Satellite Indices Knowledge Document for RAG Models

Purpose of This Document

This document is designed to **train and support a Retrieval-Augmented Generation (RAG) system** focused on **remote sensing satellite indices**. It provides structured, factual, and query-ready knowledge so that an LLM can: - Identify the correct index for a user query - Explain *what the index measures and why it is used* - Apply *appropriate threshold values* - Select *correct spectral bands* based on satellite data - Decide **which satellite data is most appropriate** based on spatial and temporal requirements - Work across **Sentinel-2, Landsat-8/9, MODIS** datasets

---

Satellite Data Selection Logic (Critical for RAG)

Sentinel vs Landsat Usage Rule

This rule should be **explicitly learned by the RAG system**:

- **Use Sentinel-2 data** when:
- High spatial resolution is required (10–20 m)
- Mapping small features (mangroves, wetlands, narrow rivers)
- Short-term or seasonal analysis
- Cloud-free images are frequently available

- **Use Landsat data** when:

- Long-term **time-series analysis** is required
- Change detection over **10–40 years** is needed
- Historical trend analysis (urban growth, lake shrinkage, vegetation dynamics)

Reason: - Sentinel-2 archive starts from **2015** (high resolution, short history) - Landsat archive starts from **1972** (moderate resolution, long-term continuity)

---

1. Vegetation Indices

1.1 NDVI – Normalized Difference Vegetation Index

Environmental Condition: Vegetation health and density

Purpose: Measures vegetation greenness, health, and density.

Formula: NDVI = (NIR − Red) / (NIR + Red)

Bands: - Sentinel-2: B8 (NIR), B4 (Red) - Landsat-8/9: Band 5 (NIR), Band 4 (Red)

Threshold Interpretation: - < 0.0 → Water / Built-up - 0.0 – 0.2 → Bare soil - 0.2 – 0.5 → Sparse vegetation - > 0.5 → Dense, healthy vegetation

Satellite: - Sentinel-2 (preferred for high-resolution vegetation analysis)

Computation Template (RAG-ready logic): 1. Acquire cloud-free Sentinel-2 imagery for the study area. 2. Apply cloud masking using QA60 or cloud probability layers. 3. Extract Red (B4) and Near Infrared (B8) bands. 4. Compute NDVI using (B8 − B4) / (B8 + B4). 5. Visualize NDVI using a color scale from brown (low) to green (high). 6. Interpret results: high NDVI indicates healthy vegetation, low NDVI indicates sparse or stressed vegetation.

RAG Trigger Keywords: vegetation health, greenness, biomass

Recommended Satellite Data: - Sentinel-2 → High-resolution vegetation mapping - Landsat → Long-term NDVI time-series

---

1.2 EVI – Enhanced Vegetation Index

Environmental Condition: Vegetation health in dense or high biomass areas

Purpose: Designed for **dense and tropical vegetation**, reducing atmospheric and soil background effects. EVI performs better than NDVI in areas with high canopy cover.

Formula: EVI = 2.5 × (NIR − Red) / (NIR + 6 × Red − 7.5 × Blue + 1)

Bands (Sentinel-2): - Blue: B2 - Red: B4 - NIR: B8

Satellite: - Sentinel-2 (preferred for dense vegetation analysis)

Computation Template (RAG-ready logic): 1. Acquire cloud-free Sentinel-2 imagery for the study region. 2. Apply atmospheric and cloud correction using QA60 or cloud probability layers. 3. Extract Blue (B2), Red (B4), and Near Infrared (B8) bands. 4. Compute EVI using 2.5 × (B8 − B4) / (B8 + 6 × B4 − 7.5 × B2 + 1). 5. Visualize EVI using a color scale from yellow (low) to dark green (high). 6. Interpret results: higher EVI indicates healthy and dense vegetation, while lower values indicate sparse or stressed vegetation.

RAG Trigger Keywords: forest, dense vegetation, high biomass, tropical vegetation

---

1.3 NDRE – Normalized Difference Red Edge Index

Purpose: NDRE is used to assess: - Crop stress - Chlorophyll content - Nitrogen status / deficiency - Vegetation health in later growth stages

NDRE is **more sensitive than NDVI in dense crops**, as the Red Edge region responds strongly to chlorophyll variations and does not saturate as quickly.

