

Simulación y modelación matemática de los parámetros fisicoquímicos del agua del Lago de Atitlán

Gaspar Yatáz¹ y Cristian Valdez²

¹Departamento de Matemática, Universidad del Valle de Guatemala

¹gyataz@uvg.edu.gt

²Departamento de Física, Universidad del Valle de Guatemala

RESUMEN: Se presenta un estudio de los parámetros fisicoquímicos del agua del Lago de Atitlán. Se analizaron cinco sitios de muestreo con información histórica recolectada por la Autoridad para el Manejo Sustentable de la Cuenca del Lago de Atitlán desde 2014 hasta 2017, con la cual se simularon y modelaron los parámetros más importantes. El propósito del estudio fue describir la interacción de los parámetros fisicoquímicos con la transparencia, explicar la influencia de los parámetros más importantes con la transparencia en cada sitio y pronosticar la transparencia para el año 2018. Se observó que en los sitios Centro WG y San Pedro todos los factores fisicoquímicos poseían una interacción mayor para el modelado de la transparencia; los parámetros temperatura, clorofila α , salinidad y oxígeno disuelto estaban fuertemente relacionados con la transparencia en la mayoría de los sitios y que el sitio Bahía Santiago tendría la transparencia media más baja con 4.98 ± 0.05 (EE) metros, mientras que San Pedro tendría la transparencia media más alta con 6.78 ± 0.07 (EE) metros para el año 2018. La transparencia es un factor físico importante y práctico para determinar la calidad del Lago de Atitlán; resulta ser un buen candidato como predictor para los parámetros que requieren de mediciones sofisticadas como temperatura, clorofila α y sólidos disueltos totales.

PALABRAS CLAVE: Parámetros fisicoquímicos, transparencia, pronóstico, Lago Atitlán, simulación y modelación.

Simulation and mathematical modeling of the water physicochemical parameters of Lake Atitlan

ABSTRACT: A study of the physicochemical parameters of the Lake Atitlan water is presented. Five sampling sites with historical information collected by the Authority for the Sustainable Management of the Lake Atitlan Basin since 2014 to 2017 were analyzed, and the most important parameters were simulated and modeled. The purpose of the study was to describe the interaction of the physicochemical parameters with the water transparency, to explain the influence of the most important parameters with the transparency in each site and to forecast transparency for the year 2018. It was observed that in the Centro WG and San Pedro sites all physicochemical factors had a greater interaction for modeling transparency; the parameters temperature, chlorophyll α , salinity and dissolved oxygen were strongly related to transparency in most of the sites and the Bahia Santiago site will have the predicted lowest average transparency with 4.98 ± 0.05 (SE) meters, while San Pedro will have the predicted highest average transparency with 6.78 ± 0.07 (SE) meters for the year 2018. Transparency is an important and practical physical factor to determine the water quality of Lake Atitlan; it's a good candidate as a predictor for parameters that require sophisticated measurements such as temperature, chlorophyll α and total dissolved solids.

KEY WORDS: Physicochemical parameters, transparency, forecast, Lake Atitlan, simulation and modeling.

Introducción

Se estudia la interacción de los factores fisicoquímicos del Lago de Atilán, variables independientes x_1, x_2, \dots, x_p donde p es tal que $1 \leq p \leq 12$, con la transparencia del agua del Lago de Atilán (variable dependiente Y). Se define para la primera parte del estudio que Y puede expresarse como

$$Y = f(X) + \epsilon \quad (1)$$

Siendo Y, ϵ escalares; tal que $X = (x_1, x_2, \dots, x_p)$ es un vector de $1 \times p$, en general para n observaciones se puede expresar Y como un vector de $n \times 1$ y X como una matriz de $n \times p$, y ϵ como un vector de error de $n \times 1$. Con ello se deduce un modelo \hat{Y} que aproxime Y con la teoría de aprendizaje estadístico (Hastie *et al.*, 2008; Seber y Lee, 2003). Las dos razones fundamentales que motivan estimar f es la predicción e inferencia, útiles para explicar la transparencia del agua del Lago de Atilán con los factores incidentes. Obtener un modelo Y que explique adecuadamente la transparencia del agua del Lago de Atilán, requiere de una predicción precisa, que acople las variables explicativas de la mejor forma; sin embargo, en la realidad envuelve complejidad en su desarrollo, para ello se utilizan técnicas de la estadística matemática para simplificar y conducir con claridad el proceso de la deducción del modelo (Hastie, Tibshirani y Friedman, 2008). El estudio parte en analizar la data disponible, registrada por el Departamento de Investigación y Calidad Ambiental de la Autoridad para el Manejo Sustentable de la Cuenca del Lago de Atilán y su Entorno (DICA-AMSCLAE); pero principalmente sobre el tratamiento de datos que se aplica a los parámetros de interés y que se define como la simulación mensual. En el mejor de los casos, si se tiene que ϵ en promedio es cero, la predicción para Y adopta la forma

$$\hat{Y} = \hat{f}(X) \quad (2)$$

siendo \hat{f} la estimación de f . La exactitud de \hat{Y} como una predicción para Y depende de dos conceptos importantes: el error reducible y el error irreducible. En la teoría de predicción se sabe que \hat{f} no es un estimador perfecto de f dado que involucra un error en su estimación; sin embargo, este error es reducible. Sea \hat{f} una estimación tal que

$$\hat{y} = \hat{f}(X) \quad (3)$$

con \hat{f} y X fijos. Considere el valor esperado de $(y - \hat{y})^2$, esto es

$$\begin{aligned} E(y - \hat{y})^2 &= E[f(X) + \epsilon - \hat{f}(X)]^2 \\ &= [f(X) - \hat{f}(X)]^2 + Var(\epsilon) \quad (4) \end{aligned}$$

con $Var(\epsilon)$ la varianza del término de error. La importancia del cálculo anterior es resaltar que $[f(X) + \epsilon - \hat{f}(X)]^2$ involucra un error reducible, mientras que el error de $Var(\epsilon)$ es irreducible (Hastie *et al.*, 2008). Para la estimación de y , del modelo de regresión lineal múltiple, se aplica el método de mínimos cuadrados con el objetivo de minimizar el error reducible involucrado. Además se desea describir cómo la transparencia del Lago de Atilán cambia como función de x_1, x_2, \dots, x_p , principalmente se quiere responder a las siguientes preguntas: ¿Cuáles de los parámetros fisicoquímicos están asociados con la transparencia?, ¿Cuál es la relación entre la transparencia con cada parámetros fisicoquímico?, ¿Puede la relación entre la transparencia y cada parámetro fisicoquímico ser descrita resumidamente usando un modelo lineal, o es la relación más complicada?. Luego se plantea al conocer las relaciones entre los parámetros fisicoquímicos importantes con la transparencia, explicar cómo estas actúan cuando aumentan un cierto porcentaje. La simplicidad del modelo Cobb-Douglas (Ioan y Ioan, 2015; Charnes, Cooper y Schinnar, 1976; Brown, 1970), ayuda a explicar las elasticidades de los parámetros fisicoquímicos del agua del lago de Atilán, ya que con ellas se tiene una estimación sobre la reducción o aumento en la transparencia.

