

Análisis espacio-temporal de florecimientos algales en un lago-cráter tropical usando datos MODIS (2003-2020)

Lizette Zareh Cortés-Macías^{1*}, Juan Pablo Rivera-Caicedo², Jushiro Cepeda-Morales^{1,3}, Óscar U. Hernández-Almeida⁴, Ricardo García-Morales⁵, Pablo Velarde-Alvarado⁶

¹ Posgrado en Ciencias Biológico-Agropecuarias, Universidad Autónoma de Nayarit, Km. 9 Carretera Tepic-Compostela, CP. 63780 Xalisco, Nayarit, México.

² CONACYT-UAN, Secretaría de Investigación y Posgrado, Universidad Autónoma de Nayarit, Ciudad de la Cultura Amado Nervo, Tepic CP. 63155, Nayarit, México.

³ Unidad Especializada en Percepción Remota Satelital de Ecosistemas continentales y Oceánicos (PERSEO), Centro Nayarita de Innovación y Transferencia de Tecnología A. C., Calle 3 S/N. C.P. 63173, Tepic, Nayarit, Universidad Autónoma de Nayarit, México.

⁴ Laboratorio de Oceanografía Biológica, Secretaría de Investigación y Posgrado, Universidad Autónoma de Nayarit, Ciudad de la Cultura S/N, C.P. 63155 Tepic, Nayarit, México.

⁵ CONACYT-Centro de Investigaciones Biológicas del Noroeste SC, Unidad Nayarit (UNCIBNOR+), 63173 Tepic, Nayarit, México.

⁶ Unidad Académica de Ciencias Básicas e Ingenierías, Universidad Autónoma de Nayarit, Ciudad de la Cultura S/N, C.P. 63155 Tepic, Nayarit, México.

Resumen: El lago-cráter de Santa María del Oro en el Estado de Nayarit, México, presenta Florecimientos Algales (FA) de manera cíclica anual. El florecimiento y posterior decaimiento de estas poblaciones de algas crea cambios de color en el agua, generalmente en la primera mitad del año. Este trabajo evalúa algoritmos de clasificación supervisada que permitan identificar estos cambios usando datos de los productos MOD09GQ y MYD09GQ del sensor MODIS en el período de enero 2003 a diciembre 2020. A partir de una revisión de FA registrados en la literatura y análisis estadísticos de gráficos de dispersión, se construyó una base de datos de información espectral y etiquetas del estado de color del lago para evaluar diferentes algoritmos de clasificación. El mejor clasificador fue Random Forest con una precisión de 83,3%. El análisis temporal y la evaluación espacial de la incidencia de los florecimientos mostraron que mayo, abril y marzo son los meses con mayor presencia de cambios de color en el lago relacionados a FA. En el análisis espacial se encontró que la mayor incidencia de florecimientos se da en la región sureste del lago y las mayores cantidades de eventos ocurrieron en los años 2011, 2008 y 2012, respectivamente. Se determina la influencia del fenómeno El Niño-Oscilación del Sur (ENSO) en la incidencia de FA en el lago-cráter debido al patrón temporal entre las anomalías en los FA y el índice multivariado de El Niño-Oscilación del Sur, donde el mayor número de eventos de FA se presentaron en las fases frías del ENSO.

Palabras clave: MODIS, algoritmos de clasificación de aprendizaje automático, florecimientos algales nocivos, lago turquesa.

Spatio-temporal analysis of algal blooms in a tropical crater-lake from MODIS data (2003-2020)

Abstract: The crater lake of Santa María del Oro in Nayarit State presents Algal Blooms (AB) in a cyclical annual manner. The blooming and subsequent decline of these populations creates color changes in the water, generally

To cite this article: Cortés-Macías, L.Z., Rivera-Caicedo, J.P., Cepeda-Morales, J., Hernández-Almeida, O.U., García-Morales, R., Velarde-Alvarado, P. 2023. Spatio-temporal analysis of algal blooms in a tropical crater-lake from MODIS data (2003-2020). *Revista de Teledetección*, 62, 39-55. <https://doi.org/10.4995/raet.2023.19673>

* Corresponding author: lizette.cortes@uan.edu.mx

in the first half of the year. This work evaluated supervised classification algorithms that allow these changes to be identified using data from the MOD09GQ and MYD09GQ products of MODIS sensor in the period from January 2003 to December 2020. Based on a review of AB recorded in the literature and statistical analysis of dispersion graphs, a database of spectral information and lake color state labels was created to evaluate the different classification algorithms. The best classifier was Random Forest with an accuracy of 83.3%. The temporal analysis and spatial evaluation of the blooms incidence showed that May, April and March are the months with the greatest presence of color changes related to AB in the lake. The spatial analysis found that the highest incidence of blooms occurs in the southeast region of the lake and the the greatest number of events occurred in the years 2011, 2008 and 2012, respectively. The influence of the El Niño-Southern Oscillation (ENSO) phenomenon on the incidence of algal blooms in the crater lake is determined by the temporal pattern between the anomalies in the AB and the Multivariate ENSO Index, where the greater number of AB events occurred in the cold phases of the ENSO.

Key words: MODIS, Machine Learning classification algorithms, Harmful algal bloom, Turquoise lake.

1. Introducción

La eutrofización es el proceso mediante el cual la producción primaria de un cuerpo de agua se incrementa debido al aporte de nutrientes (German *et al.*, 2020). En cuerpos de agua continentales (CAC) el enriquecimiento de nutrientes implica cambios significativos en la abundancia y composición de las comunidades fitoplanctónicas que lo habitan. Generalmente este aporte de nutrientes favorece a las poblaciones de cianobacterias (Tomaselli *et al.*, 2004), ya que presentan ventajas adaptativas sobre otras especies de algas en aguas con niveles altos de eutrofización. Cuando estos crecimientos poblacionales causan incrementos desmedidos de biomasa se les conoce como florecimientos algales (FA).

La presencia y permanencia de los FA en los CAC genera alteraciones severas en estos ecosistemas. Deterioran la calidad del agua generando así, problemas de salud pública y llegando a limitar la disponibilidad del recurso. Esta problemática se intensifica cuando se trata de florecimientos de cianobacterias, ya que algunas especies producen cianotoxinas, convirtiéndolas en las principales responsables de eventos de intoxicación en CAC (Paerl y Millie, 1996; Pizzolon, 1996).

Los CAC son elementales para el desarrollo y funcionamiento de nuestras ciudades (Lynch *et al.*, 2016; Dörnhöfer y Oppelt, 2016), cumplen funciones importantes para los ecosistemas terrestres debido a que proveen de hábitat a un amplio rango de especies y son un componente esencial del ciclo biogeoquímico del agua, carbono y nitrógeno (Carpenter *et al.*, 2011; Moss, 2012). Debido a esto es importante implementar programas de caracterización y monitoreo de

FA que nos permitan describir su frecuencia e intensidad. Tradicionalmente los FA se estudian mediante la realización de muestreos rutinarios de concentración de clorofila (Chl-*a*) en el agua para estimar la biomasa del fitoplancton, en conjunto con la identificación taxonómica de la especie dominante. Sin embargo, los muestreos *in-situ* presentan limitaciones en la accesibilidad, baja cobertura espacial y baja continuidad temporal (Masocha *et al.*, 2018). Es en este aspecto donde las técnicas de teledetección se presentan como una alternativa complementaria para el monitoreo y estudio de los FA en CAC, pues proporcionan mediciones temporal y espacialmente explícitas y abren la posibilidad a realizar estudios de series de tiempo.

Las primeras estimaciones de concentración de Chl-*a* satelital fueron desarrolladas en la década de 1970 para aguas oligotróficas oceánicas donde se emplearon algoritmos con datos de la región azul/verde del espectro (Klima y Roe, 1972; PiSHAROTY, 1973; Hovis y Leung, 1997). Desde entonces el avance de la capacidad computacional y la puesta en órbita de más sensores con mejores resoluciones espaciales y radiométricas han permitido la exploración de nuevas técnicas para la estimación de la concentración Chl-*a*. El sensor MODIS (*Moderate Resolution Imaging Spectroradiometer*) lanzado en 1999 y 2002 a bordo de los satélites Terra y Aqua, provee de información espectral diaria con 36 bandas en diferentes resoluciones espaciales (250 m, 500 m y 1 km). MODIS ha probado tener el potencial para monitorear parámetros de calidad en CAC (Wang *et al.*, 2018; Shi *et al.*, 2020; Jia *et al.*, 2019; Ananias *et al.*, 2022), entre estos la temporada, duración y magnitud de los FA (Shi *et al.*, 2017, 2019).

Por su resolución espacial de 250 m, los productos MOD09GQ y MYD09GQ de MODIS han sido ampliamente usados para estudiar FA en CAC (Li *et al.*, 2019; Shi *et al.*, 2015; Gitelson *et al.*, 2008). Germán *et al.* (2016) emplearon estos productos para desarrollar un modelo de estimación de Chl-*a* aplicado a un embalse de 14 km². El modelo obtuvo un R² de 0,69 y fue aplicado a una serie temporal de datos de reflectividad del período 2001-2014. Posteriormente Germán *et al.* (2017) emplearon el algoritmo Harmonic ANalysis of Time Series (HANTS) (Verhoef, 1996) para modelar la línea base de patrones estacionales de FA en la serie temporal y detectar anomalías en ellos. En Germán *et al.* (2020) evalúan sus resultados de acuerdo a cuatro diferentes enfoques de definición de un FA, dos de ellos por límite fijo (Carlson, 1977 y Tett, 1987). Concluyen que los métodos de umbral fijo tendieron a sobreestimar los FA, demostrando la importancia de entender las condiciones locales y considerar los patrones estacionales al desarrollar modelos para lagos eutróficos.