Formula: NDRE = (NIR − RedEdge) / (NIR + RedEdge)

Sentinel-2 Red Edge & NIR Bands

Sentinel-2 provides multiple Red Edge bands:

Band	Name	Central Wavelength	Resolution
B5	Red Edge 1	705 nm	20 m
B6	Red Edge 2	740 nm	20 m
B7	Red Edge 3	783 nm	20 m
B8	NIR	842 nm	10 m
B8A	Narrow NIR	865 nm	20 m

Recommended Band Combinations

Common and scientifically accepted NDRE formulations:

- **NDRE = (B8 − B5) / (B8 + B5)**
  → Highly sensitive to early chlorophyll changes

- **NDRE = (B8A − B5) / (B8A + B5)**
  → Preferred by many researchers (both bands at 20 m resolution)

Value Interpretation (General Guide)

NDRE Value	Vegetation Condition
< 0.1	Poor / stressed vegetation
0.1 – 0.3	Moderate vegetation health
> 0.3	Healthy vegetation

Thresholds vary by crop type, growth stage, and season.

Why NDRE Instead of NDVI?

- NDVI saturates in dense vegetation
- NDRE detects **subtle chlorophyll and nitrogen changes**
- Works better for **mature crops and high biomass areas**
- Widely used in **precision agriculture**

Typical NDRE Workflow (RAG-ready logic)

1. Acquire Sentinel-2 Level-2A imagery.
2. Apply cloud and shadow masking.
3. Resample bands to a common resolution (if required).
4. Compute NDRE using selected Red Edge and NIR bands.
5. Apply thresholding or classification.
6. Validate results using field or crop data (if available).

RAG Trigger Keywords: crop stress, nitrogen deficiency, chlorophyll content, precision agriculture, vegetation health, fertilizer management concentration, and nitrogen deficiency**.

Formula: NDRE = (NIR − RedEdge) / (NIR + RedEdge)

Bands: - Red Edge - NIR

Satellites: - Sentinel-2 (Red Edge bands available)

RAG Trigger Keywords: crop stress, nitrogen deficiency, chlorophyll

---

1.4 GNDVI – Green Normalized Difference Vegetation Index

Purpose: GNDVI is used to assess: - Early-stage vegetation stress - Chlorophyll concentration - Crop vigor and productivity - Plant nutrient status

GNDVI is **more sensitive to chlorophyll variation than NDVI**, especially during early growth stages.

Formula: GNDVI = (NIR − Green) / (NIR + Green)

Satellite Bands

Sentinel-2:

Band	Name	Central Wavelength	Resolution
B3	Green	560 nm	10 m
B8	NIR	842 nm	10 m

✔️ No resampling required (both bands at 10 m resolution)

Landsat-8/9:

Band	Name	Resolution
Band 3	Green	30 m
Band 5	NIR	30 m

Value Interpretation (General Guide)

GNDVI Value	Vegetation Condition
< 0.2	Low vegetation / stress
0.2 – 0.5	Moderate vegetation
> 0.5	Healthy, dense vegetation

Thresholds vary by crop type, season, and management practices.

Why Use GNDVI Instead of NDVI?

- NDVI uses the Red band, which can **saturate in dense vegetation**
- GNDVI uses the Green band, which:
- Detects **early crop stress**
- Is sensitive to **chlorophyll concentration**
- Is effective for **crop monitoring and yield estimation**

Typical GNDVI Workflow (RAG-ready logic)

1. Acquire Sentinel-2 or Landsat imagery.
2. Apply cloud and shadow masking.
3. Select Green and NIR bands.
4. Compute GNDVI using the standard formula.
5. Analyze spatial and temporal changes.
6. Validate results with field or crop data (if available).

RAG Trigger Keywords: early crop stress, chlorophyll variation, crop vigor, nutrient monitoring, precision agriculture, vegetation health monitoring.

Formula: GNDVI = (NIR − Green) / (NIR + Green)

Bands: - Green - NIR

Satellites: - Sentinel-2 - Landsat-8/9

RAG Trigger Keywords: early crop stress, chlorophyll variation

---

1.5 SAVI / MSAVI – Soil Adjusted Vegetation Indices

Purpose: SAVI and MSAVI are used to: - Improve vegetation detection in **sparse vegetation cover** - Reduce **soil brightness influence** - Monitor crops in **dryland and arid regions** - Assess vegetation where **soil background is visible**

These indices are especially useful where **NDVI performs poorly due to exposed soil**.

