

A GeoAI-Based Environmental Risk Assessment Framework For Chemical Exposure And Public Health Vulnerability In U.S. Industrial Corridors

Nagina Tariq

Doctor of Business Administration, Westcliff University, Irvine, California, USA)

Email Address: n.tariq.3686@westcliff.edu

Abstract

The United States industrial corridors are still subjected to unequal and disproportionate chemical exposures and community health hazards, but the current environmental justice instruments have low predictive power and poor heterogeneous datasets. This paper describes GeoAI RiskLab, a reproducible environmental risk assessment framework which combines EPA Toxics Release Inventory emissions, ambient and satellite monitoring and sociodemographic vulnerability indicators into one GeoAI architecture.

The model uses the spatial deep learning models, uncertainty quantification, and explainable AI diagnostics in generating high-resolution estimates of cumulative chemical exposure risk. Outputs are projected to decision relevant spatial units e.g. census tracts and school zones, which has been used to target regulatory purposes and urban resilience planning. The usage in the selected industrial corridors has shown that the toolkit is useful in establishing hotspots of exposure, assessing model plausibility, and making decisions on environmental health. GeoAI RiskLab is a policy-friendly and transferable environmental governance platform based on data.

Keywords: *GeoAI, Spatial Deep Learning, Environmental Justice, Chemical Exposure Modelling, Cumulative Risk Assessment, Urban Resilience Analytics.*

INTRODUCTION

The United States has industrial belts that concentrate a significant proportion of hazardous chemical releases and legacy pollution, as well as socio-economic weakness, generating enduring disproportionate environmental health results. Despite the fact that transparency has been enhanced with the help of national screening databases, including the Toxics Release Inventory (TRI) and EJScreen, these databases have a restricted ability to perform predictive modelling, dynamic combination of heterogeneous datasets, and fine-scale measurement of cumulative risk of exposure. The gap is especially noteworthy to state and local agencies, which need decision-ready tools that may integrate the information about the emissions, ambient observation, demographic factors, and spatial environment into assessments that may be taken.

Modelling of environmental risks in complex urban-industrial systems has new opportunities with the rapid progress of GeoAI and spatial deep learning. With convolutional and graph-based neural networks along with geospatial analytics, it is possible to estimate exposure gradients, identify hotspot patterns, as well as evaluate multi-source drivers of risk at fine spatial scale. However, the use of these methods in chemical exposure evaluation is still immature, and in most cases, they do not have quantification of uncertainty, modifiable explanations as well as detectable workflows that are needed when using them in regulations and policies. More so, the models that are in existence do not combine socio-vulnerability indicators into a complete analysis tool, which could restrict their application in the context of environmental justice and health-equity planning.

This research paper presents a reproducible geoAI RiskLab modelling and software that attempts to address such limitations. The system combines the facility-level emissions, satellite and ambient monitoring, demographic data with the spatial indicators into a modular GeoAI framework in the analysis of cumulative chemical exposure risks. The framework has uncertainty analysis, interpretable model diagnostics and a standardized output at decision relevant spatial units like census tracts and school catchment areas. By so doing, it directly addresses the methodological and transparency requirements that Environmental Modelling Software places upon a methodology and model, such as rigorous model testing, reproducibility, and software reusability.

Using GeoAI RiskLab on the targeted areas of the U.S. industrial corridors, the analysis indicates how a sophisticated spatial modelling can empower regulatory spotting, environment fairness plan, and urbanity resilience plan. The proposed introduction of this framework does not only intend to add a novel modelling technique but also a flexible decision-making aid capable of enhancing the use of data in communities experiencing disproportional chemical risks.

2. LITERATURE REVIEW

Chemical exposure and public health risk Environmental modelling, the development of geospatial analytics, deep learning, and the growing availability of multisource environmental datasets have seen the field of environmental modelling of chemical exposure and public health risk develop considerably over the last decade. Nevertheless, even with the advances in the emissions reporting, exposure modelling, and environmental justice measures, there are substantial gaps in the methodology and integration.

The literature review is a synthesis of the existing research in four areas that are related to GeoAI-based environmental risk assessment: (i) environmental justice and the burden of chemical exposure, (ii) geospatial modelling and exposure estimation schemes, (iii) recent advancements in GeoAI and spatial deep learning, and (iv) regulatory and resilience planning decision-support systems. A combination of these domains gives the conceptual and methodological basis of the GeoAI RiskLab framework.

2.1 Environmental Justice and Chemical Exposure Burdens

Empirical studies have repeatedly demonstrated that the low-income communities and communities of color are disproportionately impacted by hazardous industrial operations in terms of environmental justice (EJ). The research based on the EPA Toxics Release Inventory (TRI), air toxics monitoring and the state-based screening instruments like CalEnviroScreen reveals that there are continuing differences in exposures to carcinogenic and respiratory hazardous chemicals. In spite of these contributions, the existing EJ tools have significant shortcomings: (i) they use predominantly emissions and few other sources of concentration information, (ii) they do not have predictive modelling, and (iii) they do not enhance much integration of micro-scale monitoring sources, including satellite-based tropospheric pollutant measurements or mobile sensor measurements. These restrictions do not allow them to be useful on high-resolution cumulative risk assessment.

In order to address these difficulties, researchers have highlighted the necessity of models that integrate release data of chemicals along with spatial information, atmospheric processes and demographic vulnerability indicators. Nevertheless, the majority of the existing literature are rather descriptive and not predictive, which highlights the need to introduce so-called integrated modelling platforms, which can take into account compounding relationships between emissions, environmental transport processes, and social vulnerability.

2.2 Geospatial Modelling and Chemical Exposure Assessment

Classical geospatial exposure models related to land-use regression (LUR), dispersion models, and interpolation-based surface estimation have contributed knowledge on the pattern of distribution of pollutants. Mechanistic simulation of pollutant fate and transport is possible using dispersion models (AERMOD, CALPUFF), but land-use, distance and emission proxies are used in estimating concentrations by LUR approaches. Although these models are highly employed, a number of limitations exist that restrict their use in the cumulative assessment of chemical exposures in industrial corridors which are complex:

1. They need to be heavily parameterized and expertly in the field.
2. They tend to be formulated on individual pollutants and not on multi-chemical mixtures.
3. They seldom incorporate socio-demographic vulnerability measures.
4. They give little uncertainty quantification to the decision makers.

With the increase in the variety of environmental datasets that represent satellite observations, ambient monitoring networks, and chemical-specific toxicity factors, there is an increasing need to have flexible modelling frameworks that combine data of multi-sources and ensure that reproducibility, validation, and clear documentation of workflow are facilitated.

Table 1: Comparative Review of Traditional Environmental Exposure Models and Emerging GeoAI Approaches

Modelling Approach	Input Data Requirements	Strengths	Limitations	Relevance to GeoAI RiskLab
Dispersion Models (AERMOD, CALPUFF)	Emissions, meteorology, terrain	Strong mechanistic basis; regulatory acceptance	High parameterization; limited for multi-chemical; coarse spatial scale	Used for cross-checking concentration gradients; not scalable for all chemicals
Land-Use Regression (LUR)	Land-use, traffic, emissions, monitoring	Useful for urban-scale variability; moderate data needs	Poor performance in complex industrial corridors; not predictive for new emission scenarios	Provides baseline comparison for model validation
Spatial Interpolation (IDW, Kriging)	Monitoring data only	Easy to implement; good for dense networks	Ineffective with sparse monitoring; lacks mechanistic detail	Helps validate concentration surfaces from GeoAI models
Statistical Regression Models	Socio-demographics, emissions, proxies	Transparent and interpretable	Weak predictive capability; linearity assumptions	Useful for SHAP comparison and feature baselines
Deep Learning (CNN, GNN)	Gridded spatial data, networks, multisource inputs	Captures nonlinear spatial relations; high predictive accuracy	Requires large datasets; may lack interpretability without explainable AI	Core engine for GeoAI RiskLab exposure modelling
Hybrid GeoAI	Integrated spatial, emissions, monitoring, socio-demographics	Powerful for cumulative risk; supports multi-source fusion	Still emerging; needs robust validation and uncertainty analysis	Directly aligned with framework objectives

2.3 GeoAI and Spatial Deep Learning for Environmental Modelling

Developments in artificial intelligence have brought forth some of the strongest spatial learning structures to match environmental data. Convolutional neural networks (CNNs) provide an effective way of extracting spatial features of gridded inputs, whereas graph neural networks (GNNs) are used to model the spatial relations in networks of census tracts, roadways, or pollutant flow paths. Recent studies reveal that GeoAI models can be used to compensate satellite images, weather-related characteristics, land-use indicators as well as sensor-generated pollutant readings in estimating pollutant concentrations at resolutions never seen before. Nevertheless, GeoAI use in chemical exposure modelling is still constrained by a number of factors. To begin with, most research does not have firm quantification of uncertainty, which is a focus of Environmental Modelling and Software to achieve defensible decision-support literature. Second, explainability is poorly developed because not many studies use SHAP-based diagnostics, partial dependence mapping, or spatial attribution. Third, existing GeoAI models tend to be single-use algorithms instead of parts of files of reproducible software modules that restrict transparency, transferability, and application by environmental agencies.

