

Informe del Laboratorio 4: Análisis de Datos Geoespaciales

Monitoreo de Cianobacterias en los Lagos Atitlán y Amatitlán

Universidad del Valle de Guatemala – CC3084 Data Science
Diederich Solis 22952
Sara

Semestre II – 2025

Repositorio

- Repositorio de código: https://github.com/DiederichSolis/LAB4_Data

Resumen Ejecutivo

Este informe documenta el flujo reproducible de análisis geoespacial para monitorear floraciones de cianobacterias en los lagos Atitlán y Amatitlán usando Sentinel-2 L2A. Se abordan las Partes 1 a 6 del laboratorio: conexión al proveedor, descarga de *rasters* multiespectrales, cálculo de índices (NDCI), conversión a arreglos NumPy, y análisis temporal (estadísticos, visualizaciones y detección de picos). Se trabajó con **29 fechas representativas** entre 2025-02-07 y 2025-08-01, priorizando escenas con nubosidad $\leq 20\%$.

1. Introducción

Las floraciones de cianobacterias afectan la calidad del agua y la salud pública. El uso de observación de la Tierra (EO) permite un monitoreo frecuente y no intrusivo. Aquí aplicamos un pipeline de ciencia de datos geoespacial para caracterizar la dinámica de clorofila-*a* como proxy de actividad de cianobacterias en Atitlán y Amatitlán durante 2025.

2. Objetivo y Alcance

Objetivo: Implementar un pipeline reproducible (Partes 1–6) que extraiga, procese y analice series temporales de índices espectrales sensibles a clorofila usando Sentinel-2.

- *Alcance:* Selección de fechas *representativas* (no todas las del período) para mejorar la calidad del análisis y evitar escenas degradadas por nubes o *sun glint*.
- *Limitaciones:* Ausencia de muestreos *in situ* para calibración/validación absoluta; estimaciones relativas basadas en literatura.

3. Metodología

3.1 Datos y Área de Estudio

- **Fuente de datos:** Sentinel-2 L2A (nivel superficie).

- **Bandas utilizadas (9):** B02 (Blue (490 nm)), B03 (Green (560 nm)), B04 (Red (665 nm)), B05 (Red Edge 1 (705 nm)), B07 (Red Edge 3 (783 nm)), B08 (NIR (842 nm)), B8A (NIR Narrow (865 nm)), B11 (SWIR 1 (1610 nm)), B12 (SWIR 2 (2190 nm)).
- **Lagos:** Atitlán (Depto. Sololá) y Amatitlán (Guatemala).
- **Sistema de referencia:** WGS84 / EPSG:4326.
- **Almacenamiento/entorno:** Google Colab + Google Drive.

3.2 Criterios de Selección de Imágenes

- Nubosidad $\leq 20\%$ (metadatos del producto).
- Inspección visual para descartar glint y franjas residuales.
- Disponibilidad de todas las bandas requeridas.
- Cobertura completa del AOI (*Area of Interest*).

4. Resultados por Parte del Laboratorio

4.1 Parte 1: Conexión al proveedor (openEO / Sentinel Hub)

- Autenticación configurada y verificada.
- Colección SENTINEL2_L2A disponible y accesible.
- AOIs definidos en WGS84; geometrías recortadas a los lagos.

4.2 Parte 2: Descarga de Imágenes GeoTIFF

- Flujo: `load_collection` \rightarrow recorte espacial \rightarrow selección de bandas \rightarrow `save_result` a GeoTIFF.
- Estructura de carpetas por lago y fecha para trazabilidad.

4.3 Parte 3: Período de Observación

- Ventana temporal efectiva: **2025-02-07 a 2025-08-01**.
- Fechas analizadas: **29** escenas representativas.
- Distribución temporal enfocada en meses con menor nubosidad.