En México el fitoplancton continental incluye aproximadamente 1025 especies, de las cuales el 18.2%, pertenece al grupo de las cianobacterias (Oliva-Martínez *et al.*, 2014). De acuerdo Oliva-Martínez *et al.* (2014), las clases de fitoplancton con mayor presencia en las aguas dulces del país son: *Bacillariophyceae*, *Cyanophyceae* y *Zygnematomyceae*, con 327, 187 y 105 especies registradas, respectivamente. El lago-cráter de Santa María del Oro (SAMAQ) ubicado en el occidente del Cinturón Volcánico Mexicano, presenta FA de manera cíclica anual registrados generalmente en la primera mitad del año (Sosa-Nájera *et al.*, 2010). Durante el florecimiento de 2015, se identificaron tres especies de cianobacterias: *Limnographis robusta*, *Microcystis aeruginosa*, y *Microcystis smithii* (Salazar-Alcaraz, 2018).

El objetivo de este estudio es caracterizar la dinámica espacial y temporal de los FA que se han presentado en el lago-cráter de Santa María del Oro (SAMAQ) en el período 2003-2020 usando los productos MOD09GQ y MYD09GQ de MODIS. Para esto se evaluaron los algoritmos de clasificación: Análisis Discriminante, K-vecinos cercanos, Naive Bayes, *Random Forest* y Redes Neuronales usando una base de datos generada de manera manual a partir de campañas de campo y reportes técnicos. Además, se realiza el análisis de

la correlación de los FA identificados por técnicas de clasificación con el índice multivariado de El Niño/Oscilación del Sur (MEI).

2. Materiales y Métodos

2.1. Área de estudio

El lago-cráter de Santa María del Oro - SAMAQ se encuentra a 750 metros sobre el nivel del mar en la zona Occidental del Cinturón Volcánico Mexicano (Figura 1) (Armienta *et al.*, 2008). Se trata de un lago endorreico de edad pleistocénica con un diámetro aproximado de 2 km, una superficie de área de 3,7 km² y una profundidad media de 58 m (Sosa-Nájera *et al.*, 2010). SAMAQ es una zona importante para el abastecimiento de agua, la acuicultura, agricultura y el turismo (Serrano *et al.*, 2002).

2.2. Datos MODIS

Los datos utilizados para este estudio se descargaron de la plataforma: *Application for Extracting and Exploring Analysis Ready Samples (AppEEARS)* (<https://lpdaacsvs.cr.usgs.gov/appeears/>) aplicando una reproyección a coordenadas geográficas Datum WGS84 y EPSG: 4326. Debido a su serie temporal y considerando la superficie relativamente pequeña de nuestra área de estudio se emplearon los productos MOD09GQv006/MYD09GQv006 del sensor MODIS, éstos entregan reflectancia (Rrs) superficial diaria para las bandas del rojo - B1 (620 nm a 670 nm) e Infrarrojo cercano - B2 (841 nm a 876 nm) a una resolución espacial de 250 m. Los datos cuentan con corrección atmosférica y de efectos de dispersión (Vermote y Wolfe, 2015). Si bien, el efecto de la atmósfera en la radiancia recibida por sensores remotos es significativamente mayor cuando se trata de CAC (Moses *et al.*, 2017) y es recomendable la aplicación de un método de corrección atmosférica adecuado, consideramos que los productos MOD09GQv006/MYD09GQv006 son aceptables para realizar trabajos de clasificación ya que, aunque los productos MODIS han sido desarrollados para aguas oceánicas, su uso en aguas continentales para detección de florecimientos algales es ampliamente aceptado (Hu *et al.*, 2010; Huang *et al.*, 2014; Shi *et al.*, 2019).

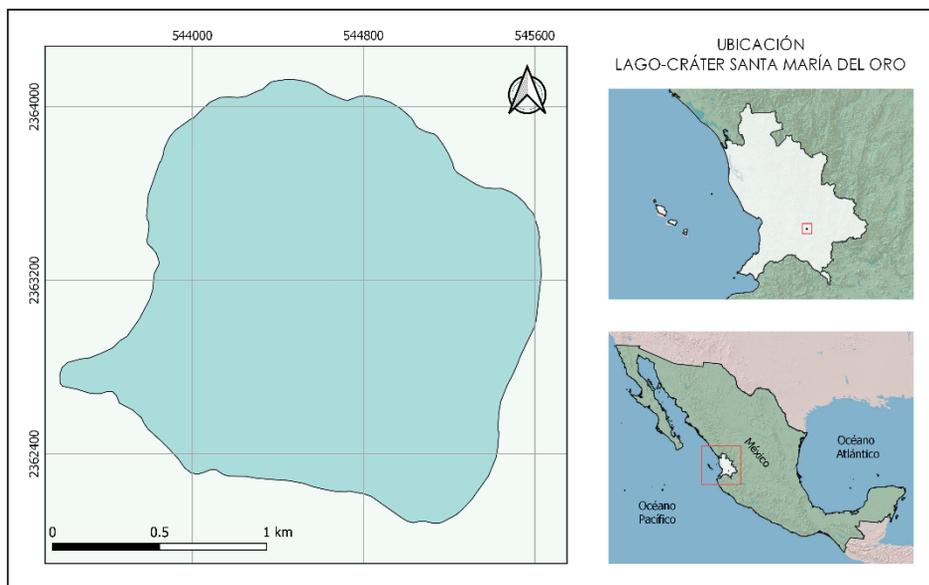


Figura 1. Ubicación del área de estudio: lago-cráter de Santa María del Oro (SAMAQ), en el occidente del Cinturón Volcánico Mexicano en Nayarit (coordenadas UTM, zona 13).

2.3. Análisis de datos

El procesamiento de datos se realizó en tres etapas que comprenden: 1. Filtrado de datos, 2. Extracción de características y 3. Generación de base de datos (BD). Los datos descargados de *AppEARS* en formato *NetCDF*, corresponden a matrices $m \times n \times t$ donde m y n corresponden a las coordenadas espaciales de longitud y latitud respectivamente y la componente t corresponde a la escala temporal diaria desde el 2003/01/01 a 2020/12/31. Para el filtrado de datos se utilizó la banda *QA*. Esta banda almacena información en formato de 16 bit y a partir de la combinación bit genera etiquetas de calidad. En este estudio solo se utilizaron los píxeles donde los bits 0-1, 1-4 y 8-11 de la banda *QA* eran cero asegurando la máxima calidad de los datos. Para descartar los píxeles con mezcla de tierra y agua de las orillas del lago se empleó una máscara.

El vector de características para la generación de la BD se conformó por los parámetros que se muestran en la Tabla 1. Donde, Banda 1 (B1) y Banda 2 (B2) son los datos espectrales obtenidos por el sensor en las regiones RED y NIR, respectivamente. *Simple ratio* 1 (SR1) se refiere a la relación entre los valores de la banda NIR sobre los de la banda RED y *Simple ratio* 2 (SR2) a los

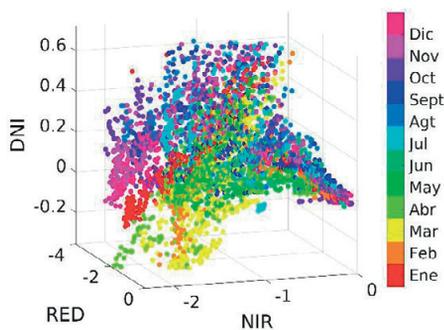
datos RED sobre los NIR. El índice de diferencia normalizada (IDN) es una división entre los valores de la resta de RED a NIR y la suma de RED a NIR. Al agregar estos parámetros clasificadores engrosamos la BD asociada a cada clase para crear una mayor diferenciación entre éstas con el fin de mejorar el entrenamiento de nuestro algoritmo. Además, para dar una mayor dispersión a los datos, estos se elevaron a logaritmo de base 10.

Tabla 1. Variables clasificadoras.

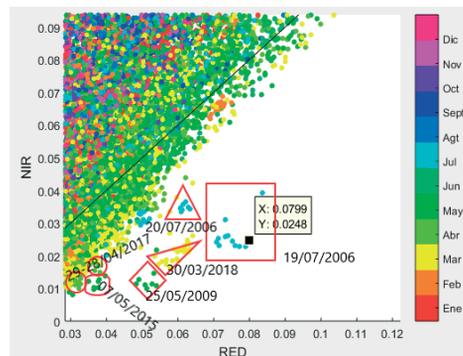
Variable	Valor	Acronimo
Banda 1	620-670 nm	RED
Banda 2	841-876 nm	NIR
Simple ratio 1	RED/NIR	SR1
Simple ratio 2	NIR/RED	SR2
Índice de diferencia normalizada	$(NIR-RED)/(NIR+RED)$	DNI

2.4. Determinación de clases

Se definieron 6 clases para entrenar el algoritmo. Estas clases incluyen los aspectos ópticos del lago que abonan al objeto de estudio: 1. Florecimiento Algal (FA), 2. Turquesa (TQ), 3. Homogéneo (HM); y tres clases más como control para reducir errores en la clasificación: 4. Nubes (NB), 5. Efecto borde de la escena (BR) y 6. Efecto



a) Dispersión de los datos de Aqua para 2018 en gráfico 3D, con dimensiones RED- NIR-IDN.



b) Acercamiento a región con clústers correspondientes a TQ. Dispersión de datos Aqua en 2D con dimensiones RED-NIR

Figura 2. Visualización del total de la BD MODIS mediante gráficos de dispersión 2D y 3D empleados para detectar mediante la regionalización de valores asociados fechas de eventos de las diferentes clases. La barra de colores indica el mes al que pertenecen los puntos que indican los valores *por pixel*.

