Adour-Garonne of the Agence de l'Eau Adour-Garonne. Since the precise location of the abstraction points is not provided (in accordance with Vigipirate security regulations, which place them at the center of the municipality), a 100 m buffer was created around the river network to include most of the points. Those located further away, and therefore too imprecise, were not considered in subsequent analyses. We used Thiessen polygons to assign vulnerability scores to river segments, based on the use of the abstraction points falling within each segment. Where two points with different uses falls within the same polygon, the most conservative score was retained.

For the **human health vulnerability analysis**, we again used the Corine Land Cover map, together with the IRIS database provided by BRGM. The IRIS database ("Ilots Regroupés pour l'Information Statistique") is the smallest census unit defined by INSEE (2023) and contains detailed demographic information at the sub-municipal level. It provides data on population density for each area as well as information on vulnerable groups, which in our analysis included children aged 0–15 and elderly people over 75 years.

For the **protected areas vulnerability analysis**, we used three different datasets: National Nature Reserves (normalisation value 1; Muséum national d'Histoire naturelle [MNHN], 2024), Regional Nature Reserves (normalisation value 0.8; Région Occitanie, 2024), and Sensitive Natural Areas (ENS, local ecological sensitive areas) (normalisation value 0.5; Observatoire Régional de l'Environnement Occitanie [OREO] / OpenIG, 2022). These datasets were merged into a single raster, and each category was assigned the corresponding normalisation value in order to generate the final protected areas receptor map.

3.5 Total risk

The resulting risk maps extend over a defined area, depending on the pathways and receptors considered. For classification and prioritization purposes, these spatially distributed values need to be aggregated to a single representative value for each PCS. For each pathway (groundwater, surface water, direct contact, and air), both maximum and mean risk values were calculated. When aggregating the pathways to obtain the total risk, we used the maximum values, as they provide a more conservative representation of the overall exposure potential across the affected area.

4. Results and Discussion

The Regional Risk Assessment was applied to 2128 PCSs in Toulouse area, for which pathway-receptor specific risks, as well as total risk, were calculated. The resulting total risk values span over three orders of magnitude, ranging from very low contributions ($<10^2$) to high-potential risk exceeding 10^5, highlighting a strong heterogeneity in the potential impact on environment and human health related to ongoing and past activities.

The distribution of total risk (Fig. 3) is strongly right-skewed. The median value of total risk is 1.96×10^4, with an interquartile range between 1.26×10^4 (25th percentile) and 4.00×10^4 (75th percentile). The upper tail is characterized by substantially higher values, with the 95th percentile reaching 8.75×10^4, and extreme values reaching 1.3×10^5, indicating that a relatively limited subset of PCSs drives the highest fraction of the cumulative regional risk. This pattern supports the relevance of the risk-based prioritization approach as a screening-level of a regional scale assessment.

To better understand how exposure pathways contribute across the risk spectrum, PCSs were grouped into quartiles of total risk, and the relative contribution of each pathway was evaluated using maximum-risk metrics (Fig. 4). Groundwater consistently represents the dominant pathway in all classes, contributing on average about 38% of the summed pathway-specific risk. Direct contact pathways affecting human health (i.e., soil ingestion and dermal contact) represents the second most relevant pathway (approximately 24–28%), followed by the contribution of particulate air exposure, mainly due to inhalation of outdoor air contaminants affecting human health (~20%). Surface water contributes a smaller share, but increasing from the lowest to higher risk classes (from about 12% to nearly 18%).

The consistently high contribution of the groundwater pathway across all risk classes reflects the high permeability of regional soils, and low slopes which enhances vertical contaminant migration. Under these hydrogeological conditions, groundwater acts as an efficient transport pathway even for sites with moderate hazard. The relative stability of the groundwater contribution across quartiles suggests that hydrogeological vulnerability exerts a structural control on risk distribution, while variations in total risk between classes are mainly influenced by differences in hazard intensity and the combined effect of additional exposure pathways and receptors.

To better understand the relative role of sources compared with exposure processes, the correlation between the aggregated hazard scores and the total risk was analysed across sites by using R software (R Core Team, 2026). The resulting trend, shown in Figure 5a, presents a generally linear relationship, with higher hazard sites exhibiting a great variability in total risk. Residual analysis indicates that this variability is approximately proportional to the hazard level (Fig. 5b).

The absence of information on site dimension and duration in the Toulouse case study introduces uncertainty in the hazard estimation, as those factors are relevant in influencing the potential magnitude of contaminant release. Their exclusion prevents differentiation between sites that share similar activity types but significantly differ in spatial extension and/or operational duration. This may lead to overestimation of hazard for smaller sites and underestimation for larger ones, leading to homogenisation of hazard scores within the same activity class. This exclusion, interestingly point out that, when stratifying by NACE macro-sector codes, the slopes of the regression lines differ, suggesting that the proportionality between hazard and risk varies by industrial sector. According to Figure 5a, industrial manufacturing activities (NACE section C) show the highest average and maximum total risk values, and account for a large share of the high-risk tail of the distribution. Sites associated with water and waste management (section E) also exhibit elevated average risk levels, consistent with their intrinsic link to contaminant handling, storage, and disposal. In contrast, commercial and service-related activities (e.g., sections G and S) display lower average risk values, although some individual sites still reach moderate-to-high risk levels, underlining the importance of not excluding these sectors a priori from regional screening exercises. This pattern may reflect the spatial clustering of different NACE-coded activities: some sectors tend to be located within or close to densely urbanized areas where environmental connectivity and exposure pathways amplify risk, whereas others are situated in locations where these factors are less influential.

Moreover, these findings suggest that the overall risk is driven by effective exposure pathways and environmental connectivity rather than solely by source characteristics and hazard. In other words, while hazard is a key driver of total risk, elevated regional risk emerges primarily where high source hazard coincides with favourable transport of contaminants and receptors' exposure conditions. Overall, the NACE-based analysis (Figs. 5a and 5b, Tab. 2) confirms that sectoral information can be effectively used as a proxy for hazard potential within regional-scale risk frameworks and can support the prioritization of inspection and detailed investigation campaigns. Our model has focus on regional risks. However, PCSs with high sum of hazards and relatively low total regional risk may have remarkable local risk for soil health.

As a further implementation, soil health can be considered as an additional, separate receptor in the RRA methodology, and described using indicators related to key soil functions in line with the SML, such as pH, Soil Organic Carbon content, texture and soil structure, land use, conductivity (water infiltration potential), biological activity. In line with the framework, if a high hazard PCS is located near to a high-quality soil health area, this area should be prioritised for investigation and remediation, since the impairment of high-quality soil health and associated soil functions should be avoided. Similarly, the inclusion of indicators such as ecological connectivity and landscape fragmentation to better represent ecosystem resilience could enhance the ecological realism of future developments of the proposed framework, although these metrics are increasingly incorporated in regional ecological risk frameworks but still not commonly integrated into regional screening methodologies for PCSs.

An additional relevant development of the proposed framework may concern its potential integration with dynamic monitoring and resilience strategies, both supporting the identification and tracking of PCSs over time and including modelling of future environmental scenarios. In relation to climate change, changes in precipitation patterns and increasing frequency of extreme events may significantly influence contaminant mobility, leaching processes, and transport pathways. Integrating such factors would enhance the capacity of the framework to support forward-looking risk assessment and improve its relevance for long-term environmental planning and management.

