UNIVERSITÉ DU LUXEMBOURG

Remote Sensing

Remote sensing (RS) data refers to the acquisition of information about the Earth's surface and atmosphere from a distance, typically using platforms such as satellite, aircrafts, drones (Unmanned Aerial Vehicles, UAVs), or Balloons.

Importance of RS Data

RS data are crucial for environmental data analytics for several reasons:

- **Access to Challenging Areas**: RS data provide valuable information from regions that are difficult or impossible to reach for in situ observations, such as dense forests, polar regions, mountainous areas, and oceanic expanses.
- **Global Reach**: Satellites orbit the Earth, providing consistent and comprehensive coverage of the entire globe, including areas with limited human presence or infrastructure.
- **Economies of Scale**: Once a satellite is launched, it can cover vast areas of the Earth's surface regularly at a fraction of the cost compared to extensive ground-based campaigns.
- **Repeatability and Consistency**: Satellites provide frequent revisits over the same area, allowing for consistent monitoring over time without the need for repeated deployment of field teams.
- **Long-Term Records**: Archives of satellite data extend back several decades, offering a historical perspective on environmental changes and trends.
- **On-Demand Data from Aerial Platforms**: Drones and aircraft can be deployed as needed to capture data at specific times, providing flexibility in temporal resolution.
- **Non-Intrusive Data Collection**:
  - **Minimal Environmental Impact**: RS data collection does not disturb the environment, making it a non-intrusive method for monitoring ecosystems, wildlife habitats, and natural resources.
  - **Safety and Convenience**: RS methods eliminate the risks associated with fieldwork in hazardous or remote locations.
- **Vertical Profiling with Balloons**: Balloons can provide vertical profiles of the atmosphere, offering unique insights into atmospheric conditions and changes.

Nature of RS Data

RS data, in its raw or processed form, is most represented as **raster** geospatial data. However, in some cases, RS data can also take the form of point clouds, particularly when using technologies like LiDAR (Light Detection and Ranging) or photogrammetry, which capture 3D spatial data by measuring points in space. Additionally, RS data can be processed and transformed into vector data, representing geographic features as points, lines, or polygons, depending on the application.

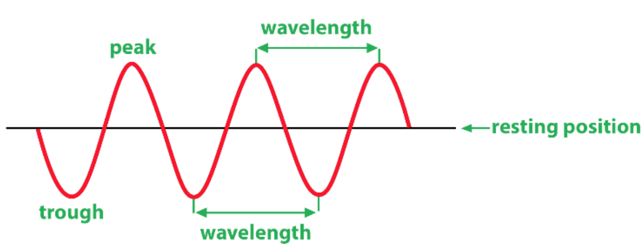
Key RS Terminology

To understand remote sensing data, you need to be familiar with a couple of key terms:

Electromagnetic Radiation

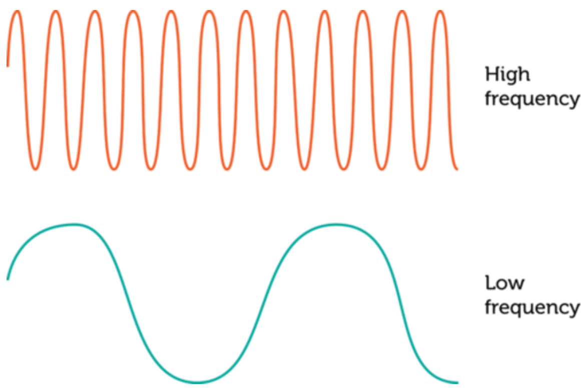
Electromagnetic radiation (EMR) is the energy emitted and propagated through space in the form of waves. While it is commonly associated with visible light, electromagnetic radiation encompasses a much broader spectrum. This spectrum includes, in order of increasing wavelength and decreasing frequency, radio waves, microwaves, infrared radiation, visible light, ultraviolet radiation, X-rays, and gamma rays.

- **Wavelength**: A wavelength is the distance between consecutive peaks of a wave, typically measured in meters or its subunits like nanometers (nm). In remote sensing, different wavelengths of electromagnetic radiation provide different types of information. Shorter wavelengths (like visible light) can capture fine details of the Earth's surface, while longer wavelengths (like infrared) can penetrate clouds and provide information on thermal properties. The choice of wavelength is crucial for the specific application and the type of data required.