Sunlight (SL). Las etiquetas para las clases FA, TQ y HM se realizaron a partir de la información recopilada en estudios previos (Salazar-Alcaraz, 2018; Cortés-Macías, 2018). Las clases NB, BR y SL se seleccionaron con la ayuda de un programa informático realizado en MATLAB que genera gráficos de dispersión en 2D (Figura 2b) y 3D (Figura 2a). La visualización de la BD mediante estos gráficos nos permitió también encontrar regiones con clústeres de eventos FA y TQ de años previos como 2009 y 2006 (Figura 2b). En la construcción de la BD para las seis clases se empleó el visualizador (<https://worldview.earthdata.nasa.gov/>) para corroborar las fechas observadas en la literatura y en los gráficos de dispersión.

La Figura 3 muestra un compuesto en color verdadero del sensor *MultiSpectral Instrument* (MSI) de la misión Sentinel 2 de los cambios de color que experimenta el lago durante los primeros meses del año (verde, turquesa, azul índigo) y que responden a la presencia de un evento cíclico anual de FA. También nos da una idea del período de tiempo en que estos cambios pueden ocurrir, en este caso de 2018 en aproximadamente un mes, ilustrando así el alto dinamismo de SAMAO. Junto a cada imagen se describe la distribución de los parámetros clasificadores mediante diagramas de caja. Estos representan la base de datos en escala Log_{10} empleada para el entrenamiento de los algoritmos de clasificación

2.5. Clasificación supervisada

Se evaluaron 5 algoritmos de clasificación supervisada: 1. Análisis Discriminante (DA); 2. K-vecinos cercanos (KNN); 3. Naive Bayes (NB); 4. *Random Forest* (RF); y 5. Redes Neuronales Artificiales (ANN). Estos algoritmos están contenidos en la *Simple Classification Toolbox* (Muñoz-Marí y Camps-Valls, 2013) desarrollada por el grupo de procesamiento de imágenes y señales (ISP) de la Universidad de Valencia (España), que ha sido implementada en el módulo de clasificación del programa ARTMO (Aghababaei *et al.*, 2022).

2.5.1. Análisis discriminante - DA

Es una generalización de la discriminante lineal de Fisher (Fisher, 1936) usado en clasificación multivariada que asume que los datos de las diferentes clases tienen una distribución normal. El objeto del algoritmo es minimizar la función (Ecuación 1).

$$\hat{y} = \arg \min \sum_{k=1}^K \hat{P}(k|x) C(y|k) \quad (1)$$

donde \hat{y} es la clase que se estima al punto x , K es el número de clases, $\hat{P}(k|x)$ es la probabilidad a posteriori de que el punto x pertenezca a la clase k y $C(y|k)$ es el costo de clasificar el punto x como y cuando su verdadera clase es k .

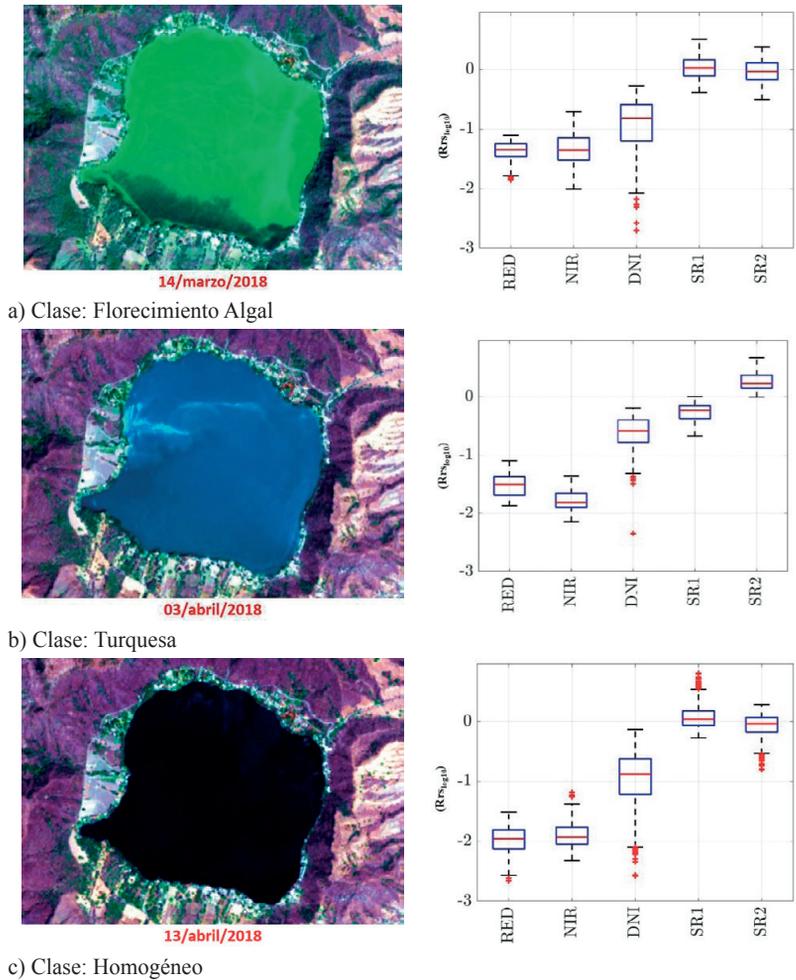


Figura 3. Compuesto de color verdadero de Sentinel 2 del área de estudio durante y después de un florecimiento algal. Fechas de las tomas: a) 14/03/2018, b) 03/04/2018 y c) 13/04/2018. A la derecha, diagramas de caja con la caracterización de valores (Rrs_{log10}) de las variables clasificadoras (eje x) en función de las clases AF, TQ y HM.

2.5.2. K-vecinos cercanos - KNN

Algoritmo no paramétrico que asigna la clase \hat{y} a un punto x por el número máximo de k a partir de una función de costo C que calcula las métricas de distancia entre el nuevo punto a clasificar y los datos usados en el entrenamiento del modelo. (Xiang *et al.*, 2008). La función de costo seleccionada fue Mahalanobis la cual utiliza la matriz de covarianza de los datos de entrenamiento.

Dado un conjunto de entrenamiento X de n puntos y la función de costo C , knn busca los k vecinos cercanos al punto y_i asignando la clase mayoritaria de los k vecinos (Ecuación 2).

$$d_{st}^2 = (x_s - y_t)C^{-1}(x_s - y_t)' \quad (2)$$

donde x_s es el vector S del conjunto de datos X , y_t vector de características que se desea clasificar y C es la matriz de covarianza.

2.5.3. Naive Bayes - NB

Este algoritmo se fundamenta en el teorema de Bayes (Flach y Lachiche, 2004). Es un clasificador probabilístico con algunas simplificaciones en sus supuestos teóricos sobre la independencia de las variables predictoras dadas las clases.

Este algoritmo calcula la probabilidad de pertenencia de un nuevo punto a cada clase del entrenamiento y le asigna la de mayor probabilidad según la Ecuación 3.

$$\hat{P}(Y = k|X_1, \dots, X_p) = \frac{\pi(Y=k) \prod_{j=1}^p P(X_j|Y=k)}{\sum_{k=1}^C \pi(Y=k) \prod_{j=1}^p P(X_j|Y=k)} \quad (3)$$

donde Y es la variable aleatoria correspondiente a la a clase, X_1, \dots, X_p son predictores aleatorios de un conjunto de datos y $\pi(Y=k)$ es la prioridad *a priori* que la clase pertenezca a k .

2.5.4. Random Forest - RF

Random Forest o Árboles aleatorios es un ensemble de varios árboles de decisión que han sido entrenados de manera paralela (Breiman, 2001). El valor de la clasificación de una nueva muestra corresponde a la clase mayoritaria que ha entregado el conjunto de los árboles. Las reglas de decisión se basaron en el índice de impureza Gini (Raileanu y Stoffel, 2004) que mide el grado de “impureza” de un nodo: $Gini(t)$ iguales a cero indican que los datos que pertenecen a una sola categoría, mientras que índices mayores que cero y con valores hasta uno indican nodos donde los datos pertenecen a más de una categoría. El índice está definido según la Ecuación 4.

$$Gini(t) = \sum_{i=1}^C P(i) * (1 - P(i)) \quad (4)$$

Donde C representa el número total de clases y P representa la probabilidad de pertenencia de un elemento a la clase i .