Because of the high density of PCSs, spatial results are presented as a risk heatmap in GIS, derived from point-based risk values and representing the spatial concentration of cumulative risk rather than discrete, classified polygons. Individual point risks are spatially aggregated, such that areas of higher intensity indicate locations where elevated risk values are more densely clustered within a defined neighbourhood. The resulting surface reflects relative risk intensity, not absolute risk at specific locations. The map in Figure 6b highlights pronounced risk hotspots in the central and eastern portions of the metropolitan area, where high PCS density, historical industrial activities, residential areas, and the presence of major surface water bodies and most vulnerable groundwater bodies coincide.

Peripheral and more rural sectors display a more diffuse and generally lower risk signal, with isolated local hotspots associated with specific industrial or infrastructural installations. This spatial configuration confirms that urban–industrial corridors act as focal areas of cumulative environmental pressure, and that RRA outputs can support spatial planning and targeted land management policies.

To provide a preliminary qualitative consistency check, the resulting prioritisation was compared with sites included in the BASOL database, the French national inventory of known polluted or potentially polluted sites, including 3700 sites that need monitoring or remediation actions. Within the Toulouse study area, 36 BASOL sites were identified, of which 14 could be reliably matched with the analysed inventory. Among the matched BASOL sites, four fall within the 90th percentile of the total risk distribution. These sites were subject to simplified or detailed risk assessments, long-term groundwater monitoring, remediation measures, up to land-use and groundwater use restrictions due to confirmed soil and groundwater contamination. Notably, one of these sites was classified as “Class 1” under the French simplified risk assessment procedure (ESR), indicating the need for priority management actions due to cadmium and chromium contamination in soils beneath the industrial facility and identifying surface waters (Garonne River used for drinking water supply) and groundwater as potential receptors. Although elevated groundwater vulnerability was consistently identified by our framework, surface waters were not highlighted as particularly vulnerable receptors for this site, suggesting that the surface water pathway modelling may represent one of the main limitations of the current application.

Among the sites ranked between the 50th and the 90th percentile of the total risk distribution, three were classified as “Class 2” (“sites requiring monitoring and/or land-use restrictions”) according to the French ESR. Investigations performed on the sites identified anomalous concentrations of contaminants in soils, groundwater, and in some cases surrounding residential areas affected by contaminant plumes. In particular, for the site where residential areas were affected by the contaminant plume, direct contact represented the highest contributing pathway in the total risk calculation, supporting the consistency of the proposed framework with the identified exposure scenario. Subsequent management and remediation measures were strongly influenced by the planned future land use, which in several cases included residential redevelopment and even sensitive land uses such as schools.

Interestingly, the only matched BASOL site ranked below the median risk value had undergone a quantitative human health risk assessment. No unacceptable risks were identified for either workers or downgradient residents, indicating good alignement between its lower relative risk ranking and the limited health concerns identified at the site.

The comparison with BASOL sites does not constitute a formal validation of contamination levels or quantitative risk, as BASOL includes sites already identified by environmental authorities as requiring monitoring or remediation actions. However, the overlap between medium-to-high risk scores and sites classified ad Class 1 and Class 2 under the French ESR supports the consistency and practical relevance of the proposed prioritisation framework as support in the identification of sites requiring further investigation, monitoring or remediation, while also confirming the importance of integrating RRA results with preliminary site investigations and land-use management strategies.

Advantages and limits

The proposed methodology supports the implementation of the SML, which requires Member States to establish harmonised soil monitoring systems and apply a risk-based, stepwise approach (Articles 12 and 43b). While the Directive does not mandate the collection of all possible anthropogenic or historical data, it emphasises the need for harmonised soil indicators, contamination screening, and prioritisation tools to guide subsequent investigations. The proposed RRA framework directly contributes to these objectives by providing a consistent and scalable screening-level methodology for identifying areas at risk.

Specifically, the methodologies support the stepwise risk assessment approach promoted in Article 43b of the SML, which requires a structured transition from broad-scale preliminary screening to more detailed investigations only when necessary. In this context, the methodology provides an effective basis for identifying areas or potential point sources that may warrant further investigation, thus optimizing resource allocation and avoiding unnecessary site-specific assessments when risk levels are low.

The regional-scale application would allow Member States and competent authorities to:

- prioritise investigation steps within a legally compliant tiered framework;
- anticipate remediation needs or preventive land management measures under Article 43b;
- facilitate strategic planning and policy decision-making at regional or national level.

The methodology, particularly the GIS application, could still need further refinement to ensure accuracy across different contexts. The main challenge lies in the availability and consistency of data, particularly when dealing with large-scale analyses. The approach may need to be adapted case by case depending on the type and quality of the data available. For this reason, even with our model available, it is essential that a GIS expert carry out the implementation. It is crucial to understand the nature of the input datasets and to verify the coherence of the obtained results.

A relevant issue in the application of the hazard assessment step to the Toulouse case study was the lack of data on site-specific attributes, such as operational duration and site area, which were originally included in the hazard assessment and therefore had to be excluded from the analysis. This limitation reduces the discriminatory power of the application study, which cannot distinguish between sites within the same NACE code despite substantial differences in site extent or operational duration and represents a source of uncertainty in the final hazard scores and risks estimations.

In the Supplementary material, Table S7 has been added to describe all the elements of uncertainty in the Toulouse case study.

Conclusions

The proposed approach aims to bridge the gap between desk-based inventories and site-specific investigations, thus it provides a consistent, information-based and spatially explicit tool for large-scale screening of potentially contaminating activities, supporting consistent prioritization across regions. In this sense, the methodology complements existing national frameworks by offering a harmonized approach for large-scale comparison and prioritization across regions and Member States.

The wide variability in total risk values demonstrates the usefulness of RRA as a screening and prioritization tool. Rather than treating all PCSs as equally urgent, the method highlights a subset of sites that mostly contribute to regional risk. The resulting risk scores should be interpreted as relative prioritisation indicators acting as a guiding filter for further site-specific analysis, and not as risk assessment indexes in the strict sense.

Such information can guide the selection of sites for detailed, site-specific risk assessment. The Regional Risk Assessment does not replace site-specific assessments indeed, but provides a transparent, spatially explicit pre-screening tool that can improve the efficiency of regional land management and the consistency of monitoring resources usage. This is particularly useful in contexts where measured environmental concentrations are often lacking or incomplete and where resources for land management and remediation are low.

A key strength of the proposed RRA framework lies in its ability to combine sector-based hazard indicators (updated with more recent NACE codes and CLP regulation) with simplified but spatially distributed pathway models and receptor characteristics. This enables the inclusion of a large number of PCSs, even in case of lacking data; the evaluation of potential contaminants transfers in the environment; and the potential impact on receptors, both human and environmental.

The results should be interpreted as relative indicators of priority and not as absolute measures of risk. Risk estimates are, in fact, based on proxy indicators of contamination potential and not on measured concentrations. Moreover, the pathway modeling relies on regional-scale and simplified environmental data, which may not capture local heterogeneities.

Despite these approximations, the approach is highly transferable to all regions where PCS inventories are available. By linking PCS data with spatially explicit pathway modeling, authorities can move from purely inventory-based approaches toward risk-informed regional planning, in line with the required management actions to Member States, following the SML principles of sustainable land management and environmental health.

Transition to Part II

The methodological framework presented in this study for the regional prioritization of PCSs represents one component of a broader and flexible approach to soil risk assessment at the territorial scale. The same conceptual and methodological structure was adapted to assess diffuse soil contamination, in line with the risk-based management principles promoted by the Soil Monitoring and Resilience Directive (EU 2025/2360).

