www.editada.org

International Journal of Combinatorial Optimization Problems and Informatics, 16(4), Sep-Dec 2025, 431-440. ISSN: 2007-1558. https://doi.org/10.61467/2007.1558.2025.v16i4.948

An exploratory analysis of hazardous substances registered in the Mexican Pollutant Releases and Transfers Register in the State of Tamaulipas

Hugo G. Reyes-Anastacio[1], Jaqueline Calderon[1], María E. Bautista-Vargas[2], Santiago Gómez-Carpizo[2]
[1] CIACYT-Faculty of Medicine, Universidad Autónoma de San Luis Potosí, México.
[2]Universidad Politécnica de Altamira, México
hugogreyesa@gmail.com, calderoj@uaslp.mx, esther.bautista@upalt.edu.mx, santiago.gomez@upalt.edu.mx

Abstract. The North American Free Trade Agreement required Mexico, the United States and Canada to measure pollutants released and transferred to the environment. In Mexico, the Ministry of Environment and Natural Resources publishes an annual report of pollutants released or transferred by public and private facilities across industrial sectors. Each facility must report substances listed in Official Mexican Standard NOM-165-SEMARNAT-2013. In this study, we used the SINAT tool to obtain records of pollutant releases by municipalities in the state of Tamaulipas. We preprocessed and analysed the substances released in Tamaulipas and mapped them to their corresponding groups based on monographs from the International Agency for Research on Cancer (IARC). These groupings were then used to identify toxic substances—such as benzene, arsenic and asbestos—classified as carcinogenic to humans (Group 1) in Tamaulipas.
Keywords: Data fusion, PRTR, IARC Groups, Tamaulipas dataset

Article Info
Received February 25, 2025
Accepted July 6, 2025

1 Introduction

Mexico measured the pollutant releases and transfers of toxic substances to the environment as part of the North American Free Trade Agreement (Jacott, Marisa et al., 2004). The pollutant-released and transferred substances listed in the Official Mexican Norm Standard NOM-165-Semarnat-2013 (The Official Daily of the Federation, 2014) made by facilities in Mexico were collected and published annually by the Ministry of Environment and Natural Resources (Semarnat, by its Spanish acronym) generating the Pollutant Release and Transfer Registry (MX-PRTR) for the 2004 to 2023 period. The MX-PRTR is public and can be consulted using the online web tool called Sinat (Semarnat, 2025).

Tamaulipas comprises forty-three municipalities (INEGI, 2020), and its municipalities have several facilities of diverse industrial sectors. These facilities reported their pollutant releases and transfer substances annually to the MX-PRTR. The Sinat allows multiple queries: i) For a complete state (i.e., Tamaulipas) in a specific year (between 2004 and 2023). The obtained results summarize the number of pollutants released and transferred for each substance and do not show the breakdown of the facilities that made the pollutant releases to the environment per municipality. ii) For a specific municipality in a specific year: list all the municipalities with reports for a selected state and summarize the results per facility. We used the search option based on municipality with custom parameters to obtain the data for all the municipalities within the Tamaulipas state for the 2004 to 2023 period.

The data obtained from the MX-PRTR did not contain additional information to allow users to identify dangerous substances, assuming that the readers can identify toxic substances by their name or even their unique number assigned by the Chemical Abstract Service (CAS Number). We searched online for some classification indicators for the MX-PRTR substances and found the ``List of Classifications; IARC Monographs on the Identification of Carcinogenic Hazards to Humans" presented by the International Agency for Research on Cancer (IARC). This list contains four groups that can classify substances: *Group 1*: carcinogenic to humans; *Group 2A:* probably carcinogenic to humans; *Group 2B*: possibly carcinogenic to humans; and *Group 3*: not classifiable for its carcinogenicity to humans.

We define the Tamaulipas state as the study area because multiple facilities from distinct industrial sectors are located and distributed across its north and south borders. In this work, we describe a method for acquiring and preparing the Tamaulipas content of the MX-PRTR data. Additionally, we describe the required process to incorporate the IARC group cancer classification label to each substance in the Tamaulipas data. Finally, we conducted an exploratory analysis of the pollutant releases and transfers registered in Tamaulipas, Mexico.

2 Related works

In the state of the art, we found some analyses of released or transferred substances in Mexico (Briseño-Bugarin et al., 2024; Razo et al., 2004; Osuna-Martínez et al., 2021) for specific zones (Orta-García et al., 2016), substances (Bhattacharya et al., 2007), or related to some cancer (Castrezana Campos, 2017). Additionally, we found that were international studies for similar datasets like the Canadian Pollutant Release and Transfer Inventory (Berthiaume, 2024; Wang et al., 2006; Taylor et al., 2020), the USA Toxic Release Inventory (Hendryx & Fedorko, 2011; Hendryx et al, 2014) and the European Release and transfer registry (E-PRTR) (Belen et al., 2014).

Other work related to the data collection method refers to the analysis of scientific documents (Dueñas et al, 2024), the use of databases acquired by equipment and instruments used at site (Hajiparvaneh et al, 2024; Roberto, J., Lee, W. Y., & Campos-Díaz, S. I., 2009), and the use of historical data (Yáñez, L., et.al., 2002; Kusters et al 2022). In the investigation, we do not collect environmental data but rather use the annual database collected by SEMARNAT.

