Introduction to Environmental Data Types and Formats

Learning Outcomes

After completing this unit, students will be able to:

- Define key types of environmental data and understand their significance in environmental studies.
- Distinguish between structured, semi-structured, and unstructured environmental data.
- Identify common data formats used in environmental science (CSV, NetCDF, HDF5, GeoJSON, shapefiles, etc.).
- Evaluate the advantages and limitations of different data formats for storage, sharing, and analysis.
- Convert and handle environmental data files using appropriate tools and libraries (e.g. Python, QGIS).

Introduction

Environmental data plays a fundamental role in understanding, monitoring, and managing natural and human-influenced systems. They provide the empirical foundation for studying climate change, air and water quality, biodiversity, land use, and many other environmental processes. Advances in sensing technologies, satellite observation systems, and computational tools have led to an unprecedented growth in the volume, variety, and complexity of environmental data. As a result, professionals and students in environmental sciences must be able to analyse data and understand the different types and formats in which environmental data are stored and exchanged.

Environmental data comes from a wide range of sources. In site measurements are collected using ground-based instruments, such as weather stations, air quality sensors, river gauges, and ecological monitoring plots. Remote sensing platforms, including satellites, drones, and aircraft, provide extensive spatial observations of the Earth’s surface and atmosphere. In addition, environmental data may be generated through numerical models and simulations, such as climate or hydrological models, which produce large multi-dimensional datasets. Each of these sources produces data with distinct structures, scales, and technical requirements, which influence how the data are stored and used.

Understanding environmental data types and formats is essential for effective data handling and analysis. Data types describe how information is organised, for example, as structured tabular data, semi-structured records, or unstructured files such as images or text. Data formats, on the other hand, define how data are encoded in files such as CSV, Parquet, NetCDF, HDF5, GeoTIFF, or GeoJSON. The choice of data format affects data accessibility, storage efficiency, interoperability between software tools, and the ability to include metadata that describe data's origin, units, spatial reference, and temporal coverage.

In environmental science, improper handling of data formats can lead to misinterpretation, data loss, or reduced reproducibility of results. For example, ignoring coordinate reference systems in spatial data may cause incorrect map overlays, while the use of overly simple formats may result in the loss of important metadata. In contrast, selecting suitable data formats enables efficient data sharing, long-term preservation, and integration of heterogeneous datasets from multiple sources. This is particularly important in interdisciplinary environmental research, where datasets from different domains must be combined and compared.

This lecture introduces the main types of environmental data and the most commonly used data formats in environmental science and geospatial applications. Provides a conceptual framework for understanding how environmental data are structured and stored, supported by practical examples from air quality monitoring, climate observations, and land use analysis. By gaining familiarity with these concepts, students will be better prepared to work with real-world environmental datasets, choose appropriate formats for analysis, and ensure that their data workflows are transparent, reproducible, and scientifically sound.

Environmental Data Types

Environmental data can be classified in several ways, depending on their structure, dimensionality, spatial and temporal characteristics, and method of collection. A clear understanding of environmental data types is essential for selecting appropriate storage formats, analysis techniques, and visualisation methods. In this section, environmental data are grouped into three broad categories based on their structural characteristics: structured, semi-structured, and unstructured data. In addition, specific attention is given to spatial and temporal dimensions, which are particularly important in environmental applications.

Structured Environmental Data

Structured data are highly organised and typically stored in tabular form, where data are arranged into rows and columns with clearly defined attributes and data types. Each row usually represents an observation, while each column corresponds to a variable or attribute. This type of data is common in environmental monitoring programmes and is often collected through standardised measurement procedures.

Examples of structured environmental data include time series of air temperature, precipitation measurements, air pollutants concentrations, water quality indicators, or soil properties measured at fixed locations. These datasets are frequently stored in formats such as CSV files, spreadsheets, or relational databases. Structured data are relatively easy to query, filter and analyse using statistical software, databases, and programming languages such as R or Python.

CSV is a simple, row-based text format that is easy to read and exchange but inefficient for large-scale analytical processing and long-term storage. In contrast, Apache Parquet is a column-based format optimized for analytical workloads, enabling faster queries on selected variables, efficient compression, and preservation of data types. As a result, Parquet is better suited for large environmental datasets and data-intensive workflows, while CSV remains useful for small datasets and data sharing.