While the present article focuses on point-source related pressures and the screening-level prioritization of individual sites for further investigation, the companion study (Part II) applies the same regional risk assessment logic to spatially distributed contamination that cannot directly attributed to discrete sources. While the present manuscript relies on data about PCSs for hazard estimation, the complementary Part II study relies on soil geochemical data, using measured soil concentrations and spatial interpolation methods to identify contamination hotspots and addressing the potential impact on receptors of diffuse contamination.

The two approaches can be used either complementarily or independently, depending on data availability and management objectives. The Potentially Contaminated Sites-oriented framework is particularly suited to guiding selection of sites for site-specific investigations and regulatory actions targeting industrial or legacy sources. Conversely, the diffuse contamination framework is better aligned with spatial planning, land-use management, and the identification of broader territorial priorities. Together, they form an integrated, flexible toolkit for supporting risk-informed soil management at the regional scale.

Figure1. Structure of the Regional Risk Assessment methodology for the prioritisation of potentially contaminated sites and related assessment steps.

Alt text: Flowchart illustrating the regional GIS-based risk assessment framework for prioritising potentially contaminated sites. The diagram shows the sequential assessment of hazard, pathways, and receptors, followed by the integration of these components into pathway-specific and total relative risk scores used to produce regional risk maps.

Figure 1. Exposure diagram for regional risk assessment, with the selected pathways linking the contamination sources to the relevant receptors.

Alt text: Conceptual exposure diagram showing the relationships between contamination sources, exposure pathways, and receptors considered in the regional risk assessment. The potentially contaminated sites are connected to human and environmental receptors through direct contact, groundwater, surface water, and air transport pathways.

Figure 3: Total PCSs risk distribution. Abbreviations: Potentially Contaminated Sites (PCSs).

Alt text: Histogram of total relative risk scores for 2128 potentially contaminated sites. Most sites have low to moderate risk scores, while a smaller number form a long right tail of higher-risk sites. Vertical dashed lines mark the first quartile, median, third quartile, and the 90th and 95th percentile thresholds.

Figure 4: Relative pathways contribution to total risk across 4 quartiles. Abbreviations: Direct Contact (DC), Groundwater (GW), Surface Water (SW).

Alt text: Grouped bar chart comparing the mean percentage contribution of four exposure pathways—direct contact, groundwater, air, and surface water—to the total risk across the four risk quartiles. Groundwater is the dominant contributor in all quartiles, with a relatively stable contribution of about 38%. Direct contact contributes more in the lower-risk quartiles and decreases slightly with increasing risk, whereas the contribution of surface water increases in the higher-risk quartiles. Air contribution remains relatively constant across all quartiles.

Figure 5a: Potentially Contaminated Sites aggregated hazard scores and total risk trend, clustered by NACE-code. Abbreviations: Statistical Classification of Economic Activities in the European Community (NACE); A: Agriculture, forestry and fishing; B: Mining and quarrying; C: Manufacturing; D: Electricity, gas, steam and air conditioning supply; E: Water supply; sewerage, waste management and remediation activities; G: Retail trade; repair of motor vehicles and motorcycles; H: Transportation and storage; S: Other service activities; V: Storage and disposal activities, including storage of chemicals, coal, flammable liquids, mining residues, dredged sediments, and gas (BRGM/BASIAS classification).

Alt text: Scatter plot showing the relationship between aggregated hazard scores and total relative risk for potentially contaminated sites, grouped by NACE activity sector. Each point represents one site, and coloured regression lines illustrate the trend for each sector. Total risk generally increases with increasing hazard scores, although the strength of the relationship varies among activity sectors.

Figure 5b: Residual analysis. Abbreviations: Statistical Classification of Economic Activities in the European Community (NACE); A: Agriculture, forestry and fishing; B: Mining and quarrying; C: Manufacturing; D: Electricity, gas, steam and air conditioning supply; E: Water supply; sewerage, waste management and remediation activities; G: Retail trade; repair of motor vehicles and motorcycles; H: Transportation and storage; S: Other service activities; V: Storage and disposal activities, including storage of chemicals, coal, flammable liquids, mining residues, dredged sediments, and gas (BRGM/BASIAS classification).

Alt text: Scatter plot of residuals against aggregated hazard scores for potentially contaminated sites, grouped by NACE activity sector. Residuals are distributed around zero across the range of hazard scores, with greater variability at higher hazard values. The figure illustrates that total risk is influenced not only by hazard scores but also by pathway and receptor characteristics included in the regional risk assessment.

Figure 6a: Total risk for 2128 PCSs in Toulouse metropole.

Alt text: Map of the Toulouse Metropolitan Area displaying the spatial distribution of total relative risk for 2,128 potentially contaminated sites. Colours represent increasing relative risk, highlighting clusters of higher-risk sites within the study area.

Figure 6b: Total risk heatmap in Toulouse metropole (zoom in central area).

Alt text: Heatmap of total relative risk in the central area of the Toulouse Metropolitan Area. The map highlights spatial hotspots where potentially contaminated sites are concentrated and where cumulative relative risk is highest.

Table 1. Selected characteristics for receptor's vulnerability evaluation.

Table 2: NACE-based activities count and percentage across total risk classes. Abbreviations: Statistical Classification of Economic Activities in the European Community (NACE); A: Agriculture, forestry and fishing; B: Mining and quarrying; C: Manufacturing; D: Electricity, gas, steam and air conditioning supply; E: Water supply; sewerage, waste management and remediation activities; G: Retail trade; repair of motor vehicles and motorcycles; H: Transportation and storage; S: Other service activities; V: Storage and disposal activities, including storage of chemicals, coal, flammable liquids, mining residues, dredged sediments, and gas (BRGM/BASIAS classification).

References

Albanese, S., De Vivo, B., Lima, A., Frattasio, G., Kříbek, B., Nyambe, I., & Majer, V. (2014). Prioritising environmental risk at the regional scale by a GIS aided technique: The Zambian Copperbelt Province case study. Journal of Geochemical Exploration, 144, 433–442. https://doi.org/10.1016/j.gexplo.2014.02.004

Agenzia per la Protezione dell'Ambiente e per i Servizi Tecnici (APAT). (2008). Criteri metodologici per l'applicazione dell'analisi assoluta di rischio ai siti contaminati (Revisione 2, 156 pp.). APAT. http://www.isprambiente.gov.it/files/temi/siti-contaminati-02marzo08.pdf

ASTM International. (2021). Standard practice for environmental site assessments: Phase I environmental site assessment process (ASTM E1527–21). ASTM International. https://doi.org/10.1520/E1527-21

Bardos, R. P., Thomas, H. F., Smith, J. W. N., Harries, N. D., Evans, F., Boyle, R., Howard, T., Lewis, R., Thomas, A. O., & Haslam, A. (2018). The development and use of sustainability criteria in SuRF-UK's sustainable remediation framework. Sustainability, 10(6), 1781. https://doi.org/10.3390/su10061781

Braud, I., Desprats, J.-F., Ayral, P.-A., Bouvier, C., & Vandervaere, J.-P. (2017). Mapping topsoil field-saturated hydraulic conductivity from point measurements. Journal of Hydrology and Hydromechanics, 65(3), 264–275. https://doi.org/10.1515/johh-2017-0017

BRGM. (2022). Capacité d'infiltration des eaux pluviales du territoire de Toulouse Métropole (CAPITOUL): Rapport final, Version 1, 24 mai 2022. BRGM/RP71904-FR.