3 Methodology

In this paper, we proposed a two-stage methodology to acquire and analyze a dataset of pollutant releases and transfers of substances listed in the NOM-165-Semarnat-2013 for the state of Tamaulipas for the 2004 to 2023 period, which includes the corresponding cancer classification group assigned by the IARC to each substance. The two stages are: *a)* data acquisition, and *b)* data preparation and fusion, which are described as follows.

3.1 Data acquisition

In the first stage, we downloaded the list of agents (substances) with their corresponding cancer classification group assigned by the IARC (*IARC* dataset) from the IARC webpage (IARC, 2025), and the Pollutant Releases and Transfers Registry (Semarnat, 2025) for the study area from the Sinat tool of the Semarnat.

We followed the process described in Algorithm 1 to download the content for all the Tamaulipas municipalities from the MX-PRTR by using the Sinat tool. To improve the queries, we searched for the municipalities' codes (*cve_mun*) assigned by the National Institute of Statistics and Geography (INEGI, by its Spanish acronym) in the study area. The INEGI is an autonomous public agency responsible for regulating and coordinating the National System of Statistical and Geographical Information, as well as for collecting and disseminating information about Mexico in terms of territory, resources, population, and economy. As a result, we found forty-three different values for the *cve_mun* set for the Tamaulipas state. As can be seen, we made 840 requests to cover all the combinations of year and municipality values (see line 9 in Algorithm 1) and using the state code (*cve_ent*) for Tamaulipas State (see line 1 in Algorithm 1) to the Sinat system. The period range can be updated, increasing the limit by one, to include 2024 when the SEMARNAT publishes the data. Algorithm 1 is generic and can be used to obtain further data, only if the data preserves the same structure. If Semarnat modifies the data structure, the user must update the script to allow data to be downloaded. However, the changes will affect the new data.

Suppose researchers are interested in obtaining the records for other Mexican states. In that case, they must update the value of *cve_ent* in line 1 and the list of municipality codes in line 2 of Algorithm 1. For example, for the Baja California Sur state, the cve_*ent* value is *03* and the *mun_list* values are [*001*,*002*, …, *007*]. The resultant dataset will be the releases and transfers reported in Baja California. This new data could be added to another dataset to increase the study area. The resultant dataset obtained and concatenated was called the Tamaulipas dataset. Table 1 describes the Tamaulipas dataset's columns.

Algorithm 1 Data acquisition process

```
cve_ent = 28
mun_list = ["001", "002", ..., "043"]
Tamaulipas_dataset = DataFrame()
for year ∈ range(2004, 2024) do
    for cve_mun ∈ mun_list do
        current_mun = "28" + cve_mun
        base_url = 'http://sinat.semarnat.gob.mx/retc/retc/consulta.php?enfe='
        url =base_url+cve_ent+'&muni='+current_mun+'&anio='+year+'&tipb=0'
        data = getData(url)
        data["anio"] = year
        data["cve_mun"] = cve_mun
        Tamaulipas_dataset.append(data)
    end for
end for
Tamaulipas_dataset.to_csv("tamaulipas_dataset.csv")
```

Table 1. Tamaulipas dataset description

Group name in Spanish	The field's name in Spanish	Group name	Fields name	Description
Datos Generales	NRA	General data	Nra	Environmental registration number (NRA, by its Spanish acronym)
	Nombre		Name	Facility name
	Estado		state	State name
	Sector		sector	Industrial sector name
Sustancias	No. CAS	Substances	cas	Refers to the Chemical Information Service (CAS) Number assigned to a substance
	Descripción		substance	Name of the substance emitted or transferred
	Unidad		unit	Measure unit
Emisiones	Agua	Releases	water	Releases of a substance to water
	Aire		air	Releases of a substance to the air
	Suelo		soil	Releases of a substance to soil
Transferencias	Reuso	Transfers	reuse	Transfers for using previously used material or waste without a transformation process
	Reciclado		recycle	Transfers to an external deposit or permanently confine waste in sites and facilities whose characteristics prevent its release into the environment and the consequent effects on the population's health and the ecosystems and their elements.
	Coprocesamiento		coprocessing	Transfers for the environmentally safe integration of waste generated by an industry or known source as an input to another production process
	Tratamiento		treatment	To carry out physical, chemical, biological, or thermal procedures, through which the characteristics of the waste are changed, and its volume or danger is reduced.
	Disposición final		finaldisposal	Substance transfer to final disposal
	Alcantarillado		sewerage	Transfer to discharge, infiltrate, deposit, or inject wastewater into a receiving body or sewer.
	Otros		others	Transfers labeled as others
	Incineración		incineration	Transfer to reduce the volume and decompose or change the physical, chemical, or biological composition of solid, liquid, or gaseous waste through thermal oxidation, in which all combustion factors, such as Temperature, retention time, and turbulence, are involved.