One of the key advantages of structured data is their simplicity and compatibility with a wide range of tools. They allow efficient aggregation, statistical analysis, and visualisation, such as calculating averages, trends, or correlations over time. However, structured data also have limitations. They are less suitable for representing complex spatial structures, multi-dimensional variables, or large raster datasets. In addition, important contextual information may be lost if metadata are not carefully documented and stored separately.

Semi-Structured Environmental Data

Semi-structured data do not conform to rigid tabular structures, but still contain organisational elements such as tags, keys, or hierarchical relationships. This type of data provides greater flexibility than structured data while maintaining a level of organisation that allows automated processing.

In environmental science, semi-structured data are commonly used to represent metadata, sensor output, and spatial features. Formats such as XML and JSON are widely used for data exchange between systems, especially in web services and application programming interfaces

(APIs). GeoJSON, a specialised extension of JSON, is frequently used to represent geographic features along with their attributes and geometry.

Semi-structured data are particularly useful when dealing with heterogeneous datasets that vary in structure or when additional attributes may be added over time. For example, environmental sensor networks often produce semi-structured outputs that include measurements, timestamps, sensor identifiers, and quality flags. The flexibility of semi-structured data enables easier integration across platforms, but it may require more complex parsing and validation compared to structured tabular data.

Unstructured Environmental Data

Unstructured data lack a predefined data model and are not easily represented in tables or hierarchical structures. This category includes images, videos, audio recordings, free-text reports, and raw remote sensing imagery. In environmental contexts, unstructured data are increasingly important due to advances in Earth observation technologies and automated data collection systems.

Examples of unstructured environmental data include satellite images, aerial photographs, drone footage, scanned field notes, and textual environmental impact assessments. These data types often require specialised processing techniques such as image analysis, natural language processing, or machine learning methods. Although unstructured data can contain rich and detailed information, they are more challenging to analyse and store efficiently.

The analysis of unstructured environmental data often depends on additional layers of interpretation or transformation. For instance, satellite images may be processed to derive vegetation indices or land cover classifications, transforming unstructured raster data into structured or semi-structured datasets suitable for further analysis. As such, unstructured data often serve as a raw input for generating more structured environmental information.

Spatial Environmental Data

Spatial data represent information that is explicitly associated with a geographic location. The location may be defined using coordinates, administrative boundaries, grids, or spatial reference systems. Spatial data are fundamental to environmental science, as most environmental phenomena are inherently linked to space.

Spatial environmental data are typically divided into vector and raster data types. Vector data represent discrete features such as monitoring stations, rivers, roads, or land parcels using points,

lines, and polygons. Raster data represent continuous phenomena such as temperature fields, elevation models, or satellite imagery using a grid of cells or pixels.

Understanding spatial data types is crucial for environmental analysis and mapping. Issues such as coordinate reference systems, spatial resolution, and scale can significantly affect the analysis results. Improper handling of spatial data can lead to misaligned layers, inaccurate measurements, or incorrect interpretations. Therefore, spatial data types require careful management and specialised tools such as Geographic Information Systems (GIS).

Temporal Environmental Data

Temporal characteristics are another defining aspect of environmental data. Many environmental processes vary over time, from short-term fluctuations in weather conditions to long-term trends in climate or ecosystem change. Temporal data may be collected at regular intervals, irregular intervals, or as continuous measurements.

Time series data are a common form of temporal environmental data. Examples include hourly air quality measurements, daily river discharge values, or monthly climate indices. Temporal resolution, time zone conventions, and data gaps are important considerations when working with time-dependent environmental data. Proper temporal alignment is essential when combining datasets from different sources.

In many cases, environmental data are both spatial and temporal, resulting in spatio-temporal datasets. These datasets can be complex and high-dimensional, such as global climate model output or satellite-derived time series. Handling spatio-temporal data requires specialised data structures and formats capable of efficiently storing and accessing large volumes of information.

Multi-Dimensional and Gridded Data

Multi-dimensional data represent environmental variables that vary across multiple dimensions, such as latitude, longitude, altitude, and time. These data are common in atmospheric and oceanographic sciences, where measurements are taken across three-dimensional space and over long time periods.

Gridded data are a specific type of multi-dimensional data where values are defined on a regular grid. Examples include climate reanalysis datasets, numerical weather prediction output, and ocean circulation models. These datasets often contain multiple variables and extensive metadata, making them significantly more complex than simple tabular data.

The complexity of multi-dimensional environmental data necessitates the use of specialised formats and software tools that can efficiently handle large arrays and preserve metadata. Understanding the nature of these data types is essential before selecting appropriate data formats and analysis workflows.