Chen, M., Cai, H., Wang, L., & Lei, M. (2022). Grid-scale regional risk assessment of potentially toxic metals using multi-source data. ISPRS International Journal of Geo-Information, 11(8), 427. https://doi.org/10.3390/ijgi11080427

Cheng, Z., Hettiarachchi, G., & Kim, K. (2021). Urban soils research: Suitma 10. Journal of Environmental Quality, 50(1), 2–6. https://doi.org/10.1002/jeq2.20191

Csavina, J., Landazuri, A., Wonaschutz, A., Rine, K., Rheinheimer, P., Barbaris, B., Conant, W., Saez, A. E., & Betterton, E. A. (2011). Metal and metalloid contaminants in atmospheric aerosols from mining operations. Water, Air, and Soil Pollution, 221(1–4), 145–157. https://doi.org/10.1007/s11270-011-0777-x

Department for Environment, Food and Rural Affairs (DEFRA). (2011). Guidelines for environmental risk assessment and management: Green Leaves III. Defra & Cranfield University, U.K.

Dragović, R., Gajić, B., Dragović, S., Đorđević, M., Đorđević, M., Mihailović, N., & Onjia, A. (2014). Assessment of the impact of geographical factors on the spatial distribution of heavy metals in soils around the steel production facility in Smederevo (Serbia). Journal of Cleaner Production, 84, 550–562. https://doi.org/10.1016/j.jclepro.2014.03.060

Duvernoy, I., Zambon, I., Sateriano, A., & Salvati, L. (2018). Pictures from the other side of the fringe: Urban growth and peri-urban agriculture in a post-industrial city (Toulouse, France). Journal of Rural Studies, 57, 25–35. https://doi.org/10.1016/j.jrurstud.2017.10.007

Esri. (2025). ArcGIS Pro (Version 3.6) [Computer software]. Esri. https://www.esri.com/en-us/arcgis/products/arcgis-pro

European Chemicals Agency (ECHA). (2017). Guidance on the Application of the CLP Criteria. European Chemicals Agency.

European Environment Agency (EEA). (2005). Provisional Risk Assessment Methodology for the Prioritisation of Contaminated Sites (PRA.MS). European Environment Agency.

European Food Safety Authority (EFSA). (2013). Guidance on the risk assessment of plant protection products on bees. European Food Safety Authority.

European Commission. (2008). Regulation (EC) No 1272/2008 on classification, labelling and packaging of substances and mixtures (CLP). Official Journal of the European Union.

European Commission. (2023). Proposal for a directive. Soil monitoring and resilience (Soil Monitoring Law) (COM(2023) 416 final). European Commission.

European Parliament & Council. (2025). Directive (EU) 2025/2360 of 12 November 2025 on soil monitoring and resilience (Soil Monitoring Law). Official Journal of the European Union. https://eur-lex.europa.eu/eli/dir/2025/2360/oj/eng

Food and Agriculture Organization of the United Nations (FAO). (2018). Global Symposium on Soil Pollution. FAO.

Hammond, E. B., Coulon, F., Hallett, S. H., Thomas, R., Hardy, D., Kingdon, A., & Beriro, D. J. (2021). A critical review of decision support systems for brownfield redevelopment. Science of the Total Environment, 785, 147132. https://doi.org/10.1016/j.scitotenv.2021.147132

Health Canada. (2010a). Federal contaminated site risk assessment in Canada. Part I: Guidance on human health preliminary quantitative risk assessment (PQRA) (Version 2.0). Health Canada.

Health Canada. (2010b). Federal contaminated site risk assessment in Canada. Part II: Health Canada toxicological reference values (TRVs) and chemical-specific factors (Version 2.0). Health Canada.

Hunsaker, C. T., Carpenter, D. E., & Messer, J. J. (1990). Ecological risk assessment: Conceptual framework and case study. Environmental Management, 14(5), 633–648. https://doi.org/10.1007/BF02394714

Institut national de la statistique et des études économiques - INSEE (2023). Population en 2020 – Recensement de la population – Base infracommunale (IRIS). Institut national de la statistique et des études économiques. https://www.insee.fr/fr/statistiques/7704076

International Organization for Standardization (ISO). (2018). Soil quality — Guidance on the determination of background values. International Organization for Standardization.

Jiang, Y., Guo, X., Ye, Y., Xu, Z., Zhou, Y., Xia, F., & Shi, Z. (2024). Spatiotemporal assessment and scenario simulation of the risk potential of industrial sites at the regional scale. Science of the Total Environment, 906, 167537. https://doi.org/10.1016/j.scitotenv.2023.167537

Landis, W. G. (2005). Regional scale ecological risk assessment: Using the relative risk model. CRC Press.

Landis, W. G., & Wiegers, J. K. (1997). Design considerations and a suggested approach for regional and comparative ecological risk assessment. Human and Ecological Risk Assessment, 3(3), 287–297.

Lerner, D. N., & Harris, B. (2009). The relationship between land use and groundwater resources and quality. Land Use Policy, 26S, S265–S273.

Li, K., Wang, J., & Zhang, Y. (2022). Heavy metal pollution risk of cultivated land from industrial production in China: Spatial pattern and its enlightenment. Science of the Total Environment, 828, 154382. https://doi.org/10.1016/j.scitotenv.2022.154382

Météo-France. (2023). Climatological data for the Toulouse region [Data set]. Météo-France. https://meteofrance.com

Minolfi, G., Albanese, S., Lima, A., Tarvainen, T., Fortelli, A., & De Vivo, B. (2018). A regional approach to the environmental risk assessment: Human health risk assessment case study in the Campania region. Journal of Geochemical Exploration, 184(B), 400–416. https://doi.org/10.1016/j.gexplo.2016.12.010

Moskal, B., & Berthrong, S. (2018). Novel soil barrier systems potentially protect urban growing beds from legacy soil contamination and improve soil health. Urban Agriculture & Regional Food Systems, 3(1), 1–9. https://doi.org/10.2134/urbanag2018.06.0003

Muséum national d'Histoire naturelle (MNHN). (2024). Réseau des Réserves Naturelles Nationales (RNN) [Data set].

Observatoire Régional de l'Environnement Occitanie (OREO) / OpenIG. (2022). Espaces Naturels Sensibles (ENS) Occitanie [Data set].

Pérez, A. P., & Rodriguez Eugenio, N. (2018). Status of local soil contamination in Europe (JRC Technical Report). Publications Office of the European Union. https://doi.org/10.2760/093804

Pizzol, L., Zabeo, A., Klusáček, P., Giubilato, E., Critto, A., Frantál, B., Martinát, S., Kunc, J., Osman, R., & Bartke, S. (2016). Timbre Brownfield Prioritization Tool to support effective brownfield regeneration. Journal of Environmental Management, 166, 178–192.

Pizzol, L., Critto, A., Agostini, P., & Marcomini, A. (2011). Regional risk assessment for contaminated sites Part 2: Ranking of potentially contaminated sites. Environment International, 37(8), 1307–1320.

Pouyat, R., Szlávecz, K., Yesilonis, I., Wong, C., Murawski, L., Marra, P., Casey, R. E., & Lev, S. (2015). Multi-scale assessment of metal contamination in residential soil and soil fauna: A case study in the Baltimore–Washington metropolitan region, USA. Landscape and Urban Planning, 142, 7–17. https://doi.org/10.1016/j.landurbplan.2015.05.001

Prabhakar, G., Sorooshian, A., Toffol, E., Arellano, A. F., & Betterton, E. A. (2014). Spatiotemporal distribution of airborne particulate metals and metalloids in a populated arid region. Atmospheric Environment, 92, 339–347. https://doi.org/10.1016/j.atmosenv.2014.04.044

Punia, A. (2021). Role of temperature, wind, and precipitation in heavy metal contamination at copper mines: A review. Environmental Science and Pollution Research, 28, 4056–4072. https://doi.org/10.1007/s11356-020-11580-8

R Core Team. (2026). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Région Occitanie. (2024). Réserves Naturelles Régionales (RNR) [Data set].

