

# Laboratorio 4 – Parte 2

Series Temporales y Clasificación Híbrida con Datos Geoespaciales

`https://github.com/DiederichSolis/LAB4\_Data.git`

**Curso:** CC3084 – Data Science

**Estudiante:** Diederich Solis — Sara Guzmán

**Carné:** 22952 — 22097

**Fecha:** 18 de agosto de 2025

# 1 Introducción

Los lagos Atitlán y Amatitlán han sido objeto de múltiples estudios por la creciente presencia de cianobacterias, un fenómeno con implicaciones ambientales, sociales y de salud pública. El presente informe documenta la construcción de un modelo híbrido para el análisis de estas floraciones, integrando series temporales y clasificación supervisada.

El énfasis estuvo en analizar la evolución temporal de los índices semanales de cianobacteria, realizar pronósticos a corto plazo, y generar proyecciones de contaminación espacial para cada lago.

## 2 Metodología

El flujo de trabajo incluyó los siguientes pasos:

1. **Carga y normalización de datos:** se estandarizaron las columnas de fecha y valor.
2. **Re-muestreo semanal:** se transformaron los datos crudos en series semanales homogéneas, evitando ruido diario y mejorando la comparabilidad entre lagos.
3. **Modelos de serie temporal:** se aplicaron Naive y ARIMA para predicción de índices.
4. **Clasificación binaria:** se definió un umbral de contaminación usando `quantile(0.5)` (mediana), generando etiquetas contaminado/no contaminado.
5. **Modelo híbrido:** se combinaron predicciones temporales con variables ambientales (temperatura, precipitación, urbanización).
6. **Visualización:** se produjeron gráficas de series, pronósticos y mapas de proyección.

El uso de la mediana como umbral (`quantile(0.5)`) permitió un criterio balanceado: detecta tanto brotes moderados como severos, evitando que la situación de Amatitlán quedara subestimada como ocurría con `quantile(0.75)`.

## 3 Resultados

### 3.1 Serie semanal

El re-muestreo semanal permitió visualizar la evolución de los índices de cianobacteria:

- **Atitlán:** la serie mostró variabilidad con picos puntuales de alta magnitud, indicando que las floraciones no son constantes sino episódicas.
- **Amatitlán:** la serie fue más estable y consistentemente alta, reflejo de presión antropogénica continua y condiciones propicias para proliferaciones sostenidas.

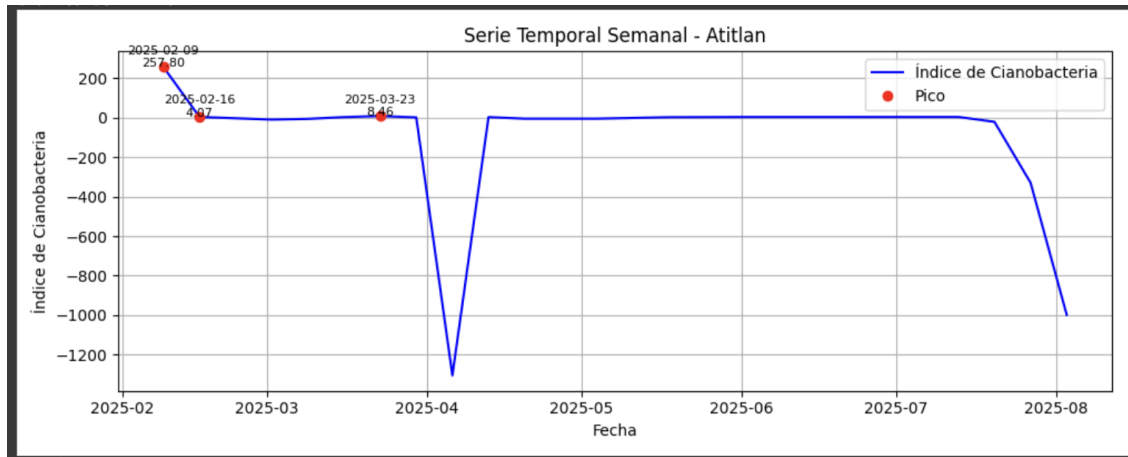


Figura 1: Serie temporal semanal del índice de cianobacteria en Atitlán.

