UNIVERSITÉ DU LUXEMBOURG

# Environmental Science Basics I: Water, Air and Soil Pollution

*This section provides an overview of key concepts in environmental science and engineering. This is not an exhaustive text; rather, we will cover some of the fundamental concepts that form the foundation of the problems addressed in environmental data analytics.*

## What do we mean by the "environment"?

Environmental data analytics involves collecting, processing, and analyzing data related to the environment to gain insights and inform decision-making.

The environment refers to the natural world on Earth and its various components. It includes the biotic environment, which consists of all living organisms, and the abiotic environment, which encompasses the non-living physical surroundings that support life. The abiotic environment includes components such as:

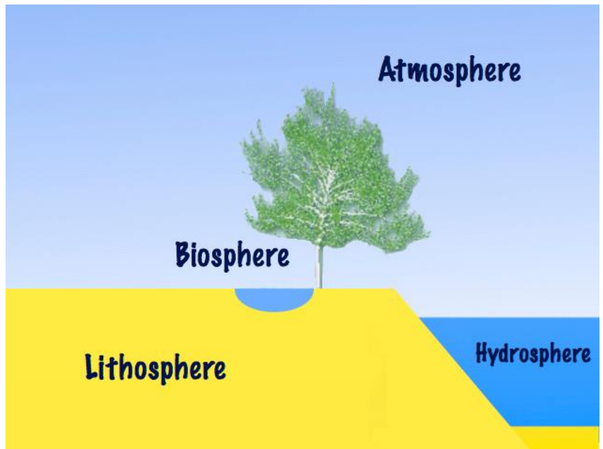
- **Atmosphere:** The layer of gases surrounding the Earth.
- **Hydrosphere:** All water bodies, including oceans, rivers, lakes, and groundwater.
- **Lithosphere:** The solid outer part of the Earth, including rocks and soil.

The **biosphere** is the system that encompasses both the biotic (living organisms) and abiotic (non-living physical and chemical) components of the environment.

The biosphere is called a "system" because it consists of interconnected and interdependent components—both biotic (living organisms) and abiotic (non-living elements)—that interact and function together in a complex, dynamic way to sustain life on Earth.

Environmental data analytics can focus on understanding natural conditions, such as weather patterns and ecosystem dynamics, as well as on environmental issues. Environmental issues generally refer to problems caused by human activities or natural processes that negatively impact the natural environment, including pollution, deforestation, climate change, loss of biodiversity, and resource depletion.

**Note that the boundary between natural conditions and human-induced environmental issues is often blurred**, as changes in natural conditions can exacerbate human-caused environmental problems. Conversely, human-induced issues, can significantly alter natural conditions (such as in global warming), creating a complex and interconnected relationship between the two.




*Atmosphere, Hydrosphere, Lithosphere, and Biosphere.*

Environmental data analytics can answer a wide range of key questions that are crucial for understanding and managing environmental issues. Here are some categories of questions and examples within each category.

## Environmental Pollution

A major environmental issue is **pollution**, which refers to the alteration of the physical, chemical, or biological characteristics of soil, water, or air in a way that negatively impacts ecosystems, human health, and natural processes.

***1) Physical pollution*** refers to harmful or undesirable changes in the physical characteristics of soil, water, or air. This type of pollution can include:

- **Thermal pollution** in water, where industrial processes release heated water into natural water bodies, raising the water temperature and affecting aquatic life.
- **Turbidity in water**, which occurs when suspended particles, such as sediment, increase the cloudiness of water, reducing sunlight penetration and negatively impacting aquatic ecosystems and photosynthesis.
- **Particulate matter (PM)** in the air, such as dust, smoke, or other fine particles, which can reduce air quality, harm human health, and decrease visibility.
- In soil, **physical pollutants** like plastic debris can alter the soil structure, affecting water absorption, nutrient availability, and overall ecosystem health.