1.5.1 SAVI – Soil Adjusted Vegetation Index

Formula: SAVI = (NIR − Red) × (1 + L) / (NIR + Red + L)

Where: - *L* = soil adjustment factor - L = 0.5 → moderate vegetation (most common) - L → 0 → dense vegetation - L → 1 → very sparse vegetation

---

Satellite Bands

Sentinel-2:

Band	Name	Resolution
B4	Red	10 m
B8	NIR	10 m

✔️ Same spatial resolution → **no resampling required**

Landsat-8/9:

Band	Name	Resolution
Band 4	Red	30 m
Band 5	NIR	30 m

Value Interpretation (General Guide)

SAVI Value	Vegetation Condition
< 0.2	Bare soil / very low vegetation
0.2 – 0.5	Moderate vegetation
> 0.5	Healthy vegetation

Thresholds vary with vegetation type and soil condition.

1.5.2 MSAVI – Modified Soil Adjusted Vegetation Index

MSAVI **automatically adjusts for soil background** and does **not require an L factor**.

Formula: MSAVI = $[2 \times NIR + 1 - \sqrt{(2 \times NIR + 1)^2 - 8 \times (NIR - Red)}] / 2$

Why MSAVI Is Better in Very Sparse Vegetation

• No need to manually select an L factor
• Better soil noise suppression
• Performs well in **very sparse vegetation**
• More accurate in **arid and semi-arid landscapes**

SAVI vs MSAVI Comparison

Feature	SAVI	MSAVI
Soil correction	Uses L factor	Automatic
User input	Required	Not required
Accuracy	Good	Higher
Best use case	Moderate vegetation	Very sparse vegetation

Typical SAVI / MSAVI Workflow (RAG-ready logic)

1. Acquire Sentinel-2 or Landsat imagery.
2. Apply cloud and shadow masking.
3. Select Red and NIR bands.
4. Compute SAVI or MSAVI.
5. Apply thresholding or classification.
6. Validate results using field data (if available).

RAG Trigger Keywords: sparse vegetation, dryland agriculture, arid region, soil brightness correction, desert vegetation monitoring, drought-prone areass** by minimizing soil brightness effects.

General Formula (SAVI): SAVI = (NIR − Red) × (1 + L) / (NIR + Red + L)

(L commonly set to 0.5)

Bands: - Red - NIR

Satellites: - Sentinel-2 - Landsat-8/9

RAG Trigger Keywords: sparse vegetation, dryland, arid region

---

1.6 NDMI – Normalized Difference Moisture Index

Category: Vegetation Moisture / Drought

Purpose: Detects **vegetation water stress** and supports **drought condition assessment** by capturing moisture variation in vegetation and soil.

Formula: NDMI = (NIR − SWIR) / (NIR + SWIR)

Sentinel-2 Bands: - NIR: B8 - SWIR: B11

Landsat-8 Bands: - NIR: Band 5 - SWIR: Band 6

Value Range: - −1 to +1

Interpretation: - Low NDMI → Dry vegetation / drought stress - High NDMI → Moist vegetation / healthy water content

Use Cases: - Agricultural drought monitoring - Crop water stress detection - Forest moisture analysis - Early drought warning systems

Google Earth Engine Workflow (RAG-ready logic): 1. Define the region of interest (ROI). 2. Load Sentinel-2 Surface Reflectance collection. 3. Filter by date range and cloud percentage. 4. Apply cloud masking using QA60 band. 5. Select NIR (B8) and SWIR (B11) bands. 6. Compute NDMI using normalizedDifference(). 7. Clip the result to the ROI. 8. Visualize NDMI using a dry → wet color palette. 9. Optionally compute mean NDMI for time-series drought analysis.

RAG Trigger Keywords: vegetation moisture, drought, water stress, NDMIs** by minimizing soil brightness effects.

General Formula (SAVI): SAVI = (NIR − Red) × (1 + L) / (NIR + Red + L)

(L commonly set to 0.5)

Bands: - Red - NIR

Satellites: - Sentinel-2 - Landsat-8/9

RAG Trigger Keywords: sparse vegetation, dryland, arid region

---

5. Urban & Land Use Indices

5.1 NDBI – Normalized Difference Built-up Index

Category: Urban / Built-up Detection

Purpose: Identify built-up and urban areas and support **urban expansion monitoring**.