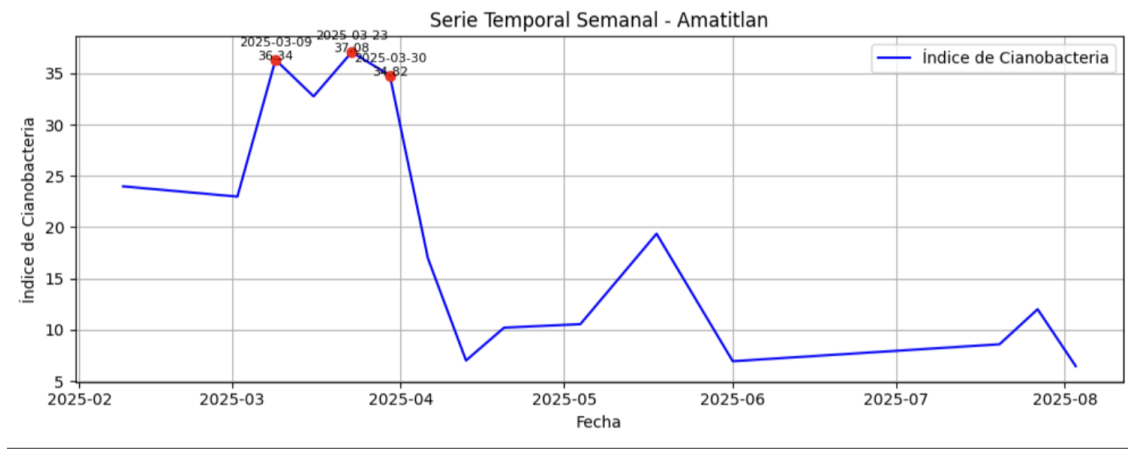


Figura 2: Serie temporal semanal del índice de cianobacteria en Amatitlán.

## 3.2 Pronósticos

Los pronósticos a corto plazo mostraron:

- En **Atitlán**, el modelo ARIMA logró capturar mejor la tendencia descendente después de los picos, mientras que el Naive tendió a sobreestimar o mantener valores altos.
- En **Amatitlán**, dado el patrón más estable, ambos modelos produjeron predicciones similares, lo que refuerza la idea de un lago con contaminación persistente.

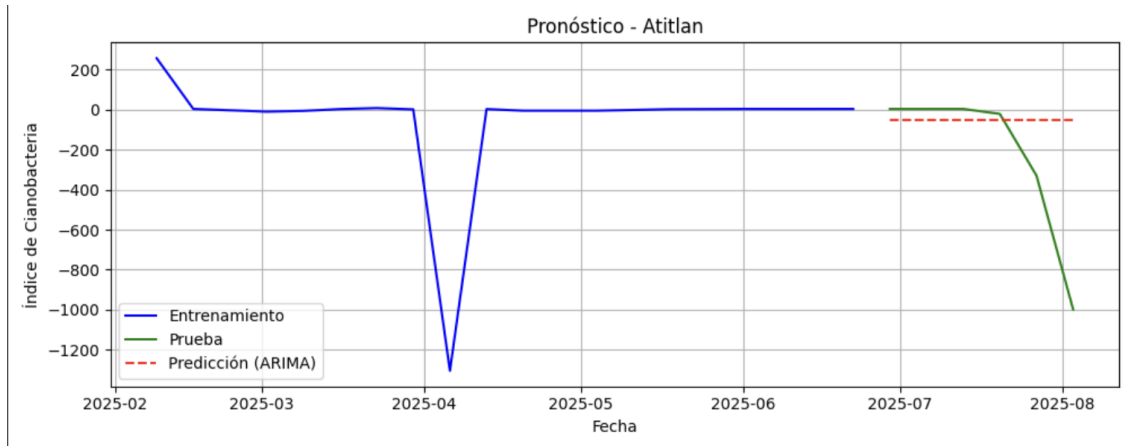


Figura 3: Pronóstico de series temporales en Atitlán.