*Turbidity in water: an example of physical pollution.*

*Plastic debris in soil: another example of physical pollution.*

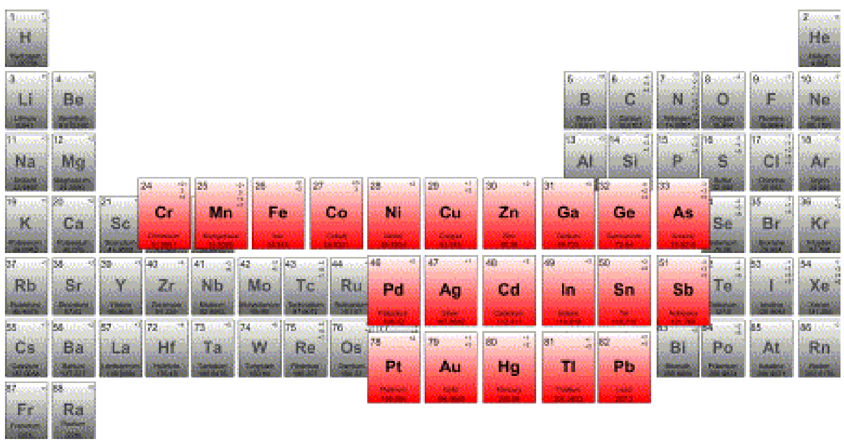
***2) Chemical pollution*** refers to the introduction of harmful or toxic chemicals into the environment, which can alter the chemical balance of soil, water, or air. Examples of chemical pollution include:

- **Heavy metals in water and soil**, such as lead, mercury, and arsenic, which can leach into water sources from industrial waste or mining activities. These metals can causing poison living organisms. Note that heavy metals are a group of metallic elements with relatively high atomic weights and densities (typically greater than 5 g/cm³) that are toxic or harmful at low concentrations.

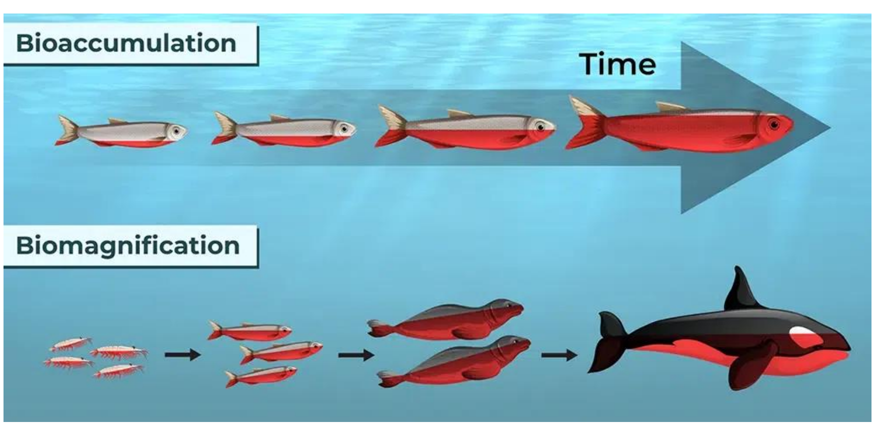
- **Nitrogen oxides (NOx) and sulfur oxides (SOx) in air**, primarily released from vehicle emissions and industrial processes, contribute to air pollution. These gases lead to acid rain, smog formation, and respiratory issues in humans.
- **Pesticides in water and soil**, which are commonly used in agriculture to control pests but can seep into water bodies and soils, contaminating the environment and harming wildlife, aquatic life, and human health through bioaccumulation and biomagnification.

***Bioaccumulation*** *is the process by which toxic substances, such as chemicals or heavy metals, build up in the tissues of organisms over time, often at higher concentrations than in the surrounding environment.*

***Biomagnification*** is the process by which the concentration of toxic substances increases at each successive level of a food chain, leading to higher toxicity in top predators.



*Heavy metals are a key source of pollution in water and soil.*



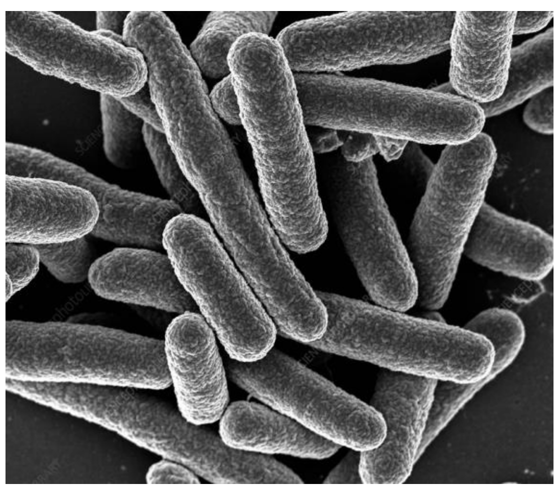

*Bioaccumulation vs. Biomagnification.*

**3) Biological pollution** refers to the introduction of harmful or invasive living organisms into the environment, which can disrupt ecosystems, harm human health, or degrade natural resources. Examples of biological pollution include:

- **Microbial contamination of water**, where harmful bacteria, viruses, or parasites, often from sewage, agricultural runoff, or untreated wastewater, leading to waterborne diseases and public health risks.
- **Invasive species**, which are non-native organisms introduced to an ecosystem, and outcompete native species for resources, disrupt food chains, and cause ecological imbalances.
- **Pathogenic microbes in soil**, which can spread through agricultural practices or contaminated water, affecting plant health, crop yields, and soil fertility.

*A **pathogen** is a microorganism, such as a virus, bacterium, or fungus, that causes disease in its host.*

Pathogens in water are typically quantified using **E. coli** as an indicator. **E. coli** (Escherichia coli) is a type of bacteria commonly found in the intestines of humans and animals, with some strains capable of causing food poisoning or serious infections. It is commonly used as an indicator of water quality because its presence suggests fecal contamination, as E. coli typically enters water through human or animal waste, signaling the potential presence of harmful pathogens and poor sanitation.



*E. coli Bacteria under the microscope.*

## Quantifying Pollution

Pollution in water, air, or soil is quantified either directly by measuring the **concentration** of the pollutant—defined as the amount of a substance present in a specific volume or mass of water, air, or soil (e.g., mg/L for water)—or through **indirect indicators**, which are proxy measurements suggesting the presence of pollutants.

Common units of pollution concentration include:

- **Milligrams per liter (mg/L)** – used for measuring pollutants in water, such as dissolved chemicals, **nutrients**, or heavy metals.
- **Micrograms per liter (µg/L)** – often used for measuring **trace pollutants** in water, like pesticides or trace metals.
- **Parts per million (ppm)** – used for both water and air pollutants, indicating the number of pollutant molecules per million molecules of air or water.
- **Parts per billion (ppb)** – for very low concentrations of pollutants, such as in air or drinking water.
- **Grams per kilogram (g/kg)** is commonly used for measuring pollution concentration in soil. This unit expresses the amount of a pollutant, such as heavy metals or organic contaminants, per kilogram of soil,

An example of an indirect indicator is ***biological oxygen demand (BOD)*** in water, which measures the amount of oxygen consumed by microorganisms during the decomposition of organic matter. High BOD levels indicate organic pollution, such as sewage or agricultural runoff, even though the exact pollutants (e.g., nutrients, organic compounds) aren't directly measured.

Similarly, ***chlorophyll levels*** in water bodies are often used as an indirect indicator of nutrient pollution and algae blooms caused by excess nitrogen and phosphorus.

A ***nutrient*** is a substance that provides essential elements for the growth and metabolism of living organisms, such as nitrogen, phosphorus, and potassium, which are vital for plant and animal development.

***Nutrient pollution*** is the excessive introduction of nutrients, particularly nitrogen and phosphorus, into water bodies from sources like agricultural runoff and sewage, which leads to an overabundance of food for certain organisms (mainly microorganisms and algae). This imbalance can overwhelm ecosystems, causing harmful effects such as algal blooms and water quality degradation.

An ***algae bloom*** is a rapid increase in the population of algae in water bodies, often triggered by nutrient pollution, which can deplete oxygen levels, harm aquatic life, and produce harmful toxins.