Se define la función Cobb-Douglas como

$$f : D \subset R_+^n \rightarrow R^+$$

$$f(x_1, x_2, \dots, x_n) = Ax^{\alpha_1} \cdot x^{\alpha_2} \cdot \dots \cdot x^{\alpha_n} \quad (5)$$

Con $\alpha_1, \alpha_2, \dots, \alpha_n$ constantes positivas (Ioan y Ioan, 2015). Las constantes α_i , con $i = 1, 2, \dots, n$ para un n fijo, suelen llamarse elasticidades del i -ésimo factor de entrada, que juegan un papel importante en la aplicación. Por lo que cualquier función de producción que adopte la forma de la función Cobb-Douglas se dice que es una función de producción de Cobb-Douglas. Se deducen algunos hechos útiles, al aplicar el \ln en ambos lados a (5) se tiene

$$\ln(f(x_1, x_2, \dots, x_n)) = \ln(Ax^{\alpha_1} \cdot x^{\alpha_2} \cdot \dots \cdot x^{\alpha_n}) \quad (6)$$

$$= \ln(A) + \sum_{i=1}^n \alpha_i \ln(x_i) \quad (7)$$

haciendo $y = \ln(f)$, $\alpha_0 = \ln(A)$, $Z_i = \ln(x_i)$ para $i = 1, 2, \dots, n$ se obtiene una simplificación de (5)

$$y = \alpha_0 + \sum_{i=1}^n z_i \alpha_i \quad (8)$$

que es básicamente una regresión lineal, precisamente un modelo log-lineal. Con (8) se pueden estimar los valores α_i con el método de mínimos cuadrados, dado que y, x_1, x_2, \dots, x_n poseen valores observados. Si $\sum_{i=1}^n \alpha_i = 1$, entonces se estará hablando de una función de producción inelástica, mientras que si $\sum_{i=1}^n \alpha_i < 1$ o $\sum_{i=1}^n \alpha_i > 1$ es una función elástica. Considere la derivada parcial de (7) sobre $\ln(x_i)$

$$\frac{\partial \ln(f)}{\partial \ln(x_i)} = \frac{x_i}{f} \frac{\partial f}{\partial x_i} \quad (9)$$

en la literatura $\frac{\partial f}{\partial x_i}$ suele llamarse la productividad marginal (Christensen y Lau, 1973). Una consideración a la teoría es imponer a (5) la condición de elasticidad:

$$\sum_{i=1}^n \alpha_i = 1 \quad (10)$$

que en conjunto forman un problema de programación no lineal; Ioan y Ioan (Ioan y Ioan, 2015) abordan las condiciones suficientes y necesarias para considerar, y transformar (5) en un problema de programación no lineal de optimización, el cual puede resolverse con algoritmos computacionales ya establecidos, como la técnica de programación cuadrática (Turlach, 2015). El primer enfoque del desarrollo de los modelos Cobb-Douglas es el de obtener las constantes de elasticidad, considerando que el problema de estudio posee otros parámetros influyentes de los cuales no se cuenta con información para realizar aprendizaje estadístico, siguiendo ese contexto el valor de A de tecnología de producción es distinto de la unidad. En el segundo enfoque, se considera a las variables explicativas, seleccionadas por la relación fuerte y media con la transparencia, como los factores que influyen en la transparencia; este estudio permite explicar cuál de los parámetros fisicoquímicos son los que determinan la transparencia en cada sitio de muestreo (Charnes, Cooper y Schinnar, 1976). Con ello se busca tener una visión más amplia sobre la dinámica de la transparencia y su relación con factores fisicoquímicos en los sitios de estudio.

Por último, se aplican las técnicas de series de tiempo a los parámetros fisicoquímicos. Las series de tiempo son una herramienta fundamental para el estudio de los procesos estocásticos; la absorción de elementos de estacionalidad, autoregresión y estacionalidad de un fenómeno son fundamentales para su predicción (Cowpertwait y Meltcalfe, 2008; Cryer y Chan, 2008). Un proceso, x_t , se le dice *ARIMA* (p,d,q) si

$$\nabla^d x_t = (1 - B)^d x_t \quad (11)$$

es *ARIMA* (p,q). En general, el modelo se escribirá como

$$\phi(B)(1 - B)^d x_t = \theta(B)w_t \quad (11)$$

si $E(\nabla^d x_t) = \mu$, entonces será de la forma

$$\phi(B)(1 - B)^d x_t = \alpha + \theta(B)w_t$$

donde $\alpha = \mu(1 - \phi_1 - \dots - \phi_p)$ (Shumway y Stoffer, 2006). Existen algunos pasos para el ajuste de modelos arima de las series de tiempo. Estos pasos envuelven colocar los datos en una gráfica respecto del tiempo de los fenómenos de estudios e identificar anomalías, una posible transformación a la data, identificación de posibles dependencias de los órdenes del modelo y los parámetros de estimación (Shumway y Stoffer, 2006; Cowpertwait y Meltcalfe, 2008; Cryer y Chan, 2008). Por ejemplo, si la variabilidad en la data crece con el tiempo, será necesario aplicar una transformación adecuada para estabilizar la varianza, una de esas transformaciones puede ser la de Box-Cox. Con los modelos arima se desean pronosticar los parámetros fisicoquímicos basándose en la selección de modelos con el criterio de AIC.

El deterioro del estado ecológico del Lago de Atitlán ha generado mucha inquietud, principalmente debido a la reducción de su transparencia en los últimos años (Dix *et al.*, 2012). Los brotes de cianobacterias en la superficie del Lago en los años 2009 y 2015 (Figura 1b y Figura 1c) han generado un despertar sobre la situación del estado ecológico del Lago. Los registros de los parámetros fisicoquímicos del Lago de Atitlán son limitados; en particular se obtienen de 61 a 120 (dependiendo el sitio de estudio) observaciones en promedio con información completa de los parámetros de interés (en el sentido del estudio de perfiles o columnas de agua) de los años 2014, 2015, 2016 y 2017 de la data proporcionada por la AMSCLAE. La transparencia es un indicador inmediato de la calidad del agua del Lago de Atitlán. El propósito del este estudio es describir la interacción de los parámetros fisicoquímicos con la transparencia, explicar cuáles de los parámetros son influyentes o determinantes en la transparencia en cada sitio de estudio, y pronosticar la media de la transparencia para el año 2018.

Materiales y métodos

Con la data registrada por la AMSCLAE, se tomaron observaciones que contenían información completa de los parámetros: temperatura (T), nitratos (NO_3^-), fosfatos (PO_4), fósforo total (PT), clorofila α ($Chl - \alpha$), sólidos disueltos totales (TDS), oxígeno disuelto (OD), saturación de oxígeno (sO_2), radiación (PAR), salinidad (Sal), potencial de hidrogeno (pH), conductividad ($Cond$) y transparencia ($Secchi$) para perfiles (desde la superficie hasta una profundidad de 20, 30, 60 y 100 metros) de cada sitio (dependiendo de la información del lugar). Los parámetros fueron registrados desde el 2014 hasta el 2017;

se resalta que las observaciones para el año 2015, eran limitadas. En este estudio la transparencia hace referencia a la medición del disco Secchi.

A. Tratamiento de datos

Definición A.1 Forma conveniente de la desviación estándar. Se define como sd_dirac a la desviación estándar de forma conveniente, tal que es cero si no existe desviación estándar y $sd_dirac = sd$ en otro caso, sd es la desviación estándar.

Definición A.2 Función generadora aleatoria. Se define por función generadora aleatoria a

$$F_i = \frac{f(x)}{2}, \quad i = 1, 2, \dots, 13 \quad (13)$$

donde $f(x)$ es la función generadora aleatoria de la distribución normal con media igual a 0, desviación estándar igual a 1 y con n igual a la cantidad de filas de la simulación; n es un parámetro de $f(x)$ que limita la cantidad de números que pueden generarse.

Definición A.3 Función de simulación mensual. Se define por función de simulación mensual a

$$V_i = M_{V_i} + F_i \cdot SD_{V_i}, \quad i = 1, 2, \dots, 13 \quad (14)$$

Donde V_i es el i -ésimo parámetro de estudio, M_{V_i} es la media mensual del i -ésimo parámetro, F_i es la i -ésima función generadora aleatoria de la distribución normal y SD_{V_i} es la desviación estándar mensual del i -ésimo parámetro calculado con sd_dirac .

Algoritmo A. 4 Simulación de parámetros.

1. Filtrar y limpiar la data hasta obtener m observaciones con información completa de los parámetros V_i de estudio.
2. Agrupar de manera mensual las mediciones de los parámetros V_i , de todos los años, luego calcular la media mensual M_{V_i} y desviación estándar mensual SD_{V_i} .
3. Con la función de simulación mensual realizar la simulación (diaria) de los años 2014, 2015, 2016 y 2017.

Los valores que se obtienen de los parámetros de estudio se denominarán valores de entrenamiento. Todos los análisis se realizarán con los valores de entrenamiento.

B. Aplicación de regresión lineal múltiple

El objetivo central de la aplicación de regresión lineal múltiple es verificar si

$$Secchi = \beta_0 + \sum_{i=1}^{12} x_i \beta_i \quad (15)$$

es un buen predictor para la transparencia del Lago de Atitlán, además de saber cuáles de los parámetros conocidos, dada el conjunto de valores de entrenamiento, están relacionados. Identificar dichas relaciones generará un marco global para estudiar dichos parámetros de manera más puntual. Para abordar sistemáticamente dichas ideas se plantean los siguientes algoritmos.