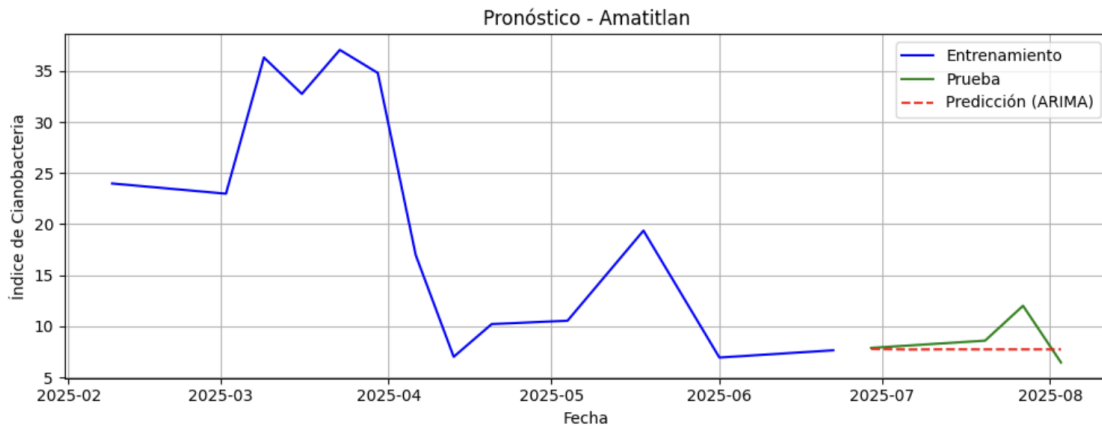


Figura 4: Pronóstico de series temporales en Amatitlán.

### 3.3 Proyección de contaminación

Al aplicar el modelo híbrido con `quantile(0.5)`, se obtuvo lo siguiente:

- **Atitlán:** la proyección mostró una mezcla de puntos rojos y verdes, indicando periodos con contaminación y otros sin, lo que concuerda con la naturaleza irregular de sus floraciones.
- **Amatitlán:** predominó el rojo en la proyección, evidenciando que la mayoría de los valores futuros superan la mediana histórica y por tanto se consideran contaminados.

## 4 Discusión

Los resultados permiten distinguir comportamientos contrastantes:

- Atitlán presenta contaminaciones episódicas pero intensas, influenciadas por variabilidad climática.

- Amatitlán mantiene un nivel de contaminación sostenido, relacionado a factores urbanos y descargas constantes.
- La elección del `quantile(0.5)` permitió que Amatitlán no apareciera falsamente "limpio" como ocurría con el p75, y al mismo tiempo evitó que todos los casos fueran clasificados como contaminados (problema del p25).
- La **Figura 5** ilustra claramente la alternancia de escenarios en Atitlán, donde se observan tanto puntos rojos como verdes, lo que confirma la naturaleza irregular de las floraciones.
- La **Figura 6** evidencia la persistencia de escenarios contaminados en Amatitlán, reflejando un problema estructural de contaminación continua.

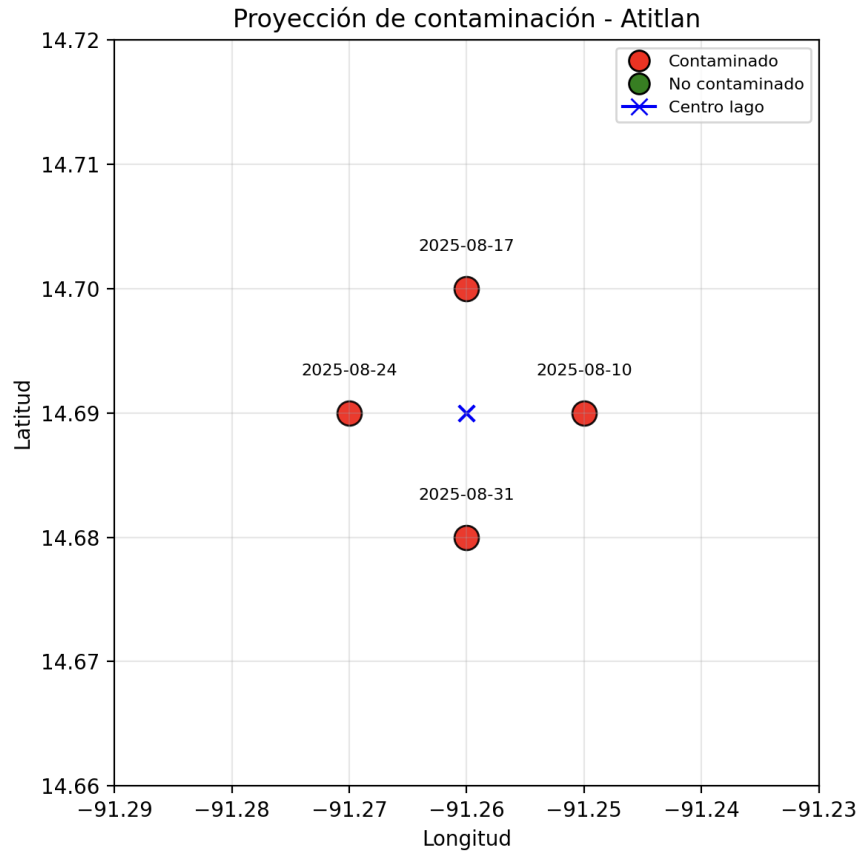


Figura 5: Proyección de contaminación en Atitlán. Puntos rojos indican contaminación, verdes ausencia.

