

Universidad del Valle de Guatemala
FACULTAD DE CIENCIAS Y HUMANIDADES
DEPARTAMENTO DE MATEMÁTICA



**Modelos estocásticos para el pronóstico de la transparencia
del agua del Lago de Atitlán, Sololá, Guatemala**

TRABAJO DE GRADUACIÓN PRESENTADO POR
Gaspar Yatáz Pop
PARA OPTAR AL GRADO ACADÉMICO
DE LICENCIADO EN MATEMÁTICA

GUATEMALA
2018

**Modelos estocásticos para el pronóstico de la transparencia
del agua del Lago de Atitlán, Sololá, Guatemala**

Universidad del Valle de Guatemala
FACULTAD DE CIENCIAS Y HUMANIDADES
DEPARTAMENTO DE MATEMÁTICA

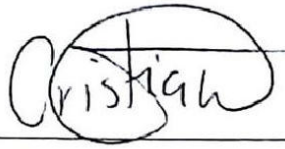


**Modelos estocásticos para el pronóstico de la transparencia
del agua del Lago de Atitlán, Sololá, Guatemala**

TRABAJO DE GRADUACIÓN PRESENTADO POR
Gaspar Yatáz Pop
PARA OPTAR AL GRADO ACADÉMICO
DE LICENCIADO EN MATEMÁTICA

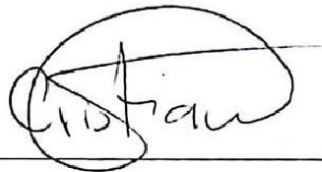
GUATEMALA
2018

Vo.Bo.:

A handwritten signature in cursive script, appearing to read "Cristian", written over a horizontal line.

Lic. Cristian E. Valdez

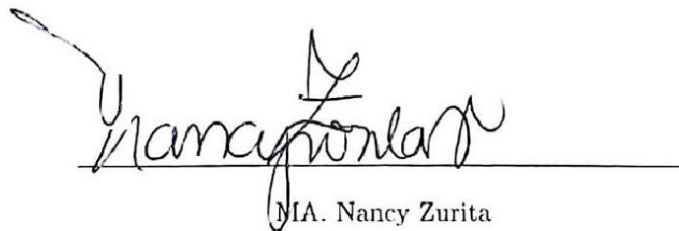
Tribunal Examinador:

A handwritten signature in cursive script, appearing to read "Cristian", written over a horizontal line.

Lic. Cristian E. Valdez

A handwritten signature in cursive script, appearing to read "Dorval Carías", written over a horizontal line.

Lic. Dorval Carías

A handwritten signature in cursive script, appearing to read "Nancy Zurita", written over a horizontal line.

MA. Nancy Zurita

Fecha de aprobación:

Guatemala, 9 de febrero de 2018

Prefacio

El Lago de Atitlán no sólo representa belleza en su inmanencia, es una fuente fundamental para los que existen en su cuenca. Mucha cosmovisión y raíces adornan su origen. Y para quienes nacimos en su cercanía, es parte de nuestro sentir.

La dinámica ecológica alterada por los que vivimos al rededor de dicho cuerpo de agua, cada vez posee mayor degradación en su estado. Las limitaciones primordiales ante la iniciativa de realizar análisis estadístico de alta calidad y estándares, a los parámetros que determinan la calidad del agua del Lago de Atitlán se resume en la poca información existente registrada basada en diseños experimentales, por ende estudiar la dinámica del Lago de Atitlán resulta ser más complejo, pues existen factores importantes los cuales su medición es casi nula.

Muchas líneas de pensamiento, es a partir de una, proviene de describir con las matemáticas la estructura de lo que nos rodea. Es así para el Lago de Atitlán.

A mi familia de vida y de sangre; que me inspiraron y contribuyeron para condensar mis sueños. A la vida, existencia y sus elementos que permiten que éste fluir entre espacios e impermanencia en cada presente uno se encuentre a sí mismo. Al Creador y Formador de lo que existe.

A mi asesor, profesores influyentes y la Universidad del Valle de Guatemala por haberme brindado el espacio para iniciarme en tan bello camino del pensamiento matemático, que es de por sí un todo, que surge de la nada.

Agradezco a Cristian Valdez por su asesoriamiento en la elaboración del presente trabajo de graduación, a Dorval Carías y Nancy Zurita por los comentarios y observaciones que le dieron una mejor estructura a las ideas que expongo en el presente trabajo de graduación. Agradezco a Fátima Reyes por su asistencia en la obtención de licencia de investigación en AMSCLAE para la facilitación de la data estudiada.