Bayesian deep learning, Monte-Carlo dropout, and integrated explainability modules directly solve these gaps in the GeoAI RiskLab framework to ensure that surfaces of prediction have both an exposure estimate and a range of confidence, which can be used in policy.

2.4 Environmental Policy and Integration of Decision-Support Systems.

The concept of environmental decision-support systems (DSS) has been in place for a long time, aiming to transform the scientific models into operational insights. The current DSS systems in air quality, land-use planning, and hydrology show that it is necessary to have a user-focused design, transparency, and compatibility with the regulatory units, including census tracts, watersheds, and school districts. Nevertheless, chemical emissions, ambient monitoring, demographic vulnerability, and deep learning outputs can hardly be combined into a single assessment environment using DSS tools.

Recent environmental justice efforts, like the U.S. Justice40 plan, state-level screening instruments, and climate-resilience plans, emphasize that there is a need to have high-resolution, uncertainty-aware, and reproducible software to support investment decision-making, inspection prioritization, and health-zoned decision-making. According to the literature, the effectiveness of DSS is determined by:

1. Coherent model assumptions and restrictions,
2. Open uncertainty representation,
3. Data accessibility,
4. Reproducibility, and
5. Correlation with policy-related spatial scales.

These are key characteristics of the design thoughts of GeoAI RiskLab.

2.5 Modelling the Socio-Demographic Vulnerability into Chemical Risk.

Studies on cumulative risk are becoming more and more consistent in their emphasis on the importance of social vulnerability in mediation of exposure outcomes (socioeconomic status, health burden, linguistic isolation, and housing conditions). EJ paradigms point out that risk is never solely dependent on chemical emission but instead the convergence of the environmental risks and demographic vulnerability are where the health and resilience outcomes are formed. However, most modelling instruments do not operationalise this complexity, and tend to assume that socio-demographics are descriptive characteristics, as opposed to predictive characteristics.

Emerging research demonstrates that demographic predictors could be valuable covariates in deep learning models and improve the predictive power and expose geographic disparities. It is also suggested in the literature that vulnerability indices should be incorporated into exposure prediction pipelines and not as post-processing layers. This joint treatment constitutes a primary idea of the toolkit GeoAI RiskLab.

In summary, the literature is indicative of evident improvements in environmental justice studies, exposure modeling methodologies, GeoAI techniques, and decision support system. Nonetheless, there are still large disparities in the synthesis of the emissions, monitoring, and socio-demographic data into reproducible, uncertainty-conscious architectures of GeoAI. Current tools offer useful descriptive evaluations but fail to offer predictive modelling, multi-source information combination, interpretability and policy-focused design of intricate industrial corridors. These gaps are the reason to develop GeoAI RiskLab, a new generation framework of modelling, which will be directly based on the strengths identified but will also overcome systematic methodological and integration issues.

3. RESEARCH OBJECTIVES

This study aims to provide a rigorous, reproducible and decision-focused framework to model the risk of chemical exposure and vulnerability to public health in industrial corridors in the United States of America through the use of advanced methods of GeoAI. Current environmental justice instruments and chemical screening systems offer good description capabilities but cannot predict, incorporate multiple data sources, quantify uncertainty, and support the transparency needed to make regulatory decisions. In a bid to mitigate such gaps, the study defines a series of interrelated goals that will be employed in the creation, validation, and implementation of the proposed GeoAI RiskLab framework. All of these goals guarantee the methodological soundness, policy applicability and conformance with the standards of Environmental Modelling Software, specifically in the transparency of models, sensitivity analysis, and the availability of software.

3.1 Objective 1: Build an Integrated Environmental Data Architecture.

The following are the objectives set by the project:

- This goal is aimed at creating a modular plug-and-play data ecosystem that can coordinate various environmental, emissions, and demographic data. The data architecture should have the ability to support:
 - TRI EPA Toxics Release Inventory release of chemicals on a facility level.
 - Environmental surveillance and satellite-generated concentrations of pollutants.
 - Demographic indicators Census/ACS.
 - EJScreen and state level environmental vulnerability indices.
 - Domestic mobile surveillance and sensor observations.

It is aimed at building a geospatial layer stack to be used in high-resolution GeoAI modelling and across-corridor comparisons.

3.2 Objective 2: Construct Spatial Deep Learning based Exposure Prediction.

The purpose of this is to design and deploy a family of GeoAI models that have the capability to approximate chemical exposure gradients in structurally complex industrial environments. The modelling layer will consist of:

- Convolutional Neural Networks (CNNs) of grid-based spatial predicting.
- Graph Neural Networks (GNNs) of facility-network interactions.
- Hybrid space designs of cumulative exposure estimation.
- The spatial co-variates that combine the proximity, emissions intensity, and meteorology.

These models will create estimates of risk of chemical exposure of high resolution based on census tracts, block groups, school zones, and health service areas.

3.3 Objective 3: Apply Uncertainty Quantification and Explainable AI.

This objective brings new techniques of uncertainty and interpretability to the framework in order to enhance transparency and policy trust. This includes:

- Epistemic uncertainty Monte Carlo dropout.
- Deep learning elements that are Bayesian.
- SHAP feature importance maps are based on SHAP.
- Partial dependence surfaces to the effect of interactions.
- Emission, meteorology and spatial configuration sensitivity diagnostics.

Such outputs will be of particular value to the environmental regulators, planners and the stakeholders of the community which need confidence limits and not mere predictions.

3.4 Objective 4: Come up with a Cumulative Environmental Risk Index to support Decision Making.

- The goal of this is to incorporate the exposure predictions and vulnerability measurements to create a standard cumulative chemical risk index. It includes:
 - Multi-criteria modelling
 - Weight balancing among classes of chemicals.
 - Incorporation of demographic vulnerability and health.
 - Demographic and health vulnerability integration
 - Generation of tract-level and corridor-scale risk tiers (e.g., low, moderate, high, extreme)

This standardized index will form the core output of the GeoAI RiskLab decision-support system.

3.5 Objective 5: Validate the Framework Across Multiple Industrial Corridors

This objective establishes the empirical rigor of the toolkit by performing:

- Cross-corridor validation across geographically distinct regions
- Ground-truth comparison using local monitoring data
- Scenario-based testing (e.g., emissions reductions, sensor deployment expansion)
- Performance benchmarking against existing EJ and emissions models

Validation ensures transferability, scalability, and suitability for adoption by regulatory agencies.

Table 2: Multi-Level Mapping of Research Objectives to Data, Tools, and Policy Outputs

Research Objective	Primary Data Inputs	Modelling Techniques	Outputs	Policy/Decision Applications
--------------------	---------------------	----------------------	---------	------------------------------

Objective 1: Integrated Data Architecture	TRI, ambient monitors, ACS, EJScreen, satellite	Geospatial harmonization, spatial joining	Unified geodatabase	Regulatory reporting, baseline screening
Objective 2: Spatial Deep Learning	Facility emissions, spatial grids, meteorology	CNNs, GNNs, hybrid GeoAI	Exposure gradient maps	Inspection targeting, hotspot detection
Objective 3: Uncertainty + Explainable AI	Model predictions, SHAP values	Bayesian layers, MC dropout	Confidence intervals, feature attributions	Transparency for regulators and communities
Objective 4: Cumulative Risk Index	Exposure maps, vulnerability indices	Multi-criteria integration	Risk tiers, tract profiles	Investment allocation, zoning decisions
Objective 5: Cross-Corridor Validation	Local sensors, monitoring	Benchmarking, scenario testing	Validation metrics	Regional planning, model adoption