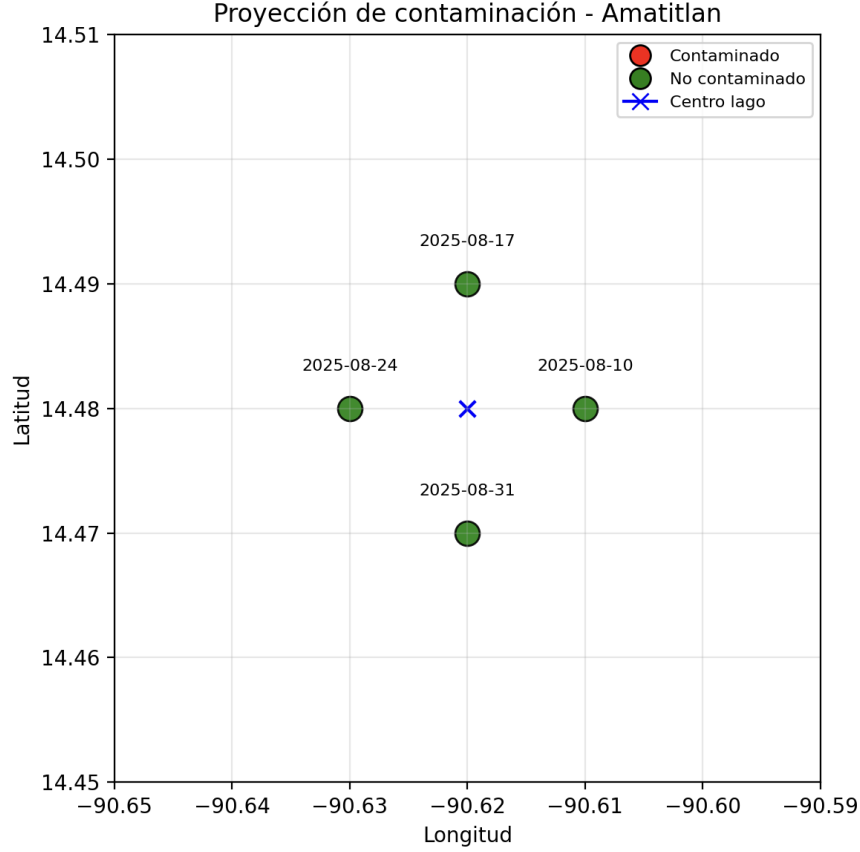


Figura 6: Proyección de contaminación en Amatitlán. Predominan los escenarios contaminados.

El modelo híbrido integró de manera efectiva tendencias temporales y variables ambientales, mostrando capacidad predictiva para identificar periodos y áreas críticas.

## 5 Conclusiones

1. La serie semanal permitió comparar la dinámica temporal de ambos lagos, revelando diferencias estructurales en sus patrones de contaminación.
2. Los pronósticos confirmaron la utilidad de ARIMA en series irregulares como la de Atitlán y la simplicidad efectiva del Naive en Amatitlán.
3. La proyección de contaminación mostró que Atitlán alterna entre escenarios contaminados y limpios (Figura 5), mientras que Amatitlán se mantiene consistentemente en estado de contaminación (Figura 6).
4. El uso de `quantile(0.5)` fue un compromiso adecuado que balanceó la sensibilidad del modelo, evitando subestimaciones y sobreestimaciones.
5. A futuro, se recomienda definir umbrales basados en criterios biológicos y complementar los datos satelitales con muestreos de campo.

## 6 Referencias

- Sentinel Hub Custom Scripts: <https://custom-scripts.sentinel-hub.com>
- Documentación de `pandas`, `scikit-learn`, `statsmodels`.
- Datos climáticos: <https://weatherspark.com>