2.5.5. Redes Neuronales Artificiales - ANN

Se utilizó una red neuronal pre-alimentada (*feed-forward* en inglés) donde las conexiones entre las unidades no forman un ciclo, la información se mueve únicamente hacia adelante. Para su diseño y entrenamiento se utilizó la función *patternnet* de la *Deep Learning Toolbox* de Matlab. Su topología es de cuatro capas: 1. Nodos de entrada, 2. Capas ocultas, 3. Capas de salida y 4. Salidas. Sus características de entrenamiento son:

- Algoritmo de entrenamiento: *Scaled Conjugate Gradient* (SCG) con tasa de convergencia superlineal (Roodschild *et al.*, 2019).
- Algoritmo de evaluación de rendimiento: *Cross-Entropy*, la cual se utiliza para ajustar los pesos de la red neuronal en la capa *output* (Wang *et al.*, 2022).
- Función de activación en la capa oculta: Sigmoidea simétrica (Goodfellow *et al.*, 2016)

- Función de activación de la capa de salida: *Soft max* (Goodfellow *et al.*, 2016)

2.6. Métricas de evaluación

El desempeño de los algoritmos se evaluó con las métricas de *Overall Accuracy* (OA) (Ecuación 5). Para el análisis a nivel de clases se calcularon las métricas de *Precision* (ecuación 6) como estimador de relevancia y *Recall* (ecuación 7) como métrica de sensibilidad de los algoritmos de clasificación (Congalton y Green, 2019). Se dividió la base de datos generada en entrenamiento/validación con una proporción 70/30 respectivamente usando una partición estratificada. Los datos de entrenamiento se utilizaron para ajustar los hiperparámetros de los algoritmos de Redes Neuronales Artificiales y *Random Forest* usando el método *kfold-n* con particiones estratificadas con un n de 10.

$$Overall Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \quad (5)$$

$$Precision = \frac{T_p}{T_p + F_p} \quad (6)$$

$$Recall = \frac{T_p}{T_p + F_n} \quad (7)$$

donde T_p es el número de positivos verdaderos, T_n es el número de negativos verdaderos, F_p es el número de falsos positivos y F_n el número de falsos negativos.

2.7. Caracterización espacio-temporal de FA en SAMAO

La caracterización de la distribución espacial y temporal de la dinámica de los eventos de FA entre 2003-2020 se realizaron en dos escalas: 1, mensual y 2, anual. Se definió como criterio para catalogar un evento de FA que al menos el 30% de los píxeles del lago se hayan clasificado. La caracterización espacial se analizó a partir de la generación de la cartografía de la frecuencia de ocurrencia de eventos FA a nivel de píxel. La distribución temporal se realizó descomponiendo la serie de los eventos acumulados, se calcularon las estadísticas descriptivas de media, desviación estándar y la regresión lineal. Se calcularon los componentes de la serie: tendencia, ciclo estacional y anomalías (Hamilton, 2020).

La influencia de la condición ENSO en la frecuencia de los FA se analizó conforme a la correlación entre las anomalías de la serie histórica de 17 años y el índice bimestral multivariado de El Niño/Oscilación del Sur (ENSO) versión 2 *MEI.v2* (Wolter y Timlin, 1993; Kobayashi *et al.*, 2015). Este índice es la serie de tiempo de la Función Ortogonal Empírica (EOF). Se estima a partir de 12 temporadas bimensuales superpuestas y combina cinco variables: presión al nivel del mar, temperatura de la superficie del mar, componentes zonales y meridionales del viento a nivel superficial y la radiación de onda larga saliente sobre la cuenca del Pacífico tropical (30 °S-30 °N y 100 °E-70 °W).

3. Resultados y discusión

3.1. Base de datos

Para entrenar los algoritmos de clasificación se construyó una base de datos con 3529 observaciones (BDclass), donde cada observación corresponde a un píxel etiquetado de acuerdo a los criterios de determinación de clases. La BDclass se encuentra equilibrada, cada clase cuenta con 588 observaciones, los *boxplots* en la Figura 4 muestran el comportamiento de los valores obtenidos en los índices clasificadores (eje Y) correspondientes a cada clase (eje X).

Las reflectividades en las bandas RED y NIR presentan comportamiento semejante entre sí, con

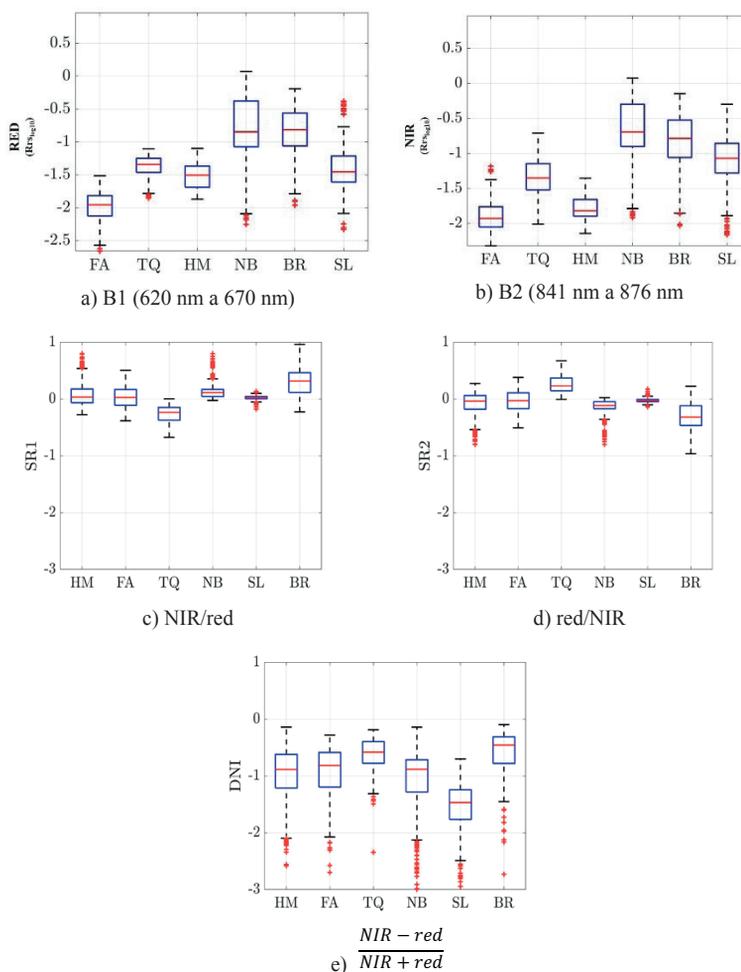


Figura 4. Diagramas de bigotes de la distribución de las clases en función de cada variable clasificadora. En el eje ‘x’ las clases: HM, FA, TQ, NB SL y BR. En el eje ‘y’ las reflectividades NIR y RED (Rrs_{log10}) y los índices SR1, SR2 y DNI. Los pies de figura a-b indican el ancho de las bandas, los pies de c-e detallan las fórmulas empleadas para obtener los índices.

una dispersión alta en las clases de control y baja en las clases de color del agua. Estos resultados son los esperados debido al alto albedo connatural de las nubes y los efectos que engloban las clases NB, SL y BR, mientras que las clases FA, TQ y HM tienen baja dispersión en sus valores debido al carácter sumamente absorbente del agua que contribuye con solo el 20% de la radiación total que mide el sensor (Moses *et al.*, 2017).

3.2. Clasificación

Se llevó a cabo la evaluación del rendimiento de los algoritmos de clasificación. Los algoritmos de DA, KNN y NB se evaluaron a partir de la función de optimización (*hyperparameters*). Para la evaluación de DA se optimizaron los parámetros: Delta y Gamma, y se obtuvo un valor *OA* del 65,5%. Para KNN se ajustaron los parámetros de número de vecinos y métricas de distancia obteniendo un

81,3% de *OA*. Para NB se ajustaron los parámetros de tipo de distribución (normal, kernel) y los pesos datos a las muestras, éste obtuvo un 76% de *OA*. Para ANN se evaluaron diferente número de nodos en la capa oculta: 2, 4, 6, 10, 15, 20, 25 y 30, donde los mejores resultados fueron de 70,8% *OA* obtenidos con 30 nodos. De modo que la topología de ANN consistió en 5 entradas, 30 capas ocultas, 6 capas de salida y 6 salidas. Para *Random Forest* se evaluaron diferentes números de árboles entre 5 y 400, el mejor desempeño se obtuvo con un bosque de 20 árboles con valores de *OA* de 83,3%. En la Tabla 2 se muestran los resultados de las métricas *Overall Accuracy*, *Precision* y *Recall* para el conjunto de algoritmos evaluados, y en la Figura 5 se detalla mediante matrices de confusión el desempeño de éstos. Imprevistamente, los mayores errores en la clasificación ocurrieron para la clase TQ en la evaluación de los 5 algoritmos, con

KNN

1	162	14	9	3	1	3	84.4%	15.6%
2	9	140	16	4	10	15	72.2%	27.8%
3	6	18	167		1	2	86.1%	13.9%
4	2	10		155	14	12	80.3%	19.7%
5	6	12	3	14	154	3	80.2%	19.8%
6	3	16	2	7	1	165	85.1%	14.9%
	86.2%	66.7%	84.8%	84.7%	85.1%	82.5%		
	13.8%	33.3%	15.2%	15.3%	14.9%	17.5%		
	1	2	3	4	5	6		

(a) K-Vecinos cercanos

NB

1	156	11	17	2	1	5	81.2%	18.7%
2	11	135	18	6	11	13	69.6%	30.4%
3	10	26	156		2		80.4%	19.6%
4	3	12		141	21	16	73.1%	26.9%
5	7	15		18	145	7	75.5%	24.5%
6	8	19	2	12	8	145	74.7%	25.3%
	80.0%	61.9%	80.8%	78.8%	77.1%	78.0%		
	20.0%	38.1%	19.2%	21.2%	22.9%	22.0%		
	1	2	3	4	5	6		

(b) Naive Bayes

NN

1	157	5	23			7	81.0%	18.2%
2	15	117	34			7	60.3%	39.7%
3	10	20	164				84.5%	15.5%
4	4	16		136	26	11	70.5%	29.5%
5	6	25	1	15	145		75.5%	24.5%
6	17	43	4	22	6	102	52.6%	47.4%
	75.1%	51.8%	72.6%	78.6%	78.8%	72.3%		
	24.9%	48.2%	27.4%	21.4%	21.2%	27.7%		
	1	2	3	4	5	6		

(c) Redes Neuronales Artificiales

AD

1	138	8	31			15	71.9%	28.1%
2	12	87	41			7	44.8%	55.2%
3	7	15	172				88.7%	11.3%
4	3	14		109	46	21	56.5%	43.5%
5	6	35		34	117		60.9%	39.1%
6	3	29	7	13	6	136	70.1%	29.9%
	81.7%	46.3%	68.5%	69.9%	66.5%	62.1%		
	18.3%	53.7%	31.5%	30.1%	33.5%	37.9%		
	1	2	3	4	5	6		

(d) Análisis discriminante

Figura 5. Matrices de confusión en la evaluación de los algoritmos KNN, NB, ANN y DA.

mayor número de falsos positivos en la clase HM con KNN y NB, y en la clase BR con RF y DA.