3.2 Data preparation and fusion

We observed three different values in the unit column of the Tamaulipas dataset (kg/año, ton/año, and g/año, which means kg/year, ton/year, and g/year). To homogenize these values, we transformed all the values from tons or grams to kilograms. We transformed the values of the released and transferred substances into the Tamaulipas dataset based on the column unit following the process described in Algorithm 2. In this stage, we read the Tamaulipas dataset and iterate over all the rows, looking for the *ton/año* and *g/año* values in the unit column. We found 4,982 records with the *ton/año* unit and ten with the *g/año* unit. Based on the unit column value, we applied the required conversion to *kg/año* for all the release and transfer values (columns previously described in
Table 1). As a result, we obtained and stored an updated version of the Tamaulipas dataset with the homogenized measure unit values.

As our final preprocessing stage, we searched for coincidences between the *IARC* and Tamaulipas datasets by using the common fields (CAS and CAS No.) to incorporate the group value of the *group* field from the *IARC* to the Tamaulipas dataset, following the process described in Algorithm 3. The IARC considers the following four groups: *Group 1 Carcinogenic to humans; Group 2A Probably carcinogenic to humans; Group 2B Possibly carcinogenic to humans; and Group 3 Not classifiable as to its carcinogenicity to humans.* We added the *Not considered* group for the substances that appear in the Tamaulipas dataset but are not listed in the IARC monographs.

3.3 Prototype description

To produce the Tamaulipas dataset, we developed a set of scripts. We used Python to generate the scripts used for data acquisition, complemented by the Pandas library to obtain the data per year and municipality. We used *pandas* and *NumPy* libraries for unit conversion and data fusion processes. We developed scripts using the matplotlib and sunburst libraries for the visualization process.

Algorithm 2 Unit homogenization process

```
Tamaulipas_dataset = read_csv("tamaulipas_dataset.csv")
for row ∈ Tamaulipas_dataset do
    if row["unit"] == "ton/año" then
        row["unit"] = "kg/año"
        Tamaulipas_dataset[row] = convert_ton_to_kg(row)
    else
        if row["unit"] == "g/año" then
            row["unit"] = "kg/año"
            Tamaulipas_dataset[row] = convert_g_to_kg(row)
        end if
    end if
end for
Tamaulipas_dataset.to_csv("tamaulipas_dataset_v1.csv")
```

Algorithm 3 IARC group labeling process

```
Tamps_data = read_csv("tamaulipas_dataset_v1.csv")          ▷ Tamaulipas dataset
iarc_dataset = read_csv("iarc_vol_1_to_137.csv")
Tamps_data["iarc_group"] = "Not considered"          ▷ Add a default value to all the records
Tamps_cas_list = Tamps_data["cas"].unique()
for cas ∈ Tamps_cas_list do
    iarc_subs = iarc_dataset.loc[iarc_dataset["CAS No."] == cas]
    if size(iarc_subs) > 0 then          ▷ If the CAS Number appears in both datasets
        Tamps_data.loc[Tamps_data["iarc_group"], "iarc_group"] = iarc_subs["group"]
    end if
end for
Tamps_data.to_csv("tamaulipas_dataset_v2.csv")
```

The obtained Tamaulipas dataset was uploaded to the Zenodo storage service to make it available. It can be consulted by using the following link: https://doi.org/10.5281/zenodo.15588115. We uploaded two versions of the dataset, one for the column names in Spanish and another for the column names in English (see Table 1).

4 Exploratory analysis results

The Tamaulipas dataset contains substances released and transferred to the environment. Each record represents a pollutant release or transfer of a substance reported by a specific facility (in kilograms) for a specific year. As we mentioned previously, we acquired data for the 2004 to 2023 period. This dataset comprises 14,026 records of pollutant releases and transfers made by twenty-one industrial sectors and distributed across nineteen of the forty-three municipalities of Tamaulipas.

Table 2 describes the list of municipalities with pollutant releases. As you can see from the column Records, the Reynosa, Altamira, Matamoros, Nuevo Laredo, and Ciudad Madero records represent 89% of the Tamaulipas dataset's records. The other fourteen municipalities only represent 11% of the records.

Table 2. List of municipalities of Tamaulipas and the number of records reported to the Tamaulipas dataset

CVE_MUN	Name	Records	CVE_MUN	Name	Records
003	Altamira	3251	025	Miguel Alemán	1
004	Antiguo Morelos	3	027	Nuevo Laredo	1274
007	Camargo	48	032	Reynosa	4850
009	Ciudad Madero	499	033	Río Bravo	183
016	Hidalgo	27	035	San Fernando	31
017	Jaumave	32	037	Soto la Marina	7
019	Llera	11	038	Tampico	98
021	El Mante	138	040	Valle Hermoso	482
022	Matamoros	2719	041	Victoria	300
			043	Xicoténcatl	72

We grouped the data by the *Sector* column values to identify the names of the industrial sectors that registered the pollutant releases and transfers. As a result, we found that twenty-one industrial sectors have records in the study area. Table 3 describes the sectors' names in Spanish and English and includes a label to identify the twenty-one sectors.