Common Environmental Data Formats

Environmental data formats define how data are encoded, stored, and exchanged between systems. The choice of data format has a significant impact on data accessibility, storage efficiency, interoperability, and long-term usability. In environmental science, data formats must often support large volumes of data, spatial and temporal dimensions, and detailed metadata. This subchapter introduces the most commonly used data formats in environmental and geospatial applications, highlighting their characteristics, advantages, and limitations.

Tabular Data Formats

Tabular formats are among the most widely used data formats in environmental science due to their simplicity and broad compatibility. These formats store data in rows and columns and are commonly used for structured environmental datasets such as time series measurements and summary statistics.

The most common tabular format is the Comma-Separated Values (CSV) format. CSV files are plain-text files in which each row represents a record and the values are separated by commas or other delimiters. CSV is easy to create, read, and exchange across different software platforms. It is widely used to share air quality measurements, meteorological observations, and water quality data. However, CSV files lack standardised support for metadata, data types, and spatial reference information, which can lead to ambiguity or loss of data if not properly documented.

Spreadsheet formats such as Microsoft Excel files are also frequently used, particularly in the early stages of data collection or for small datasets. Although spreadsheets provide user-friendly interfaces and basic data validation features, they are less suitable for large datasets, automated processing, or reproducible workflows. Hidden formatting, multiple sheets, and manual edits can introduce errors and reduce transparency.

Columnar and Analytical Data Formats

Apache Parquet is a columnar, open-source data storage format optimized for analytical workloads and large-scale data processing. Unlike traditional row-based tabular formats such as CSV, Parquet stores data column by column, which significantly improves performance for queries that access only a subset of variables. This makes Parquet particularly suitable for large environmental datasets with many attributes and observations.

In environmental science, Parquet is increasingly used for storing high-volume tabular data such as long-term air quality measurements, climate model outputs converted to tabular form, and aggregated sensor data. The format supports efficient compression and encoding, reducing storage requirements while maintaining fast read performance. Parquet also preserves data types and schema information, addressing many limitations of plain text formats.

Although Parquet is not a native GIS format, it plays an important role in environmental data pipelines where data are processed in cloud platforms, data warehouses, or distributed computing environments before being visualized or spatially analysed in GIS software. Data stored in Parquet format can be converted to spatial formats or connected to GIS tools through intermediate processing steps using Python, SQL-based engines, or spatial data libraries.

Scientific and Multi-Dimensional Data Formats

Scientific data formats are designed to efficiently store large, multi-dimensional datasets and to include rich metadata describing the data structure, units, and coordinate systems. These formats are particularly important in climate science, meteorology, oceanography, and atmospheric research.

NetCDF (Network Common Data Form) is one of the most widely used scientific data formats in environmental science. It supports multi-dimensional arrays and self-describing metadata, making it suitable for storing spatio-temporal data such as climate model outputs, reanalysis datasets, and satellite-derived variables. NetCDF files can contain multiple variables and dimensions within a single file, allowing efficient data access and subsetting.

HDF5 (Hierarchical Data Format version 5) is another powerful format for storing complex environmental datasets. Provides a hierarchical structure that allows data to be organised into groups and datasets, similar to a file system within a file. HDF5 is highly scalable and efficient, which makes it suitable for very large datasets. However, its complexity may pose challenges for beginners, and specialised libraries are often required for reading and writing HDF5 files.

Spatial Vector Data Formats

Vector data formats are used to represent discrete geographic features, such as points, lines, and polygons. These formats are essential for mapping environmental monitoring locations, administrative boundaries, rivers, roads, and land use features.

The ESRI Shapefile has been a standard vector format in GIS for many years. A shapefile consists of multiple associated files that together store geometry, attributes, and spatial reference information. Although shapefiles are widely supported, they have several limitations, including restrictions on file size, attribute name length, and data types.

GeoJSON is a modern, open, text-based format for encoding geographic features using JSON. It is particularly popular for web-based mapping and data exchange as a result of its simplicity and readability. GeoJSON supports attributes and geometry in a single file and integrates well with web technologies. However, GeoJSON files can become large and inefficient for complex or extensive datasets.

Spatial Raster Data Formats

Raster data formats are used to represent continuous environmental phenomena, where values are stored in a grid of cells or pixels. Common examples include satellite imagery, digital elevation models, land surface temperature maps, and vegetation indices.