Índice general

Resumen	XIII
I. Introducción	1
II. Fundamentación matemática	5
A. Teoría de Muestreo	5
B. Procesos Estocásticos	8
C. Optimización	10
III. Motivaciones de la teoría y al problema	15
A. Regresión lineal múltiple	15
B. Cobb-Douglas	16
C. Series de Tiempo	16
D. Problema	17
IV. Regresión lineal múltiple	19
A. Modelo lineal y mínimos cuadrados	19
B. Distribución, correlación y prueba estadística	25
V. Series de Tiempo	33
A. Modelos autoregresivos y de media móvil	33
B. Ecuaciones en diferencia	39
C. Función de autocorrelación y autocorrelación parcial	43
D. Predicción	45
E. Modelos ARIMA	48
F. Modelos ARIMA multiplicativos estacionales	51
VI. Cobb Douglas	55
VII. Pronóstico de la transparencia del agua del Lago de Atitlán	59
A. Tratamiento de datos	60
B. Aplicación de regresión lineal múltiple	61

C. Aplicación de Cobb-Douglas	62
D. Aplicación de series de tiempo	62
VIII. Resultados	65
IX. Conclusiones	93
X. Recomendaciones	95
XI Bibliografía	97
XII. Anexo	99
A. Propiedades de mínimos cuadrados	99
B. Valor esperado y covarianza	100
C. Estimación y distribución	100
D. Autocorrelaciones	101

Resumen

La presente investigación se realizó con base en tres objetivos principales: describir la interacción de los factores físico-químicos con la transparencia del agua del Lago de Atitlán en distintos sitios de muestreo seleccionados, clasificar las relaciones más importantes de los factores físico-químicos con la transparencia hasta explicar su influencia y pronosticar la transparencia para el año 2018. Para cumplir el primer objetivo se desarrolló la teoría de regresión lineal múltiple partiendo de las ideas que exponen Hastie *et al.* (2008), Seber y Lee (2003). Mientras que para el segundo objetivo se realizó una breve exposición de ideas que engloban las propiedades de la función de Cobb-Douglas, basándose principalmente en los trabajos de Ioan y Ioan (2015), Charnes, Cooper y Schinnar (1976). Y para el tercer objetivo se desarrolló la teoría de los modelos ARIMA, partiendo desde la construcción de modelos AR, MA, ARMA, ARIMA y SARIMA; exponiendo algunas ideas fundamentales en el desarrollo y análisis de las series de tiempo tomando como base Shumway y Stoffer (2006). En el desarrollo de las teorías se buscó detallar las pruebas presentadas en los textos base de ésta investigación. Finalmente para la parte de aplicación se centró en la simulación de data; en los cálculos de los modelos de regresión lineal múltiple, modelos Cobb-Douglas y de modelos ARIMA. El código empleado en la simulación, en los cálculos, gráficas y análisis estadístico fue desarrollado con el software R.

I Introducción

En el presente trabajo de graduación se estudia la interacción de los factores físico-químicos del Lago de Atitlán, variables independientes x_1, x_2, \dots, x_p , p es tal que $1 \leq p \leq 12$, con la transparencia del agua del Lago de Atitlán, variable dependiente Y . Se define para la primera parte del trabajo que Y puede expresarse como

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

siendo Y , ϵ escalares, tal que $X^T = (x_1, x_2, \dots, x_p)$ es un vector de $1 \times p$, en general para n observaciones se puede expresar \mathbf{Y} como un vector de $n \times 1$ y \mathbf{X} como una matriz de $n \times p$, y ϵ como un vector de error de $n \times 1$. Se deduce un modelo \hat{Y} que aproxime Y con la teoría de aprendizaje estadístico que plantean 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 Atitlán con los factores incidentes.

Obtener un modelo Y que explique adecuadamente la transparencia del agua del Lago de Atitlán, 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 abordarán técnicas de la estadística matemática que simplifiquen y conduzcan con claridad el proceso de deducción del modelo; partiendo de la data disponible registrada por la Autoridad para el Manejo Sustentable del la Cuenca del Lago de Atitlán y su Entorno - AMSCLAE; pero principalmente sobre el tratamiento de datos que se define como la simulación muestral.

En el mejor de los casos, si se tiene que ϵ en promedio es cero, una predicción Y es de la forma

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

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, basta con aplicar apropiadamente una técnica de aprendizaje estadístico para lograr una mejor estimación para f . No obstante ϵ en Y no es casi siempre cero, y no es posible efectuar su predicción con X^T . Por lo que la variabilidad que introduce ϵ afecta la exactitud de predicción de \hat{Y} , este contexto

define el error irreducible; dado que \hat{f} no puede usarse para predecir ϵ .

Sea \hat{f} una estimación tal que

$$\hat{\mathbf{y}} = \hat{f}(\mathbf{X})$$

y que \hat{f} , \mathbf{X} son fijos. Considere el valor esperado de $(\mathbf{y} - \hat{\mathbf{y}})^2$:

$$E(\mathbf{y} - \hat{\mathbf{y}})^2 = E[f(\mathbf{X}) + \epsilon - \hat{f}(\mathbf{X})]^2 = [f(\mathbf{X}) - \hat{f}(\mathbf{X})]^2 + Var(\epsilon)$$

con $Var(\epsilon)$, la varianza del termino de error. La importancia del cálculo anterior es resaltar que $[f(\mathbf{X}) - \hat{f}(\mathbf{X})]^2$ involucra un error reducible, mientras que el error de $Var(\epsilon)$ es irreducible.