Sandre. (2025). Atlas des masses d'eau souterraines [Data set]. Service d'Administration Nationale des Données et Référentiels sur l'Eau. https://www.sandre.eaufrance.fr/atlas

Tenailleau, Q., Lanier, C., Prud'homme, J., Cuny, D., Deram, A., & Occelli, F. (2024). Distance-based indicators for evaluating environmental multi-contamination and related exposure: how far should we go? Environmental Science and Pollution Research, 31(38), 50642–50653.

Toulouse Métropole. (2025). Open environmental and meteorological data portal [Data set]. Toulouse Métropole. https://data.toulouse-metropole.fr/explore/dataset/01-station-meteo-toulouse-meteopole/information/

U.S. Environmental Protection Agency (USEPA). (1991). Risk assessment guidance for Superfund: Volume I — Human health evaluation manual (Part B, Development of risk-based preliminary remediation goals). U.S. Environmental Protection Agency.

U.S. Environmental Protection Agency (USEPA). (1996). Soil screening guidance: Technical background document. U.S. Environmental Protection Agency. https://www.epa.gov/superfund/soil-screening-guidance

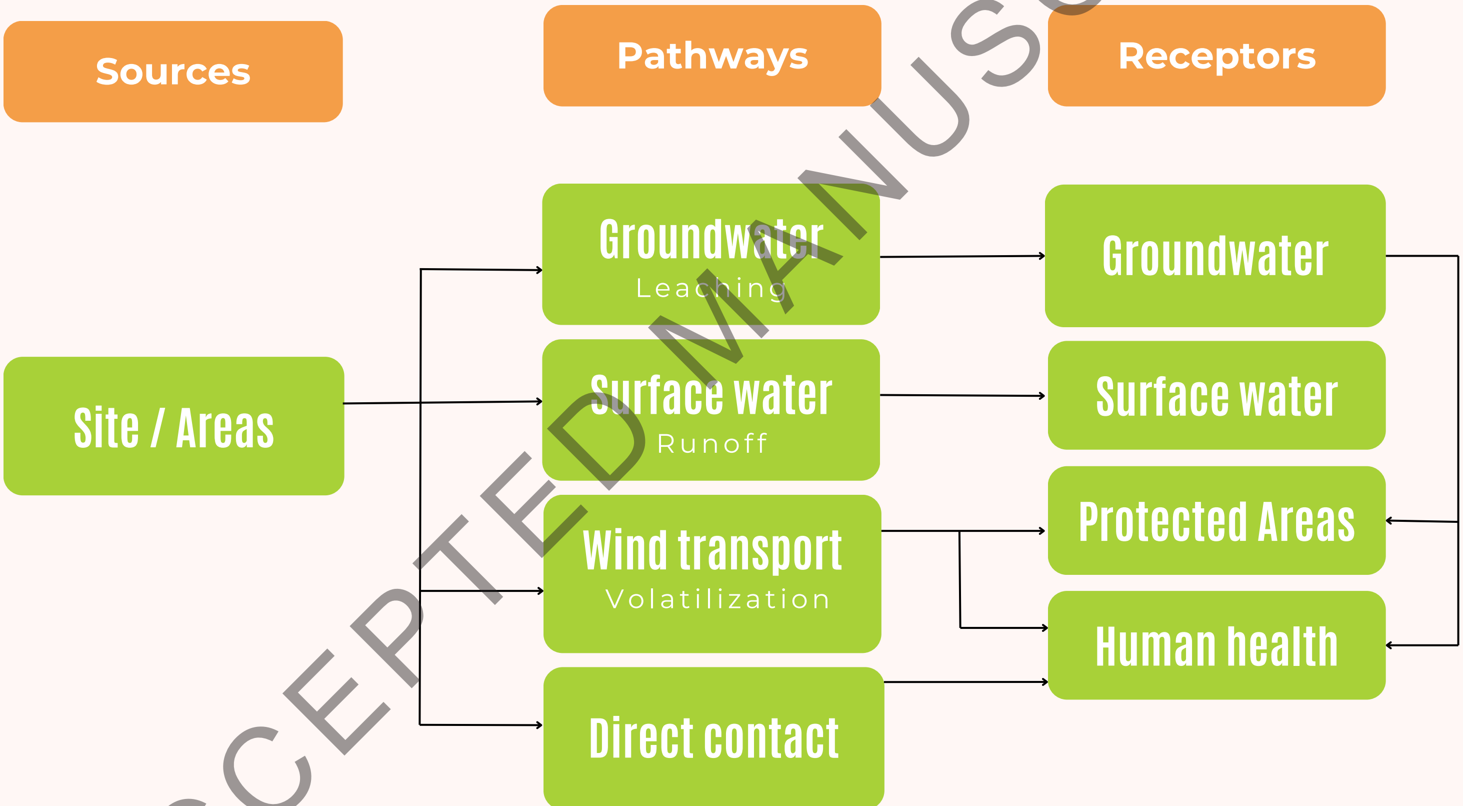
U.S. Environmental Protection Agency (USEPA). (2020). Regional screening levels (RSLs) – User's guide. U.S. Environmental Protection Agency. https://www.epa.gov/risk/regional-screening-levels-rsls