Tabla 2. Métricas del desempeño de los algoritmos clasificadores.

Algoritmo	Overall		
	Accuracy	Precision	Recall
Análisis Discriminante	65,49	65,80	65,50
K-vecinos cercanos	81,36	81,60	81,40
Naive Bayes	76,01	76,60	76,00
Random Forest	83,35	83,60	83,30
Redes Neuronales Artificiales	70,84	71,50	70,90

El algoritmo KNN obtuvo resultados *OA* de apenas 1,99% menores a RF, por lo que se considera tener potencial para clasificar los estados ópticos de SAMAO, mientras que DA tuvo errores mayores al resto de algoritmos. RF ha probado ser eficaz como clasificador en estudios de FA previos (Pal, 2005; Shaik y Srinivasan, 2019). Ananias et al. (2022) integraron RF a una cadena automatizada de detección de FA mediante datos MODIS, donde en la fase de evaluación el algoritmo obtuvo un acierto del 95%. La Figura 6a muestra las simulaciones realizadas con diferentes bosques conformados por entre 5 y 400 árboles. De acuerdo a Breiman (2001) el desempeño de RF mejora con el incremento de árboles. Sin embargo, en nuestros resultados indican que RF con un tamaño de bosque de 20 árboles presentó el mayor valor de *OA*. Los resultados de precisión en la matriz

de confusión de RF (Figura 6b) para las clases FA, TQ y HM fueron de 82,8%, 71,9% y 83,9%, respectivamente. Las clases de control NB, BR y SL por su parte obtuvieron 87,4%, 88,1% y 87,3% de precisión.

3.3. Caracterización espacial y temporal

Se realizó la caracterización espacial y temporal 2003-2020 presentando los resultados obtenidos para la clase FA. En orden de mostrar la variabilidad espacial de los florecimientos los resultados se representan a través de matrices de SAMAO (Figura 7). Estas corresponden a la máscara de píxeles empleada para la extracción de datos. La barra de valores que va de 0 a 250 indica el número acumulado de ocasiones en que cada píxel fue clasificado como FA en la serie tiempo. Los eventos de FA varían mensualmente iniciando el incremento de febrero (100 eventos) hasta el máximo en mayo (250 eventos), descendiendo hasta julio (<50 eventos). De julio a diciembre las frecuencias de FA se mantienen bajas. Este comportamiento temporal corrobora la periodicidad cíclica del los FA en SAMAO, asociados a la estación de primavera. En el análisis espacial, los resultados indican que los píxeles con mayor incidencia de FA se ubican en la región sureste del lago, lo cual asociamos como una respuesta al régimen diurno del viento, el cual es uno de los principales componentes en la distribución espacial

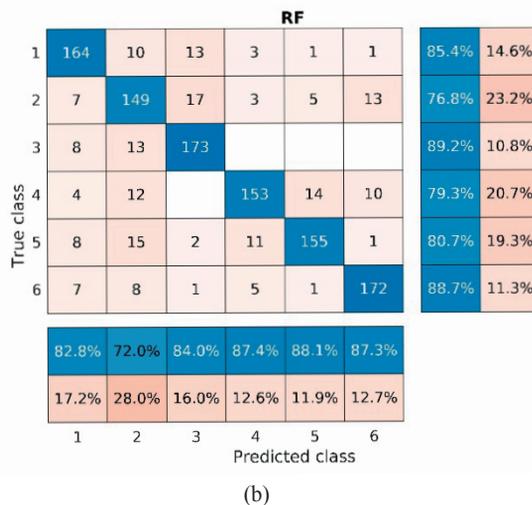
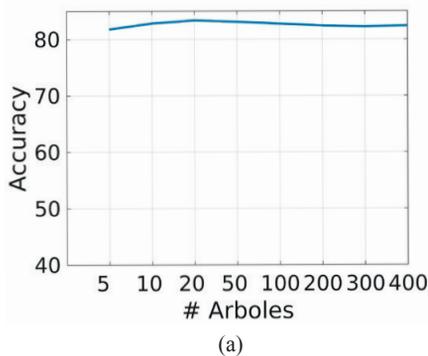


Figura 6. (a) Precisión de Random Forest en relación al número de árboles, y (b) Matriz de confusión de la evaluación de Random Forest con un bosque conformado por 20 árboles.

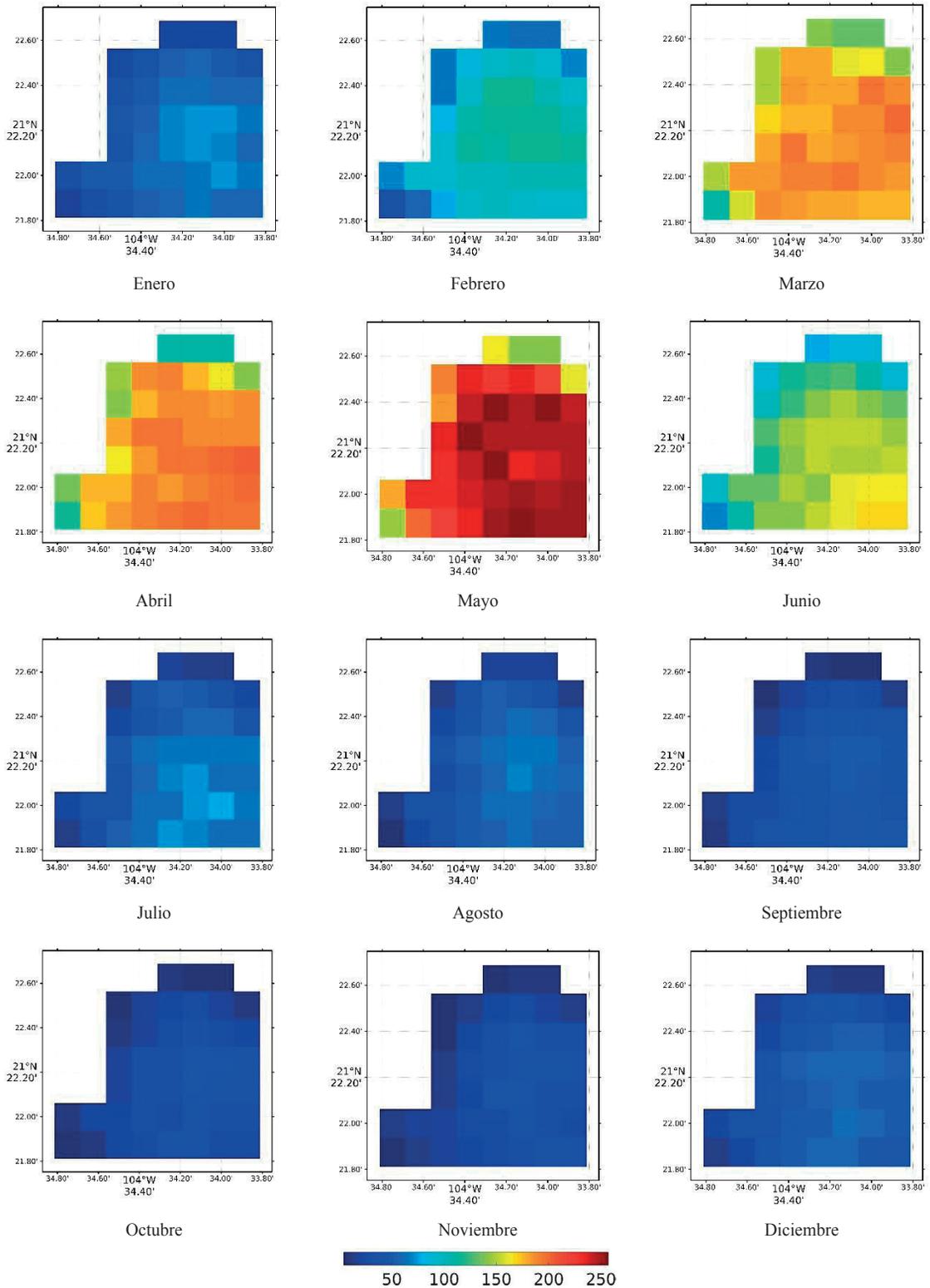


Figura 7. Incidencia espacial de FA en SAMAO por periodos mensuales para la serie de tiempo 2003-2020. La barra de valores indica el número acumulado de ocasiones en que cada píxel fue clasificado como FA en la serie de tiempo.