Wavelength.

- **Frequency:** Frequency is the number of times a wave repeats or cycles in one second, measured in hertz (Hz). It represents how often the crests of a wave pass a given point in a set amount of time. For electromagnetic waves, higher frequency means more energy and typically shorter wavelengths, as seen in X-rays and gamma rays, while lower frequency corresponds to less energy and longer wavelengths, such as in radio waves. Frequency is an essential property in understanding wave behavior across the electromagnetic spectrum.

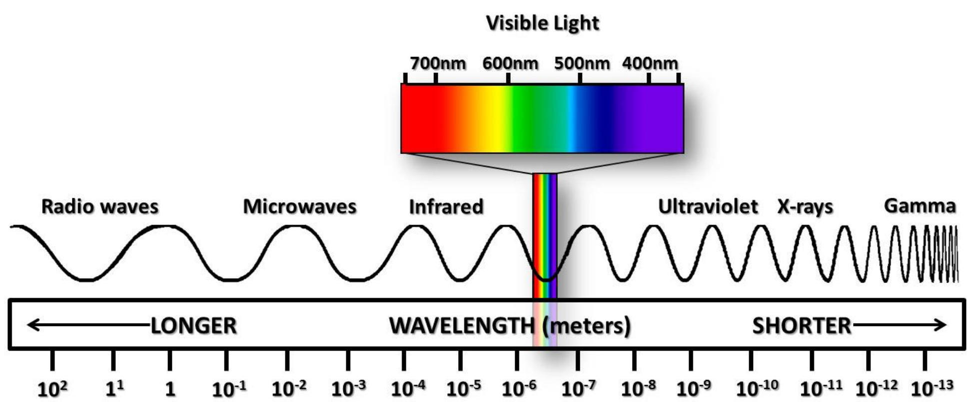


Frequency.

In the context of remote sensing, EMR is essential as it serves as the medium through which sensors detect information about the Earth's surface and atmosphere. This radiation can be reflected, absorbed, or emitted by objects, and sensors capture these interactions to generate data. Different types of sensors are designed to detect various parts of the electromagnetic spectrum, providing insights into the physical and chemical properties of the observed objects.

Electromagnetic Spectrum

The electromagnetic spectrum is the complete range of electromagnetic radiation, encompassing all wavelengths and frequencies. It includes various types of radiation such as gamma rays, X-rays, ultraviolet (UV), visible light, infrared (IR), microwaves, and radio waves. In remote sensing, different parts of the electromagnetic spectrum are utilized to gather different data about the Earth's surface and atmosphere. Each type of radiation interacts with matter in distinct ways, allowing sensors to capture a wide range of information. Understanding the electromagnetic spectrum is fundamental to selecting the appropriate sensors and wavelength bands for specific remote sensing applications, ensuring the accurate and effective collection of environmental data.



The Electromagnetic Spectrum.

Wavelength Bands

Wavelength bands are specific ranges of wavelengths within the electromagnetic spectrum that sensors can detect. These bands are categorized based on their ability to capture particular types of information. For example, the visible band (400-700 nm) is used to capture images similar to what the human eye sees, while the near-infrared band (700-1400 nm) is used to monitor vegetation health. Each band provides unique data, allowing for comprehensive analysis when combined.

Single-band (or single-channel) Imaging

Single-band (or single-channel) imaging captures data in only one specific band or wavelength of the electromagnetic spectrum. Examples of single-band imaging include the following:

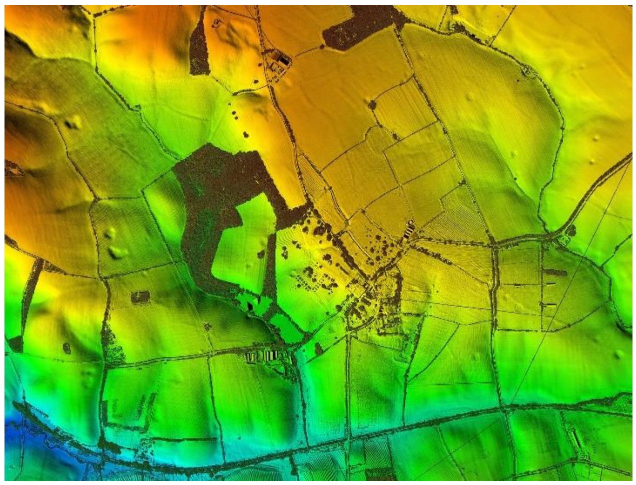
- ***Thermal Infrared Imaging:*** Thermal infrared (TIR) imaging measures emitted radiation in the thermal infrared region of the spectrum, typically between 3 and 14 micrometers. This type of imaging is used to detect heat and temperature variations, making it valuable for applications like urban heat island studies, geothermal exploration, and monitoring of vegetation stress and water bodies.