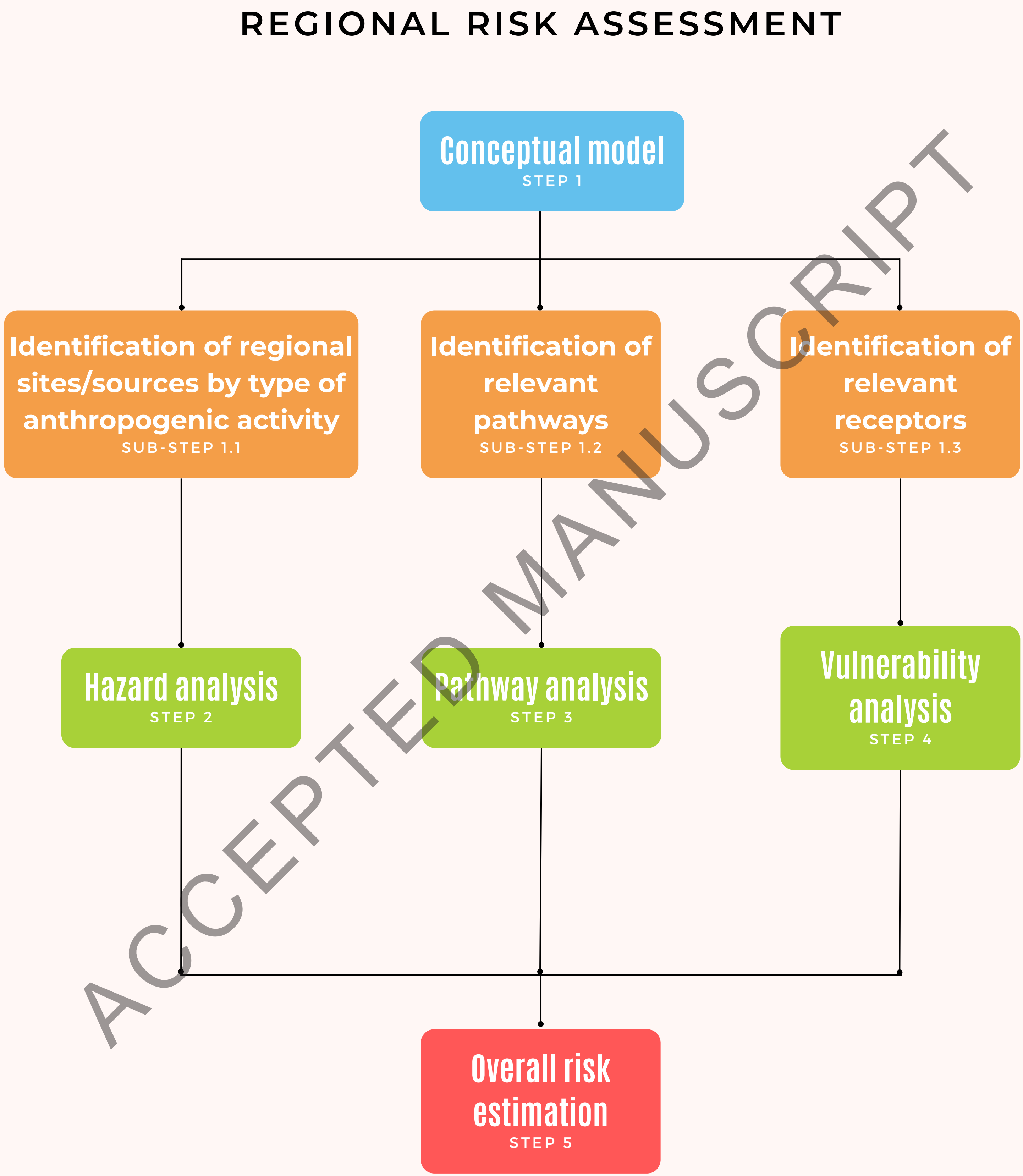
Van-Camp, L., Bujarrabal, B., Gentile, A.-R., Jones, R. J. A., Montanarella, L., Olazabal, C., & Selvaradjou, S.-K. (2004). Reports of the Technical Working Groups established under the Thematic Strategy for Soil Protection. European Commission, Joint Research Centre.

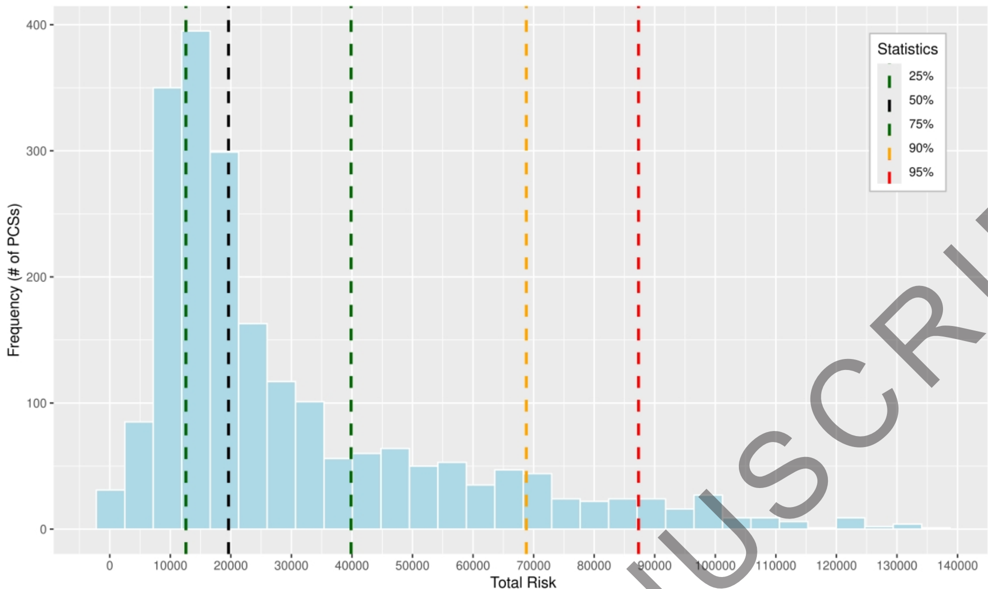
Zabeo, A., Pizzol, L., Agostini, P., Critto, A., Giove, S., & Marcomini, A. (2011). Regional risk assessment for contaminated sites Part 1: Vulnerability assessment by multicriteria decision analysis. Environment International, 37(8), 1295–1306.

World Bank Group, & Technical University of Denmark. (2023). Global Wind Atlas 3.0 [Data set]. https://globalwindatlas.info

EXPOSURE DIAGRAM

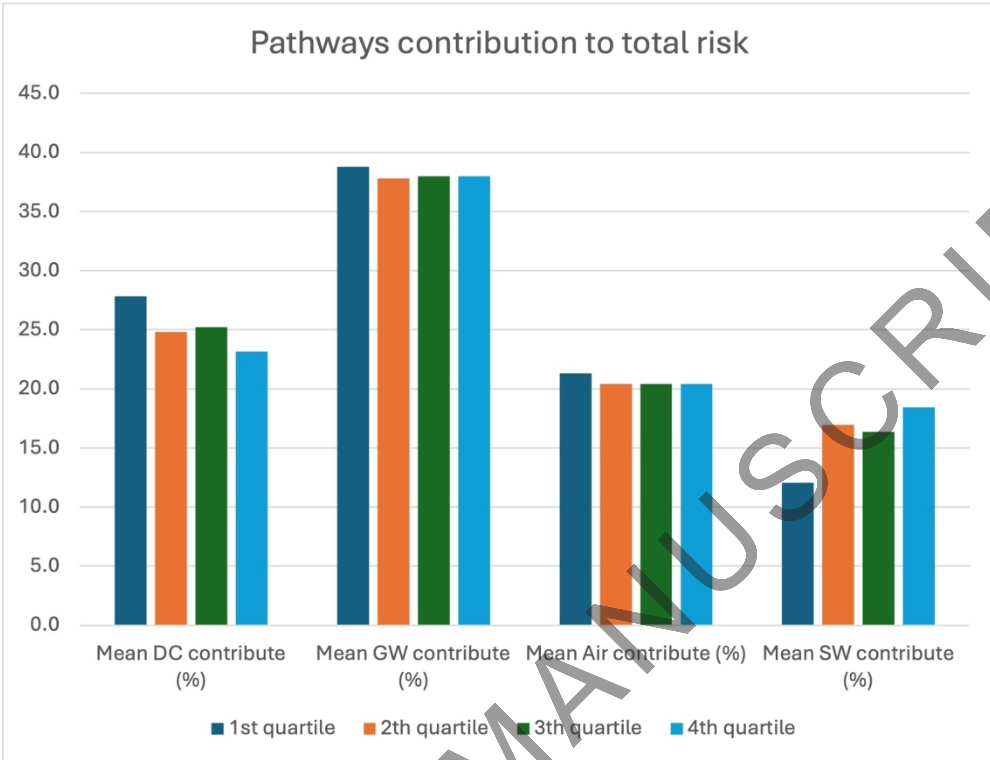



REGIONAL RISK ASSESSMENT
Conceptual model
STEP 1
Identification of regional sites/sources by type of anthropogenic activity
SUB-STEP 1.1
Identification of relevant pathways
SUB-STEP 1.2
Identification of relevant receptors
SUB-STEP 1.3
Hazard analysis
STEP 2
Pathway analysis
STEP 3
Vulnerability analysis
STEP 4
Overall risk estimation
STEP 5