Algoritmo B.1 Análisis de modelo.

1. Plantear el modelo

$$Secchi = \beta_0 + \sum_{i=1}^{12} x_i \beta_i$$

donde β_0 es el intercepto $x_1 = T, x_2 = Cond, x_3 = Sal, x_4 = TDS, x_5 = NO_3^-, x_6 = PO_4, x_7 = PT, x_8 = Chl-\alpha, x_9 = PAR, x_{10} = OD, x_{11} = sO_2, y x_{12} = pH$.

2. Aplicar el criterio de AIC al modelo.
3. Luego al modelo con menor AIC encontrado obtener su R^2 , y p -valor de significancia.
4. Aplicar la prueba de Breusch-Pagan al modelo encontrado.
5. Aplicar la prueba de Durbin-Watson a los residuos.
6. Aplicar la prueba de Kolmogorov-Smirnov a los residuos.
7. Aplicar el criterio Vif para multicolinealidad.
8. Graficar análisis de residuos.
9. Identificar puntos atípicos e influyentes en la data de entrenamiento.

Algoritmo B.2 Correlaciones.

1. Graficar las correlaciones entre los parámetros.
2. Calcular la matriz de correlaciones.
3. Clasificar las correlaciones entre:

relación débil si $0.4 \leq \lambda < 0.49$, relación media si $0.5 \leq \lambda < 0.59$ y fuerte si $0.6 \leq \lambda < 1$, con λ la correlación entre dos parámetros de estudio.

4. Elegir relaciones importantes para modelar.

C. Aplicación de Cobb-Douglas

De los modelos con menor AIC y las relaciones entre parámetros de estudio que se pueden establecer con el Algoritmo B.2, se pueden plantear modelos tipo Cobb-Douglas (Charnes, Cooper y Schinnar, 1976; Christensen, Jorgenson y Lau, 1973; Fletcher y Lu, 1968), con el fin de entender la dinámica de un parámetro de estudio que está descrita en términos de otros; principalmente sobre su respuesta ante cualquier aumento en el nivel de sus variables explicativas, es decir determinadas por las elasticidades de los factores de modelado. Con ello se plantea el siguiente algoritmo.

Algoritmo C.1 Elasticidades.

1. Plantear un modelo Cobb-Douglas con factores (variables explicativas) determinados por el Algoritmo B.1 hasta el paso 2 o por el Algoritmo B.2 a partir de los tipos de relaciones tales que cumplan las condiciones al modelo.
2. Aplicar logaritmo en ambos lados del modelo de Cobb-Douglas planteado.
3. Encontrar los α_i (elasticidades) del modelo log lineal con el método de mínimos cuadrados.

D. Aplicación de series de tiempo

A partir de la data de entrenamiento para la transparencia, se desea saber cómo este parámetro físico del Lago de Atitlán se comportará durante el año 2018, precisamente se desea encontrar un modelo arima que describa dicho comportamiento, buscando que tenga el menor AIC posible, en cada sitio de interés. Para ese objetivo se presenta el algoritmo de construcción de modelos arima.

Algoritmo D.1 Modelos arima para transparencia.

1. Colocar los valores de entrenamiento de la transparencia del agua del Lago de Atitlán en una gráfica.
2. Identificar anomalías en la gráfica.
3. Colocar en una gráfica los valores de ACF y PACF.
4. Identificar el orden de diferencia d .
5. Identificar el orden de autoregresividad p .
6. Identificar el orden de media móvil q .

7. De no tener claridad sobre el comportamiento de los de ACF, PACF con los diferentes lags; aplicar una transformación y luego una diferencia a los valores de entrenamiento de la transparencia.

8. De los posibles modelos, seleccionar el mejor modelo con el criterio de AIC.

9. Realizar la prueba de Box-Pierce a los residuos.

Los cuadros que se exponen en la siguiente sección, contienen los resultados más importantes de la investigación realizada; que principalmente es un producto de la aplicación de los algoritmos desarrollados en esta sección. Las simulaciones, cálculos y figuras fueron realizadas con el software R. El código que se desarrolló e implantó en R de este estudio se basa en las librerías, paquetes y funciones creadas por otros autores. Para la parte de simulación mensual se utilizaron tidyverse, ggplot2 y GGally (Wickham y Grolemund, 2016; Wicklam, 2009; Schloerke et al., 2011). Para las gráficas se utilizó plotly (Sievert et al., 2017). Para la regresión lineal múltiple se emplearon lmtest, car, DAAG, faraway, reshap2 y corplot (Hothorn et al., 2018; Fox et al., 2017; Braun y Maindonald, 2015; Faraway, 2016; Wicklam, 2017). Mientras que para las elasticidades del modelo de Cobb-Douglas se empleó quadprog (Turlach, 2015). Y para los modelos arima se usaron forecast, xts y hexbin (Hyndman et al., 2017; Shumway y Sttoter, 2006; Ulrich et al., 2017; Koh y Maechler, 2018).

Resultados y discusión

Con la data proporcionada por la AMSCLAE se estudiaron los sitios Bahía Santiago (W14), Bahía San Lucas (WD), Panajachel (WP), Centro WG (WG) y San Pedro (WA) (Figura 1a). Los modelos de regresión lineal múltiple encontrados para los sitios de muestreo en el Lago de Atitlán con el criterio de AIC, describen de manera global la transparencia esto es debido al p-valor de los modelos que se aprecia en el Cuadro 1; sin embargo, los modelos encontrados para Bahía Santiago y Bahía San Lucas explican un 68.48% y 60.61% de variabilidad de la transparencia, mayores a los otros tres modelos, esto es con un 95% de confianza. En los sitios Centro WG y San Pedro todos los parámetros estudiados excepto sólidos disueltos totales son componentes importantes en el modelado de la transparencia, como se aprecia en el Cuadro 1 y Cuadro 2; en dichos sitios existe mayor dinámica fisicoquímica con transparencia dado que los modelos analizados con los criterios AIC, AICc y BIC muestran que los parámetros explicativos para transparencia poseen un mayor grado de eficiencia y calidad. Los modelos de regresión lineal múltiple encontrados para la transparencia no son buenos predictores para la transparencia del Lago de Atitlán, debido a que no cumplen todas las condiciones de Gauss-Markov (Cuadro 3). Dicho de otra manera, los estimadores lineales no sesgados que se encontraron no son de mínima varianza; invitan a pensar que dichos parámetros requieren de

Cuadro 1. Constantes de determinación y valores de AIC de los modelos de regresión lineal múltiple

Sitio	R cuadrático	R ajustado	p-valor	AIC inicial	AIC final
Bahía Santiago	0.684	0.682	<2.2E-16	-2476.04	-2481.12
Bahía San Lucas	0.606	0.602	<2.2E-16	-631.46	-632.45
Centro WG	0.443	0.438	<2E-16	-193.00	-193.00
Panajachel	0.406	0.402	<2E-16	-451.00	-454.00
San Pedro	0.408	0.403	<2E-16	-303.00	-303.00

Cuadro 2. Coeficientes lineales de los modelos de regresión

Parámetro	Bahía Santiago	Bahía San Lucas	Centro WG	Panajachel	San Pedro
Temperatura (°C)	0.20	-0.22	-0.09	-0.08	0.15
Conductividad ($\mu S/cm$)	-0.001	0	-0.005	-0.010	0.14
Salinidad (mg/L)	-2.93	-74.80	-2.83	-36.77	3.59
Sólidos Disueltos totales (mg/L)	-7.61	53.15	0	0	6.58
Oxígeno Disuelto %	0	-0.22	8.77	0.03	-0.13
Saturación de oxígeno %	0	-0.012	-0.009	0	-0.005
pH	0	-0.08	-0.28	0.34	0.32
Nitratos ($\mu g/L$)	0.038	0.050	-0.003	0.012	-0.016
Fosfatos ($\mu g/L$)	0.004	-0.037	0.005	0.002	0.014
Fósforo total ($\mu g/L$)	-0.003	0.004	-0.016	-0.016	-0.004
Clorofila α ($\mu g/L$)	-0.09	-0.18	-0.57	-0.12	-0.20
PAR ($\mu Em^{-2} s^{-1}$)	-0.0002	-0.0004	0.0001	0	-0.0001