The right image shows the same area as the left image captured with an aerial thermal infrared sensor.

- ***Microwave and Radar Sensing:*** Microwave and radar remote sensing involve the use of microwaves (longer wavelengths) to penetrate clouds, vegetation, and, in some cases, the ground. Synthetic Aperture Radar (SAR) is a popular microwave remote sensing technology that provides detailed topographic maps and can be used for applications like monitoring deforestation, soil moisture estimation, and disaster management (e.g., detecting oil spills or assessing earthquake damage).

Example of a Synthetic Aperture Radar (SAR) image.

- ***LiDAR (Light Detection and Ranging):*** LiDAR uses laser pulses to measure distances and generate high-resolution, three-dimensional maps of the Earth's surface. It is especially effective for detailed topographic mapping, vegetation structure analysis, and urban planning. LiDAR is widely used in forestry, archaeology, and autonomous vehicle navigation. LiDAR typically operates in the near-infrared (NIR) region of the electromagnetic spectrum, with common wavelengths ranging from 800 nm to 1550 nm.



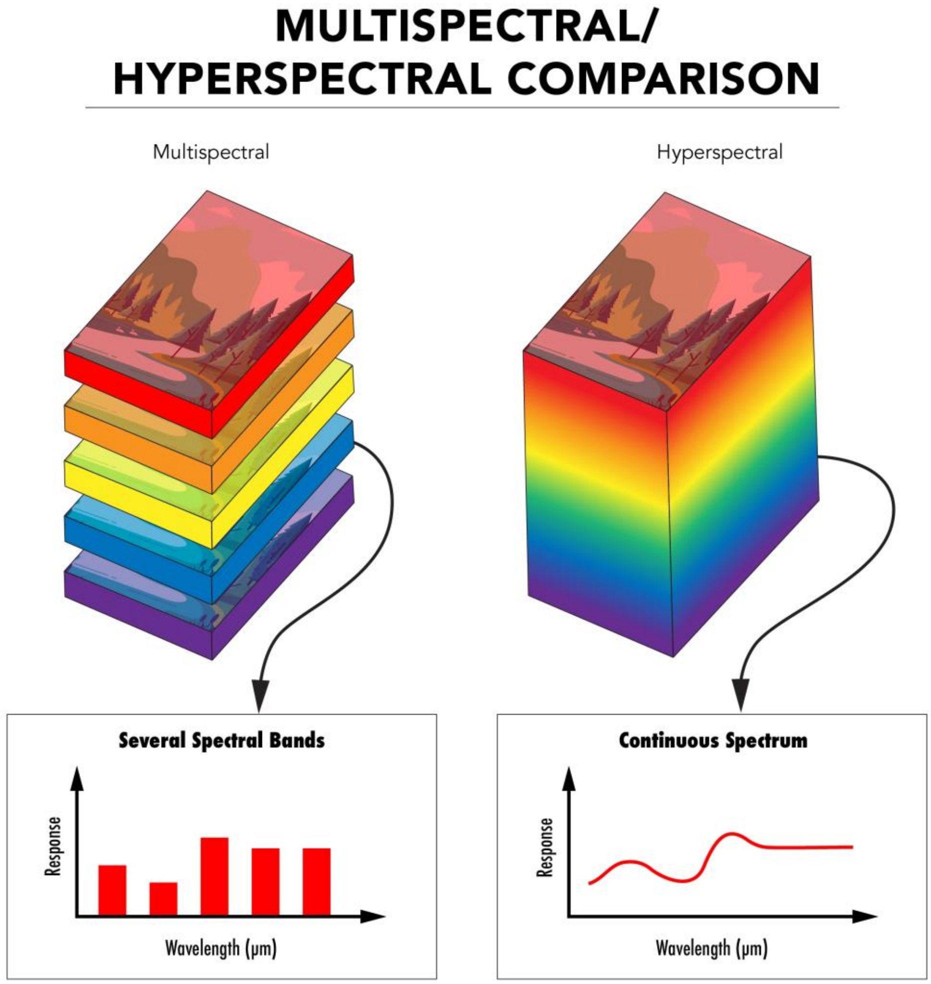
Example of a LiDAR image.