Table 3. Industrial sector names

Label	In Spanish	In English	Label	In Spanish	In English
A	Alimenticio y/o de consumo humano	Food and human consumption	L	Madera y Productos	Wood-based products
B	Artículos y productos compuestos de diferentes materiales	Articles and products composed of varied materials	M	Metalúrgica (incluye la siderúrgica)	Metallurgy (including steel)
C	Artículos y productos metálicos	Metal articles and products	N	Otros	Others
D	Artículos y productos plásticos	Plastic articles and products	O	Petróleo y petroquímica	Petroleum and petrochemicals
E	Asbesto	Asbestos	P	Pinturas y tintas	Paints and inks
F	Automotriz	Automotive	Q	Química	Chemistry
G	Bebidas y tabaco	Beverages and tobacco	R	Servicios de salud y de asistencia	Health and care services
H	Celulosa y papel	Pulp and paper	S	Textiles, fibras e hilos	Textiles, fibers, and threads
I	Cemento y Cal	Cement and Lime	T	Tratamiento de residuos peligrosos	Hazardous waste treatment
J	Equipos y artículos electrónicos, eléctricos y domésticos	Electronic, electrical, and household equipment and items	U	Vidrio	Glass
K	Generación de energía eléctrica	Electrical Power Generation			

Fig. 1 presents the number of records made by the industrial sector (we used the label value instead of the complete sector names) for each year from 2004 to 2023. As can be seen, the sectors that reported more records were Chemistry (Q), electronic, electrical, and household equipment and items (J), automotive (F), petroleum and petrochemicals (O), Electrical Power Generation (K), and Metal articles and products (C).

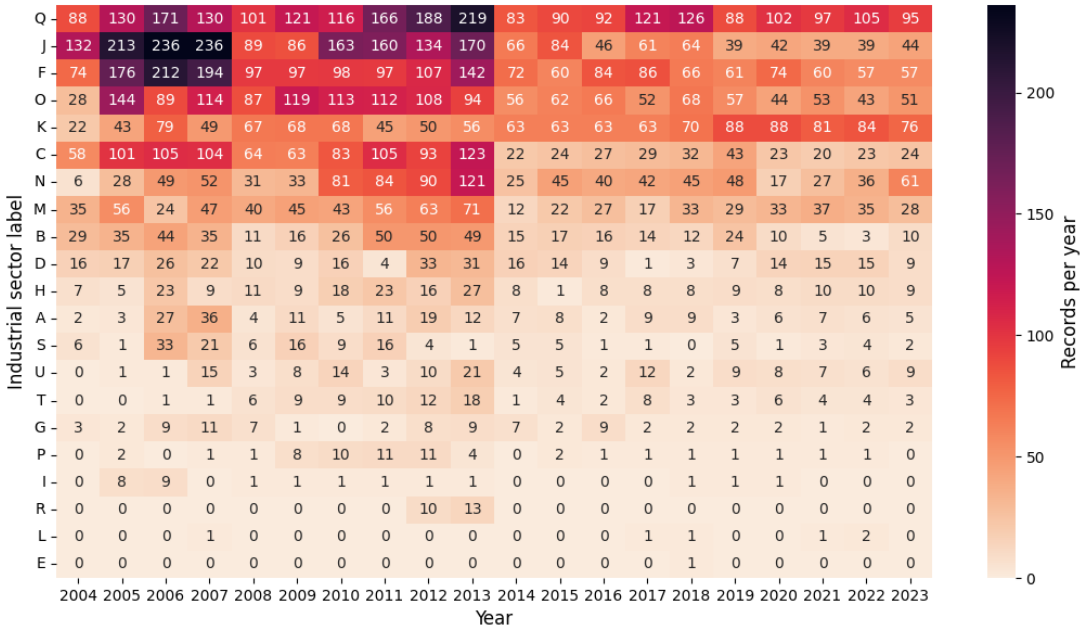

Fig. 1. Records made by the industrial sector for the 2004 to 2023 period.

We analyzed the Tamaulipas dataset and found seventy-four different registered substances; the list of substances is described in Table 4. The CAS Number can identify these substances. However, the MX-PRTR incorporates additional values into the column *cas*. These values represent the chemical symbol of the substance, or a categorical value assigned by the Semarnat, and were used to describe compounds and soluble compounds. We preserved these values in the Tamaulipas dataset to maintain consistency with the MX-PRTR.

As mentioned, we incorporated the IARC group label into the Tamaulipas dataset substances. During this process, we set the IARC group of the substances to their compounds. We grouped the substances using the IARC group column values to summarize the number of labeled substances per group. Fig. 2 shows the percentage of substances for each IARC group released into the study area. As can be seen, 27.03% of the substances are not listed in the IARC monographs. However, 56.78% of the substances were classified with the IARC 1, 2A, and 2B groups. These substances' pollutant releases and transfers, especially those in the IARC group 1, require a deeper analysis to study their effects on health in the study area.