Cuadro 3. Pruebas estadísticas a los modelos de regresión

Sitio	Breusch-Pagan	Durbin-Watson	Kolmogorov-Smirnov	Multicolinealidad
Bahía Santiago	1.38E-12	<2.2E-16	<2.2E-16	No tiene
Bahía San Lucas	<2.2E-16	<2.2E-16	<2.2E-16	Tiene
Centro WG	<2E-12	<2E-16	<2.2E-16	Tiene
Panajachel	<2E-12	<2.2E-16	<2E-16	No tiene
San Pedro	<2E-12	<2E-16	<2E-16	No tiene

Cuadro 4. Relaciones importantes de los parámetros fisicoquímicos con la transparencia

Sitio	Relaciones
Bahía Santiago	(Secchi: T, NO ₃ -, PO ₄ , Chl - α)
Bahía San Lucas	(Secchi: T DS, OD, sO ₂ , PO ₄)
Centro WG	(Secchi: T, Cond, Sal, OD, sO ₂ , NO ₃ -, PO ₄ , Chl - α)
Panajachel	(Secchi: pH, Sal, Cond, PT)
San Pedro	(Secchi: PT, Chl - α , OD, T DS, Sal, T)

su distribución. La transparencia posee relaciones fuertes con temperatura, clorofila α , salinidad y oxígeno disuelto, mientras que relaciones medias con sólidos disueltos totales, nitratos, fosfatos, fósforo total y conductividad como se aprecia en el cuadro (Cuadro 4). La correlación existente entre clorofila α con nitratos y fosfatos en Bahía Santiago es de -0.55 y -0.45 respectivamente, en Bahía San Lucas -0.45 y -0.55, en Centro WG es de -0.26 y -0.46, para Panajachel -0.46 y -0.41; mientras que en San Pedro -0.45 y -0.43. Implicando que a medida que aumenta la profundidad en cada sitio (por perfil), existe un notable consumo de nitratos y fosfatos a través del fitoplancton.

Cuadro 5. Elasticidades de los modelos Cobb-Douglas, con el primer enfoque

Parámetro	Bahía Santiago	Bahía San Lucas	Centro WG	Panajachel	San Pedro
Temperatura (°C)	0.39	0	0.22	0	0.29
Conductividad ($\mu S/cm$)	0	0	0.03	1.08	0
Salinidad (mg/L)	0	0	0.05	1.76	0.13
Sólidos Disueltos totales (mg/L)	0	0.20	0.	0	0.20
Oxígeno Disuelto %	0	0.26	0.10	0	0.02
Saturación de oxígeno %	0	0.18	0.05	0	0
pH	0	0	0	0.56	0
Nitratos (ug/L)	0.04	0	0.04	0	0
Fosfatos (ug/L)	0.02	0.01	0	0	0
Fósforo total (ug/L)	0	0	0.03	0.002	0.014
Clorofila α (ug/L)	0.06	0	0.06	0	0.08
PAR ($\mu Em^{-2} s^{-1}$)	0	0	0	0	0
A	15.79	28.10	16.11	130.32	1.6

Cuadro 6. Elasticidades de los modelos Cobb-Douglas, con el segundo enfoque

Parámetro	Bahía Santiago	Bahía San Lucas	Centro WG	Panajachel	San Pedro
Temperatura (°C)	0.98	0	0.69	0	0.62
Conductividad ($\mu S/cm$)	0	0	0.27	0.27	0
Salinidad (mg/L)	0	0	0	0.55	0.11
Sólidos Disueltos totales (mg/L)	0	0.99	0	0	0.26
Oxígeno Disuelto %	0	0	0.02	0	0
Saturación de oxígeno %	0	0	0.002	0	0
pH	0	0	0	0.188	0
Nitratos (ug/L)	0.02	0	0.02	0	0
Fosfatos (ug/L)	0.00008	0.000008	0.0001	0	0
Fósforo total (ug/L)	0	0	0	0.00006	0
Clorofila α (ug/L)	0	0	0	0	0.00009
PAR ($\mu Em^{-2} s^{-1}$)	0	0	0	0	0

En los modelos de Cobb-Douglas encontrados (Cuadro 5) para modelos abiertos, a partir de las elasticidades se evidencia que en Bahía Santiago la temperatura es un parámetro influyente pues, un aumento del 10% de temperatura y clorofila α , la transparencia se reduce un 3.9% y 0.65% respectivamente. En la Bahía de San Lucas un aumento del 10% de sólidos totales disueltos, oxígeno disuelto y saturación de oxígeno la transparencia se reduce un 2.09%, 2.60% y 1.9%. En el Centro WG un aumento del 10% de temperatura, oxígeno disuelto y clorofila α , la transparencia se reduce un 2.2%, 1.01% y un 0.6% respectivamente. Mientras que en Panajachel un decremento del 10% de conductividad, salinidad y pH la transparencia se

incrementa un 10.8%, 17.67% y 5.6% respectivamente. Y en San Pedro, un aumento del 10% de temperatura, salinidad y sólidos totales disueltos la transparencia se reduce en un 2.9%, 1.37% y 2.02% respectivamente. En el Cuadro 5 se supone que A representa los otros parámetros fisicoquímicos que inciden en la calidad del agua del lago de Atitlán de los cuales no se cuenta con información. Por lo tanto, en el sitio Panajachel la disminución de los parámetros conductividad, salinidad y pH en porcentaje, aumenta considerablemente la transparencia; de ahí que la transparencia en Panajachel es muy elástica. Si se considera que los modelos de Cobb-Douglas encontrados en el Cuadro 6, modelos cerrados con la condición de $A=1$ y $\sum_{i=1}^n \alpha_i=1$,

Cuadro 7. Modelos arima de pronóstico de transparencia con criterios de selección de los modelos arima.

Sitio	Model	ar1	ar2	ar3	ar4	ma1	ma2	ma3	ma4
Bahía Santiago	(3,0,3)	-0.11	0.13	0.92	0	0.45	0.30	-0.54	0
Bahía San Lucas	(1,0,1)	0.95	0	0	0	-0.08	0	0	0
Panajachel	(3,0,2)	1.42	-0.82	0.36	0	-0.83	0.37	6.00	0
Centro WG	(2,0,2)	0.37	0.16	0	0	0.16	-0.38	0	0
San Pedro	(4,0,4)	0.09	1.68	0.06	-0.85	0.46	-1.27	-0.48	0.45

es decir los parámetros usados para la estimación de estos modelos representan todas las variables que inciden en la transparencia del Lago de Atitlán, muestran que: en Bahía Santiago la temperatura es un parámetro que determina la transparencia, en Bahía San Lucas los sólidos disueltos totales determinan la transparencia, mientras que en el Centro WG la temperatura y conductividad son factores influyentes, para Panajachel la salinidad es más influyente seguido de la conductividad y el pH. Y en San Pedro la temperatura, los sólidos disueltos totales y salinidad son los factores más influyentes en la transparencia.

El modelo arima, para el pronóstico de la transparencia, de Bahía Santiago posee el menor AIC, AICc y BIC (903, 904 y 945 respectivamente). Los modelos arima encontrados, que se exponen en el Cuadro 7, describen modelos estacionarios todos con un p-valor mayor a 0.9 dado por la prueba de Box Pierce. Del pronóstico de transparencia para el año 2018 para los sitios estudiados: Bahía Santiago, Centro WG, Panajachel, Bahía San Lucas y San Pedro tendrán una media en transparencia de 4.984 ± 0.05 (EE), 6.07 ± 0.07 (EE), 6.085 ± 0.09 (EE), 6.642 ± 0.06 (EE) y 6.787 ± 0.07 (EE) metros respectivamente. Notando que Bahía Santiago es el más bajo y San Pedro el más alto. Se espera que la mejor predicción para transparencia por sitio sea el de Bahía Santiago, debido a que posee el menor AIC, AICc y BIC. Se resalta que la temperatura es un parámetro interesante de estudio con la transparencia; es un parámetro que influye en el comportamiento de la transparencia, precisamente ha sido un factor influyente en la transparencia en tres de los cinco sitios estudiados. Por lo que sugiere plantearse cómo la temperatura a través del tiempo ha tenido impacto en la transparencia y si el deterioro del estado ecológico del Lago de Atitlán está ligado al cambio climático, principalmente en la aceleración del deterioro de su estado ecológico. La transparencia es una variable importante y práctica para determinar la calidad del agua del Lago de Atitlán; dada las relaciones encontradas en el Cuadro 4, resulta ser un buen candidato como predictor para los parámetros que requieren de mediciones sofisticadas como clorofila α , fosfatos, nitratos, temperatura y sólidos disueltos totales; esto puede facilitar alternativas para la toma de datos, además de acoplar diseños experimentales para el estudio global y uniforme de los sitios de muestreo del Lago de Atitlán. Una

ilustración sobre el análisis realizado con los algoritmos anteriores para los sitios de estudio puede apreciarse desde de la Figura 2a hasta la Figura 2h que corresponden al sitio Centro WG.