Multispectral Imaging

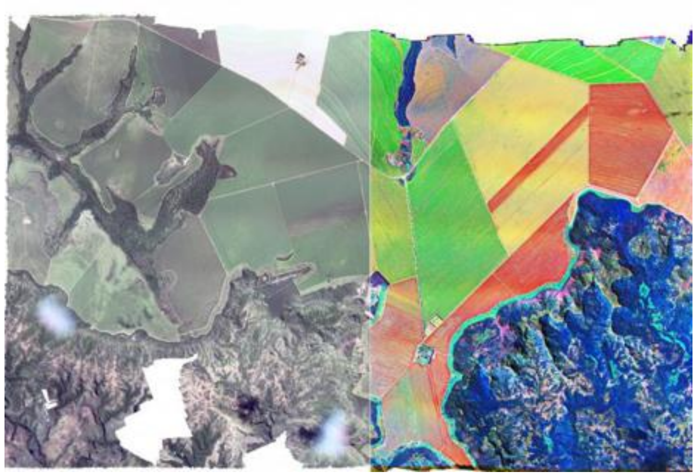
Multispectral remote sensing involves capturing data from multiple wavelength bands across the electromagnetic spectrum. Typically, these sensors capture data from a limited number of broad bands (e.g., red, green, blue, and near-infrared). Multispectral imaging is useful for applications like land cover classification, vegetation analysis, and water quality monitoring. It provides a balance between detail and coverage, making it widely used in various remote sensing applications.

Hyperspectral Imaging

Hyperspectral remote sensing takes the concept of multispectral imaging further by capturing data from hundreds of narrow, contiguous wavelength bands across the electromagnetic spectrum. This high spectral resolution allows for the detailed identification of materials and substances based on their unique spectral signatures. Hyperspectral sensors are particularly valuable in applications requiring precise material discrimination, such as agricultural assessments. The rich data provided by hyperspectral imaging enables more accurate and detailed analysis compared to multispectral imaging.



Conceptual illustration of Hyper-spectral (left) vs multi-spectral (right) images.



Examples of Hyper-spectral (left) vs multi-spectral (right) satellite images from the same location.

Types Of Platforms

RS sensors can be mounted on various types of platforms, which are vehicles or structures that carry the sensors to capture data about the Earth's surface. A platform can be stationary or mobile, depending on the sensor's needs. Satellites are spaceborne platforms that orbit the Earth, offering large-scale coverage for monitoring global phenomena over time. Airplanes, or manned aircraft, operate at lower altitudes, providing higher-resolution imagery over specific regions. Drones, or unmanned aerial vehicles (UAVs), are more flexible and can fly at even lower altitudes, making them ideal for detailed, localized mapping and monitoring tasks. Balloons, either tethered or free-floating, are cost-effective for atmospheric or environmental studies and can hover in place or travel short distances while collecting data. Each platform serves different purposes based on the desired scale, resolution, and area of interest.

Platform	**Spatial Resolution**	**Temporal Resolution**	**Coverage**
Satellite	High to low*	Consistent data over extended period, Ranges from daily to monthly; some satellites provide multiple observations per day	In many instances global or continental
Aerial	High to very high*	On-demand, but sometimes consistent, In the latter case frequency depending on flight schedules.	Variable, often in the order of 100 to 1000s of km^2
Drone	Very high*	On-demand, but sometimes consistent	Variable, often in the order of 10s to 100s of km^2
Balloon	High*	Variable, can provide long-duration observations of the same area	Variable, often in the order of 10s to 100s of km^2

Low Spatial Resolution: > 1 kilometer (1000 meters) per pixel.

Medium Spatial Resolution: 10 to 1000 meters per pixel.

High Spatial Resolution: 1 to 10 meters per pixel.

Very High Spatial Resolution: < 1 meter per pixel.