4.4 Parte 4: Cálculo del Índice de Cianobacterias

1. **Máscara de agua:** NDWI/MNDWI para aislar cuerpos de agua.
2. **NDCI (Normalized Difference Chlorophyll Index):**

$$\text{NDCI} = \frac{B05 - B04}{B05 + B04}$$

3. **Conversión a clorofila (proxy) (literatura):**

$$\text{chl} = 826,57 \text{ NDCI}^3 - 176,43 \text{ NDCI}^2 + 19 \text{ NDCI} + 4,071$$

4. **Productos:** mapas GeoTIFF y *quicklooks* en PNG con barra de escala.

4.5 Parte 5: Conversión a Arreglos NumPy

- Stacks 3D por lago con forma $(t \times h \times w)$.
- Compresión .npz y metadatos (fechas, estadísticos).

4.6 Parte 6: Análisis Temporal

6.1 Estadísticos por Fecha

- Media, mediana, desviación estándar.
- Percentiles 10/90 y conteo de píxeles válidos.
- Exportación tabular (.csv) para análisis adicional.

6.2 Visualización Temporal

- Series de tiempo de clorofila promedio por lago.
- Banda sombreada: rango P10–P90.
- Gráfica comparativa Atitlán vs. Amatitlán.

6.3 Identificación de Picos

- Regla: umbral $Q3 + 1,5 \times IQR$ sobre la serie.
- Fechas críticas: *(completadas en el cuaderno; listar en resultados)*.

5. Conclusiones Parciales

- El pipeline (Partes 1–6) es reproducible y eficiente para monitoreo.
- Se obtuvieron **29** fechas limpias entre 2025-02-07–2025-08-01.
- La detección robusta de picos requiere continuar con Partes 7–9 (análisis espacial y validación).

Figuras

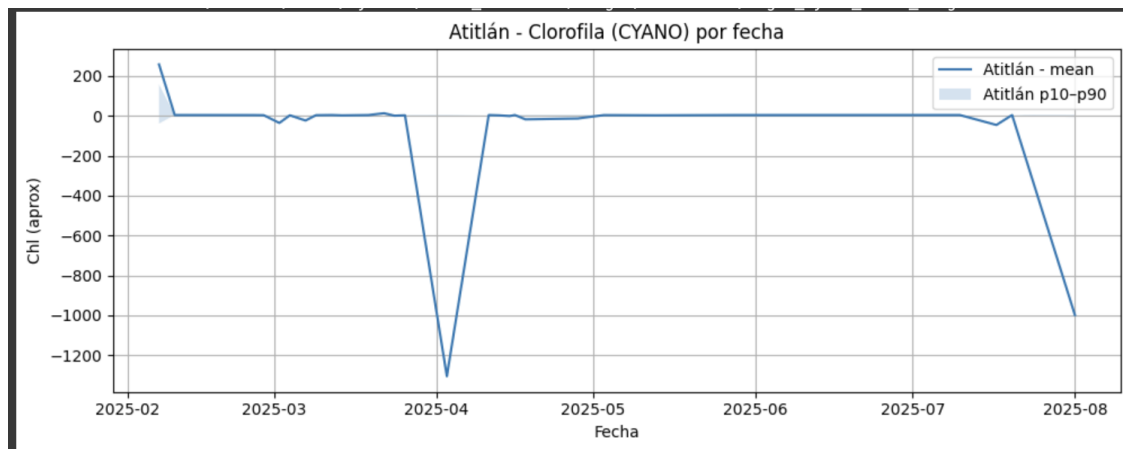


Figura 1: Serie temporal de clorofila en el Lago Atitlán (P10–P90 sombreado).