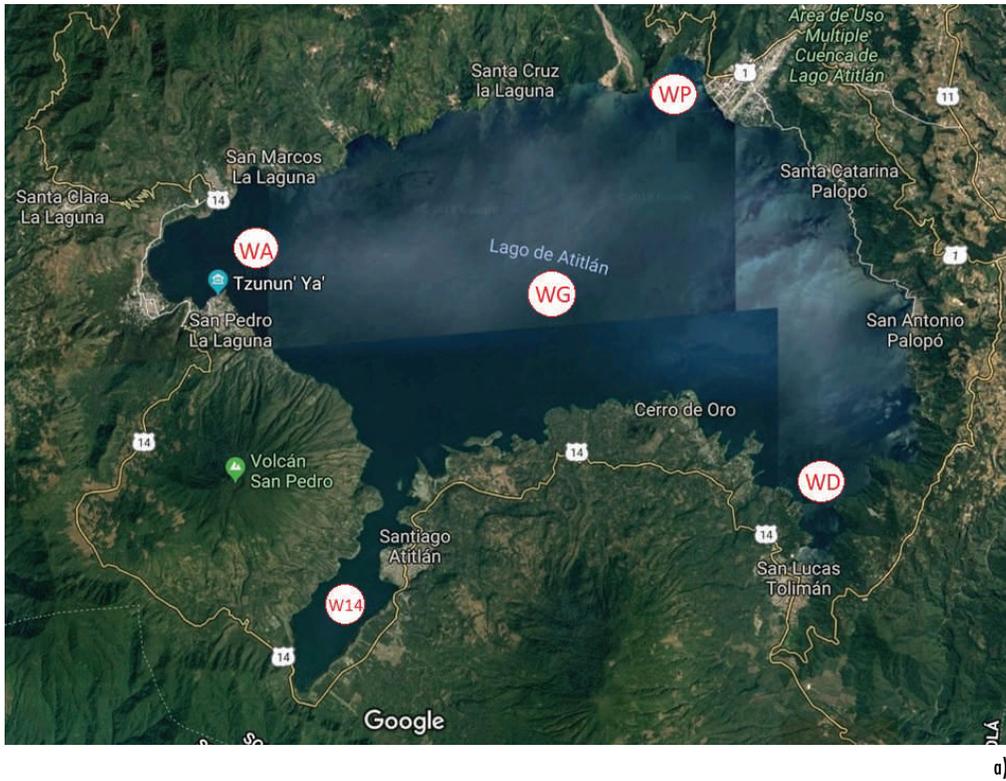
Conclusiones

En los sitios Centro WG y San Pedro todos los factores fisicoquímicos poseen una interacción mayor para el modelado de la transparencia, los parámetros temperatura, clorofila α , salinidad y oxígeno disuelto están fuertemente relacionados con la transparencia en la mayoría de los sitios; existe un notable consumo de nitratos y fosfatos a través del fitoplancton en los perfiles de agua en cada sitio a través del tiempo, debido a las correlaciones encontradas de nitratos y fosfatos con clorofila α . El sitio Bahía Santiago tendrá la transparencia media más baja con 4.98 ± 0.05 (EE) metros, mientras que San Pedro tendrá la transparencia media más alta con 6.78 ± 0.07 (EE) metros para el año 2018. Los modelos de regresión lineal múltiple encontrados para la transparencia, poseen estimadores lineales no sesgados que no son de mínima varianza. La temperatura es un parámetro importante de estudio en la transparencia del agua del Lago de Atitlán, debido a que ha sido un factor influyente en la transparencia en tres de los cinco sitios estudiados. El modelo arima de pronóstico del sitio Bahía Santiago posee mayor calidad y eficiencia comparado con los otros modelos. El estado ecológico del Lago de Atitlán no es un problema homogéneo; como se evidencia con la transparencia (Figura 1d) no es posible definir, inferir y generalizar la calidad del agua del Lago de Atitlán sólo con un sitio de estudio.

Agradecimiento

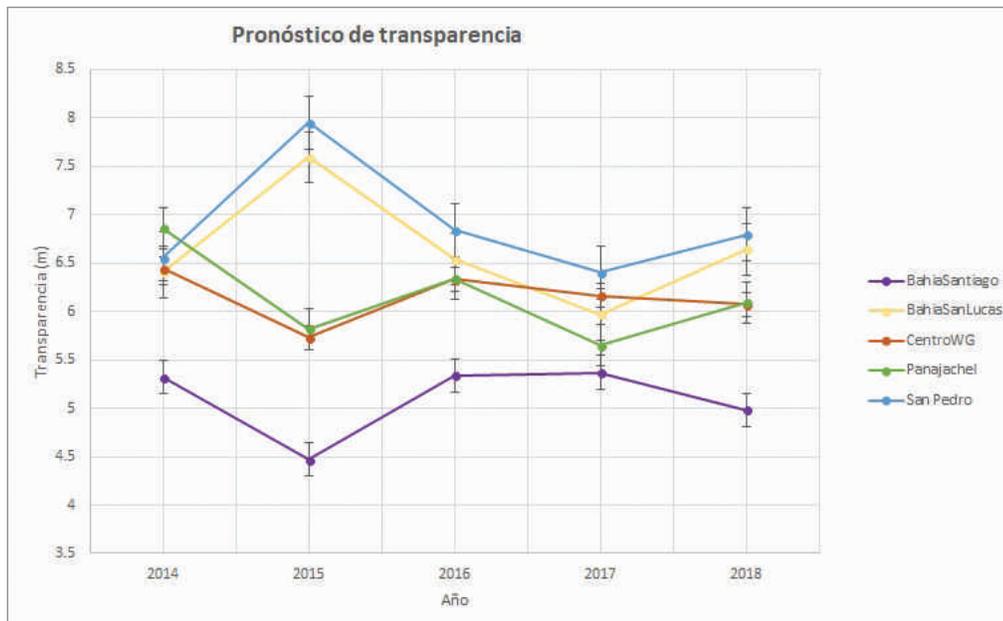
Agradecemos a la Universidad del Valle de Guatemala por crear un espacio de aprendizaje e iniciación en tan bello camino del pensamiento matemático, que es de por sí un todo, que surge de la nada; a Fátima Reyes del DICA - AMSCLAE por su asistencia en la obtención de licencia de investigación para la facilitación de la data estudiada y a Hugo Villavicencio por las sugerencias de búsqueda de información relativa al Lago de Atitlán.

Figura 1. a-d. a) Sitios de estudios, elaborado con Google imágenes/CNES/AIRBUS/LANDSAT, 2018; b) estado del Lago de 2009, elaborado con ASTER/NASA, 2009; c) estado del Lago, 2015, elaborado con LANDSAT/SERVIR&GLOBAL, 2015 y d) pronóstico de la media de transparencia de los sitios estudiados, con base en los datos de AMSCLAE.