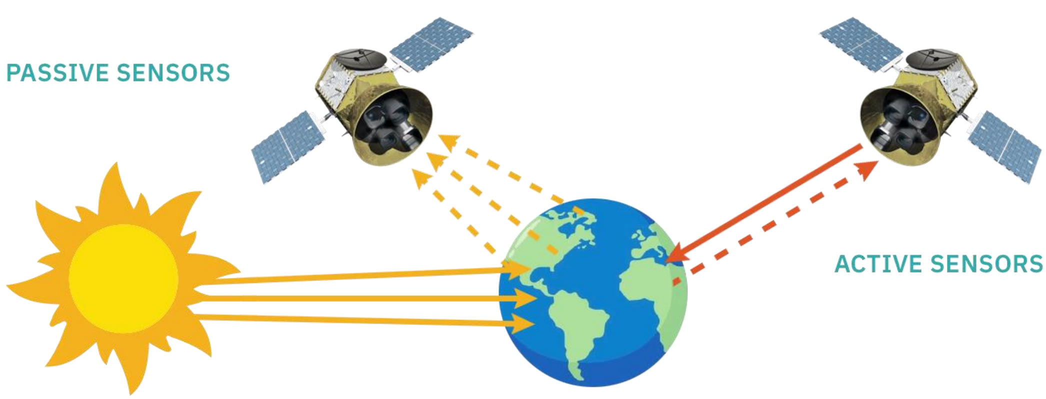
Types of Sensors

RS platforms, such as satellites, aircraft, drones, or balloons, can host various types of sensors that collect various types of data about the Earth's surface and atmosphere. These sensors can be classified into two main categories: passive and active.

Passive sensors detect natural energy that is emitted or reflected by the observed objects. The most common example of a passive sensor is a camera, which captures visible light reflected from the Earth's surface. Other examples include radiometers and spectrometers, which measure thermal radiation and specific wavelengths of light, respectively.

Active sensors, on the other hand, emit their own energy and measure the amount of that energy reflected back from the target. This allows them to collect data regardless of natural light conditions. Examples of active sensors include LiDAR (Light Detection and Ranging), which uses laser pulses to create high-resolution maps of surface topography, and RADAR (Radio Detection and Ranging), which uses radio waves to measure distances and detect objects.

Note that the same platform (satellites, aircraft, drones, or balloons) can host multiple sensors at the same time.

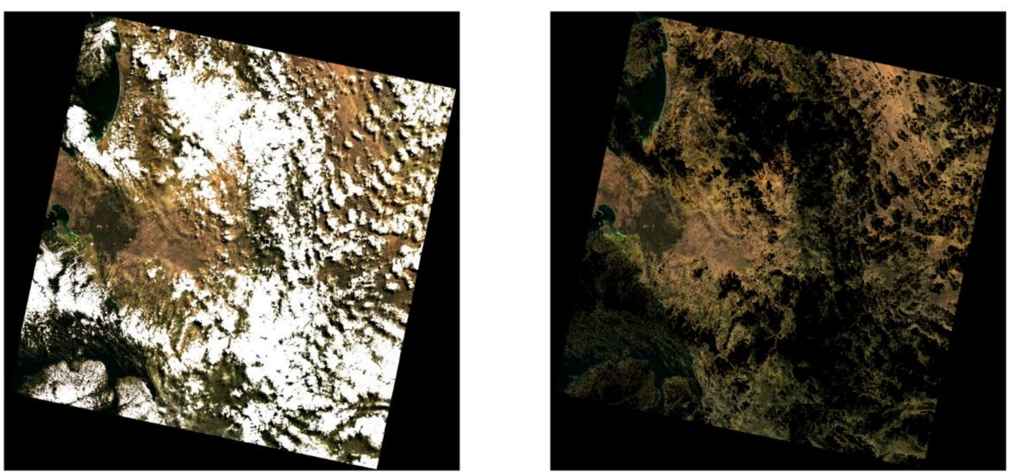


Passive vs. active sensors.