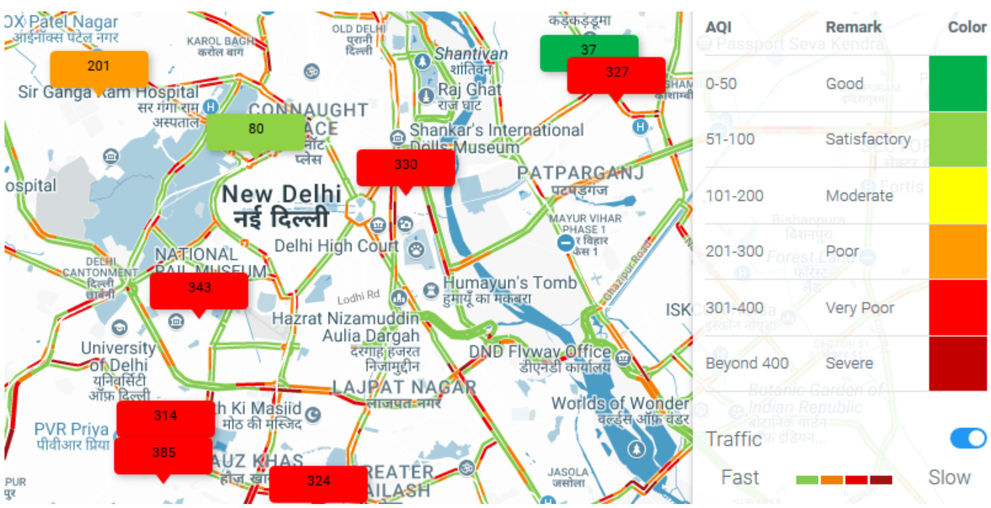
*Algal bloom.*

## Composite Indices

In environmental data analytics, some indices are designed to combine the effects of multiple pollutants, providing a more comprehensive assessment of pollution levels.

*Example from Air Pollution:*

- **Air Quality Index (AQI)**: The AQI combines the concentrations of multiple air pollutants, including particulate matter (***PM2.5 and PM10***), ozone (O3), sulfur dioxide (SO2), nitrogen dioxide (NO2), and carbon monoxide (CO). It provides a single score that indicates the overall air quality and its potential health effects, ranging from "Good" to "Hazardous." It is a very common index to report air quality in cities worldwide.




*AQI map of Delhi.*

## Sources of Pollution in Soil and Water

**Sources of pollution in soil and water c**an be categorized into two main types: **point sources** and **non-point sources**:

1. **Point sources**: These are specific, identifiable locations where pollutants are discharged directly into water bodies or soil. Examples include:
   - **Industrial facilities**: Factories and power plants that discharge waste products like chemicals, heavy metals, or heated water into rivers or lakes through pipes or outlets.
   - **Wastewater treatment plants**: Facilities that treat wastewater before releasing treated or sometimes untreated effluent into water bodies.
   - **Oil spills**: Accidental or deliberate discharges of oil from pipelines, ships, or storage facilities directly into oceans, rivers or soil.
2. **Non-point sources**: These are diffuse, widespread sources of pollution that are harder to trace to a single discharge point, often carried by runoff from the land. Examples include:
   - **Agricultural runoff**: Rainwater washing fertilizers, pesticides, and manure from farmlands into nearby streams, rivers, or lakes.
   - **Urban runoff**: Rainfall flowing over streets, parking lots, and other impermeable surfaces, carrying pollutants such as oil, grease, trash, and heavy metals into storm drains and water bodies.
   - **Atmospheric deposition**: Pollutants like nitrogen oxides and sulfur oxides from vehicle exhaust and industrial emissions that are carried by wind and settle into water bodies through rain or snow.

## Sources of Pollution in Air

Sources of air pollution can be categorized into two main types: **mobile sources** and **stationary sources**:

1. **Mobile sources**: These are sources of air pollution that move from place to place, primarily associated with transportation. Examples include:
   - **Vehicles**: Cars, trucks, buses, and motorcycles that release pollutants like nitrogen oxides (NOx), carbon monoxide (CO), particulate matter (PM), and ***volatile organic compounds (VOCs)*** from fuel combustion.
   - **Aircraft**: Planes contribute to air pollution by emitting carbon dioxide (CO2), nitrogen oxides (NOx), and particulate matter during flight, especially at higher altitudes.

- **Ships and boats**: Marine vessels release sulfur oxides (SOx), nitrogen oxides (NOx), and particulate matter from burning fuel.
- **Trains**: Locomotives, especially those using diesel engines, emit CO2, nitrogen oxides (NOx), and particulate matter.