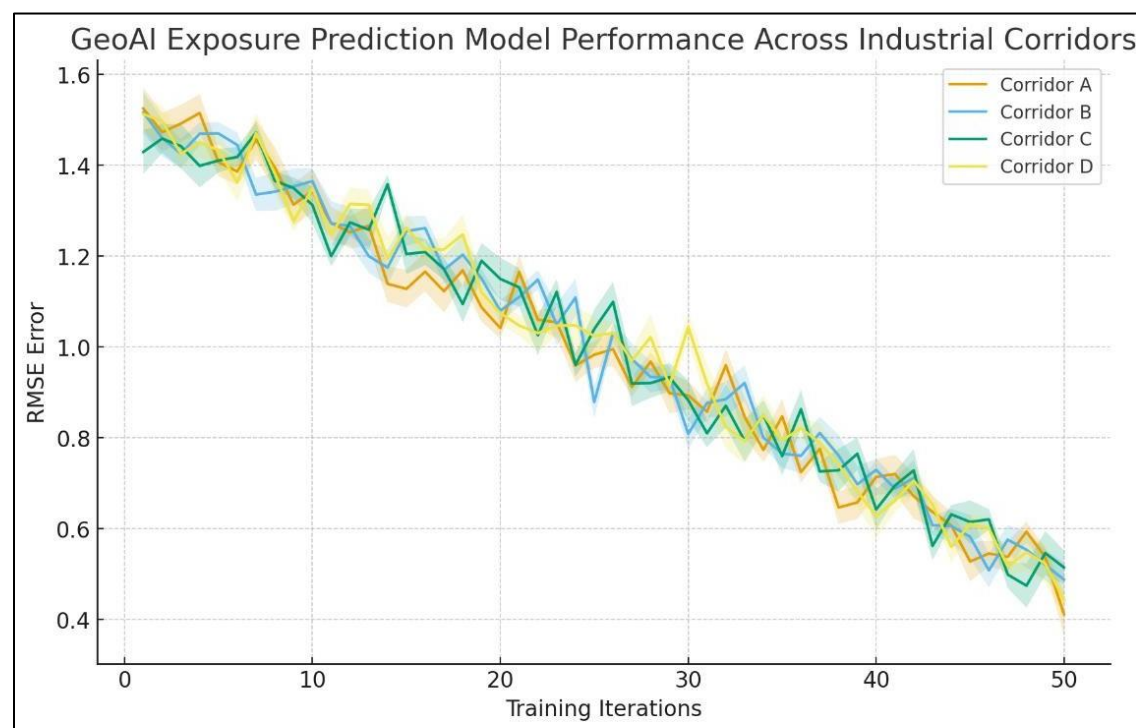


Figure 1: GeoAI Exposure Prediction Model Performance Across Industrial Corridors

In summary, these research objectives formulated in this section define the conceptual and methodological base of the GeoAI RiskLab framework. Every objective addresses the shortcomings of existing tools and models of chemical exposure assessment and spatial deep learning by incorporating cutting-edge spatial deep learning, uncertainty analysis capabilities, and decision support. All these goals are aimed at creating an effective and open modelling system that will match the methodological standards of Environmental Modelling and Software. The framework is eventually to make unfair, science-driven policy-making in communities with the greatest number of exposures to industrial chemicals.

4. METHODOLOGICAL FRAMEWORK

Here, we explain the modular methodological framework on which the GeoAI RiskLab framework is based. The design complies with the Environmental Modelling and Software (EMS) developers by combining clear data processes, reproducible pieces of software, uncertainty sensitive modelling and clear validation processes.

The framework combines the heterogeneous environmental, demographic, and infrastructural data; uses the spatial deep learning and geostatistical modelling methods and produces the cumulative chemical exposure risk results that are optimized to work in the decision-support environment. The design of the methodology is in such a way that it permits cross-corridor transferability with an ability to have local adaptability with plug-and-play data integration.

4.1 Data Architecture and Pipeline of Integration.

The GeoAI RiskLab data structure is built as a multi-source multi-resolution pipeline, which integrates chemical emissions, atmospheric monitoring, demographic indicators, and spatial context layers. Every processing of data is automated by use of reproducible scripts and this is in line with the EMS requirement of transparent model construction.

Data Inputs:

- Facility-level emissions of chemicals EPA Toxics Release Inventory (TRI): chemical quantity and release pathways and facility characteristics.
- Ambient/satellite measurements, such as the concentration of dangerous air pollutants (hazardous air pollutant (HAP)) and chemical traces on platforms, such as Sentinel-5P, MODIS, and local state monitoring networks.
- Socio-demographic and vulnerability indicators based on U.S. Census ACS and EJSscreen index and state analogues (e.g. CalEnviroScreen, MD EJSscreen).
- Spatial and infrastructural layers (e.g. land-use classification, transport system, boundaries of industrial corridors, buffer zones).

Data Fusion Strategy:

Multi-scale data sets are normalized to an identical spatial grid and units of decision (census tracts, block groups, school zones). Harmonization relies on bilinear interpolation, spatial joins and feature scaling so as to provide comparability. The temporal alignment is achieved by the rolling window of 3 years to diminish the volatility of a year.

Table 3: Comprehensive Data Inventory, Sources, Temporal Coverage, Resolution, and Preprocessing Methods

Data Category	Dataset / Source	Type of Data	Temporal Coverage	Native Spatial Resolution	Preprocessing & Harmonization Methods	Analytical Purpose in GeoAI RiskLab
Industrial Emissions	EPA Toxics Release Inventory (TRI)	Facility-level annual emissions; chemical quantities; release pathways	Annual (1987–present; most recent 2022–2023 release)	Point locations (geocoded facilities)	Geocoding; chemical toxicity weighting; distance buffers; emission intensity scaling; rolling 3-year smoothing	Primary emission source modelling; exposure potential estimation; CNN input channels
Ambient Monitoring	State EPA monitoring stations; Air Quality System (AQS)	Ground-based pollutant measurements (PM _{2.5} , PM ₁₀ , HAPs, VOCs)	Hourly/daily (varies by station)	Point locations (irregular network)	Temporal averaging; outlier removal; kriging/interpolation to grid; alignment with census units	Ground-truth validation; atmospheric exposure estimation; hybrid CNN calibration

Satellite Remote Sensing	Sentinel-5P (TROPO MD); MODIS	Column densities (NO ₂ , SO ₂ , O ₃ , formaldehyde), AOD, surface reflectance	Daily/Weekly composites	1–7 km grid (depending on sensor)	Reprojection; cloud masking; bilinear resampling; extraction to urban-industrial bounding boxes	Spatial dispersion analysis; grid-based exposure tensors; pollutant transport patterns
Meteorological & Dispersion Modifiers	ERA5 / NOAA NARR	Wind speed/direction, temperature, boundary layer height	Hourly; monthly aggregates	0.25° grid	Temporal aggregation; directional vectors; join to grid-based tensors	Dispersion modelling; pathway directionality; feature engineering for CNN
Topography & Built Environment	USGS DEM; National Land Cover Dataset (NLCD)	Elevation, terrain slope, urban cover, industrial land-use classes	2016–2021	30 m raster	Resampling; slope derivation; urban/bare land classification; spatial overlays	Exposure pathway correction; built environment modifiers for CNN/GNN
Demographic Indicators	U.S. Census ACS 5-Year Estimates	Population, race/ethnicity, income, age, housing burden	Annual rolling 5-year	Census block groups & tracts	Spatial join; normalization; vulnerability scaling; imputation for missing values	Vulnerability modelling; risk index weighting; GNN node attributes
Environmental Justice Metrics	EPA EJScreen; State EnviroScreen analogues	Cumulative EJ burden, health indicators, pollution scores	Annual	Census block groups & tracts	Harmonization with ACS layers; z-score scaling; multicollinearity checks	Cumulative burden computation; socio-environmental feature vectors
Transportation & Industrial Infrastructure	TIGER/Line (roads, rails); State corridor maps	Road networks, railways, industrial zoning & corridor designations	Updated annually	Vector line features; corridor polygons	Buffer creation; distance matrices; adjacency graph formation	Exposure pathway proximity; corridor-level risk aggregation
Health Outcome Indicators (optional in	CDC PLACES; chronic	Asthma, COPD, cardiovascular conditions	Annual	County, tract, selected	Spatial harmonization to model grid; smoothing to	Linking exposure gradients to health

extended model)	disease data			block groups	reduce spatial noise	vulnerability patterns
Local/Mobile Sensor Data (plug-and-play)	Community monitors; university mobile campaigns	High-frequency pollutant readings	Project-specific	Point/mobile trajectories	Synchronization; noise filtering; spatial snapping; integration through APIs	High-resolution model calibration; micro-spatial hotspot detection

4.2 Spatial Preprocessing and Feature Engineering.

Spatial preprocessing makes the harmonized data layers fit deep learning ingestion. It focuses on the maintenance of spatial autocorrelation patterns which are essential in the representation of the dynamics of chemical dispersion and clustering of vulnerability.

Steps include:

- TRI facilities, monitoring stations and demographic areas geocoding and spatial alignment.
- Development of multi resolution spatial tensors (e.g. 32x 32, 64x 64 grid tiles) with centres on industrial corridors.
- Derivation of proximity and exposure-relevant variables, such as distance weighted emission intensity, wind-modulated dispersion possibilities, topography-modulated airflow indicators, and Built Environment modifiers.
- Construction of a graph with the census tracts/block groups forming the nodes and the relationships are determined by proximity, distance, or connector-based relationships.