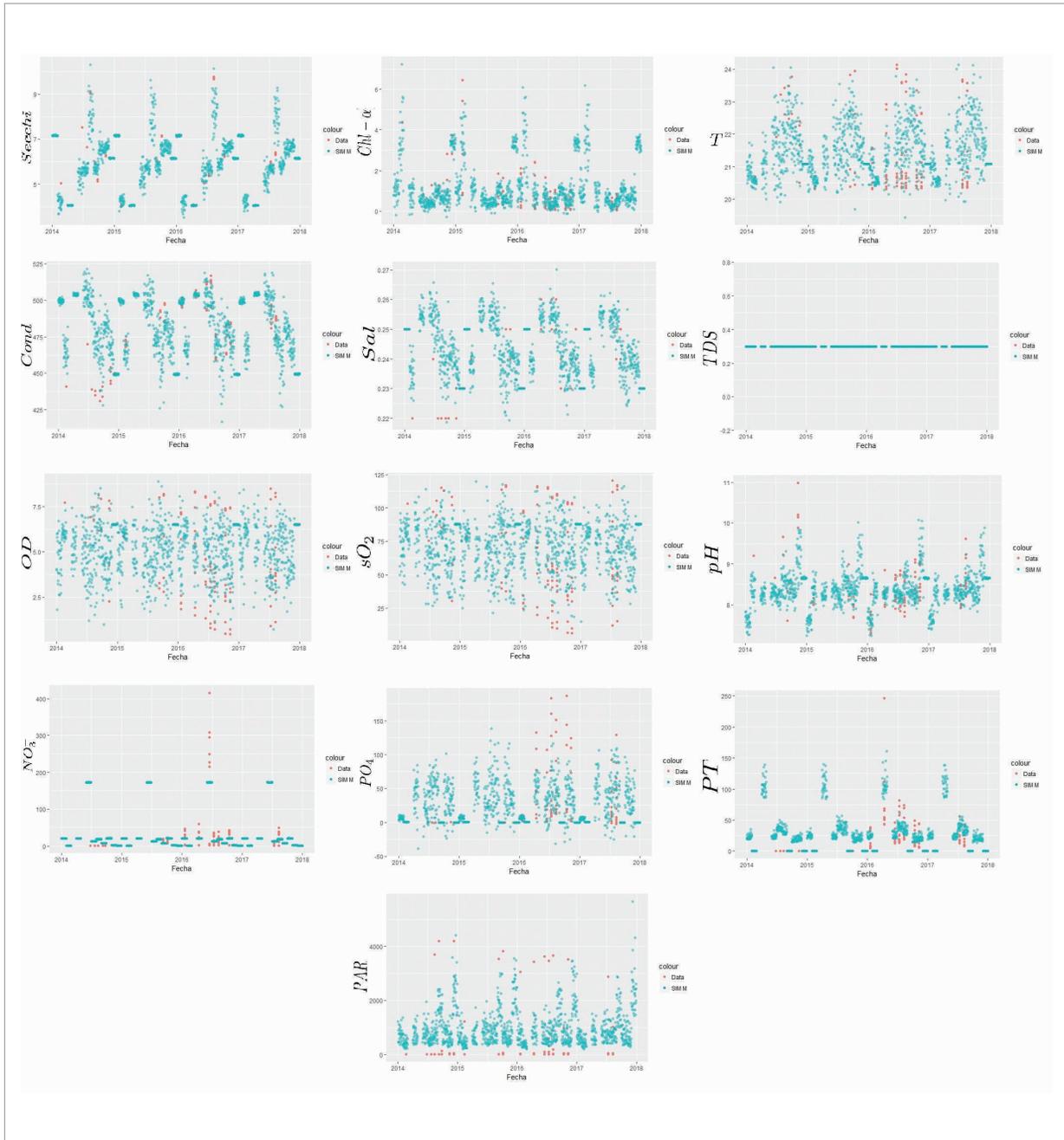


c)

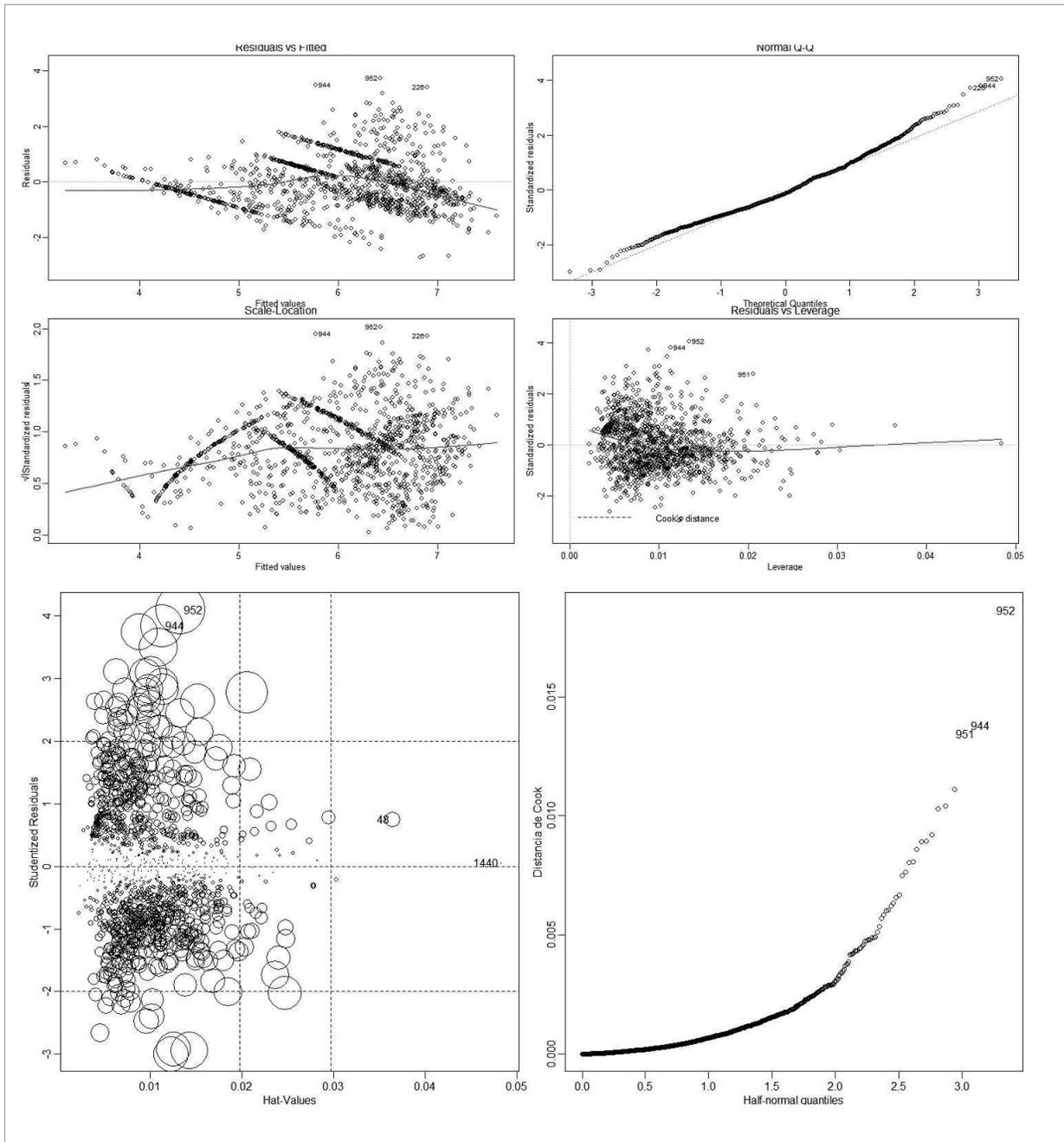


d)

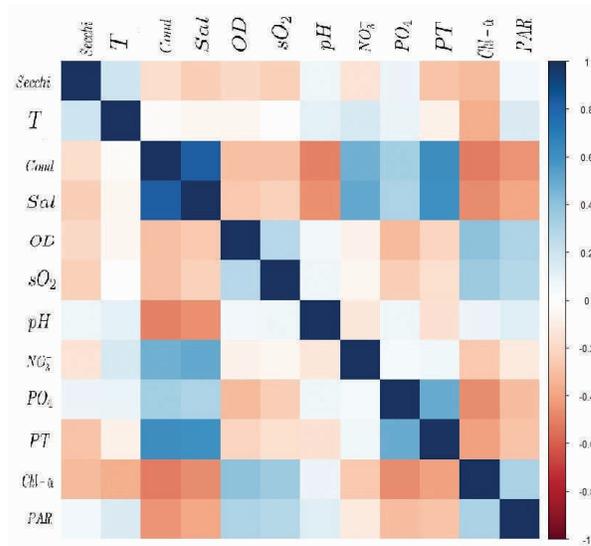
Figura 2. a-h. a) Simulación vs real; b) análisis de residuos; c) matriz de correlaciones; d) matriz de correlaciones con dispersión; e) tendencia de transparencia; f) diferencia de transparencia; g) ACF y PACF de transparencia y h) pronóstico 2018 y su diagnóstico. Las figuras son del sitio Centro WG, con base en los datos de AMSCLAE.