As shown in Fig. 2, by analyzing the Tamaulipas dataset, we found that seventeen substances were classified as IARC *group 1 carcinogenic for humans*, six substances classified as IARC *group 2A probably carcinogenic to humans,* nineteen substances classified as IARC *group 2B possibly carcinogenic to humans,* twelve substances classified as IARC *group 3 not classifiable as to its carcinogenicity to humans* and finally, we classified the last twenty substances as *Not considered*. These labels allow users to find substances related to public health problems in their municipalities. Local governments can use the geospatial data in the Tamaulipas dataset to improve their environmental policies according to the IARC group of substances detected in their territories.

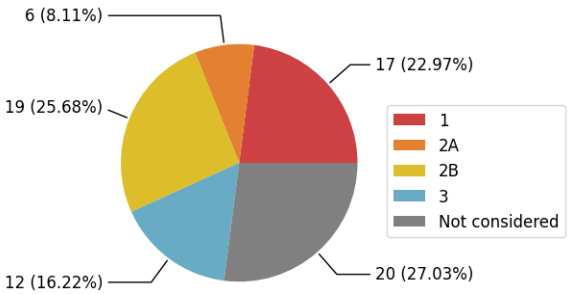

Fig. 2. Percentage of substances by IARC group registered in the Tamaulipas dataset.

Table 4. List of substances released or transferred in Tamaulipas registered in the MX-PRTR

CAS Number	Substance name	CAS Number	Substance name
50-00-0	Formaldehyde	107-13-1	Acrylonitrile
1336-36-3	Polychlorinated biphenyls	S/C6	Lead compounds
7440-38-2	Arsenic respirable dust, fumes and vapors	Ni	Nickel soluble compounds
50-32-8	Benzo[a]pyrene	79-34-5	1,1,2,2-Tetrachloroethane
75-01-4	Vinyl chloride	67-66-3	Chloroform
71-43-2	Benzene	75-45-6	Chlorodifluoromethane
1332-21-4	Asbestos (all forms, including actinolite, amosite, anthophyllite, chrysotile, crocidolite, tremolite)	108-95-2	Phenol
79-01-6	Trichloroethylene	101-68-8	4,4'-Methylenediphenyl diisocyanate
As	Arsenic soluble compounds	7439-97-6	Mercury respirable dust, fumes and vapors
S/C1	Arsenic compounds	S/C4	Mercury compounds
S/C2	Cadmium compounds	9016-87-9	Polymethylene polyphenyl isocyanate
Cd	Cadmium soluble compounds	Hg	Mercury soluble compounds
542-88-1	Bis(chloromethyl)ether; chloromethyl methyl ether (technical-grade)	108-88-3	Toluene
7440-43-9	Cadmium respirable dust, fumes and vapors	7440-47-3	Chromium respirable dust, fumes and vapors
106-99-0	1,3-Butadiene	1330-20-7	Xylenes
75-07-0	Acetaldehyde is associated with the consumption of alcoholic beverages	S/C3	Chromium compounds
87-86-5	Pentachlorophenol (see also Polychlorophenols)	Cr	Chromium soluble compounds
100-42-5	Styrene	2551-62-4	Sulfur hexafluoride
71-55-6	1,1,1-Trichloroethane	10024-97-2	Nitrous oxide
107-02-8	Acrolein	124-38-9	carbon dioxide
302-01-2	Hydrazine	306-83-2	2,2-Dichloro-1,1,1-trifluoroethane
62-53-3	Aniline (see also Aniline hydrochloride)	7758-98-7	Copper(II) sulfate pentahydrate
75-09-2	Dichloromethane (Methylene chloride)	HFCs	Hydrofluorocarbons (HFCs)
207-08-9	Benzo[k]fluoranthene	74-82-8	Methane
107-06-2	1,2-Dichloroethane	10102-44-0	Nitrogen dioxide
106-46-7	Para-Dichlorobenzene	10049-04-4	chlorine dioxide
193-39-5	Indeno[1,2,3-cd]pyrene	1717-00-6	Dichlorofluoroethane
Pb	Lead soluble compounds	PCDDs	Polychlorinated dibenzodioxins (PCDDs)
110-86-1	Pyridine	S/C8	Hydrofluorocarbons (HFCs)
26471-62-5	Toluene diisocyanates	S/C10	Dioxins
7440-02-0	Nickel, metallic	7783-06-4	Hydrogen sulfide
7439-92-1	Lead respirable dust, fumes and vapors	92-52-4	Biphenyl
S/C5	Nickel compounds	PCDFs	Polychlorinated dibenzofurans (PCDFs)
118-74-1	Hexachlorobenzene	57-12-5	Cyanide anion
205-99-2	Benzo[b]fluoranthene	105-39-5	Ethyl Chloroacetate-13C2
56-55-3	Benz[a]anthracene	S/C11	Furans
108-05-4	Vinyl acetate	58-90-2	2,3,4,6-Tetrachlorophenol

Table 5 presents the list of substances classified as IARC *group 1* with a visual indicator for the municipalities' codes that report them (the municipalities' names were described previously in Table 2). To determine which IARC group 1 substances reported more releases, we sum the releases and transfers per type for the 2004 to 2023 period and sort from the highest to the lowest. Fig. 3 presents the breakdown per type of release and transfers for the top five substances with IARC group 1. As can be seen, the pollutant releases were made to the *Air* (Formaldehyde, Benzene, Vinyl Chloride, and 1,3 Butadiene) and, in some cases, to the soil. The Asbestos substance only registered *final disposal* transfers.