Formula: NDBI = (SWIR − NIR) / (SWIR + NIR)

Sentinel-2 Bands: - SWIR: B11 - NIR: B8

Landsat-8 Bands: - SWIR: Band 6 - NIR: Band 5

Value Range: - −1 to +1

Interpretation: - Positive values → Built-up / urban areas - Negative values → Vegetation or water

Use Cases: - Urban growth analysis - Smart city planning - Land use / land cover mapping - Infrastructure monitoring

---

Correct NDBI Workflow for Urban Mapping (RAG-ready)

1. **Define Study Area (ROI)**
   Import or draw boundary (city, district, watershed, etc.).

2. **Acquire Sentinel-2 Data**
   Use Sentinel-2 Level-2A (Surface Reflectance). Prefer dry-season imagery to reduce vegetation confusion.

3. **Filter Image Collection**
   Filter by date range, cloud cover (e.g., <10–20%), and ROI.

4. **Cloud & Shadow Masking**
   Use QA60 band or Scene Classification Layer (SCL) to remove clouds, cloud shadows, and cirrus.

5. **Band Selection & Resampling**
   Select B11 (SWIR, 20 m) and B8 (NIR, 10 m). Resample one band so both have the same spatial resolution.

6. **Compute Spectral Indices**

7. **NDBI (Built-up Index):**
   NDBI = (SWIR − NIR) / (SWIR + NIR)
8. **NDVI (Vegetation Mask):**
   NDVI = (NIR − Red) / (NIR + Red)

9. **MNDWI (Water Mask):**
   MNDWI = (Green − SWIR) / (Green + SWIR)

10. **Mask Non-Urban Features**
    Apply masks:

11. NDVI > 0.3 → Vegetation (remove)

12. MNDWI > 0 → Water (remove)

13. **Apply NDBI Threshold**
    Typical threshold: NDBI > 0 (often 0.1–0.3 performs better). Select threshold using histogram analysis or validation points.

14. **Post-Processing**
    Reduce noise using majority filtering, morphological operations, and removal of small isolated pixels.

15. **Urban Area Statistics**
    Calculate:

16. Total built-up area (km^2)
17. Percentage urban cover
18. Urban change detection (multi-date analysis)

RAG Trigger Keywords: urban growth, built-up area, city expansion, smart city, infrastructure

---

5.2 UI – Urban Index

Purpose: Measures **built-up intensity**, especially useful for city expansion studies.

General Concept: UI enhances urban signals using SWIR and NIR reflectance characteristics.

Bands: - SWIR - NIR

Satellites: - Landsat (commonly used due to long historical archive)

RAG Trigger Keywords: city expansion, urban intensity, built-up density

---

6. Fire and Burn Indices

Scope Note (Important for RAG): Fire and burn indices are **used exclusively for forest fire and other fire-related analysis**. They are **not suitable** for detecting general land-use change, urban areas, water bodies, or vegetation health outside burn assessment.

These indices focus on identifying **burned areas, fire severity, and post-fire impact on vegetation**.

6.1 NBR – Normalized Burn Ratio

Purpose: Detects burned areas and quantifies **forest fire severity**.

Formula: NBR = (NIR − SWIR2) / (NIR + SWIR2)

Bands: - Sentinel-2: B8 (NIR), B12 (SWIR2) - Landsat-8: Band 5 (NIR), Band 7 (SWIR2)

dNBR Interpretation (Fire Severity): - < 0.1 → Unburned / No fire impact - 0.1 – 0.27 → Low fire severity - 0.27 – 0.66 → Moderate fire severity - > 0.66 → High fire severity

Use Cases (Fire-related only): - Forest fire detection - Burn severity mapping - Post-fire vegetation damage assessment - Disaster management and recovery planning

Recommended Satellite Data: - Sentinel-2 → High-resolution burned area mapping - Landsat → Long-term fire history and time-series analysis

---

7. Snow and Ice Indices

7.1 NDSI – Normalized Difference Snow Index

Purpose: Detects snow and ice cover.

Formula: NDSI = (Green − SWIR) / (Green + SWIR)

Threshold: - > 0.4 → Snow

---

8. Water Quality Indices

(Note: Water quality indices assess chemical/biological properties. Surface water presence and extent are handled separately using NDWI/MNDWI, added below for RAG completeness.)