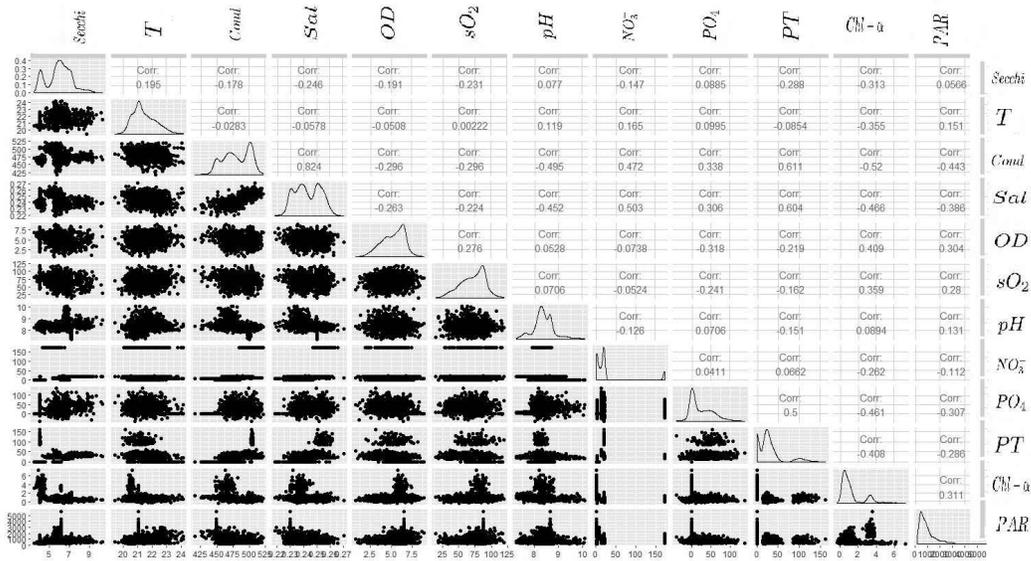
a)



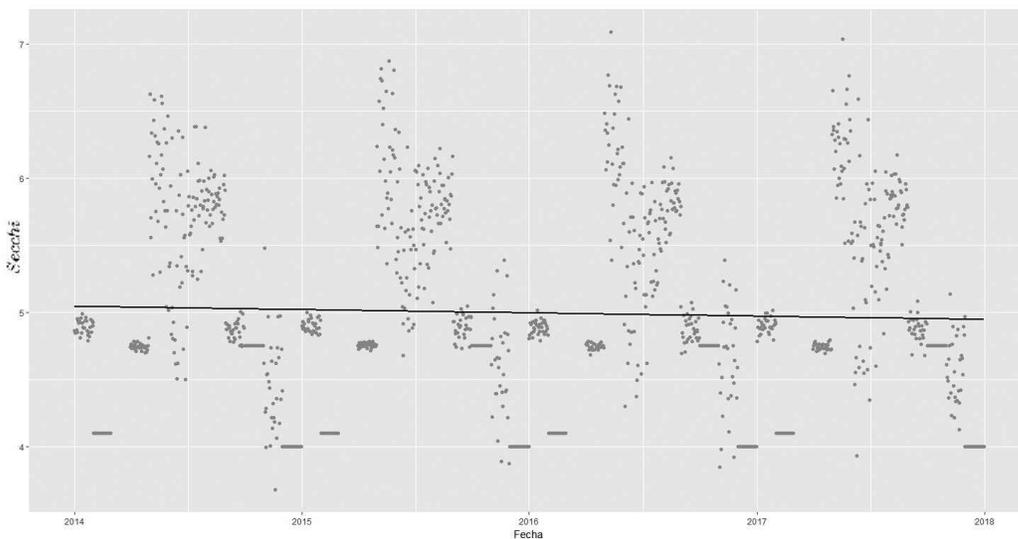
b)



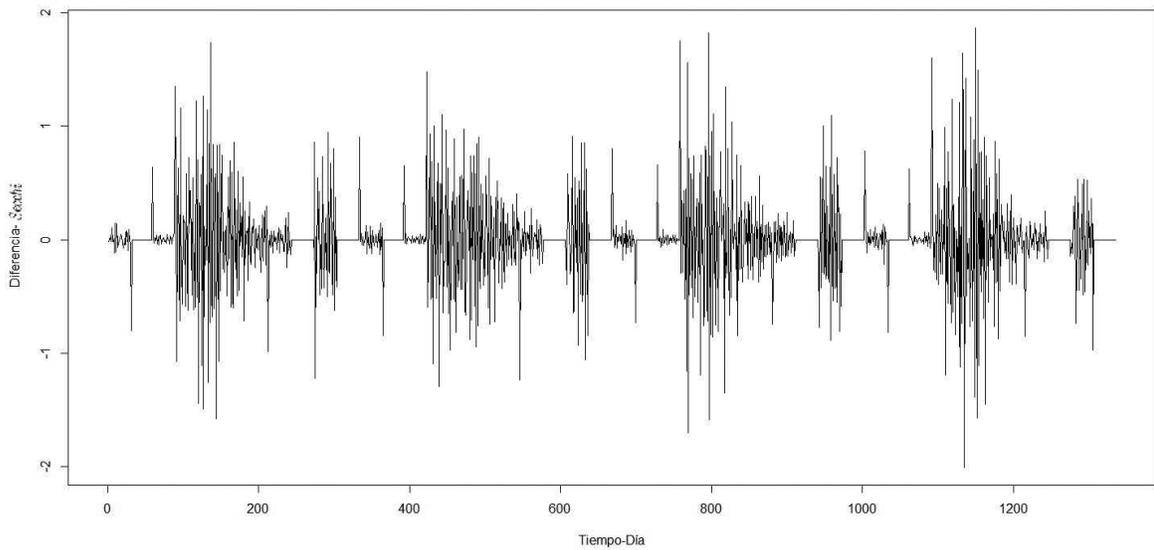
c)



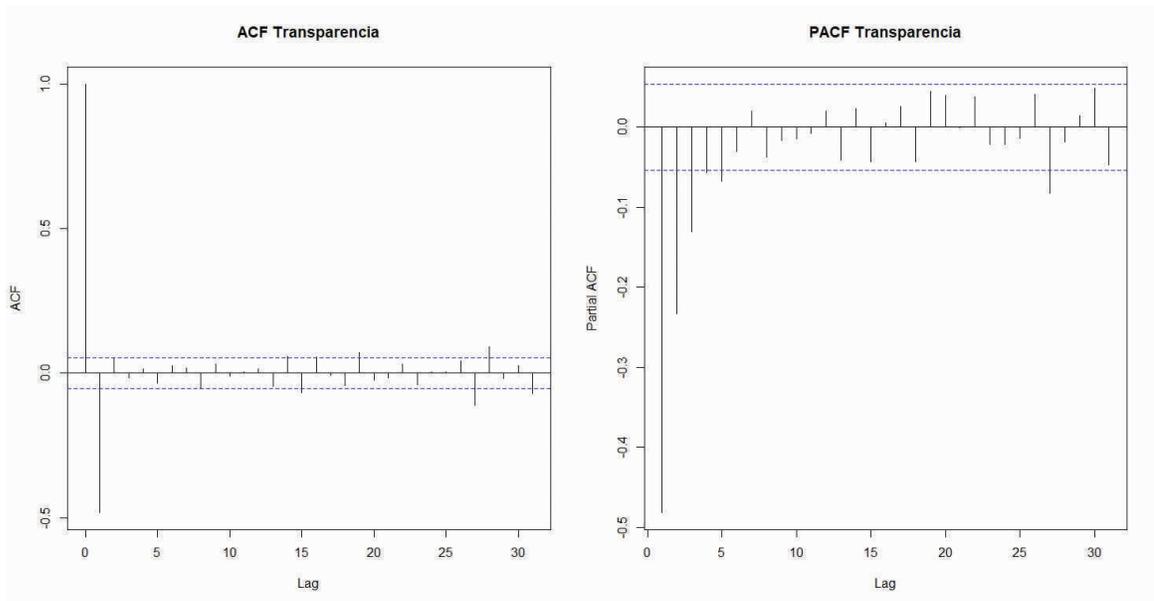
d)