The Formaldehyde, the most released substance, is classified as group 1 by the IARC group 1 into the Tamaulipas dataset. As mentioned, we added the municipality code (CVE_MUN) and the state code (CVE_ENT) to the Tamaulipas dataset. With this data, we can summarize the releases to air registered by the Tamaulipas municipalities. Fig. 4 presents a map of the sum of pollutants released into the air of Formaldehyde per municipality for the 2004 to 2023 period. As can be seen, the south and north municipalities reported more releases of Formaldehyde. The five municipalities with the most pollutant releases to air of Formaldehyde are: Altamira (727,994.66 Kg), Valle Hermoso (405,776.7 Kg), Ciudad Madero (64695.93 Kg), Nuevo Laredo (57,675.26 Kg), and Reynosa (7,566.81 Kg).

Table 5. Substances classified as IARC group 1 in the Tamaulipas dataset and the municipalities code (CVE_MUN) that reports them

CAS Number	Substance name	IARC Group	CVE_MUN																		
			003	004	007	009	016	017	019	021	022	025	027	032	033	035	037	038	040	041	043
79-01-6	Trichloroethylene	1	●								●		●								
1336-36-3	Polychlorinated biphenyls	1	●											●							
542-88-1	Bis(chloromethyl)ether; chloromethyl methyl ether (technical-grade)	1									●										
Cd	Cadmium soluble compounds	1	●			●				●	●		●	●	●			●		●	●
S/C1	Arsenic compounds	1	●		●	●				●	●		●	●	●	●			●	●	●
87-86-5	Pentachlorophenol (see also Polychlorophenols)	1									●										
As	Arsenic soluble compounds	1	●			●				●	●		●	●	●			●			●
S/C2	Cadmium compounds	1	●		●	●				●	●		●	●	●	●			●	●	●
71-43-2	Benzene	1	●			●			●	●	●		●	●						●	
75-01-4	Vinyl chloride	1	●																		
106-99-0	1,3-Butadiene	1	●																		
7440-43-9	Cadmium respirable dust, fumes and vapors	1	●		●	●	●	●		●	●		●	●	●	●		●	●	●	●
1332-21-4	Asbestos	1	●								●		●	●							
50-32-8	Benzo[a]pyrene	1	●																		
50-00-0	Formaldehyde	1	●			●			●	●	●		●	●					●	●	
75-07-0	Acetaldehyde is associated with the consumption of alcoholic beverages	1	●			●			●	●			●	●					●	●	
7440-38-2	Arsenic respirable dust, fumes and vapors	1	●		●	●	●	●	●	●	●		●	●	●	●		●	●	●	●

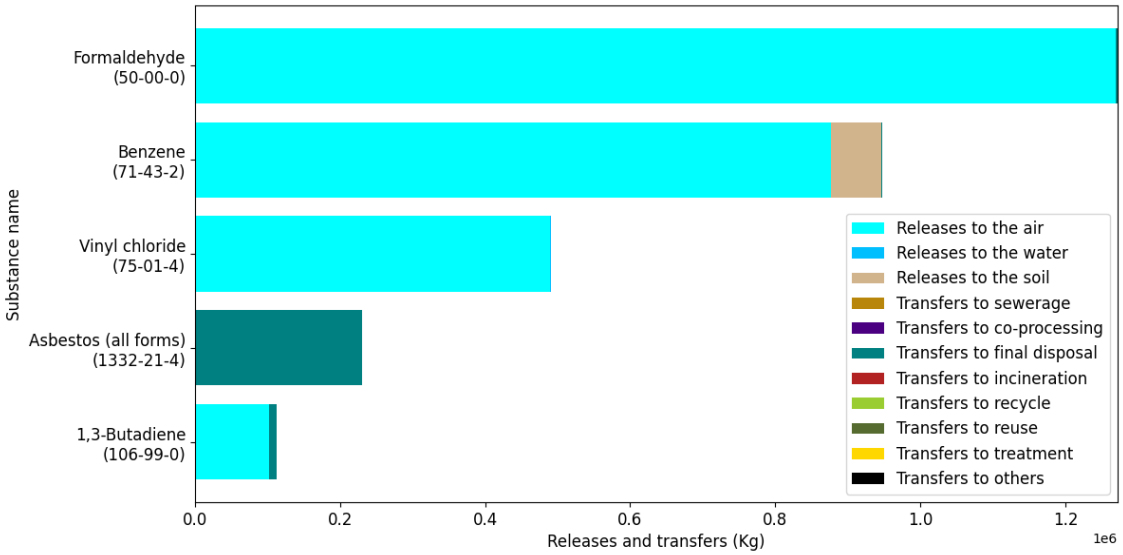

Fig. 3. Breakdown for type of pollutant releases and transfers for the top five substances labeled as IARC group 1 in the Tamaulipas dataset.