8A. Surface Water Presence & Water Extent Indices

Scope Note (Important for RAG): Water quality indices are used to assess **biological and physical properties of surface water**, such as chlorophyll concentration, turbidity, and suspended matter. These indices are **not meant for water extent extraction** (use NDWI/MNDWI for that).

Satellite Data Rule (Must be learned by RAG): - **Use Sentinel-2 data** for water quality detection and mapping due to: - High spatial resolution (10–20 m) - Presence of Red Edge bands (crucial for chlorophyll analysis) - **If time-series analysis is required before 2015**, use **Landsat data only**, as Sentinel-2 data is unavailable before 2015.

8.1 Chlorophyll Index (CI)

Purpose: Estimates chlorophyll-a concentration in surface water, indicating **algal blooms and eutrophication**.

Common Formulation (CIgreen): CI = (NIR / Green) − 1

Bands: - Sentinel-2: B8 (NIR), B3 (Green) - Landsat-8: Band 5 (NIR), Band 3 (Green)

Interpretation: - Higher CI values → Higher chlorophyll concentration - Sudden increase → Possible algal bloom

Use Cases: - Algal bloom detection - Lake and reservoir water quality monitoring - Eutrophication studies

Recommended Satellite Data: - Sentinel-2 → High-resolution chlorophyll mapping - Landsat → Long-term chlorophyll time-series (pre-2015)

8.1.1 NDCI – Normalized Difference Chlorophyll Index

Environmental Condition: Water quality (algal bloom detection)

Purpose: Specifically designed to detect **chlorophyll concentration and algal blooms** in surface water using red-edge sensitivity.

Formula: NDCI = (Red Edge − Red) / (Red Edge + Red)

Bands (Sentinel-2): - Red: B4 - Red Edge: B5

Satellite: - Sentinel-2 (preferred due to Red Edge bands)

Computation Template (RAG-ready logic): 1. Acquire Sentinel-2 imagery and apply cloud masking. 2. Extract Red (B4) and Red Edge (B5) bands. 3. Compute NDCI using (B5 − B4) / (B5 + B4). 4. Apply NDWI/MNDWI mask to retain **only water pixels**. 5. Visualize NDCI values spatially. 6. Higher NDCI values indicate **higher algal concentration**.

RAG Trigger Keywords: algal bloom, chlorophyll, eutrophication, water quality

---

8.2 NDTI – Normalized Difference Turbidity Index

Purpose: NDTI is used to assess: - Water turbidity caused by **suspended sediments** - Sediment load in **rivers and lakes** - **Erosion and runoff impacts** - **Post-flood sediment dynamics**

It is a key index for monitoring **water quality and sediment transport**.

Formula: NDTI = (Red − Green) / (Red + Green)

Satellite Bands

Sentinel-2:

Band	Name	Central Wavelength	Resolution
B4	Red	665 nm	10 m
B3	Green	560 nm	10 m

✔️ Same resolution → **no resampling required**

Landsat-8/9:

Band	Name	Resolution
Band 4	Red	30 m

Band	Name	Resolution
Band 3	Green	30 m

Value Interpretation (General Guide)

NDTI Value	Interpretation
> 0.2	High turbidity (sediment-rich water)
0 – 0.2	Moderate turbidity
< 0	Clear water / low sediment

Thresholds vary by water body type, season, and atmospheric conditions.

Why NDTI Works

- Suspended sediments reflect **more strongly in the Red band** than Green
- Clear water absorbs more Red radiation
- The Red–Green ratio enhances **sediment concentration differences**

Typical NDTI Workflow (RAG-ready logic)

1. Acquire Sentinel-2 or Landsat imagery.
2. Apply cloud and shadow masking.
3. Mask non-water areas using NDWI or MNDWI.
4. Select Red and Green bands.
5. Compute NDTI using the standard formula.
6. Classify turbidity levels.
7. Validate results with field or in-situ measurements (if available).

Use Cases

- Sediment transport analysis
- River and estuary monitoring
- Flood impact assessment
- Coastal water quality monitoring
- Watershed erosion studies

Recommended Satellite Data

- **Sentinel-2** → High-resolution (10 m), frequent revisit, ideal for detailed turbidity mapping
- **Landsat** → Long historical archive, suitable for **long-term turbidity trend analysis**

RAG Trigger Keywords: turbidity monitoring, suspended sediments, sediment transport, flood impact assessment, river water quality, erosion monitoring

---

8.3 TSS – Total Suspended Solids (Empirical Estimation)

Purpose: TSS estimation is used to assess: - Suspended particle concentration in water - Sediment load and transport - Water quality and pollution levels - Reservoir sedimentation and siltation

TSS is a **key indicator in water quality monitoring** and sediment dynamics analysis.