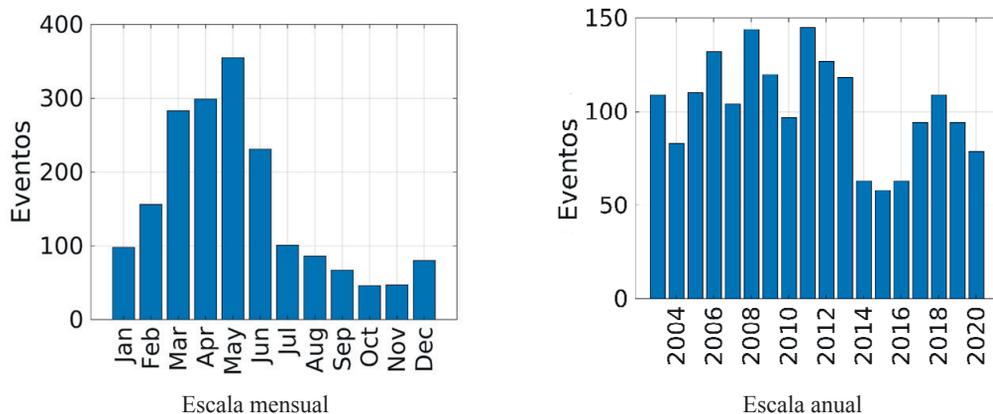


Figura 8. Ocurrencia de eventos FA a escala mensual (a) y anual (b) durante el período 2003-2020.

superficial de las poblaciones de algas en los CAC (Hsiao, 1988). En este sentido, las características orográficas de SAMAO definen un patrón en las fluctuaciones del viento donde, la brisa diurna se mueve al oeste por la noche y al este por la tarde con una rotación en sentido antihorario. Alrededor del mediodía, en las horas aproximadas al paso del sensor, Serrano *et al.* (2002) encontraron que se forma un giro anticiclónico en la parte sur del lago lo que puede explicar por qué estos píxeles han sido los de mayor incidencia de FA detectados.

Para este trabajo se determinó como un *evento de FA* a aquellas imágenes donde el área de píxeles clasificada fuese superior al 30% de la cobertura del lago. La Figura 8a muestra el acumulado de eventos de FA obtenidos entre los años 2003-2020 a escala mensual. Se observa un patrón estacional donde mayo, abril y marzo son, en ese orden, los meses con la mayor acumulación de eventos, lo que coincide con los resultados encontrados en 2015 por Cortés-Macías, (2018) y Salazar-Alcaraz *et al.*, (2021), donde el florecimiento tuvo lugar de enero a abril. Los meses con menor incidencia de FA encontrados ocurren en enero y de julio a diciembre cuando el lago presenta alta estratificación (Sosa-Nájera *et al.*, 2010). Germán *et al.* (2016) encontraron un comportamiento periódico en los FA con picos en verano y valles en invierno empleando una BD MODIS del período 2001-2014. Shi *et al.* (2019) evaluaron un modelo NIR-RED con datos de entre 2013-2017 de MODIS y reportaron también ciclos estacionales con picos anuales de FA en julio-agosto.

La Figura 9 muestra el análisis de la descomposición de la serie temporal de los eventos FA durante el período 2003-2020 agrupados mensualmente. La Figura 9a muestra la distribución observada del número de eventos mensuales que se presentaron en el período de estudio; el ciclo estacional de los FA se extrajo de la serie temporal (Figura 9b) donde el patrón temporal presenta un marcado ciclo anual, con un máximo en mayo y un mínimo en octubre indicando que es la escala temporal dominante en el comportamiento de los FA. Complementariamente, la variabilidad interanual entre los valores observados (Figura 9a) y el ciclo anual (Figura 9b), indica la influencia de fenómenos de diferentes escalas actuando sobre SAMAO. La Figura 9c indica la tendencia de los eventos la cual se ha ajustado a una recta del tipo $y=mx+b$, donde la pendiente m tienen un valor de $-0,00052$ y el corte con el eje y de $392,53$. El valor de la pendiente nos indica que se observa una tendencia negativa muy baja en la serie. La Figura 9d nos muestra las anomalías: una resta del ciclo estacional a la serie completa, dichas anomalías abarcan un rango entre -10 a 11 .

Como se observa en la Figura 8b, los años 2011, 2008 y 2012 son los de mayor número de eventos FA. Estos años coinciden con períodos de La Niña (fase fría). Mientras que los años 2015, 2016 y 2014, que corresponden a períodos El Niño (fase cálida), tienen los menores registros de eventos clasificados como FA. En la Figura 10 se observa que existe una tendencia inversa de las anomalías de FA obtenidas en este trabajo con el

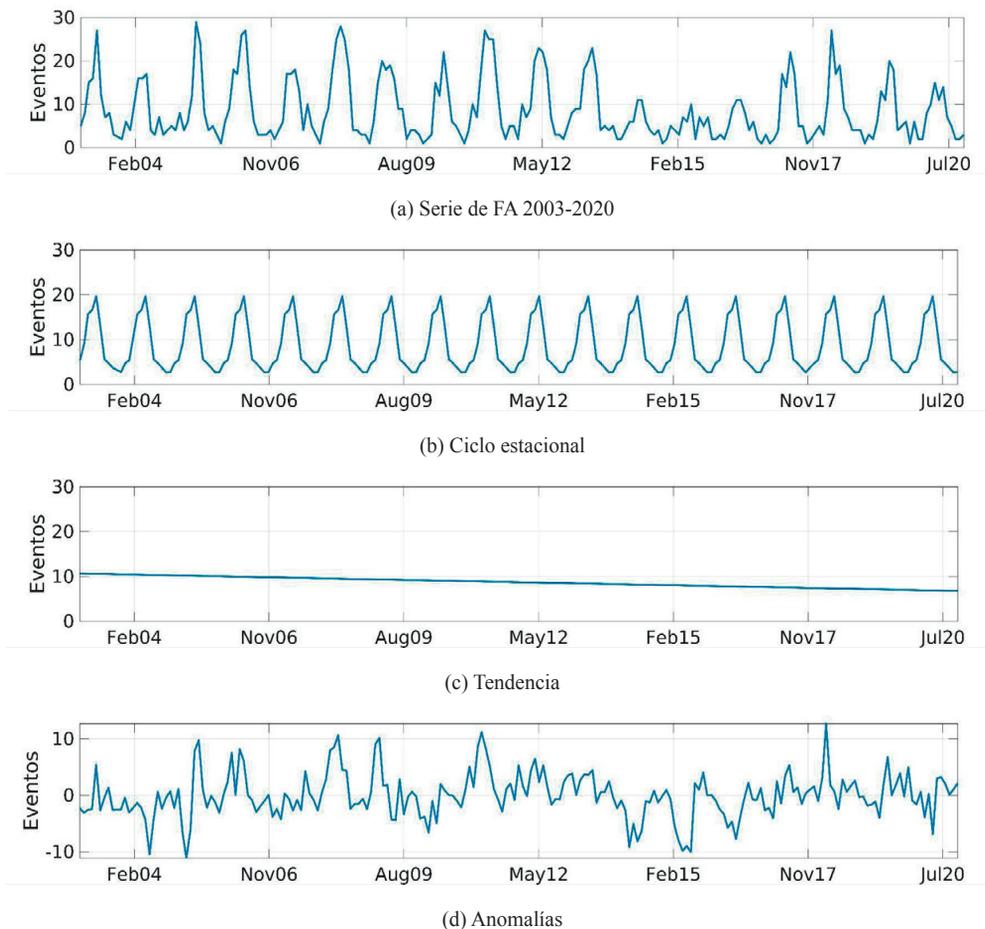


Figura 9. Descomposición de la serie de tiempo de FA obtenidos a partir del algoritmo de clasificación RF empleando datos MODIS, lago SAMAO, 2003-2020.

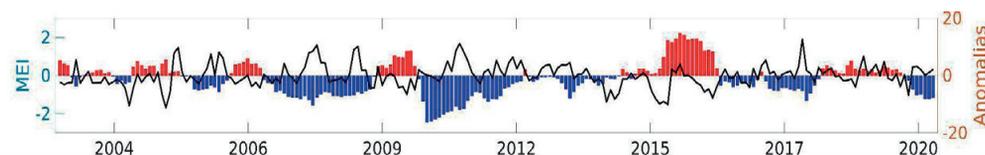


Figura 10. Índice ENSO en correlación con las anomalías en la ocurrencia de eventos FA (período 2003-2020).

índice multivariado de El Niño/Oscilación del Sur (MEI). De modo que las anomalías positivas de FA ocurren durante períodos fríos y anomalías negativas de FA ocurren en períodos cálidos. En la Figura 10, los períodos cálidos (barras en color rojo) o fríos (barras en color azul) que sobrepasan el umbral de $+0,5^{\circ}\text{C}$ o $-0,5^{\circ}\text{C}$ indican la ocurrencia de El Niño o La Niña respectivamente, y las anomalías de eventos FA se representan con una línea sólida. La Figura 11 indica la correlación cruzada entre el índice MEI y las anomalías de

los eventos FA. En el eje x se indica el desplazamiento de la serie por unidad de tiempo donde se evalúan desfases de -20 a $+20$ meses. En el eje y muestra la correlación de las series según su desplazamiento. Esta gráfica nos muestra que la correlación es de $-0,37$ sin ningún desplazamiento. La correlación negativa confirma la influencia de eventos interanuales sobre los FA en SAMAO. La relación del fenómeno ENSO con la productividad primaria en CAC ha sido documentada en Germán *et al.* (2016), donde distinguen oscilaciones en

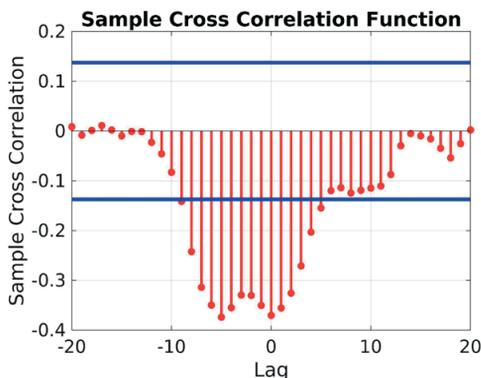


Figura 11. Análisis de correlación cruzada entre el índice MEI y las anomalías de los eventos FA.

los FA que atribuyen a el fenómeno de El Niño. Además, señalan que la tendencia del volumen y cantidad de FA va en aumento lo que difiere de los resultados encontrados en este trabajo donde, aunque con valores bajos, la tendencia va en decrecimiento. Esta disminución en la tendencia puede deberse al desplazamiento de especies en la comunidad biológica del lago hacia especies más pequeñas no formadoras de natas (*scum*) y por lo tanto definiendo diferentes propiedades ópticas, lo que requiere de una nueva caracterización y modelación. Sin embargo, los estudios de identificación de especies de fitoplancton en SAMAO indican un cambio en la comunidad que, a su vez, genera cambio en las propiedades ópticas del lago.