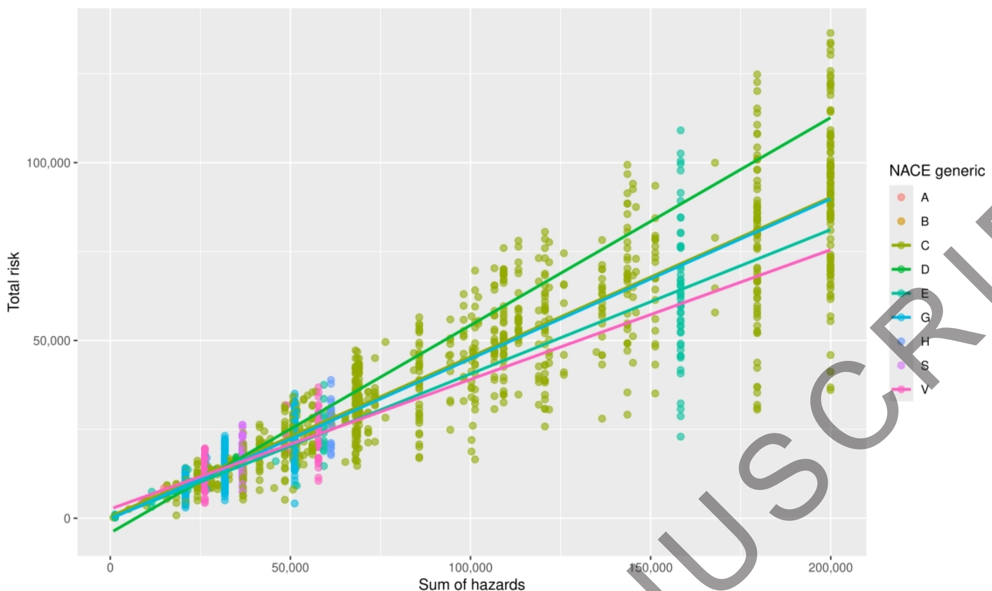
Total PCSs risk distribution. Abbreviations: Potentially Contaminated Sites (PCSs).

645x387mm (236 x 236 DPI)



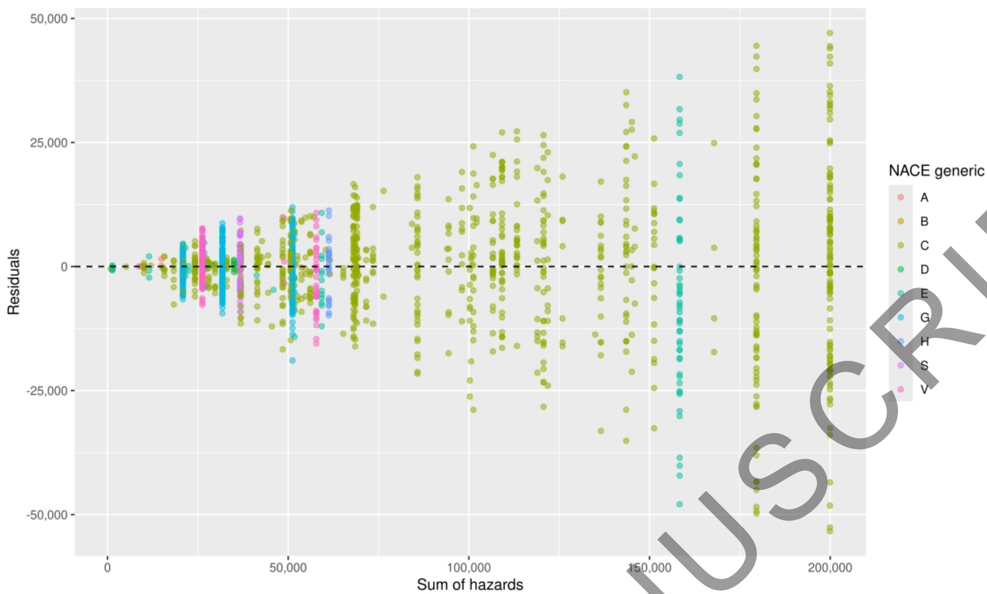
Relative pathways contribution to total risk across 4 quartiles. Abbreviations: Direct Contact (DC), Groundwater (GW), Surface Water (SW).

144x110mm (300 x 300 DPI)



Potentially Contaminated Sites aggregated hazard scores and total risk trend, clustered by NACE-code. Abbreviations: Statistical Classification of Economic Activities in the European Community (NACE); A: Agriculture, forestry and fishing; B: Mining and quarrying; C: Manufacturing; D: Electricity, gas, steam and air conditioning supply; E: Water supply; sewerage, waste management and remediation activities; G: Retail trade; repair of motor vehicles and motorcycles; H: Transportation and storage; S: Other service activities; V: Storage and disposal activities, including storage of chemicals, coal, flammable liquids, mining residues, dredged sediments, and gas (BRGM/BASIAS classification).

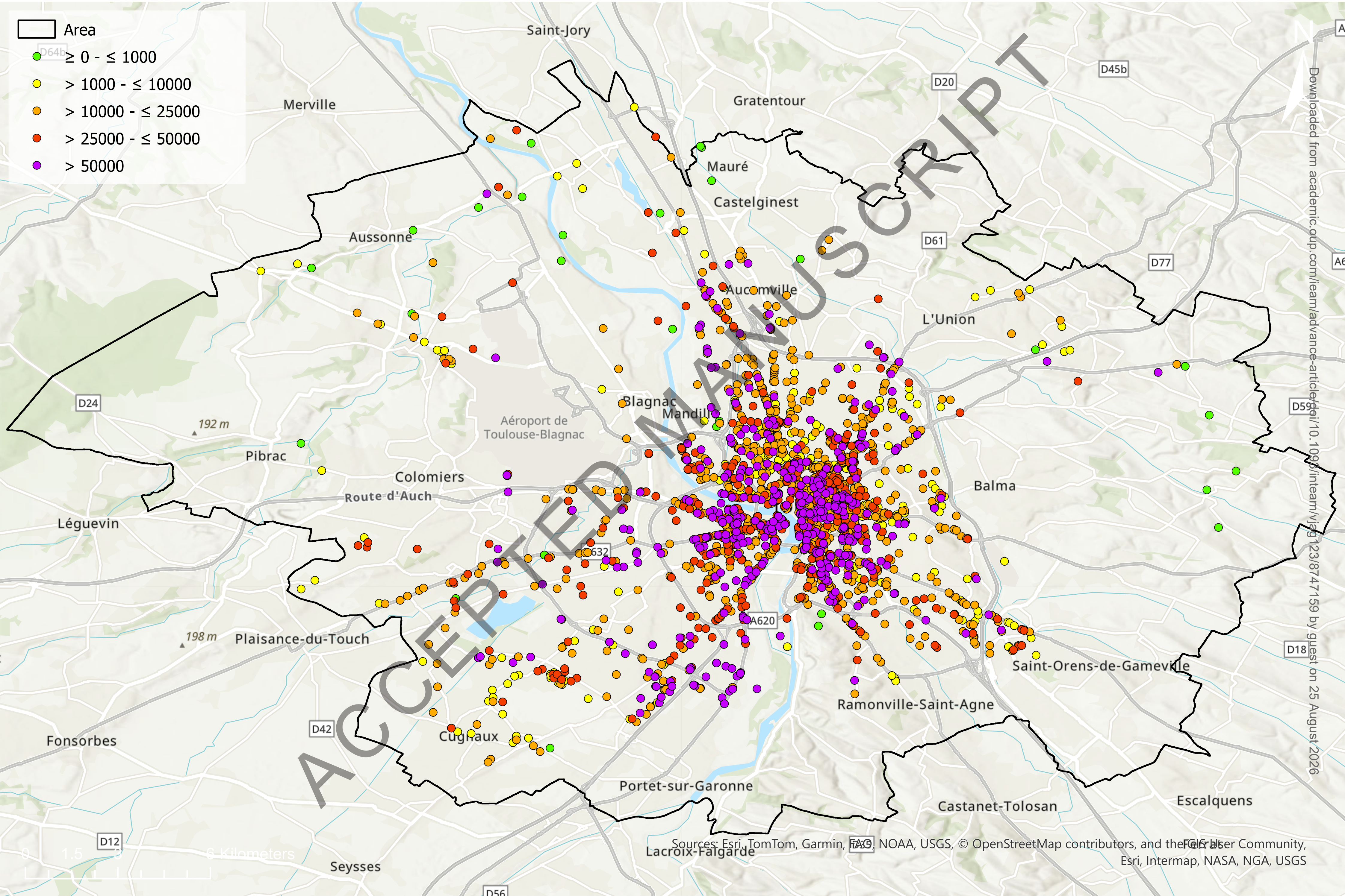
645x387mm (236 x 236 DPI)