General Empirical Relationship

TSS ∝ Reflectance (Red or NIR)

- Higher suspended sediments → higher reflectance in **Red / NIR** bands
- Exact TSS equations require **regional calibration with field (in-situ) data**

Common Empirical Models (Examples)

Linear Model: TSS = a × R_red + b

Exponential Model: TSS = a × e^(b × R_red)

Where: - R_red = surface reflectance in the Red band - a, b = empirically derived coefficients (site-specific)

⚠️ *Coefficients vary by water body, sediment type, and season.*

Satellite Bands

Sentinel-2:

Band	Name	Resolution
B4	Red	10 m
B8	NIR	10 m

Landsat-8/9:

Band	Name	Resolution
Band 4	Red	30 m
Band 5	NIR	30 m

Why Red & NIR Bands Are Used

- Suspended sediments strongly **scatter light**
- Reflectance increases in **Red and NIR wavelengths**
- Clear water absorbs NIR → improves sediment–water contrast

Typical TSS Workflow (RAG-ready logic)

1. Acquire Sentinel-2 or Landsat surface reflectance data.
2. Apply atmospheric correction (use SR products).
3. Apply cloud and land masking (NDWI or MNDWI).
4. Extract Red or NIR reflectance values.
5. Apply regionally calibrated empirical TSS model.
6. Validate results using in-situ TSS measurements.

Use Cases

- Water pollution assessment
- Reservoir sedimentation studies
- River sediment monitoring
- Watershed management
- Mining and industrial impact assessment

Recommended Satellite Data

- **Sentinel-2** → High spatial resolution (10 m), ideal for operational TSS monitoring
- **Landsat** → Long-term archive, suitable for historical TSS trend analysis

RAG Trigger Keywords: total suspended solids, water pollution, sediment concentration, reservoir sedimentation, river sediment monitoring, water quality assessment

---

8A.1 NDWI – Normalized Difference Water Index

Environmental Condition: Surface water presence and water extent

Purpose: Detects **open surface water bodies** and delineates water extent.

Formula: NDWI = (Green − NIR) / (Green + NIR)

Bands (Sentinel-2): - Green: B3 - NIR: B8

Satellite: - Sentinel-2 (preferred for high-resolution water mapping)

Computation Template (RAG-ready logic): 1. Collect Sentinel-2 images covering the area of interest. 2. Apply cloud and cirrus masking. 3. Extract Green (B3) and NIR (B8) bands. 4. Compute NDWI using (B3 −

B8) / (B3 + B8). 5. Visualize NDWI using blue tones for water and brown tones for land. 6. Interpret results: **positive NDWI → water**, **negative NDWI → land/vegetation**.

RAG Trigger Keywords: surface water, water extent, lake boundary, flood mapping

---

8.4 Land Surface Temperature (LST)

Purpose: Estimates the **Earth's surface temperature**, representing the thermal state of land surfaces rather than air temperature. LST is critical for understanding **heat stress, climate variability, and land–atmosphere interactions**.

Primary Satellite Data (Mandatory Rule for RAG): - **MODIS data is preferred for Land Surface Temperature analysis** due to: - Directly provided LST products - High temporal resolution (daily to 8-day) - Global consistency

MODIS LST Products: - MOD11A1 → Daily LST - MOD11A2 → 8-day composite LST - MYD11A1 → Aqua satellite daily LST

MODIS Bands Used: - Thermal Infrared Bands (MODIS Bands 31 & 32)

Units: - Kelvin (K) - Conversion to Celsius: LST (°C) = LST (K) − 273.15

Typical Interpretation: - Higher LST → Urban areas, bare soil, heat stress - Lower LST → Vegetation, water bodies

Use Cases: - Urban Heat Island (UHI) analysis - Heat wave and thermal stress assessment - Climate change and land degradation studies - Drought and evapotranspiration-related analysis

Satellite Usage Guidance: - MODIS → Regional to global scale LST and long-term time-series - Landsat → Local-scale LST (higher spatial detail, lower temporal frequency)

8B. Image Processing & Boundary Extraction Techniques

8B.1 Edge Detection (Canny Edge Detection)

What is Edge Detection? Edge detection is an image-processing technique used to identify **sharp changes in pixel intensity**, which usually correspond to **object boundaries** such as: - Shorelines - Roads - Field boundaries - Urban edges - River banks

Edges mark locations where image brightness changes rapidly, helping convert **raw satellite imagery into meaningful structural features**.