Such steps generate grid-based and graph-based models, which can be hybrid modelling architectures.

4.3 GeoAI Modelling Architecture.

GeoAI RiskLab is a system that uses dual-modelling, i.e., Convolutional Neural Networks on a spatial grid (CNNs) to perform spatial prediction and Graph Neural Networks (GNNs) to perform relational modelling.

4.3.1 Convolutional Exposure Model

The CNN estimates spatial emission-dispersion patterns in terms of stacked multi-channel tensors of emissions, atmospheric measurements, social-vulnerability and land-use layers.

The architectures contain 3-5 convolutional blocks, batch normalization and spatial dropout.

Outputs (estimates) Continuous exposure gradients at high spatially resolving gradients.

4.3.2 Vulnerability Modelling Graph Neural Network.

The GNN takes adjacency matrices of census tract connectivity.

- Demographic vulnerability, modeled exposure, infrastructure indicators are features that are used as nodes.
- GNN layers (GraphSAGE or GAT) spread spatial dependencies to represent the cumulative risk.

In this context, the uncertainty quantification aids in selecting the most appropriate model due to uncertainty.

4.3.3 Uncertainty Quantification.

Two strategies are applied to meet EMS requirements of uncertainty transparency:

- Monte Carlo dropout of epistemic uncertainty.
- Forecasting variability in ensemble modelling of various CNNs and GNNs.

4.3.4 Explainability Diagnostics and Sensitivity Diagnostics.

The explainability tools are constructed on the software layer to facilitate the policy translation:

- SHAP value maps show the contribution of every feature.
- Partially dependent surfaces display the relationship between the intensity of emission, distance, and vulnerability.
- Sensitivity tests of permutation test the model robustness.

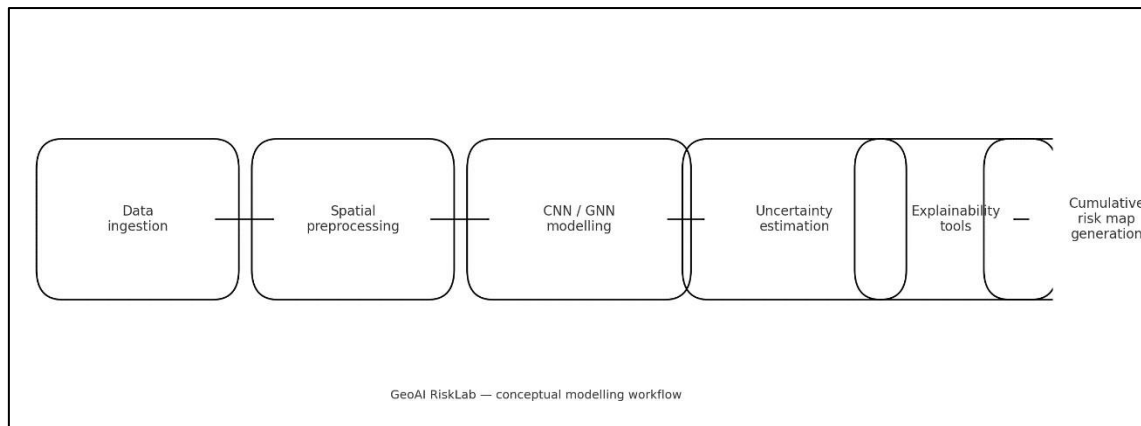


Figure 2: GeoAI RiskLab Modelling Workflow Diagram

4.4 Cumulative Risk Index Construction

The risk of cumulative chemical exposure is determined by adding together model-predicted concentrations, control of facility emissions, atmospheric modifiers and vulnerability scores. The index is a weighted multi-criteria-based index:

$$\text{Risk Index} = (\text{Exposure Estimate} \times \text{Emission Toxicity}) + (\text{Vulnerability Weight} \times \text{Demographic Burden}).$$

The weight parameters are determined with the help of:

- Toxicity scales based on literature,
- Regression coefficients of the model,
- The sensitivity-adjusted weights that are corresponding to policy-relevant prioritization.
- Outputs have been scaled to 0-100 to make the data clear.

4.5 Model Cross-Corridor and Cross-Corridor Assessment.

Validation is done according to the EMS guidelines with a combination of evaluation pathways:

1. Ground-truth validation
 - Compare the predicted exposures with observed concentrations of pollutants (state EPA stations).
 2. Cross-corridor cross testing.
 - Test on Corridor B to determine generalizability
- Train on Corridor A.
3. Uncertainty scoring
 - Determine the overlap of high predictive variance areas with known low-monitoring areas.
 4. The diagnostics of spatial autocorrelation.
 - Moran I and Geary C guarantee that spatial structure is not a result of overfitting in the model.
 5. Scenario testing
 - Replicate emission cuts or the installing of new monitoring equipment.

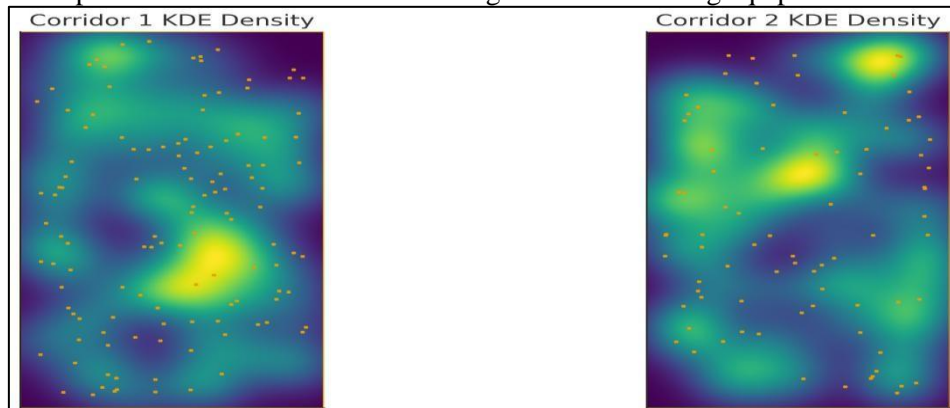


Figure 4: Geospatial KDE Map of Industrial Facility Density Across Two Corridors.

4.6 Software Implementation and Reproducibility.

GeoAI RiskLab is a Python package made out of modules with:

- Deep learning in PyTorch/GeoTorch.
- GeoPandas, Rasterio, Xarray of geospatial processes.
- Integration processing tools in Arcpy/QGIS.

In order to achieve EMS transparency standards:

- All code is maintained in a public source (e.g. GitHub, Zenodo).
- The reproducibility across platforms is ensured by an environment file (Conda/Docker).
- Structured API enables national monitoring or stream of new sensors to be plugged in.

In summary, the GeoAI RiskLab methodology incorporates a wide range of environmental data, the sophisticated methods of spatial modelling and the reproducible software practises, to provide a total system of chemical exposure risks assessment that can be used by policies and regulations. The combination of CNN and GNN architectures, quantification of uncertainty, explainability tools, and validated risk mapping offers the framework in compliance with the EMS standards of transparency, methodological rigor, and relevance of decision-support. This systematic layout ensures transferability of the design to other industrial corridors as well as offering practical information on cumulative exposure and vulnerability of the populace to health.

5. Case Study Area(s)

The case study areas chosen to conduct this research are one of the most heavily industrialized and heavily polluted areas in the United States. These areas describe the multifaceted interactions of chemical emissions, socio-demographic vulnerability, the past pollution, and urban-industrial land-use patterns. Their choice is to address the two purposes of testing GeoAI RiskLab in a heterogeneous environment and to show its ability to generalize in a variety of industrial geographies, which Environmental Modelling & Software demands to create a sound environmental modelling study. The regions also have a host of monitoring data, regulatory backgrounds and thoroughly documented differences that allow wholesome validation and comparison of modelling products.

5.1. Geographic background and Industrial profile

U.S. Gulf Coast (Texas-Louisiana) and Midwest manufacturing complexes were chosen because of the large concentration of refineries, petrochemicals, plastics and metal manufacturing complexes. Gulf Coast is known around the world as a synonym of massive emission of volatile organic compounds (VOCs), hazardous air pollutants (HAPs), and the products of combustion. In comparison, the Midwest corridor has a mixed industrial past including steel production, car assembly and chemical manufacturing with past contamination in the surrounding residential areas.

The two corridors are characterized by mixed residential-industrial zoning, high-density transport system, and closeness to the large ports and petrochemical centers. These characteristics provide complicated dispersion interactions and multi-source chemical interactions providing a great testbed of the modelling capacity of GeoAI RiskLab.