4. Conclusiones

Los productos diarios de MODIS proporcionan series temporales largas con las cuales se logró extraer la señal del ciclo anual y la variabilidad interanual (2003-2020) de los FA. Los productos MOD09GQv006/MYD09GQv006 con resolución de 250 m por píxel permitieron realizar estudios históricos de florecimientos algales en cuerpos de agua continentales con superficies relativamente pequeñas ($\approx 4 \text{ km}^2$).

El ciclo anual de los florecimientos algales en el lago-cráter de SAMAO presenta su máxima frecuencia entre mayo y un mínimo en octubre, donde la mayor incidencia se observa en la región sureste del lago. Por su parte, el análisis interanual de las anomalías en la frecuencia de florecimientos algales presentó una distribución temporal

inversa a las anomalías del índice MEI, lo que indica la influencia del ENSO en la ocurrencia de FA. Se observa una mayor presencia de FA durante los períodos de La Niña, mientras que durante los períodos El Niño hay un descenso en la ocurrencia media de eventos. La tendencia de largo período en la serie temporal indica que los eventos FA en SAMAO propenden a su reducción. Esta reducción pudiera ser indicativo de cambios en la respuesta de la comunidad que sostiene los florecimientos algales en el cuerpo de agua.

Para robustecer estos resultados y corroborar las observaciones de relación entre ambos fenómenos es necesario continuar monitoreando las concentraciones de Chl-*a* en SAMAO, así como obtener mediciones de temperatura y precipitación *in-situ*.

Referencias

- Aghababaei, M., Ebrahimi, A., Naghipour, A.A., Asadi, E., Pérez-Suay, A., Morata, M., & Verrelst, J. 2022. Introducing artmo's machine-learning classification algorithms toolbox: Application to plant-type detection in a semi-steppe iranian landscape. *Remote Sensing*, 14(18), 4452. <https://doi.org/10.3390/rs14184452>
- Ananias, P.H.M., Negri, R.G., Dias, M.A., Silva, E.A., & Casaca, W. 2022. A fully unsupervised machine learning framework for algal bloom forecasting in inland waters using modis time series and climatic products. *Remote Sensing*, 14(17), 4283 <https://doi.org/10.3390/rs14174283>
- Armienta, M.A., Vilaclara, G., De la Cruz-Reyna, S., Ramos, S., Cenicerros, N., Cruz, O., Arcega-Cabrera, F. 2008. Water chemistry of lakes related to active and inactive mexican volcanoes. *Journal of Volcanology and Geothermal Research*, 178(2), 249-258. <https://doi.org/10.1016/j.jvolgeores.2008.06.019>
- Breiman, L. 2001. Random forests. *Machine learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
- Caicedo, J.P.R., Verrelst, J., Muñoz-Marí, J., Moreno, J., & Camps-Valls, G. 2014. Towarda semiautomatic machine learning retrieval of biophysical parameters. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(4), 1249-1259 <https://doi.org/10.1109/JSTARS.2014.2298752>
- Carlson, R.E. 1977. A trophic state index for lakes 1. *Limnology and Oceanography*, 22(2), 361-369. <https://doi.org/10.4319/lo.1977.22.2.0361>

- Carpenter, S.R., Stanley, E.H., & Vander Zanden, M.J. 2011. State of the world's freshwater ecosystems: physical, chemical, and biological changes. *Annual review of Environment and Resources*, 36, 75-99 <https://doi.org/10.1146/annurev-environ-021810-094524>
- Congalton, R.G., & Green, K. 2019. *Assessing the accuracy of remotely sensed data: principles and practices*. CRC press <https://doi.org/10.1201/9780429052729>
- Cortés-Macías, L.Z. 2018. *Validación y calibración del algoritmo OC2 para Landsat 8 aplicado al lago cráter de Santa María del Oro, Nayarit*.
- Dörnhöfer, K., & Oppelt, N. 2016. Remote sensing for lake research and monitoring-recent advances. *Ecological Indicators*, 64, 105-122 <https://doi.org/10.1016/j.ecolind.2015.12.009>
- Eleveld, M.A., Ruescas, A.B., Hommersom, A., Moore, T.S., Peters, S.W., & Brockmann, C. 2017. An optical classification tool for global lake waters. *Remote Sensing*, 9(5), 420. <https://doi.org/10.3390/rs9050420>
- Fisher, R.A. 1936. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2), 179-188. <https://doi.org/10.1111/j.1469-1809.1936.tb02137.x>
- Flach, P.A., & Lachiche, N. 2004. Naive bayesian classification of structured data. *Machine learning*, 57(3), 233-269. <https://doi.org/10.1023/B:MACH.0000039778.69032.ab>
- German, A., Andreo, V., Tauro, C., Scavuzzo, C.M., & Ferral, A. 2020. A novel method based on time series satellite data analysis to detect algal blooms. *Ecological Informatics*, 59, 101131. <https://doi.org/10.1016/j.ecoinf.2020.101131>
- Germán, A., Tauro, C., Andreo, V., Bernasconi, I., & Ferral, A. 2016. Análisis de una serie temporal de clorofila-a a partir de imágenes modis de un embalse eutrófico. En *2016 IEEE Biennial Congress of Argentina (argencon)* (pp. 1-6). <https://doi.org/10.1109/ARGENCON.2016.7585365>
- Germán, A., Tauro, C., Scavuzzo, M.C., & Ferral, A. 2017. Detection of algal blooms in a eutrophic reservoir based on chlorophyll-a time series data from modis. En *2017 IEEE international geoscience and remote sensing symposium (IGARSS)* (pp. 4008-4011). <https://doi.org/10.1109/IGARSS.2017.8127879>
- Gitelson, A.A., Dall'Olmo, G., Moses, W., Rundquist, D.C., Barrow, T., Fisher, T.R.,... Holz, J. 2008. A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: Validation. *Remote Sensing of Environment*, 112(9), 3582-3593. <https://doi.org/10.1016/j.rse.2008.04.015>
- Hamilton, J.D. 2020. *Time series analysis*. Princeton university press. <https://doi.org/10.2307/j.ctv14jx6sm>
- Hovis, W.A., & Leung, K. 1977. Remote sensing of ocean color. *Optical Engineering*, 16(2), 158-166. <https://doi.org/10.1117/12.7972093>
- Hsiao, S.I. 1988. Spatial and seasonal variations in primary production of sea ice microalgae and phytoplankton in frobisher bay, arctic canada. *Marine Ecology Progress Series*, 275-285. <https://doi.org/10.3354/meps044275>
- Goodfellow, I., Bengio, Y., & Courville, A. 2016. *Deep learning*. MIT press.
- Hu, C., Lee, Z., Ma, R., Yu, K., Li, D., & Shang, S. 2010. Moderate resolution imaging spectroradiometer (MODIS) observations of cyanobacteria blooms in Taihu Lake, China. *Journal of Geophysical Research: Oceans*, 115(C4). <https://doi.org/10.1029/2009JC005511>
- Huang, C., Li, Y., Yang, H., Sun, D., Yu, Z., Zhang, Z.,... & Xu, L. 2014. Detection of algal bloom and factors influencing its formation in Taihu Lake from 2000 to 2011 by MODIS. *Environmental earth sciences*, 71, 3705-3714. <https://doi.org/10.1007/s12665-013-2764-6>
- Jia, T., Zhang, X., & Dong, R. 2019. Long-term spatial and temporal monitoring of cyanobacteria blooms using modis on google earth engine: A case study in taihu lake. *Remote Sensing*, 11(19), 2269. <https://doi.org/10.3390/rs11192269>
- Klima, E.F., & Roe, R.B. 1972. Report of the national marine fisheries service southeast fisheries center, pascagoula laboratory, fiscal years 1970 and 1971.
- Kobayashi, S., Ota, Y., Harada, Y., Ebata, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., Takahashi, K., 2015. The JRA-55 Reanalysis: general specifications and basic characteristics. *J. Meteor. Soc. Jpn.*, 93, 5-48. <https://doi.org/10.2151/jmsj.2015-001>
- Li, J., Gao, M., Feng, L., Zhao, H., Shen, Q., Zhang, F.,... Zhang, B. 2019. Estimation of chlorophyll-a concentrations in a highly turbid eutrophic lake using a classification-based modis land-band algorithm. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(10), 3769-3783. <https://doi.org/10.1109/JSTARS.2019.2936403>
- Lynch, A.J., Cooke, S.J., Deines, A.M., Bower, S.D., Bunnell, D.B., Cowx, I.G.,... others 2016. The social, economic, and environmental importance of inland fish and fisheries. *Environmental Reviews*, 24(2), 115-121. <https://doi.org/10.1139/er-2015-0064>
- Masocha, M., Dube, T., Nhwatiwa, T., & Choruma, D. 2018. Testing utility of landsat 8 for remote assessment of water quality in two subtropical african reservoirs with contrasting trophic states. *Geocarto International*, 33(7), 667-680. <https://doi.org/10.1080/010106049.2017.1289561>