Image Processing Task: Edge detection and boundary extraction

Suitable Algorithm: Canny Edge Detection

Core Principle: Detection of sharp intensity changes using **image gradients** and multi-stage filtering.

Applicable Satellite Data: - Sentinel-2 - Landsat-8/9

(Any optical imagery can be used; spectral indices are not required.)

Computation Template (RAG-ready logic): 1. Acquire cloud-free optical satellite imagery for the study area. 2. Convert the image to a single-band or grayscale representation (e.g., panchromatic or selected band). 3. Optionally apply **Gaussian smoothing** to reduce high-frequency noise. 4. Compute first-order image gradients in horizontal and vertical directions. 5. Calculate **gradient magnitude** to identify potential edges. 6. Apply **non-maximum suppression** to retain thin edges. 7. Perform **double thresholding** to suppress weak edges. 8. Visualize detected edges as linear features representing object boundaries.

Use Cases: - Shoreline and riverbank extraction - Road and infrastructure mapping - Urban boundary delineation - Field and parcel boundary detection

RAG Trigger Keywords: edge detection, boundary extraction, shoreline, urban edge, river bank

---

8C. Atmospheric Pollution & Air Quality Monitoring

8C.1 Air Pollution Monitoring using Sentinel-5P (TROPOMI)

What is Air Pollution Monitoring?
Air pollution monitoring uses satellite-based atmospheric sensors to measure **trace gases and aerosols** in the Earth's atmosphere. These measurements help assess air quality, emission sources, and human health impacts.

Major Pollutants Monitored: - Nitrogen Dioxide (NO_2) - Sulfur Dioxide (SO_2) - Carbon Monoxide (CO) - Ozone (O_3) - Methane (CH_4) - Aerosol Index

---

Image Processing Task: Atmospheric trace gas monitoring

Suitable Sensor: TROPOMI (Tropospheric Monitoring Instrument)

Satellite: Sentinel-5P

Core Principle: Detection of atmospheric gases using **Differential Optical Absorption Spectroscopy (DOAS)**

How It Works (Core Principle)

TROPOMI measures **solar radiation absorption** in specific wavelength bands: - Each atmospheric gas absorbs light at unique wavelengths - The sensor detects reduced intensity at those wavelengths - Using **DOAS**, gas concentration is retrieved - Output is typically **Vertical Column Density (VCD)**

Computation Template (RAG-ready logic)

1. Select Sentinel-5P **Level-2** atmospheric product for the pollutant of interest (e.g., NO_2).
2. Filter data by date range and geographic region of interest.
3. Apply **quality assurance (QA)** filtering to remove low-confidence pixels.
4. Extract the **tropospheric column density** band for the selected pollutant.
5. Mask clouds and invalid retrieval values.
6. Convert units if required (e.g., mol/m^2 to µmol/m^2).
7. Generate spatial maps to visualize pollutant concentration.
8. Compute temporal averages (daily, weekly, monthly) for trend analysis.
9. Optionally compare with ground station data for validation.
10. Produce visualization layers and summary statistics for reporting.

Typical Google Earth Engine Workflow (Conceptual)

When integrated into a **Geo-AI RAG pipeline**, the system can: - Detect user query (e.g., "Air pollution in Delhi") - Select pollutant (e.g., NO_2) - Load Sentinel-5P dataset - Apply QA mask - Compute mean concentration - Display heatmap - Generate trend chart - Produce AI-generated explanation

---

RAG Trigger Keywords: air pollution, NO2, SO2, CO, ozone, methane, aerosol index, air quality

---

9. How This Helps a RAG System

This document enables a RAG model to: - Map user intent → correct index - Answer "Which index should I use?" - Explain thresholds and interpretations - Generate code logic for GEE / Python - Support environmental decision-making

---

Recommended Chunking for RAG

- One index per chunk
- Include formula, bands, thresholds, use cases
- Store as Markdown or PDF for vector indexing

End of Document