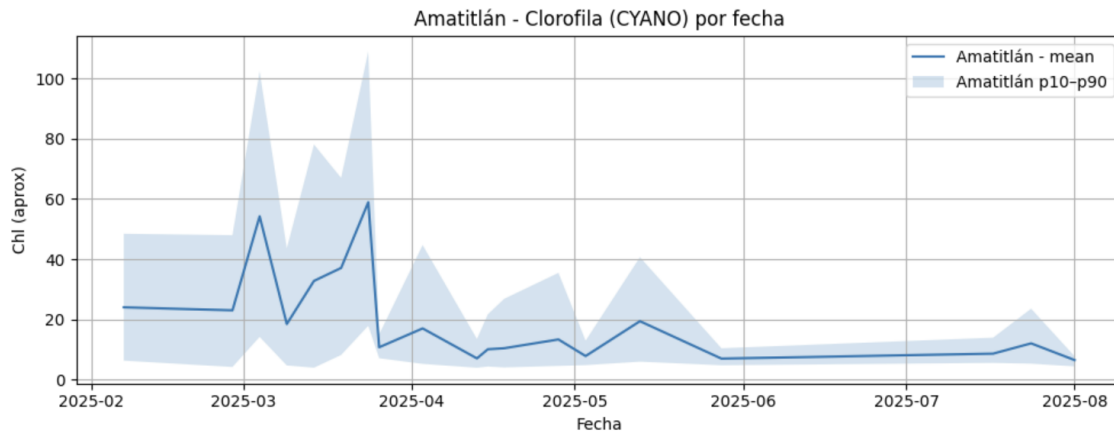


Figura 2: Serie temporal de clorofila en el Lago Amatitlán (P10–P90 sombreado).

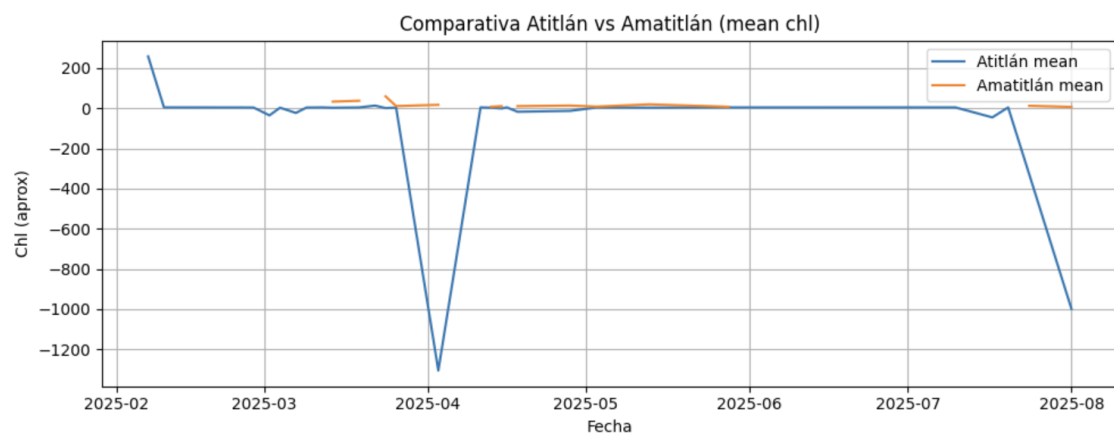


Figura 3: Comparación Atitlán vs. Amatitlán.

Apéndice A: Fechas Analizadas (29)

- | | | |
|--------------|--------------|--------------|
| ■ 2025-02-07 | ■ 2025-03-19 | ■ 2025-04-28 |
| ■ 2025-02-10 | ■ 2025-03-22 | ■ 2025-05-03 |
| ■ 2025-02-25 | ■ 2025-03-24 | ■ 2025-05-13 |
| ■ 2025-02-27 | ■ 2025-03-26 | ■ 2025-05-28 |
| ■ 2025-03-02 | ■ 2025-04-03 | ■ 2025-07-10 |
| ■ 2025-03-04 | ■ 2025-04-11 | ■ 2025-07-17 |
| ■ 2025-03-07 | ■ 2025-04-13 | ■ 2025-07-20 |
| ■ 2025-03-09 | ■ 2025-04-15 | ■ 2025-07-24 |
| ■ 2025-03-12 | ■ 2025-04-16 | ■ 2025-08-01 |
| ■ 2025-03-14 | ■ 2025-04-18 | |