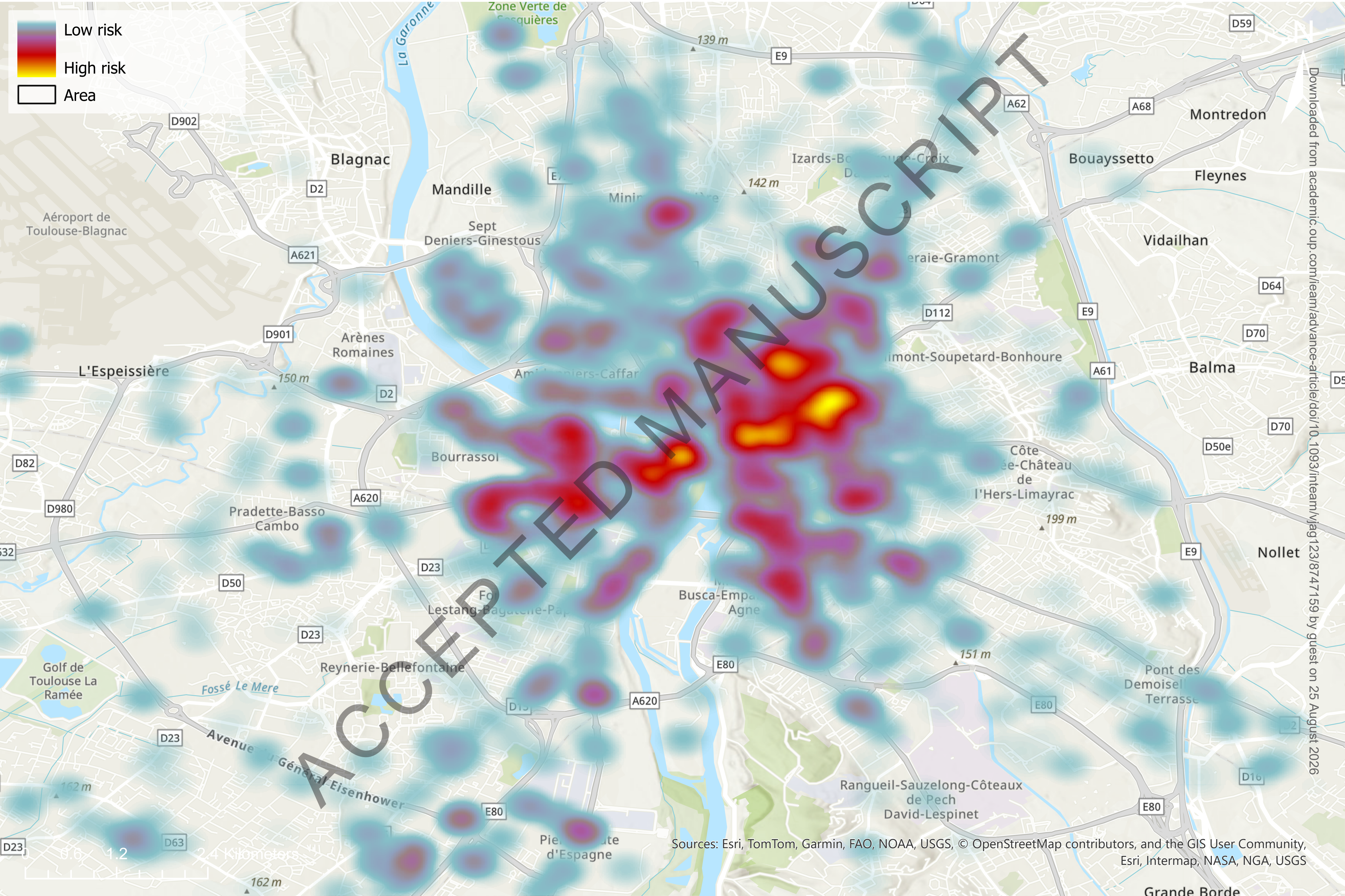
Residual analysis. Abbreviations: Statistical Classification of Economic Activities in the European Community (NACE); A: Agriculture, forestry and fishing; B: Mining and quarrying; C: Manufacturing; D: Electricity, gas, steam and air conditioning supply; E: Water supply; sewerage, waste management and remediation activities; G: Retail trade; repair of motor vehicles and motorcycles; H: Transportation and storage; S: Other service activities; V: Storage and disposal activities, including storage of chemicals, coal, flammable liquids, mining residues, dredged sediments, and gas (BRGM/BASIAS classification).

645x387mm (236 x 236 DPI)

Potentially Contaminated Sites Risk



Potentially Contaminated Sites - Total risk



Receptor	Attributes
Groundwater	Groundwater use Groundwater quality
Surface water	Surface water order Surface water quality Surface water use
Human health	Land use Population density Percentage of vulnerable population groups
Protected areas	Type of protection

Table 1. Selected characteristics for receptor's vulnerability evaluation.

Table 2: NACE-based activities count and percentage across total risk classes.

	NACE Code																		
Class	A	Percentage of A activites	B	Percentage of B activites	C	Percentage of C activites	D	Percentage of D activites	E	Percentage of E activites	G	Percentage of G activites	H	Percentage of H activites	S	Percentage of S activites	V	Percentage of V activites	Total
4rd Quartile					488	49%			46	43%									532
3rd Quartile			1	100%	285	28%	1	11%	18	17%	144	26%	23	92%	29	23%	29	10%	532
2nd Quartile					105	10%	8	89%	8	7%	217	39%	2	8%	84	67%	108	37%	530
1st Quartile	2	100%			128	13%			36	33%	198	35%			12	10%	156	53%	534
Total	2		1		1006		9		108		559		25		125		293		2128

Note. Statistical Classification of Economic Activities in the European Community (NACE); A: Agriculture, forestry and fishing; B: Mining and quarrying; C: Manufacturing; D: Electricity, gas, steam and air conditioning supply; E: Water supply; sewerage, waste management and remediation activities; G: Retail trade; repair of motor vehicles and motorcycles; H: Transportation and storage; S: Other service activities; V: Storage and disposal activities, including storage of chemicals, coal, flammable liquids, mining residues, dredged sediments, and gas (BRGM/BASIAS classification).