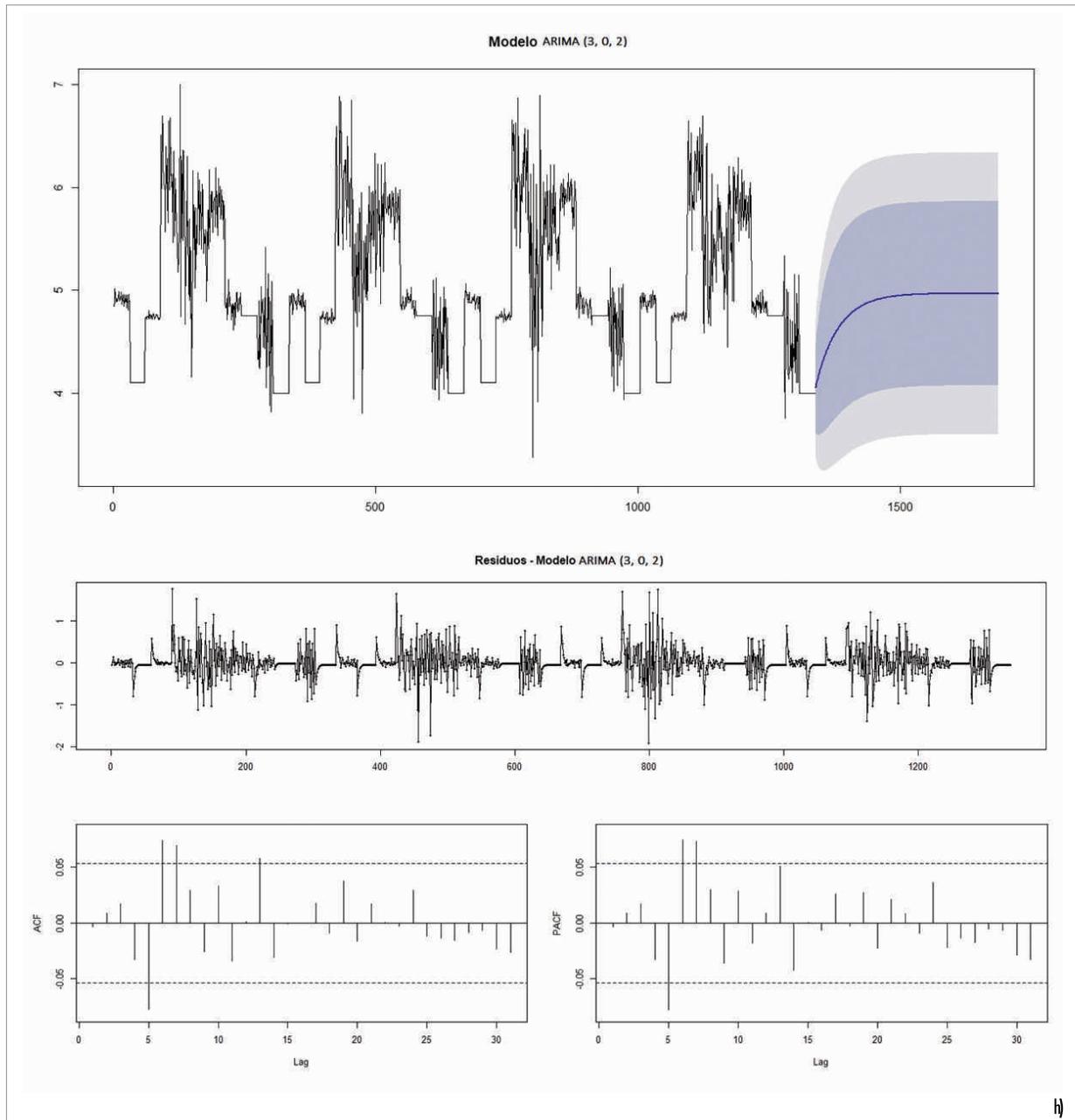
e)



f)



g)



Bibliografía

Braun, J., Maindonald, J. H. (2015) *Package DAAG*. Disponible en: <https://cran.rproject.org/web/packages/DAAG/index.html>

Brown, B. (1970) *Tests for Cobb-Douglas and CES production functions* International Economic Review 11(1): 77-83.

Carr, D., Lewin-Koh, N. Maechler, M. (2018) *Package hexbin* Disponible en: <https://cran.rproject.org/web/packages/hexbin/index.html>

Charnes, A., Cooper, W., Schinnar A. (1976) *A theorem on homogeneous functions and extended Cobb Douglas forms* Proceedings of the National Academy of Sciences 73(10): 3747-3748.

Christensen, L., Jorgenson, D., Lau, L. (1973) *Transcendental logarithmic production frontiers* The Review of Economics and Statistics 55(1): 28-45.

Cowpertwait, P. S., Meltcalfe, A. V. (2008) *Introductory Time Series with R* 2nd ed. Springer, New York.

Cryer, J. D., Chan, Kung-Sik. (2008) *Time Series Analysis, With Applications in R* 2nd ed. Springer, New York.

Dix, MA., Dix M., Orozco, M., Cabrera, D., Bocel, E., Toledo, A., Symonds, E. (2012) *El Lago Atitlán, Guatemala: su estado ecológico octubre 2009 – diciembre 2011* Revista de la Universidad del Valle de Guatemala 24: 35-50.

Faraway, J. (2016) *Package faraway* Disponible en: <https://cran.rproject.org/web/packages/faraway/index.html>

- Fletcher, L., Lu, Y. (1968) *A generalization of the CES production function* The Review of Economics and Statistics 50(4): 449-452.
- Fox, J., Weisberg, S., Adler, D. (2017) *Package car*. Disponible en: <https://cran.rproject.org/web/packages/car/index.html>
- Hastie, T., Tibshirani, R., Friedman, J. (2008) *The elements of statistical learning: Data mining, inference, and prediction* 2nd ed. Springer, New York.
- Hothorn, T., Zeileis, A., Farebrother, R. W., Cummins, C., Mitchell, D. (2018) *Package lmtree* Disponible en: <https://cran.rproject.org/web/packages/lmtree/index.html>
- Hyndman, R., O'Hara, M., Bergmeir, C., Razbashm, S., Wang, E. (2017) *Package forecast* Disponible en: <https://cran.rproject.org/web/packages/forecast/index.html>
- Ioan, C., Ioan, G. (2015) *The complete theory of Cobb-Douglas production function* Acta Universitatis Danubius. Economica 11(1): 74-114.
- Schoerke, B., Crowley, J., Cook, D., Hofmann, H., Wickham, H. (2017) *Package GGally* Disponible en: <https://cran.rproject.org/web/packages/GGally/index.html>
- Seber, G., Lee, A. (2003) *Lineal Regression Analysis* 2nd ed. Jhon Wiley & Sons, Inc., New Jersey.
- Seivert, C., Parmer, C., Hocking, T., Chamberlain, S., Ram, K. (2017) *Package Plotly* Disponible en: <https://cran.rproject.org/web/packages/plotly/index.html>
- Shumway, R., Stoffer, D. (2006) *Time Series Analysis and Its Applications: With R examples* 2nd ed. Springer, New York.
- Turlach, B. A. (2015) *Package quadprog* Disponible en: <https://cran.rproject.org/web/packages/quadprog/index.html>
- Ulrich, J., Ryan, J. A., Bennett, Ross. (2017) *Package xts* Disponible en: <https://cran.rproject.org/web/packages/xts/index.html>
- Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., Zemlea, J. (2017) *Package corrplot* Disponible en: <https://cran.rproject.org/web/packages/corrplot/index.html>
- Wickham, H., Grolemund, G. (2016) *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data* O'Reilly Media, Inc., Canada.
- Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis* Springer, New York.
- Wicklam, H. (2017) *Package reshape2*. Disponible en: <https://cran.rproject.org/web/packages/reshape2/index.html>