GeoTIFF is a widely used raster format that extends the standard TIFF format by embedding geographic metadata such as coordinate reference systems and spatial resolution. GeoTIFF files are commonly used in remote sensing and GIS applications and are supported by most geospatial software.

Other raster formats, such as NetCDF and HDF5, are also frequently used for gridded environmental data, especially when time and vertical dimensions are involved. The choice between raster formats often depends on the complexity of the data, file size, and analysis requirements.

Metadata and Data Description Standards

Metadata are critical to understanding and using environmental data correctly. Metadata describe the origin, content, structure, and quality of the data, as well as information about spatial and temporal coverage. Without adequate metadata, environmental datasets may be difficult or impossible to interpret accurately.

Several metadata standards are commonly used in environmental and geospatial applications. ISO 19115 is an international standard for geographic metadata, providing a comprehensive framework for describing spatial datasets. Dublin Core is a simpler, more general metadata standard that is used in various domains. Many scientific data formats, such as NetCDF, support embedded metadata that follow established conventions.

Proper use of metadata standards improves interoperability, reproducibility, and long-term preservation of data. It also facilitates data discovery through data portals and repositories, enabling researchers and policymakers to reuse environmental datasets effectively.

Choosing the Appropriate Data Format

The selection of an appropriate data format is a critical step in the management of environmental data. The choice depends on several factors, including the size, structure, spatial and temporal dimensions, analysis goals, and intended audience. Simple tabular formats may be sufficient for small, well-documented datasets, while complex, multi-dimensional formats are necessary for large-scale environmental modelling and analysis.

In many cases, environmental data workflows involve the conversion of data between formats to meet specific analytical or visualisation needs. Understanding the strengths and limitations of common environmental data formats enables informed decisions and reduces the risk of loss or misinterpretation of data.

Applied Use Case: Exploring Environmental Data Formats Using QGIS

Geographic Information Systems (GIS) provide a powerful environment for working with spatial environmental data from multiple sources and in different formats. QGIS is an open-source GIS platform widely used that supports a broad range of environmental data formats and analytical workflows. This applied use case demonstrates how environmental data types and formats can be explored, visualised, and converted using QGIS, highlighting practical challenges and best practices in environmental data handling.

Overview of the Use Case

The use case focusses on the integration and analysing of environmental datasets stored in different formats within a single GIS project. Students work with three representative datasets: air quality monitoring data stored in CSV format, land cover data provided as a vector data set

(shapefile or GeoJSON), and satellite-derived temperature data stored in raster format (GeoTIFF or NetCDF). These datasets differ in structure, dimensionality, and metadata content, making them suitable for illustrating key concepts introduced in previous chapters.

The objective of this use case is not advanced spatial analysis, but rather to develop practical skills in loading, inspecting, and managing diverse environmental data formats. Through hands-on interaction with QGIS, students learn how data formats influence visualisation, spatial alignment, and interpretability.

Loading Tabular Environmental Data (CSV)

The first step involves loading air quality measurements stored in a CSV file. The data set typically contains columns such as the station identifier, latitude, longitude, date, and pollutant concentration. When importing CSV data into QGIS, students must specify which columns represent spatial coordinates and select the appropriate coordinate reference system.

This step highlights important considerations when working with tabular environmental data. Incorrect coordinate definitions can result in mislocated points, while missing metadata may require assumptions about units or measurement standards. Once loaded, the CSV data are displayed as point features, allowing students to visually inspect the spatial distribution of monitoring stations.

Students are encouraged to explore attribute tables, apply basic filters, and examine temporal patterns using attribute queries. This shows how structured data can be spatially enabled and integrated into GIS workflows.

Working with Vector Environmental Data

The second data set consists of land cover information provided in a vector format, such as a shapefile or GeoJSON. This data set represents discrete spatial features and includes attribute data describing land use categories. Upon loading the vector data set into QGIS, students examine its geometry types, attribute fields, and coordinate reference system.

This part of the exercise emphasises the importance of spatial reference systems and the consistency of attributes. Students compare the coordinate reference system of the land cover dataset with that of the air quality points and reproject layers if necessary. They also explore basic symbology options to visualise land cover classes and understand how vector attributes are linked to spatial features.

Through this process, students gain insight into the strengths and limitations of vector data formats, including file structure, attribute storage, and interoperability.

Exploring Raster Environmental Data

The third data set is a raster data set representing the surface temperature derived from the satellite or another continuous environmental variable. This data set may be provided in GeoTIFF or NetCDF format. Raster data are loaded into QGIS as grid-based layers, where each cell contains a numerical value representing the measured or modelled variable.