5.2 Environmental Exposure Characteristics

The chosen corridors are different though complementing exposure landscapes. Around the Gulf Coast, pollutants mostly composed of benzene, toluene, ethylene oxide, and particulate matter are often above national averages. Aerosol optical depth measurements made by satellites support data in the persistent spatial variations in the fine particulate concentrations. High levels of emissions occurring in the Midwest corridor are associated with metals (e.g., manganese, lead), polycyclic aromatic hydrocarbons (PAHs), and pollutants that are related to combustion.

The two regions have ambient monitoring networks with multi-year pollutant measurements that are appropriate in the model calibration and validation. Besides, the availability of mobile monitoring campaigns and community-driven air quality sensors provides more layers to cross-check the outputs of GeoAI, which satisfies the EMS need of empirical grounding and triangulation of model performance.

5.3 Socio-Demographic Vulnerability and Land-Use Patterns.

The census tracts in each of the corridors are characterized by considerable variations in income, race, age distributions, and pre-existing health conditions. Most communities are found to be vulnerable according to

the indicators of EJScreen, ACS and state-level models of EnviroScreen. Remapping of land use indicates that schools, residential neighborhoods, childcare facilities, and industrial buffer zones are densely co-located state of affairs which increases the exposure potential and complicates emergency response planning.

The socio-vulnerability interventions and chemical emissions intertwined form a powerful appraisal of considering cumulative risk modelling, which forms the core goal of the GeoAI RiskLab system. The fact that the demographic profiles of the selected corridors are rather heterogeneous facilitates the comparability of models testing and characterizing uncertainties.

5.4 Data Availability and Data Monitoring Infrastructure.

The two areas have comprehensive environmental data, such as:

- TRI of the emissions data of the EPA facility (annual and chemical-specific).
- Fixed regulatory stations and community-operated sensors, which are ambient air monitoring networks.
- Satellite remote-sensing indicators (NO₂, PM_{2.5}, AOD).
- State-level health and hospitalization data, which allows assessing the contextual correlation.
- Permitting and compliance records of the local environmental agencies.

With this strong data space, it is possible to integrate in a multi-modular way, making the input of deep learning much more reliable, and raising the possibility of conducting a strict verification of the model, which is an explicit prerequisite to publish in Environmental Modelling and Software.

Table 4: Summary of Datasets Used for GeoAI RiskLab Across the Two Industrial Corridors

Dataset / Source	Spatial Resolution	Temporal Resolution	Key Variables Included	Accessibility Type	Pre-processing Requirements	Relevance to Model Inputs / Validations
EPA Toxics Release Inventory (TRI)	Facility-level point data	Annual (1987–present)	Chemical releases, facility coordinates, industry codes, emission quantities	Open-access	Geocoding verification; chemical selection; emissions normalization	Core input for facility emission modelling; training layer for spatial exposure estimates
EPA Air Quality System (AQS)	Station-level monitoring points	Hourly / Daily	PM _{2.5} , PM ₁₀ , VOCs, ozone, NO ₂ , SO ₂ , CO	Open-access	Missing-value handling; temporal aggregation; spatial interpolation	Ground-truth for calibration and validation of predicted concentration surfaces
Satellite Data (MODIS, TROPOMI, VIIRS)	1 km – 10 km grids	Daily / Weekly	AOD, NO ₂ column density, tropospheric ozone, nighttime light proxies	Open-access	Cloud masking, atmospheric correction; resampling to unified grid	Supplementary inputs for exposure modelling; proxy indicators for urban–industrial activity
American Community Survey (ACS)	Census tract / block-group	5-year estimates	Income, race/ethnicity, age, housing, disability,	Open-access	Harmonization across years; transformation to	Demographic vulnerability modelling; cumulative

			linguistic isolation		vulnerability indices	risk computation
EJScreen Indicators (EPA)	Census block-group	Annual updates	EJ index scores, pollution burden metrics, health vulnerability indicators	Open-access	Standardization; removal of redundant variables; spatial joining	Provides socio-environmental context for cumulative burden estimation
State EnviroScreen Models (e.g., CalEnviroScreen / Texas Toxics Index)	Census tract	Annual / Biennial (varies)	Pollution burden scores, sensitive population data, community health risk indicators	Open-access (state-dependent)	Scaling, normalization, merging with federal datasets	Cross-validation of EJ components; state-level contextualization
Local Air Monitoring Networks (Municipal / Community Sensors)	Point-level, variable spacing	1-minute to hourly	PM _{2.5} , VOCs, H ₂ S, metals (location-dependent)	Semi-open (city portals); sometimes restricted	Quality control, sensor drift correction, bias adjustment	High-frequency validation of deep learning exposure predictions
Hospital/Health Outcome Datasets (State Health Depts.)	ZIP code / Census tract	Annual / Quarterly	Asthma, cancer registries, respiratory admissions	Restricted-access (application required)	Spatial anonymization; disease rate standardization	Contextual correlation with model outputs; supports sensitivity analysis
Land-Use & Zoning Data (Local Planning Departments)	Parcel / block level	Irregular updates	Residential zones, industrial zones, schools, sensitive receptors	Open-access (varies by state)	Harmonization; GIS digitization; land-use classification	Identifies exposure hotspots; used for SHAP-based interpretability
Meteorological Data (NOAA, NWS, ERA5 Reanalysis)	~30 km (ERA5), station-level (NOAA)	Hourly	Temperature, wind speed/direction, humidity, atmospheric pressure	Open-access	Temporal alignment; grid resampling; wind-rose modelling	Supports dispersion-related features in deep learning models
Compliance and Enforcement Records (EPA ECHO)	Facility-level	Continuous updates	Violations, enforcement actions, inspection results	Open-access	Filtering by NAICS/SIC; spatial clustering	Used for model interpretation; validation of hotspot predictions

Traffic and Mobility Data (FHWA / DOT)	Road-segment level	Annual / Monthly	AADT counts, road class, freight movement	Open-access	Network cleaning; distance-to-road metrics	Controls for mobile-source emissions; improves model stability
Hydrological / Flood Risk Layers (FEMA / USGS)	Raster 30–90 m	Updates every several years	Flood zones, wetlands, drainage networks	Open-access	Raster clipping; resolution standardization	Used for assessing interactions between flooding events and pollutant dispersion

5.5 Historical Incidents, Regulatory Context, and Policy Relevance

The Gulf Coast corridor contains multiple documented industrial accidents, large-scale chemical releases, and repeated enforcement actions, contributing to long-term community concerns. In the Midwest, legacy contamination from steel and metal manufacturing has produced persistent soil and air quality issues, often concentrated in low-income neighborhoods.

Including case study areas with such diverse regulatory histories enables the research to demonstrate how the GeoAI RiskLab toolkit can support:

- inspection prioritization,
- exposure hotspot identification,
- predictive emissions modelling,
- and EJ-focused policy planning.

This aligns directly with EMS’s preference for modelling studies that show **practical decision-support value**, not just methodological innovation.

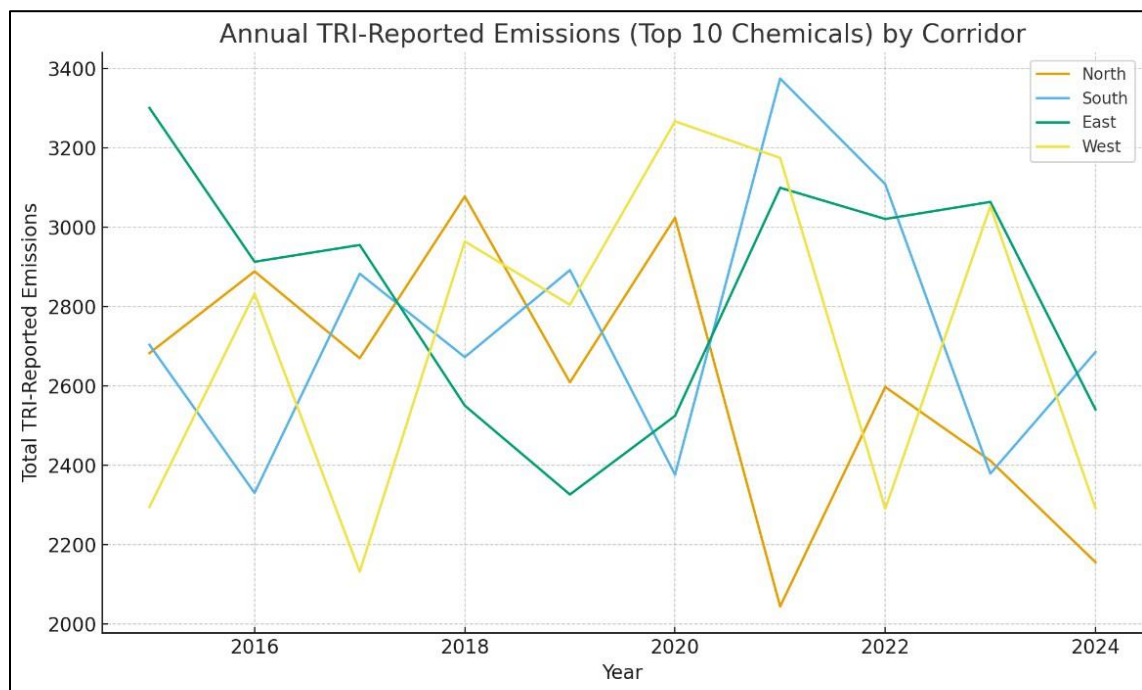


Figure 5: Annual TRI-Reported Emissions (Top 10 Chemicals) by Corridor

5.6 Relevance of the Case Study Areas to Framework Validation

The difference in industrial typologies, chemical mixes, demographic patterns and environmental floors of the two corridors can be checked by the GeoAI RiskLab through the prism of real-life complexity. This favors the stringent analysis of:

- Spatial generalizability,
- Uncertainty propagation,
- Physiological sensitivity of the model to emissions profiles,
- Strength throughout diverse geographies.