5 Conclusions

In this work, we proposed a methodology to obtain the Tamaulipas dataset that includes the pollutant substances released and transferred in the Tamaulipas state from the MX-PRTR published by the Semarnat for the 2004 to 2023 period, enriched with the IARC cancer classification group label to each substance found according to its CAS Number. We found seventy-four substances in the study area and classified these substances according to their IARC cancer classification group.

As a result, we classified: seventeen substances as *group 1*, six as *group 2A*, nineteen as *group 2B*, and twelve as *group 3*. The last twenty substances were labeled as *not considered* because they only appear in the Tamaulipas dataset but not in the IARC monographs.

We explored the Tamaulipas dataset to prove that the obtained dataset can be used to identify the pollutant releases for distinct levels. We showed an analysis based on the industrial sector, IARC group, and municipalities for the complete period. Our results show that the substance Formaldehyde (classified as IARC 1) is the highest substance released (specific to air) in the south and north borders of the Tamaulipas state for the period 2004 to 2023. We included the year column in the Tamaulipas dataset to allow a more specific search for a specific year or a determined period.

Researchers could use the obtained dataset for specific analyses or even include complementary information to enrich the data. For example, it can incorporate a specific type of cancer related to substances, like the Mammary Carcinogens Review Database published by the Silent Spring Institute. We are working on the data acquisition and processing of the complete MX-PRTR data for the rest of the thirty-one states in Mexico. However, the proposed method requires additional modifications to generate a national dataset by joining the thirty-two states in the Mexican territory. Additionally, other public datasets can be fused to the obtained Tamaulipas dataset to increase the information and improve the data analyses.

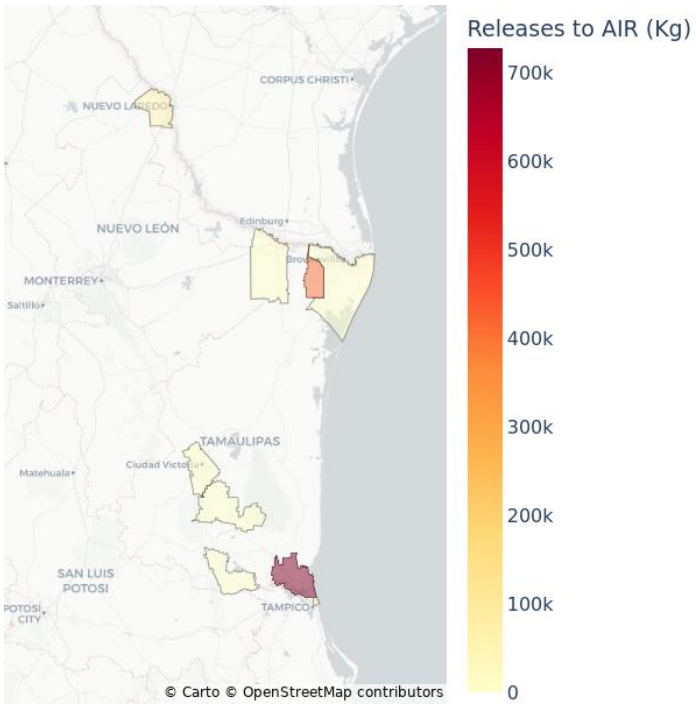

Fig. 4. Releases to air of Formaldehyde reported by the Tamaulipas municipalities for the 2004 to 2023 period.

References

Jacott, M., Reed, C., & Winfield, M. (2004). The generation and management of hazardous wastes and transboundary hazardous waste shipments between Mexico, Canada and the United States Since NAFTA: A 2004 Update. *Austin, TX: Texas Center for Policy Studies*.

de la Federación, D. O. (2014). Norma Oficial Mexicana NOM-165-SEMARNAT-2013, Que establece la lista de sustancias sujetas a reporte para el registro de emisiones y transferencia de contaminantes. *México. Fecha de publicación, 24*. https://biblioteca.semarnat.gob.mx/janium/Documentos/Ciga/agenda/DOFsr/DO3231.pdf

Secretaría de Medio Ambiente y Recursos Naturales (Semarnat). (n.d). *Pollutant Release and Transfer Registry (PRTR)*. http://sinat.semarnat.gob.mx/retc/retc/consulta.php

Instituto Nacional de Estadística y Geografía (INEGI) (2020). *División municipal. Tamaulipas*. https://cuentame.inegi.org.mx/monografias/informacion/tam/territorio/div_municipal.aspx

Briseño-Bugarín, J., Araujo-Padilla, X., Escot-Espinoza, V. M., Cardoso-Ortiz, J., Flores de la Torre, J. A., & López-Luna, A. (2024). Lead (Pb) pollution in soil: a systematic review and meta-analysis of contamination grade and health risk in Mexico. *Environments*, *11*(3), 43. https://doi.org/10.3390/environments11030043

Razo, I., Carrizales, L., Castro, J., Díaz-Barriga, F., & Monroy, M. (2004). Arsenic and heavy metal pollution of soil, water and sediments in a semi-arid climate mining area in Mexico. *Water, Air, and Soil Pollution*, *152*(1), 129-152.