Students examine raster properties such as spatial resolution, extent, and data type. Visualisation techniques, including colour ramps and value classification, are used to interpret spatial patterns. If a NetCDF dataset is used, students explore how multiple variables or time steps can be accessed within a single file.

This step illustrates how raster formats are suited for representing continuous environmental phenomena and how metadata embedded in raster files support accurate spatial interpretation.

Comparison and integration of multiple data formats

Once all datasets are loaded, students overlay vector and raster layers within the same QGIS project. This integration allows them to visually assess the relationships between air quality measurements, land cover types, and temperature patterns. Basic spatial queries, such as identifying monitoring stations within specific land cover classes, can be performed to demonstrate the practical integration of different data types.

This comparison highlights how environmental data formats complement each other and how GIS serves as a unifying platform for multi-format data analysis. Students also observe how data quality and resolution differences affect interpretation.

Data Conversion and Export

An important component of the use case is data conversion. Students practice exporting vector layers from the shapefile to GeoJSON format and raster layers from GeoTIFF to another supported format. This exercise reinforces the concept that data formats are often chosen based on specific tasks, software requirements, or data-sharing needs.

Attention is paid to preserving metadata during conversion and understanding potential limitations or data loss. Students are encouraged to document conversion steps and verify the output.

Practical Exercises — Step-by-Step QGIS Tasks

These practical exercises are designed to provide hands-on experience with environmental data types and formats using QGIS, a free and open-source Geographic Information System. Students will work with real-world environmental datasets in different formats, learning to load, visualise, integrate, and convert them.

Data Sources (Free and Open Access)

1. Air Quality Monitoring Data (CSV)

Data sets of free air quality and monitoring stations are available from the European Environment Agency (EEA) and other open data portals of government.

• European Environment Agency – Environmental Data and-maps

https://www.eea.europa.eu/Data and Maps

Students can download air quality measurements and export them as CSV files that contain station coordinates, timestamps, and pollutant concentrations.

Alternative option:

The NASA POWER plugin for QGIS provides direct access to climate and atmospheric variables (temperature, precipitation, radiation) without manual downloading.

---

2. Land Cover Data (Raster / GeoTIFF or Web Services)

ESA WorldCover provides a global land cover data set derived from Sentinel satellite imagery.

• ESA WorldCover (download or web services)

http://ESA WorldCover.org

WMTS/WMS services (can be used directly in QGIS):
• WMTS: `https://services.terrascope.be/wmts/v2`
• WMS: `https://services.terrascope.be/wms/v2`

---

3. Satellite Imagery (Raster / GeoTIFF)

Free satellite imagery suitable for environmental analysis can be obtained from:

• **Copernicus Open Access Hub** (Sentinel missions)
https://dataspace.copernicus.eu

• **USGS EarthExplorer** (Landsat archive)
https://earthexplorer.usgs.gov

Students are advised to download small subsets or tiles due to file size.

---

4. Vector Data (OpenStreetMap)

OpenStreetMap provides free vector data for roads, land use, boundaries, and infrastructure.

• OpenStreetMap

https://www.openstreetmap.org

Data can be imported into QGIS using the **QuickOSM** plugin.

Exercise 1 — Import and Visualise Air Quality CSV Data

Objective: Load tabular air quality data and visualise monitoring stations as point features.

Steps:

1. Open QGIS and select **Layer → Add Layer → Add Delimited Text Layer**.
2. Browse to the downloaded CSV file (e.g., `air_quality.csv`).
3. Set the X field to longitude and the Y field to latitude.
4. Select the coordinate reference system (EPSG:4326).
5. Click **Add** to display the points on the map.
6. Open the Attribute Table (F6) to inspect the available variables.

Discussion:
Which attributes are included in the dataset? Are units and timestamps clearly defined?

Exercise 2 — Load and Style Data

Objective: Visualise land cover data using raster layers and apply thematic styling.

Steps:

1. In QGIS, go to **Layer → Add Layer → Add WMS/WMTS Layer**.
2. Create a new connexion using the ESA WorldCover WMTS URL.
3. Select a land cover layer and add it to the project.
4. Open Layer Styling (F7) and apply a categorical colour scheme for land cover classes.
5. Add a legend to interpret land cover categories.

Discussion:
How does the distribution of land cover vary in the study area?

Exercise 3 — Load and Explore Satellite Raster Data

Objective: Load satellite raster data and explore spatial patterns.