Preprocessing Remote Sensing Data

Regardless of the sensor type and technology used, raw remote sensing data often has very limited use. This is because raw RS data typically contains various distortions and noise that must be corrected before the data can be useful. To make raw RS data usable, it needs significant preprocessing. The following includes categories of different types of remote sensing data preprocessing and brief explanations for each:

- **Georeferencing**: Georeferencing is the process of assigning geographic coordinates to raw remote sensing images to align them with a known spatial reference system. This is done using ***ground control points (GCPs)***, which are specific, accurately known locations on the Earth's surface. GCPs are typically captured in-situ using high-precision GPS to record their exact geographic coordinates (latitude, longitude, and sometimes elevation). The selected locations should be easily identifiable in remote sensing imagery (e.g., road intersections, building corners, or landmarks). Georeferencing involves matching the positions of these GCPs in the raw imagery to their real-world coordinates, allowing a coordinate to be assigned to each pixel in the RS images.
- **Cloud Masking:** In optical imagery, clouds can obscure features on the Earth's surface. Cloud masking techniques identify and remove or flag cloudy areas, allowing users to focus only on clear-sky observations in the data.



Removing clouds from a Landsat image.

- **Atmospheric Correction:** This process corrects for the ***systematic distortions*** caused by atmospheric particles, such as PM10, and gases, like **water vapor or Ozone**, which can scatter or absorb light, impacting the quality and accuracy of the imagery. By removing these distortions, the data better reflects the true surface reflectance.
- **Radiometric Correction:** This process involves adjusting the brightness or intensity values in the imagery to correct for sensor noise and varying illumination conditions (caused by factors such as changing sunlight angles and shadows). It ensures that the pixel values accurately represent the reflectance of the Earth's surface, enabling more consistent and reliable data analysis.
- **Geometric Correction**: This step corrects distortions caused by the sensor's perspective, Earth's curvature, and topographic variations. The goal is to align the image so that it accurately matches real-world coordinates and shapes, ensuring that spatial features are in their correct positions.
- **Mosaicking**: When large areas need to be covered, multiple RS images are combined into a composite image. Mosaicking ensures that adjacent images align without visible seams or inconsistencies.
- **Calibration**: Calibration ensures that the raw data from sensors is consistent and comparable over time and across different sensors. It involves adjusting data values based on known reference values, often captured using on-ground in-situ measurements. These measurements help correct sensor biases and ensure data accuracy and reliability.

RS data is often processed at different levels to make it more useful for various applications. These levels of processing typically follow a standardized hierarchy, often referred to as Level 0 to Level 4. Here's a brief explanation of each level:

Level 0: Raw Data

This is the unprocessed data directly received from the remote sensing instrument. It includes all the original measurements, often in binary format, with no corrections or calibrations applied. This data is not typically useful for end-users without significant processing.

Level 1: Georeferenced Data

At this level, the raw data is processed to correct for sensor-specific distortions and is usually georeferenced, meaning it is tied to specific coordinates on the Earth's surface. This data is **calibrated** radiometrically and geometrically but still requires further processing for most practical applications. For drone, plane, and balloon imagery, georeferencing can be more accurate due to closer proximity to the ground and often more control over flight paths. Satellite images might have more complex geometric distortions due to their higher altitudes and wider coverage areas.

Level 2: Derived Products

This level involves further processing to derive specific geophysical variables from Level 1 data. These products are often corrected for atmospheric effects and other environmental factors. Examples include surface reflectance, sea surface temperature, or vegetation indices like NDVI (Normalized Difference Vegetation Index). The data at this level is more user-friendly and is often used directly in scientific research and applications. This step is crucial for both satellite and aerial platforms, though drones and balloons might have fewer atmospheric distortions to correct.

Level 3: Mapped Products

Level 3 data products are typically derived from Level 2 data and are mapped onto a uniform grid. These products often involve spatial and/or temporal aggregation, making them more convenient for analysis over larger areas or longer periods. Examples include global monthly averages of sea surface temperature or vegetation cover maps. This level of data is widely used in climate studies, environmental monitoring, and policy-making. Mapped products that aggregate data over larger areas or longer timescales are more commonly associated with satellite data due to their broader coverage. However, drones and planes can also contribute to Level 3 products through repeated flights over specific areas.