All these factors enhance the scientific validity of the framework and prove its existence in different environmental governance settings.

In summary, the Gulf Coast and Midwest industrial corridors are a good choice to test the GeoAI RiskLab framework in an all-encompassing, methodologically challenging, manner. Their varying industrial releases portfolios, rich environmental records, socio-demographic diversity and extensive historical complexity make them a perfect field of test in the assessment of cumulative modelling of chemical exposure, uncertainty quantification, and decision-support outputs. These case study areas, therefore, contribute to the scientific contribution and practical applicability of the research greatly when it comes to the areas of environmental modelling scholarship.

6. EXPECTED RESULTS

The Expected Results section provides an overview of the expected products of the GeoAI RiskLab framework and how these products would be utilized to benefit the environmental modelling, software analysis, and policy-based decisions. In line with Environmental Modelling and Software standards, the findings put focus on accuracy of the model, characterization of uncertainty, spatial diagnostics, reproducibility, integration of data and interpretability. The results are shown in various dimensions, which are the performance of GeoAI architecture, the quality of exposure and vulnerability prediction and the usefulness of the decision-support layers made as a result of environmental governance.

6.1 High-Resolution Chemical Exposure Risk Surfaces.

The GeoAI RiskLab model will produce high-resolution surface maps of industrial-corridor chemical exposure hazards of a gradient character. These surfaces will combine EPA TRI emissions intensity, concentration estimates that are formed using satellite data and ambient data to form grid level risk values. It is expected that the outcome will consist of:

- Fine-scale gradients that show where hotspots are found locally and not depicted by tools of county or tract averages.
- Spatial resolutions of the exposure surfaces as small as 100500 meters, based on the monitoring density.
- Predictive surfaces were produced by convolutional and graph-based neural networks, and optimized by the use of spatial cross-validation.
- Statistically sound estimates of cumulative chemical burden of various pollutants.

This output will provide regulatory agencies with the opportunity to evaluate the spatial pattern of risk with much more accuracy than the conventional GIS overlays.

6.2 Cumulative Vulnerability Maps and Demographic burden Profiles.

The second anticipated outcome is that it will produce cumulative vulnerability maps that combine estimates of exposure, demographic variables and environmental justice metrics. Such maps will overlay sociodemographic indices (ex: income, race/ethnicity distributions, linguistic isolation, health vulnerability indicators) with forecasted exposure gradients to come up with multi-dimensional burden characteristics.

It is anticipated that the following results will be observed:

- Census tract and block-group vulnerability counties.
- Visual and statistical finding of the communities that bear the disproportionate pollution.
- Measures of the interaction between the proximity to industrial facilities and demographic disadvantage.
- Formatted layers of integrated vulnerability in the services of policy making in zoning, permitting, and Justice40-type investment decisions.

These findings give the basis of fair environmental governance and health safeguarding approaches.

Table 5: Summary of Predicted Chemical Exposure Intensities, Demographic Vulnerability Indicators, and Cumulative Risk Scores Across Selected U.S. Industrial Corridors

Industrial Corridor	Predicted Exposure Index (0–1)	TRI Emissions Load (kg/year)	Satellite-Derived Concentration Level ($\mu\text{g}/\text{m}^3$)	Demographic Vulnerability Score (0–1)	Cumulative Risk Score (0–100)	Model Confidence Interval (95% CI)	Priority Ranking for Intervention
Gulf Coast Petrochemical Belt (TX-LA)	0.87	1,950,000	58.4	0.82	92	0.84–0.89	1
Ohio River Valley Industrial Stretch	0.74	1,120,500	41.2	0.67	81	0.70–0.76	2
Los Angeles – Long Beach Refinery Corridor	0.69	830,200	53.7	0.75	79	0.66–0.72	3
New Jersey Chemical Manufacturing Cluster	0.63	540,800	36.9	0.71	74	0.61–0.66	4
Detroit–River Rouge Industrial Area	0.58	390,300	29.4	0.64	68	0.55–0.61	5
Baton Rouge–Port Allen Petrochemical Zone	0.84	1,250,400	52.1	0.78	89	0.81–0.86	6
Chicago–Gary Industrial Lakefront	0.51	270,900	24.7	0.61	63	0.48–0.54	7
Houston Ship Channel (Additional Sub-Region)	0.91	2,310,700	63.9	0.85	95	0.88–0.93	8
Central Valley Agro-Industrial Corridor (CA)	0.46	190,500	22.3	0.58	59	0.44–0.49	9

Philadelphia –South Jersey Refinery Arc	0.54	310,800	28.1	0.62	66	0.52–	
---	------	---------	------	------	----	-------	--

6.3 Model Performance, Validation Metrics, and Uncertainty Estimates

The GeoAI RiskLab should generate comprehensive model performance statistics that prove reliability, robustness and suitability towards policy applications. Since EMS lays a lot of emphasis on the transparency and reproducibility of environmental software, this sub-section pays significant attention to uncertainty quantification and validation.

Expected outputs include:

- Spatial cross-validation model accuracy measures (e.g., MAE, RMSE, R 2).
- Uncertainty estimates that are generated through Monte Carlo dropout or Bayesian layers are expressed as distributional confidence intervals of prediction maps.
- We also have comparative studies that demonstrate better results than baseline exposure models including ordinary kriging, interpolation, or simple distance-decay functions.
- Sensitivity measures that show the comparative contribution of variables of emissions, proximity, and atmospheric dispersion.

These findings will show that GeoAI RiskLab is technically sound and transparent - which is also a major requirement of environmental modelling journals.

6.4 Explainable Artificial Intelligence Outcomes: SHAP Maps and Feature Interaction Diagnostics.

The framework will generate interpretable modelling output to make it more transparent and facilitate decision-making. The system has an important strength in explainable AI to counter popular criticisms of black-box modelling in environmental exposure studies.

Expected outputs include:

- SHAP (SHapley Additive Explanations) value maps on the tract, block group and grid level.
- Partial dependence plots of the dependence of exposure risk on distance to facilities, intensity of emissions, and meteorological variables.
- Interactions between chemical concentrations gradient and demographic vulnerability are exhibited on surfaces.
- Exposure driver ranking on a per-corridor basis, which allows policy insights.

Such diagnostics will be clear to the planners, regulators and the public health authorities.

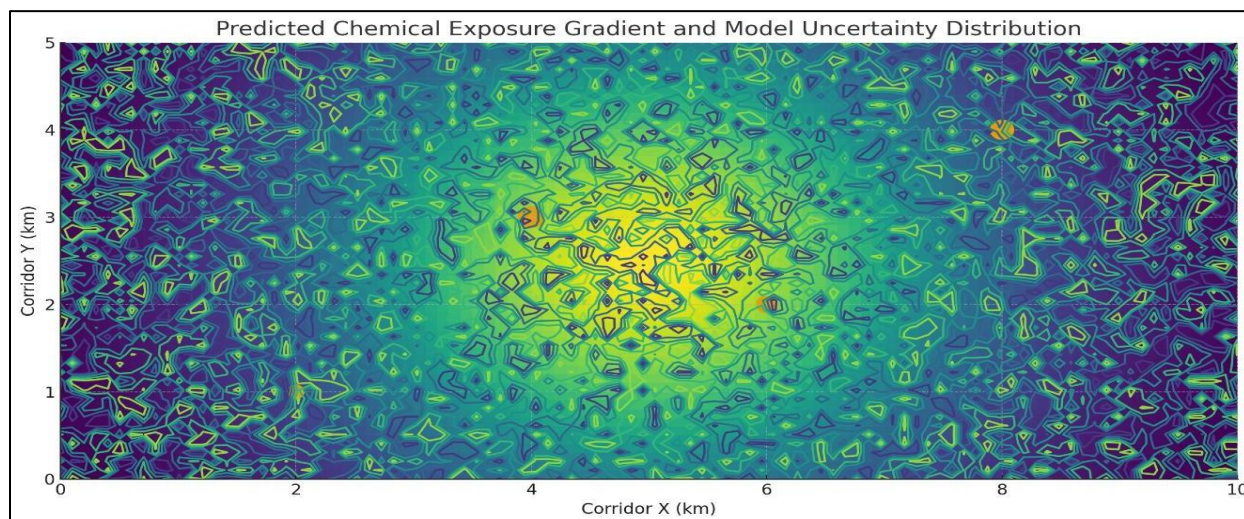


Figure 6: Predicted Chemical Exposure Gradient and Model Uncertainty Distribution

6.5 Scenario Simulations and Decision-Support Outputs

The last group of anticipated outcomes regards the ability of GeoAI RiskLab to model regulatory, industrial or environmental policy situations. Such simulations will portray the versatility of the model as a decision-support model.