Osuna-Martínez, C. C., Armienta, M. A., Bergés-Tiznado, M. E., & Páez-Osuna, F. (2021). Arsenic in waters, soils, sediments, and biota from Mexico: An environmental review. *Science of The Total Environment*, *752*, 142062. https://doi.org/10.1016/j.scitotenv.2020.142062

Orta-García, S. T., Ochoa-Martínez, A. C., Carrizalez-Yáñez, L., Varela-Silva, J. A., Pérez-Vázquez, F. J., Pruneda-Álvarez, L. G., ... & Pérez-Maldonado, I. N. (2016). Persistent organic pollutants and heavy metal concentrations in soil from the Metropolitan Area of Monterrey, Nuevo Leon, Mexico. *Archives of Environmental Contamination and Toxicology*, *70*, 452-463. https://doi.org/10.1007/s00244-015-0239-3

Bhattacharya, P., Welch, A. H., Stollenwerk, K. G., McLaughlin, M. J., Bundschuh, J., & Panaullah, G. (2007). Arsenic in the environment: biology and chemistry. *Science of the total environment, 379*(2-3),109-120. https://doi.org/10.1016/j.scitotenv.2007.02.037

Wang, S., & Mulligan, C. N. (2006). Occurrence of arsenic contamination in Canada: sources, behavior and distribution. *Science of the total Environment*, *366*(2-3), 701-721. https://doi.org/10.1016/j.scitotenv.2005.09.005

Taylor, S., Edwards, S. J., & Walker, T. R. (2020). A toxicity-based analysis of Canada's National Pollutant Release Inventory (NPRI): a case study in Nova Scotia. *Environmental Science and Pollution Research, 27,* 2238–2247. https://doi.org/10.1007/s11356-019-06933-x

Hendryx, M., & Fedorko, E. (2011). The relationship between toxics release inventory discharges and mortality rates in rural and urban areas of the United States. *The Journal of Rural Health*, *27*(4), 358–366. https://doi.org/10.1111/j.1748-0361.2011.00367.x

Hendryx, M., Luo, J., & Chen, B. C. (2014). Total and cardiovascular mortality rates in relation to discharges from toxics release inventory sites in the United States. *Environmental research*, *133*, 36–41. https://doi.org/10.1016/j.envres.2014.05.010

Berthiaume, A. (2024). Unveiling underutilized public data on Canadian industrial pollutant transfers and disposals. *Journal of the Air & Waste Management Association*, *74*(9), 664-684. https://doi.org/10.1080/10962247.2024.2377327

Beelen, R., Raaschou-Nielsen, O., Stafoggia, M., Andersen, Z. J., Weinmayr, G., Hoffmann, B., ... & Hoek, G. (2014). Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 22 European cohorts within the multicentre ESCAPE project. *The Lancet, 383(9919),* Issue 9919, 785 – 795.

International Agency for Research on Cancer (IARC). (n.d.). List of Classifications; IARC Monographs on the Identification of Carcinogenic Hazards to Humans. https://monographs.iarc.who.int/list-of-classifications.

Dueñas-Moreno, J., Mora, A., Narvaez-Montoya, C., & Mahlknecht, J. (2024). Trace elements and heavy metal(loid)s triggering ecological risks in a heavily polluted river-reservoir system of central Mexico: Probabilistic approaches. *Environmental Research*, Volume 262, Part 2, 119937. https://doi.org/10.1016/j.envres.2024.119937.

Erfan Hajiparvaneh, Hossein Alizadeh, Vahid Hosseini, Charles Robert Koch, Machine Learning Classification of Air Quality Monitoring Stations to Achieve Ambient NO2 Objectives Using Emission Scenarios and Chemical Transport Model, A*tmospheric Environment*, *Volume 331*, 2024, 120582, ISSN 1352-2310. https://doi.org/10.1016/j.atmosenv.2024.120582.

Kusters, M. S., Essers, E., Muetzel, R., Ambrós, A., Tiemeier, H., & Guxens, M. (2022). Air pollution exposure during pregnancy and childhood, cognitive function, and emotional and behavioral problems in adolescents. *Environmental research*, *214*, 113891.

Yáñez, L., Ortiz, D., Calderón, J., Batres, L., Carrizales, L., Mejía, J., ... & Díaz-Barriga, F. (2002). Overview of human health and chemical mixtures: problems facing developing countries. *Environmental Health Perspectives*, *110*(suppl 6), 901-909.

Roberto, J., Lee, W. Y., & Campos-Díaz, S. I. (2009). Soil-borne polycyclic aromatic hydrocarbons in El Paso, Texas: analysis of a potential problem in the United States/Mexico border region. *Journal of Hazardous Materials, 163*(2-3), 946-958. https://doi.org/10.1016/j.jhazmat.2008.07.089.

Reyes-Anastacio, H. G., Gómez-Carpizo, S., Bautista-Vargas, M. E., & Calderon-Hernández, J. (2025). Hazardous substances registered in the Mexican Pollutant Releases and Transfers Register in the State of Tamaulipas (0.1) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.15588116.