Level 4: Model Outputs and Analysis

This is the highest level of processed data, which often involves assimilation of Level 3 data into models and the generation of outputs that provide insight into specific phenomena. Level 4 products might include data assimilated into weather prediction models, carbon flux estimates, or predictive models for disease outbreaks. These products are often used for decision support in various fields, including agriculture, disaster management, and environmental conservation. Integrating data into models and generating advanced outputs is applicable to all platforms, though satellites often provide the large-scale data needed for global models, while drones and planes offer high-resolution inputs for more localized models.

In environmental data analytics, data scientists typically use Level 2, Level 3, and Level 4 RS data products.

Who Performs These Analyses

Typically, the organization that owns or operates the sensors (satellites, drones, planes, or balloons) is responsible for the initial processing of data through these levels:

- **Satellite Owners/Operators**: Agencies like NASA, ESA, or commercial satellite companies usually handle the processing from Level 0 to Level 4. They produce and distribute various levels of data products for public or commercial use.
- **Drone, Plane, and Balloon Operators**: These might be research institutions, government agencies, or private companies. They perform the initial processing (Level 1 and Level 2) and often provide these products to end-users who might then carry out further analysis.

Data Latency and Recency

- **Latency**: The higher the processing level, the more time it typically takes to make the data available. Level 0 data can be available almost immediately after capture, but Level 4 data, which involves integration into models and extensive analysis, can take days, weeks, or even months to produce.
- **Recency**: For the most up-to-date data, end-users might need to process raw or minimally processed data themselves. For example, if immediate surface reflectance data is needed after a drone flight, the data scientist would start from Level 0 or Level 1 and perform the necessary corrections and analyses promptly.

Most Useful Satellite Missions in EDA

In environmental data analytics, there are several major satellite missions, primarily developed by NASA and the European Space Agency (ESA), that provide vast amounts of free data for various types of analysis. In many satellite missions, multiple complementary sensors are often placed on the same satellite, and their data are jointly analyzed to provide comprehensive and useful outputs. The following are some of the most famous and commonly used satellite missions for environmental data analytics, along with their primary use cases:

Landsat Program (Landsat 1-9):

- **Year Started**: 1972 (Landsat 1), ongoing with Landsat 9 launched in 2021.
- **Data Level and Latency**: Provided as Level 1 (georeferenced) and Level 2 (atmospherically corrected) products. The typical latency for Level 1 data is around 24-48 hours after acquisition.
- **Use Case**: Monitoring land use and land cover changes, deforestation, urban expansion, and agricultural analysis. Landsat data is widely used for multi-

decade environmental change studies due to its long history (since 1972) and moderate-resolution imagery.

Sentinel-1 (Copernicus Program):

- **Year Started**: 2014 (Sentinel-1A), 2016 (Sentinel-1B).
- **Data Level and Latency**: Provided as Level 1 and Level 2. Latency for Level 1 data is typically 1-3 hours for emergency applications, and within 24 hours for other purposes.
- **Use Case**: Synthetic Aperture Radar (SAR) imagery for mapping flood extents, monitoring soil moisture, and detecting deforestation and land deformation. It is especially valuable for applications requiring all-weather, day-and-night data.

Sentinel-2 (Copernicus Program):

- **Year Started**: 2015 (Sentinel-2A), 2017 (Sentinel-2B).
- **Use Case**: High-resolution optical imagery for vegetation monitoring, water quality assessment, and land cover mapping. Sentinel-2's frequent revisit time (every 5 days) makes it ideal for tracking rapid environmental changes.
- **Data Level and Latency**: Provided as Level 1 and Level 2. Latency is typically 1-3 days for Level 1 products.

MODIS (Moderate Resolution Imaging Spectroradiometer) on **Terra** and **Aqua** satellites:

- **Year Started**: Terra (1999), Aqua (2002); both missions are still operational.
- **Data Level and Latency**: Commonly provided as Level 2 (georeferenced, calibrated data) and Level 3 (gridded products). Latency for Level 2 data is usually less than 24 hours.
- **Use Case**: Large-scale monitoring of global vegetation, sea surface temperature, atmospheric properties, and wildfire detection. MODIS provides continuous and extensive data for climate and environmental research.

SMAP (Soil Moisture Active Passive):

- **Year Started**: 2015, ongoing.
- **Data Level and Latency**: Data is provided at Level 2. Latency is generally around 2-3 days.
- **Use Case**: Measuring soil moisture levels and freeze-thaw state to support water cycle research, agricultural productivity, and drought monitoring.