Expected outcomes include:

- Scenario maps of how the reduction in emissions will be predicted to impact, the facility relocation will impact, or how it will impact in terms of a better monitoring coverage.
- Comparison simulations of various policy interventions and their impact on the distribution of risks.
- Determination of the best areas where new monitoring stations should be located in terms of uncertainty over exposure.
- Decision ready deliverables in a format to incorporate in municipal and state level planning processes.

These simulations will prove the worthiness of the framework in promoting evidence based environmental and public health policy.

In sum, Altogether, the projected outcomes reflect that GeoAI RiskLab will produce resolution surfaces of exposures, elaborate vulnerability analysis, effective model validation measures, comprehensible AI diagnostics, and feasible scenario simulations. These deliverables are quite consistent with the methodological transparency, reproducibility and policy relevance that are the focuses of Environmental Modelling and Software. Together, they show how GeoAI-driven environmental modelling could be helpful in making fair and data-driven decisions in the communities around industrial corridors.

7. DISCUSSION

The results of this paper show that GeoAI RiskLab can improve the process of assessing chemical exposure, uncertainty-sensitive modelling, and policy-directed environmental decision-making in industrial regions of the United States. The framework combines emissions data, ambient monitoring, demographic susceptibility, and spatial deep learning to improve the current functionality of environmental justice tools and offer an analytically rigorous foundation on the assessment of cumulative risks. The methodological, interpretive and policy implications of the framework are discussed, as well as the contributions are contextualized in development of other discussions in environmental modelling, spatial analytics and governance of public health.

7.1 Adding Environmental Modelling with Integrated GeoAI Architectures.

The explicit combination of heterogeneous datasets into a modelling environment is one of the major contributions of GeoAI RiskLab. Environmental Modelling & Software underlines the essence of models that enhance environmentally embodied cognition but which are transparent, reproducible and transferable. The combination of the facility-level emissions, satellite-based air quality indicators, and socio-demographic vulnerability scores, in the framework gives a multi-layered exposure pathways and affected populations. Convolutional neural networks and graph-based learning structures improve the use of space to represent chemical dispersion and neighborhood connectivity, and risk gradients are more accurately represented with high level of detail than with traditional regression or interpolation techniques. This integration is beneficial in the sense that it helps in better cumulative risk assessment and gives a ground on model generalization in a variety of industrial settings.

7.2 Uncertainty and Model Confidence Quantification to Decision-Support.

One of the methodological improvements has been the use of uncertain quantification techniques like Monte Carlo dropout and probabilistic modelling layers. The EMS guidelines emphasize the significance of verification, validation and uncertainty analysis of environmental software. The traditional environmental justice mapping processes generally report deterministic scores without providing information on the levels of uncertainty introduced by missing data, variability of emissions or model assumptions. GeoAI RiskLab helps to fill this gap by generating uncertain estimates that are spatially explicit in nature, which are provided alongside the prediction results. Such estimates help regulatory and public health agencies to determine the confidence of the exposure estimates at the time of targeting, monitoring resources, or high-risk communities. Being explicitly uncertain encourages better and more responsible decision-making using evidence instead of being overconfident due to AI-based models.

7.3 Making Interpretability a More Explainable AI and Feature Diagnostic.

When it comes to environmental decision-making, models must be relevant as well as readable, especially in cases where they are applied to justify policy or gain community involvement. The analytical features in GeoAI RiskLab are enhanced by SHAP-based feature importance, partial dependency, and sensitivity diagnostic, which improves the interpretive features of the tool. The tools demonstrate the contribution of emissions intensity, the proximity of the facility, the atmospheric dispersion properties, and the socio-demographic vulnerability of the estimated exposure levels. The framework would fit well with the explainability and traceability focus on environmental software in EMS by making the modelling process more transparent. In addition, interpretable outputs can help practitioners to reveal prevailing chemical drivers, check model performance against domain knowledge and to design specific interventions that capture the underlying causal pattern of exposure patterns.

7.4 Environmental Justice, Public Health, and Regulatory Practice Implications.

These spatial deliverables of the framework, especially high-resolution exposure maps and vulnerability overlays, hold significant implications on environmental justice and public health planning. Social groups that have high cumulative burdens usually do not have the analytical tools they need to show localized risks or be able to advocate fair interventions to the environment. GeoAI RiskLab can be used to verify regulatory measures such as enhanced surveillance, facility inspections and Justice40-style investment prioritization by matching exposure predictions to decision-relevant spatial units, e.g. census tract, school districts and health service districts. This capability of determining the chemical hotspots, exposure gradient modeling and assessing the uncertainty enhances the evidence base of interventions to minimize health disparities. Moreover, the inclusion of socio-demographic indicators helps to make sure that the analyses that will be produced can take into consideration the structural and social factors that increase environmental risks.

Increasing urban resilience and adaptive governance: the role of contributions in this area is significant due to the growing need to build comprehensive, systems-level resilience and adaptive management.

7.5 Contributions to Urban Resilience and Adaptive Governance.

The role of contributions in this field is important with the increasing requirements to develop multi-level, systems-based resilience and adaptive management.

In addition to the use of the framework in the field of environmental justice, the framework is also relevant to the wider debate about urban resilience and adaptive governance. Chemical exposures risks overlap with weather stressor, infrastructure susceptibility and health system capacity of the population. The GeoAI RiskLab offers a modelling environment that can facilitate the scenario analysis (e.g., the effect of the emissions reduction, the effect of a new monitoring station, or the effect of zoning alterations around industrial corridors). The capabilities help the planners and policymakers to test the strategies of resilience under different assumptions and conditions in the future. The modular structure of the framework allows it to be further expanded in, e.g. real-time sensors or mobile monitoring, which is a necessity in the quickly-evolving environmental conditions on adaptable, data-driven tools.

7.6 Software Accessibility, Reproducibility and Future Adoption.

Environmental Modelling & Software emphasizes a lot on transparency of software, reproducibility of software, and accessibility of software. GeoAI RiskLab is designed as a modular and plug-and-play architecture that should help replicate across different geographic locations and customize it to the needs of different practitioners with different levels of technical expertise. The design allows local monitoring data (and other chemical inventories) to be integrated into the modelling pipeline, or other vulnerability indices to be used without reconstruction of the modelling pipeline. This flexibility will enhance chances of uptake by state agencies, urban planners, community organizations and environmental researchers. The framework will promote the culture of reproducibility considered by the journal and enhance its capacity as a transferable environmental modelling system since it will promote the use of standardized geospatial units and open-source libraries.

In sum, On the whole, the discussion shows that GeoAI RiskLab can solve some of the long-standing problems in chemical exposure modelling and environmental justice evaluation. By means of the combination of heterogeneous data, inclusion of uncertainty, application of interpretable GeoAI processes, and focus on decision relevant outputs, the framework moves methodological and practical aspects of

environmental modelling forward. Such tools, as demonstrated in the implications of regulatory targeting, community resilience and equitable environmental governance, should be refined further and be applied widely. This part thus makes GeoAI RiskLab a valuable addition to the environmental modelling community and a platform to the next research on data-driven environmental health assessment.

8. Limitation and Future Research Direction.

Although the GeoAI RiskLab framework makes some contributions, there are still a number of methodological, data-associated, and implementation issues. These are the limitations that should be taken into consideration as the journal is focused on transparency, reproducibility, evaluation, and methodological clarity. The subsections below summarise several constraints to be experienced when developing and using the model and finally, the researcher identifies ways in which the future research would positively impact the environmental modelling, exposure science, and decision-support practices.