- Moore, T.S., Dowell, M.D., Bradt, S., & Verdu, A.R. 2014. An optical water type framework for selecting and blending retrievals from bio-optical algorithms in lakes and coastal waters. *Remote sensing of environment*, 143, 97-111. <https://doi.org/10.1016/j.rse.2013.11.021>
- Moses, W.J., Sterckx, S., Montes, M.J., De Keukelaere, L., & Knaeps, E. 2017. Atmospheric correction for inland waters. En *Bio-optical modeling and remote sensing of inland Waters* (pp. 69-100). Elsevier. <https://doi.org/10.1016/B978-0-12-804644-9.00003-3>
- Moss, B. 2012. Cogs in the endless machine: lakes, climate change and nutrient cycles: a review. *Science of the Total Environment*, 434, 130-142. <https://doi.org/10.1016/j.scitotenv.2011.07.069>
- Muñoz-Marí, J., & Camps-Valls, G. 2013. Simpleclass: Simple classification toolbox [Manual de software informático]. Descargado de <https://github.com/IPL-UV/simpleClass> (accessed October 21, 2020).
- Oliva-Martínez, M.G., Godínez-Ortega, J.L., & Zuñiga-Ramos, C.A. 2014. Biodiversidad del fitoplancton de aguas continentales en México. *Revista mexicana de biodiversidad*, 85, 54-61. <https://doi.org/10.7550/rmb.32706>
- Paerl, H.W., & Millie, D.F. 1996. Physiological ecology of toxic aquatic cyanobacteria. *Phycologia*, 35(sup6), 160-167.
- Pal, M. 2005. Random forest classifier for remote sensing classification. *International journal of remote sensing*, 26(1), 217-222. <https://doi.org/10.2216/i0031-8884-35-6S-160.1>
- PiSHAROTY, P. 1973. Space technology and oceanography. *MBAI Special Publication dedicated to Dr. NK Panikkar* (1), 46-51.
- Pizzolon, L. 1996. Importancia de las cianobacterias como factor de toxicidad en las aguas continentales. *Interciencia*, 21(6), 239-245.
- Raileanu, L.E., & Stoffel, K. 2004. Theoretical comparison between the gini index and information gain criteria. *Annals of Mathematics and Artificial Intelligence*, 41(1), 77-93. <https://doi.org/10.1023/B:AMAI.0000018580.96245.c6>
- Roodschild, M., Gotay Sardiñas, J., Will, A.E., & Rodríguez, S.A. 2019. Optimización de scaled conjugate gradient para frog neural networks. En XX Simposio Argentino de Inteligencia Artificial (ASAI 2019)-JAIIO 48(SALTA).
- Salazar-Alcaraz, I. 2018. Identificación y aislamiento de cianobacterias de un lago cráter tropical (MATHESIS). Universidad Autónoma de Nayarit.
- Salazar-Alcaraz, I., Ochoa-Zamora, G.G., Hernández-Almeida, O.U., Palomino-Hermosillo, Y.A., Leyva-Valencia, I., Romero-Bañuelos, C.A., & Cepeda-Morales, J. 2021. Polyphasic assessment of the bloom-forming cyanobacterial species *Limnorphis robusta* (oscillatoriaceae) and *Microcystis aeruginosa* (microcystaceae) in a mexican subtropical crater lake. *Revista mexicana de biodiversidad*, 92 <https://doi.org/10.22201/ib.20078706e.2021.92.3485>
- Serrano, D., Filonov, A., & Tereshchenko, I. 2002. Dynamic response to valley breeze circulation in santa maria del oro, a volcanic lake in mexico. *Geophysical Research Letters*, 29(13), 1-4. <https://doi.org/10.1029/2001GL014142>
- Shaik, A.B., & Srinivasan, S. 2019. A brief survey on random forest ensembles in classification model. En *International conference on innovative computing and communications: Proceedings of ICICC 2018, volume 2*(pp. 253-260). https://doi.org/10.1007/978-981-13-2354-6_27
- Shi, K., Li, Y., Li, L., Lu, H., Song, K., Liu, Z.,... Li, Z. 2013. Remote chlorophyll-a estimates for inland waters based on a cluster-based classification. *Science of the Total Environment*, 444, 1-15. <https://doi.org/10.1016/j.scitotenv.2012.11.058>
- Shi, K., Zhang, Y., Xu, H., Zhu, G., Qin, B., Huang, C.,... Lv, H. 2015. Long-term satellite observations of microcystin concentrations in lake taihu during cyanobacterial bloom periods. *Environmental Science & Technology*, 49(11), 6448-6456. <https://doi.org/10.1021/es505901a>
- Shi, K., Zhang, Y., Zhang, Y., Li, N., Qin, B., Zhu, G., & Zhou, Y. 2019. Phenology of phytoplankton blooms in a trophic lake observed from long-term modis data. *Environmental science & technology*, 53(5), 2324-2331. <https://doi.org/10.1021/acs.est.8b06887>
- Shi, K., Zhang, Y., Zhang, Y., Qin, B., & Zhu, G. 2020. Understanding the long-term tren of particulate phosphorus in a cyanobacteria-dominated lake using modis-aqua observations. *Science of The Total Environment*, 737, 139736. <https://doi.org/10.1016/j.scitotenv.2020.139736>
- Shi, K., Zhang, Y., Zhou, Y., Liu, X., Zhu, G., Qin, B., & Gao, G. 2017. Long-term modis observations of cyanobacterial dynamics in lake taihu: Responses to nutrient enrichment and meteorological factors. *Scientific reports*, 7(1), 1-16. <https://doi.org/10.1038/srep40326>

- Sosa-Nájera, S., Lozano-García, S., Roy, P.D., & Caballero, M. 2010. Registro de sequías históricas en el occidente de México con base en el análisis elemental de sedimentos lacustres: El caso del lago de Santa María del Oro. *Boletín de la Sociedad Geológica Mexicana*, 62(3), 437-451. <https://doi.org/10.18268/BSGM2010v62n3a8>
- Spyrakos, E., O'donnell, R., Hunter, P.D., Miller, C., Scott, M., Simis, S.G.,... others 2018. Optical types of inland and coastal waters. *Limnology and Oceanography*, 63(2), 846-870. <https://doi.org/10.1002/lno.10674>
- Tett, P. 1987. The ecophysiology of exceptional blooms. Rapport et Proces-verbaux des Reunions. *Conseil international pour l'Exploration de la Mer*, 187, 47-60.
- Tharwat, A., Gaber, T., Ibrahim, A., & Hassanien, A.E. 2017. Linear discriminant analysis: A detailed tutorial. *AI communications*, 30(2), 169-190. <https://doi.org/10.3233/AIC-170729>
- The MathWorks, I. 2010. Deep learning toolbox [Manual de software informático]. Natick, Massachusetts, United State. Descargado de <https://www.mathworks.com/help/deeplearning/ref/patternnet.html>
- Tomaselli, L., & cols. 2004. The microalgal cell. *Handbook of microalgal culture: Biotechnology and applied phycolgy*, 1, 3-19 <https://doi.org/10.1002/9780470995280.ch1>
- Verhoef, W. 1996. Application of harmonic analysis of ndvi time series (hants). *Fourier análisis of temporal NDVI in the Southern African and American continents*, 108, 19-24.
- Vermote, Eric, y Wolfe, Robert. 2015. MOD09GQ MODIS/Terra Surface Reflectance Daily L2G Global 250m SIN Grid V006. NASA EOSDIS Land Processes DAAC. Descargado 2021-07-22, de: <https://lpdaac.usgs.gov/products/mod09gqv006/> (Type: dataset) doi:10.5067/MODIS/MOD09GQ.006
- Wang, Q., Ma, Y., Zhao, K., & Tian, Y. 2022. A comprehensive survey of loss functions in machine learning. *Annals of Data Science*, 9(2), 187-212 <https://doi.org/10.1007/s40745-020-00253-5>
- Wang, S., Li, J., Zhang, B., Spyrakos, E., Tyler, A.N., Shen, Q., Zhang, F., Kuster, T., Lehmann, M.K., Wu, Y., Peng, D. 2018. Trophic state assessment of global inland waters using a modis-derived forel-ule index. *Remote Sensing of Environment*, 217, 444-460. <https://doi.org/10.1016/j.rse.2018.08.026>
- Wolter, K., & Timlin, M.S. 1993. Monitoring enso in coads with a seasonally adjusted principal component index. En *Proceedings of the 17th Climate Diagnostics Workshop, Norman, OK, NOAA/NMC/CAC, NSSL, Oklahoma Clim. Survey, Cimms And The School Of Meteor.*, univ. Of Oklahoma (Vol. 52).
- Xiang, S., Nie, F., & Zhang, C. 2008. Learning a mahalanobis distance metric for data clustering and classification. *Pattern recognition*, 41(12), 3600-3612 <https://doi.org/10.1016/j.patcog.2008.05.018>
- Xing, X.-G., Zhao, D.-Z., Liu, Y.-G., Yang, J.-H., Xiu, P., & Wang, L. 2007. An overview of 508 remote sensing of chlorophyll fluorescence. *Ocean Science Journal*, 42, 49-59.509 <https://doi.org/10.1007/BF03020910>