2. **Stationary sources**: These are non-moving sources that release pollutants from a specific location or facility. Examples include:
    - **Power plants**: Fossil fuel-based power plants (coal, oil, or natural gas) emit large quantities of pollutants such as sulfur dioxide (SO2), nitrogen oxides (NOx), particulate matter (PM), and carbon dioxide (CO2).
    - **Industrial facilities**: Factories, chemical plants, refineries, and manufacturing units release a variety of pollutants including VOCs, heavy metals, nitrogen oxides, and sulfur oxides.
    - **Residential heating**: Wood-burning stoves, furnaces, and boilers that emit pollutants like carbon monoxide, particulate matter, and VOCs into the air.
    - **Construction sites**: Activities at construction sites, such as demolition and the use of heavy machinery, can release dust, particulate matter, and other pollutants.

**Volatile Organic Compounds (VOCs)** are a group of organic chemicals that easily evaporate into the air, contributing to air pollution and smog formation, with harmful effects on human health and the environment. Examples include benzene, formaldehyde, toluene, and acetone.

## The Source-Pathway-Receptor Model

The **Source-Pathway-Receptor Model** is a framework used in environmental science to understand and assess how pollutants move through the environment and affect living organisms or ecosystems. It consists of three main components:

The **source** is the origin or starting point of pollution. It represents the location or activity that releases pollutants into the environment.

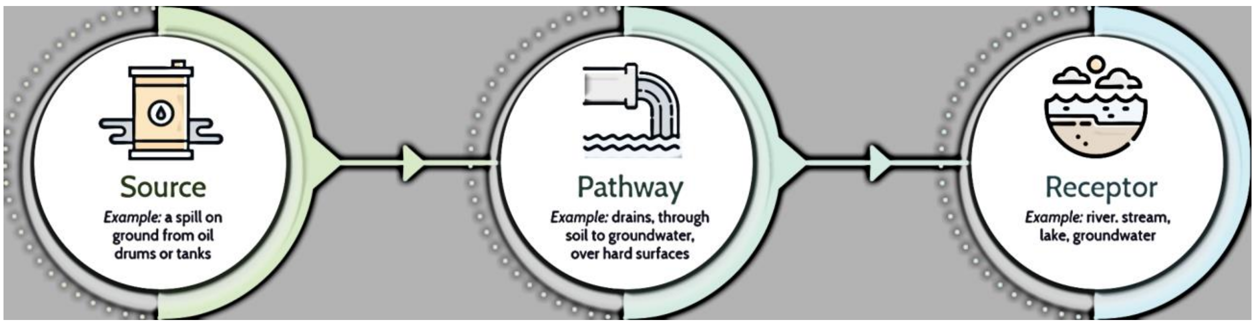
The **pathway** is the route through which the pollutant travels from the source to reach the environment and potentially the receptor. Pathways can involve:

- **Air**: Pollutants can be carried by wind or atmospheric dispersion.
- **Water**: Pollutants can move through surface water, groundwater, or stormwater systems.
- **Soil**: Pollutants may leach into the soil, move via soil erosion, or spread through groundwater.

The **receptor** is the organism, population, or ecosystem that is affected by the pollutant. Receptors can include:

- **Humans**: Impacted by air pollution, contaminated water, or food.
- **Wildlife and ecosystems**: Affected by pollutants in water bodies, soil contamination, or habitat destruction.

**Exposure** refers to the contact between the pollutant and the receptor. If there is no direct contact, then even if there is pollution, there may not be an immediate impact.




*The Source-Pathway-Receptor Model.*

The **Source-Pathway-Receptor Model** and **exposure** are key concepts in environmental data analytics because they provide a structured framework for assessing and understanding the flow of contaminants and their impacts on the environment and human health. Here's why they are important:

- It simplifies the complexity of environmental systems by breaking down the interactions into understandable units.
- It helps in pinpointing the critical stages where interventions can be applied (e.g., stopping pollution at the source, disrupting the pathway).
- Understanding exposure levels allows for accurate assessments of health risks, environmental damage, and the overall safety of a population.
- By analyzing exposure data (e.g., pollutant concentrations, duration of exposure, frequency), decision-makers can craft tailored policies, such as pollution limits or public health interventions.
- In environmental analytics, exposure data is crucial for building models that predict the future impacts of pollution or other environmental hazards.