Steps:

1. Download a Sentinel-2 or Landsat GeoTIFF image from the Copernicus or USGS portals.
2. In QGIS, go to **Layer → Add Layer → Add Raster Layer** and load the file.
3. Inspect raster properties such as resolution, extent, and coordinate system.
4. Apply colour ramps or create a false-colour composite.
5. (Optional) Use the **Raster Calculator** to compute NDVI.

Discussion:
What environmental features are visible in the raster data?

Exercise 4 — Integrate Vector and Raster Layers

Objective: Combine multiple data formats and perform basic spatial interpretation.

Steps:

1. Overlay air quality monitoring points on land cover and satellite raster layers.
2. Use the **Identify Features** tool to inspect raster values at point locations.
3. (Optional) Use Raster → Zonal Statistics to summarise raster values by land cover class.

Discussion:
Are there observable relationships between air quality measurements and land cover types?

Exercise 5 — Convert and Export Data Formats

Objective: Convert environmental data between different formats.

Steps:

1. Right-click on a vector layer and select Export → Save Features As….
2. Export the layer as **GeoJSON**.
3. For raster data, go to **Raster → Conversion → Translate (Convert Format)**.
4. Save the raster in an alternative format (e.g., GeoTIFF or GeoPackage).
5. Verify that the spatial reference and attributes are preserved.

Discussion:
Which formats are most suitable for data sharing and long-term storage?

Assessment

Question 1Which type of environmental data is best represented in a tabular structure with rows and columns?

A) Unstructured data

B) Semi-structured data

C) Structured data

D) Raster data

Correct answer: C) Structured data

Question 2Which data format is most suitable for storing multi-dimensional climate or atmospheric data with embedded metadata?

A) CSV
B) GeoJSON
C) NetCDF
D) Shapefile

Correct answer: C) NetCDF

Question 3Which QGIS data format is primarily used to represent continuous environmental phenomena such as temperature or elevation?

A) Vector (points, lines, polygons)
B) Raster

C) Tabular

D) Text-based

Correct answer: (B) Raster

Question 4What is a key advantage of using GeoJSON compared to ESRI Shapefile?

A) It supports larger file sizes than any other format

B) It stores geometry and attributes in a single, human-readable file

C) It is faster for raster analysis

D) It is required for all GIS applications

Correct answer: B) It stores geometry and attributes in a single, human-readable file

---

Question 5Why are metadata particularly important in environmental datasets?

A) They reduce file size

B) They improve visualisation speed

C) They describe data origin, units, spatial and temporal reference

D) They replace the need for data analysis

Correct answer: C) They describe data origin, units, spatial, and temporal reference

---

Practical Tasks (4–5 Exercises)

Students are assessed based on the successful completion of the following practical exercises using QGIS and open environmental datasets.

1. **Data Import and InspectionLoad a CSV dataset containing environmental measurements into QGIS, define the correct coordinate reference system, and inspect attribute fields and metadata.**
2. **Vector and Raster Data handlingLoad vector land cover data and raster satellite imagery, verify spatial alignment, and apply appropriate symbology.**
3. **Data IntegrationOverlay point, vector, and raster datasets and interpret spatial relationships between environmental variables.**
4. **Data Format Conversion**
   Convert vector data from Shapefile to GeoJSON and export raster data to an alternative supported format while preserving spatial reference information.
5. **Visualisation and Interpretation (optional / bonus)**
   Create a simple map layout with legend and title and briefly interpret the observed environmental patterns.

References / Further Reading

- Davis, Jerry. *Introduction to Environmental Data Science.* Bookdown.org, 2022. Open Access. https://bookdown.org/igisc/EnvDataSci.
- Earth-Env-Data-Science Project. *Earth & Environmental Data Science – An Introduction.* 2023. Open Educational Resource. https://earth-env-data-science.github.io/intro.html.
- Lovelace, Robin, Jakub Nowosad, and Jannes Muenchow. *Geocomputation with R.* CRC Press, 2019. Open Access edition. https://r.geocompx.org/.
- Rey, Sergio J. and Levi J. Wolf. Geo-data science with Python. Geographicdata.science, 2022. Open Textbook. https://geographicdata.science/book/intro.html.
- Bolstad, Paul, and Michael Shin. *Essentials of Geographic Information Systems.* 2nd ed., University of Minnesota Libraries Publishing, 2018. Open Textbook Library, https://open.umn.edu/opentextbooks/textbooks/67.