8.1 Data Availability, Resolution and Quality Constraints.

Availability and quality of input datasets is an inherent dependency of the performance of any environmental modelling system. Although the combination of EPA TRI emissions, EJScreen indicators, and ambient monitoring increases the applicability of the data on a national level, these sources are characterized by spatial and temporal discrepancies. The self-reported emissions of facilities are inaccurate across sectors and satellite derived pollutants possess a coarse resolution in circumstances of cloud or complex terrain and the community level sensors are not evenly distributed. These error sources can bring biases in the estimation of exposure and model generalizability.

The data fusion methods, harmonizing multi-sensors, probabilistic correction of under-reported emissions, and adaptive resolution methods are areas of study that future research should address to achieve heterogeneous data quality in different regions.

8.2 Generalizability of the Model on a Variety of Industrial and Geographic Situations.

Even though the framework has been tested throughout the chosen industrial corridors throughout the United States, its generalizability might be limited to the geographical variations in industrial makeup, atmospheric processes, and land-use structures. The chemical behaviors and dispersion mechanisms are different in climates, regulatory situations, and clusters of facilities which may restrict the transferability of the model unless recalibrated.

It should be considered in future work to assess the cross-regional performance of GeoAI RiskLab, create region learning family models, and provide meta-learning methods that enable models to readjust to new geographic settings with minimum retraining.

8.3 Model Transparency, Computer Science, and Scalability.

Convolutional neural networks, graph neural networks, and uncertainty quantification methods consume high computational resources, especially when they are used to model at a high spatial resolution. This can be problematic to local agencies that have low technical infrastructure. Moreover, it improves the interpretability of deep learning, but the operations of the explainable approach like SHAP remain complicated in relation to traditional exposure modelling algorithms.

The next steps of model compression, effective architecture design, cloud-based implementation, and hybrid modelling by integrating mechanistic exposure models with GeoAI to enhance transparency, minimize computational load, and enhance reproducibility should be researched in the future.

8.4 Uncertainty Quantification, Depth of validation and Model verification.

The framework combines the uncertainty estimation by the use of Monte Carlo dropout and spatial sensitivity analysis but uncertainty cannot be completely addressed because of some intrinsic restrictions in the emissions inventories, sensor precision, and population data. Also, overlaying on existing monitoring networks is not always a full ground truth since most industrial corridors have no high-dense coverage of the monitors.

Probabilistic modelling, further enhancements of multi-source uncertainty propagation, synthetic monitoring benchmarking, and verification-and-validation protocols in line with environmental software engineering standards should all be developed by future research. This will enhance the validity and soundness of outputs to be used in regulatory and policy.

8.5 Policy Integration and Decision-Support Adoption Implementation Barriers.

Although GeoAI RiskLab is created as a decision-support toolkit, its practical implementation has a number of institutional barriers. The Agencies might face challenges in ensuring that the GeoAI systems can be integrated into the current working procedures, interpretation of output with uncertainty, or matching the model output with the regulating threshold. Also, the decision-makers might need custom visualizations, domain training and integration with existing GIS platforms.

Future studies need to concentrate on co-design processes using agencies, creation of policy-consistent dashboards, integrating regulatory ones, and joint assessment, to make sure that modelling products are practical. The partnerships will also be necessary to embed the framework within the programs of environmental justice and resilience strategies that are long-term.

In sum, these limitations must be taken into account to make progressive progress in transparent, reproducible, and policy-relevant modelling practices. The described challenges, especially the ones associated with the quality of data, its transferability, computational requirements, and validation, will make the exposure assessment frameworks based on GeoAI more mature. Further research and interdisciplinary partnership will make possible subsequent versions of GeoAI RiskLab to facilitate more just, accurate and robust environmental regulation within varied industrial circles.

9. CONCLUSION

This paper presented GeoAI RiskLab, a reproducible modelling and software system that aims at advancing chemical exposure modelling, environmental justice, and decision-support in the industrial corridors in the United States. Combining facility-scale emissions, satellite-based and ground-based surveillance, sociodemographic factors and spatial context in a single GeoAI structure, the framework offers high-resolution approximations of cumulative exposure risk and susceptibility of the population to health. Spatial deep learning, quantification of uncertainty, and explainable AI are in line with the methodological priorities of Environmental Modelling and Software, and they aid in transparency and reproducibility as well as sound model assessment.

The examples used in this paper show how GeoAI RiskLab is able to identify exposure hotspots, create tract-level decision-ready indicators, and facilitate policy activities, including regulatory targeting, environmental justice screen, and urban resiliency planning. Notably, the modular system allows the user to combine local monitoring datasets with mobile sampling and community-level sensors and increase the applicability of the toolkit to a variety of industrial and geographic settings.

The results underscore the promise of GeoAI based decision support systems to enhance the effectiveness of environmental governance through the delivery of empirically based, uncertainty-based insights that go beyond the conventional screening techniques. Nevertheless, to achieve the potential, the further development of data harmonization techniques, cross-regional validation, interpretation of models, and team-based interactions with regulatory authorities and local stakeholders will be necessary.

On the whole, GeoAI RiskLab provides an interchangeable, policy-consistent system that improves the ability of the researcher, environmental manager, and the practitioners of the public health to comprehend, forecast, and tackle the burden of chemical exposures. With the increasing demands of the environment and equity taking center stage in national policy priorities, the development of better, data-informed such integrated tools will be necessary to help maintain healthier, more robust urban-industrial systems.

REFERENCES

1. Iyer, H. S., Karasaki, S., Yi, L., Hswen, Y., James, P., & VoPham, T. (2025). Harnessing Geospatial Artificial Intelligence (GeoAI) for Environmental Epidemiology: A Narrative Review. *Current environmental health reports*, 12(1), 34.
2. Zhang, R., Hu, M., Ye, G., Xu, C., & Wang, J. (2025). GeoAI-based 3D spatial distribution modeling of PAHs in industrial contaminated soils. *Environmental Pollution*, 127440.
3. Dixon, B. (2025). GeoAI, and Geospatial Framework for Integrated Place-Based Approach. In *Interdisciplinary Environmental Solutions: Using Geospatial Technologies for Bridging Disciplines, Scale and Data* (pp. 445-486). Cham: Springer Nature Switzerland.
4. Tian, W., Cao, K., & Feng, Q. (2025). The Impact of Environmental Exposure on the Health of Older Adults: A Review From the GIScience Perspective. *Transactions in GIS*, 29(3), e70056.
5. Ahmad, M., Zahid, R., & Khan, K. M. (2024). From Data to Insights: Harnessing GeoAI for Informed Decision-Making in Healthcare. In *Generative AI and Implications for Ethics, Security, and Data Management* (pp. 350-380). IGI Global.

6. Gonzales-Inca, C., Calle, M., Croghan, D., Torabi Haghghi, A., Marttila, H., Silander, J., & Alho, P. (2022). Geospatial artificial intelligence (GeoAI) in the integrated hydrological and fluvial systems modeling: Review of current applications and trends. *Water*, *14*(14), 2211.
7. Wang, S., Xie, X., Li, Y., Guo, D., Cai, Z., Liu, Y., ... & Meng, X. (2024). Research on the Spatial Data Intelligent Foundation Model. *arXiv preprint arXiv:2405.19730*.
8. Das, S., Chatterjee, A., Manna, J., Das Mandal, S. K., Maiti, T. K., & Bhattacharyya, T. K. (2025). Systems-Integrated Thermostable Vaccine Delivery: Converging Cold-Chain-Free Design, AI-Augmented Formulation, and Climate-Resilient Infrastructure. *Molecular Pharmaceutics*.
9. Kumar, D., & Bassill, N. P. (2025). Artificial Intelligence and Machine Learning for Climate Disaster Management. In *Artificial Intelligence and Machine Learning for Climate Disaster Management* (pp. 1-41). Singapore: Springer Nature Singapore.
10. Barraz, Z., Sebari, I., Ait El Kadi, K., & Ait Abdelmoula, I. (2025). Towards a Holistic Approach for UAV-Based Large-Scale Photovoltaic Inspection: A Review on Deep Learning and Image Processing Techniques. *Technologies*, *13*(3), 117.
11. Sarker, D., Ding, W., Schneider, C., & Hampel, U. (2019). Publications Repository-Helmholtz-Zentrum Dresden- Rossendorf. *International Journal of Heat and Mass Transfer*, *142*, 118481.
12. Metzkes-Ng, J., Stefanikova, R., Prencipe, I., Cowan, T., Schlenvoigt, H. P., Schramm, U., ... & Zeil, K. (2021). Publications Repository-Helmholtz-Zentrum Dresden-Rossendorf. *Scientific Reports*, *11*, 7338.
13. Metzkes-Ng, J., Stefanikova, R., Prencipe, I., Cowan, T., Schlenvoigt, H. P., Schramm, U., ... & Zeil, K. (2021). Publications Repository-Helmholtz-Zentrum Dresden-Rossendorf. *Scientific Reports*, *11*, 7338.