## Publicly Available Pollution Datasets

there are several publicly available pollution datasets that can be accessed for research or analysis purposes. Below are some of the notable ones:

*1. World Air Quality Index Project*

- **Description**: Provides real-time and historical air quality data for cities worldwide, including PM2.5, PM10, ozone (O3), nitrogen dioxide (NO2), and carbon monoxide (CO).
- **Access**: World Air Quality Index Project

*2. Air Quality Open Data Platform (European Environment Agency)*

- **Description**: Provides data on air pollution in Europe, covering indicators like NO2, SO2, PM10, PM2.5, and ozone concentrations, collected from monitoring stations across European countries.
- **Access**: European Environment Agency Air Quality

*3. US Environmental Protection Agency (EPA) Air Quality System (AQS)*

- **Description**: Contains ambient air pollution data collected by EPA, state, local, and tribal air pollution control agencies. The dataset includes levels of ozone, particulate matter, carbon monoxide, sulfur dioxide, and nitrogen dioxide.
- **Access**: EPA Air Quality System

*4. OpenAQ*

- **Description**: OpenAQ aggregates real-time and historical air quality data from over 100 countries. The data includes pollutants such as PM2.5, PM10, ozone, NO2, SO2, CO, and black carbon.
- **Access**: OpenAQ

*5. Water Pollution Dataset:*

**United Nations Environment Programme (UNEP) - GEMS/Water Data**

- **Description**: The GEMS/Water database provides global water quality data, including parameters such as pH, temperature, dissolved oxygen, nitrates, phosphates, heavy metals, and other pollutants. It covers lakes, rivers, and groundwater from monitoring stations worldwide.
- **Access**: GEMS/Water Data Portal

*6. Soil Pollution Dataset:*

**European Soil Data Centre (ESDAC) - Soil Pollutants and Contamination Data**

- **Description**: ESDAC provides detailed datasets on soil contamination, including data on heavy metals, organic pollutants, and industrial contaminants in European soils. The data helps analyze pollution sources like agriculture, industry, and urbanization.
- **Access**: ESDAC Soil Contamination Data

## Types of Questions Addressed by EDA

Environmental data analytics can answer a wide range of key questions that are crucial for understanding and managing environmental pollution. Here are some categories of questions and examples within each category:

1. Pollution Source Identification

- Accurately identify and quantify the contributions of various pollution sources (e.g., industrial, agricultural, transportation) to overall environmental degradation within a specific area.
- Trace pollutants back to their origin in complex environments, determining the source(s) of contamination.

*2. Pollution Prediction and Forecasting*

- Predict future pollution levels using a combination of historical data and real-time inputs.
- Identify potential pollution hotspots in air or water over the next 5–10 years, considering current trends, policies, and environmental changes.

*3. Pattern and Anomaly Detection*

- Detect unusual patterns or spikes in pollution levels, which may indicate environmental violations, accidents, or system malfunctions.

*4. Impact Assessment and Risk Prediction*

- Assess the projected risk of water contamination spreading to surrounding ecosystems under various climate change scenarios, using advanced predictive models.

*5. Optimization of Mitigation Strategies*

- Optimize pollution control measures, such as emissions reduction strategies, to maximize environmental benefits while minimizing costs.
- Determine the optimal placement of sensors or monitoring stations to improve the accuracy and efficiency of environmental data collection and monitoring.

*6. Spatio-Temporal Modeling*

- Model the spread of pollutants over time and space using spatio-temporal data and machine learning techniques.
- Develop dynamic models that integrate real-time data, continuously updating predictions on pollution dispersion.

*7. Causal Inference and Relationships*

- Infer causal relationships between specific pollution sources and observed environmental degradation, using advanced data analytics and causal modeling techniques.
- Quantify the effectiveness of various environmental policies on reducing pollution levels by leveraging predictive modeling.

*8. Climate and Pollution Interaction*

- Model the interaction between climate change factors (e.g., rising temperatures, changing rainfall patterns) and pollution levels to understand future environmental risks.
- Predict the long-term impacts of climate-induced changes on air and water quality using integrated environmental and climate data.