En el desarrollo de estimación para \mathbf{y} del modelo de regresión lineal múltiple, se fundamenta la aplicación del método de mínimos cuadrados con el objetivo de minimizar el error reducible involucrado. Además se desea entender cómo la transparencia del agua del Lago de Atitlán cambia como función de x_1, \dots, x_p , principalmente se quiere responder a las siguientes preguntas: ¿Cuáles de los factores físico-químicos, las variables explicativas, están asociadas con la transparencia del agua del Lago de Atitlán?, ¿Cuál es la relación entre la transparencia con cada factor físico-químico?, ¿Puede la relación entre la transparencia y cada factor físico-químico ser descrita resumidamente usando un modelo lineal, o es la relación más complicada?. Estas preguntas forman la parte modular del desarrollo de la teoría del modelo de regresión lineal múltiple.

Luego se plantea al conocer las relaciones entre las variables explicativas importantes con la transparencia, explicar cómo éstas actúan cuando aumentan un cierto porcentaje. La simplicidad del modelo de Cobb-Douglas, ayuda a explicar las elasticidades de los parámetros físico-químicos del agua del lago de Atitlán, ya que con ellas se tiene una estimación sobre la reducción o aumento en la transparencia. 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; éste estudio permite explicar cual de los parámetros físico-químicos son los que determinan la transparencia en cada sitio de muestreo. Buscando tener una visión más amplia sobre la dinámica de la transparencia y su relación con factores físico-químicos en los sitios de estudio.

Y por último se estudian las series de tiempo aplicadas a la transparencia del agua del Lago de Atitlán, con la teoría que se presenta en Shumway y Stoffer (2006). Las series de tiempo, son una herramienta fundamental para el estudio de los procesos estocásticos; la absorción de elementos de

estacionalidad, autoregresión y estacionariedad de un fenómeno son fundamentales para su predicción. Principalmente el estudio de la transparencia del agua del Lago de Atitlán desde el punto de vista de los modelos ARIMA; con dichos modelos se desea pronosticar 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. Los brotes de cianobacteria en la superficie del Lago en los años 2009 y 2015, como se aprecia en las figuras no. 1 y no. 2 en anexos, han generado un despertar sobre la situación del estado ecológico del Lago.

Los registros de los factores físico-químicos del Lago de Atitlán son limitados; en particular se tienen 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. Pocos son los estudios sobre la interacción de dichos factores con la transparencia. Por otro lado no existen modelos que expliquen cuáles de las variables existentes y registradas son influyentes o determinantes en la transparencia, además no se cuentan con pronósticos de la transparencia y mucho menos del estado ecológico del lago de Atitlán de los próximos años.

II Fundamentación matemática

A. Teoría de Muestreo

Definición A.1 *Al conjunto de entidades sobre las cuales se puede realizar la medición de la característica de interés se le denomina población.*

Definición A.2 *Al conjunto de mediciones de la característica de interés de todos los elementos de la población se le denomina población estadística.*

Definición A.3 *Cualquier subconjunto de una población recibe el nombre de muestra.*

Definición A.4 *Una muestra representativa de una población es aquella que la reproduce fielmente en lo relativo a la característica de interés.*

Definición A.5 *Una muestra probabilística es aquella en la cual cada unidad de la población tiene una probabilidad de selección conocida, y se emplea un método aleatorio para seleccionar las unidades específicas que se incluirán en la muestra.*

Definición A.6 *Una muestra cuyos elementos son seleccionados aleatoriamente, de tal forma que cada elemento tiene una probabilidad igual de ser seleccionado, es llamada muestra aleatoria simple.*

Definición A.7 *El conjunto de n variables aleatorias $X_1, X_2, X_3, \dots, X_n$ independientes e idénticamente distribuidas, con la misma función de probabilidad de la población, constituyen una muestra aleatoria simple de dicha población.*

Definición A.8 *Un estadístico es cualquier función de las variables aleatorias que componen una muestra y, de ser necesario, puede incluir otras cantidades conocidas.*

Definición A.9 *Estadística es el valor numérico correspondiente a la realización de un estadístico en una muestra aleatoria específica.*

Definición A.10 *Todo estadístico es una variable aleatoria y, por tanto, tiene una función de probabilidad asociada, la cual se conoce como distribución de muestreo, o distribución muestral.*

Teorema A.11 *Teorema del límite central.* Sean $X_1, X_2, X_3, \dots, X_n$ una secuencia de n variables aleatorias independientes e idénticamente distribuidas, con media y varianza finita. Entonces el promedio muestral tiene una distribución de probabilidad con media y varianza dadas por:

$$\mu_{\bar{X}} = \mu_X \quad (\text{II.1})$$

y

$$\sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n} \quad (\text{II.2})$$

dicha distribución tiende hacia una distribución normal conforme n tiende a infinito.

En la literatura clásica el teorema del límite central se expresa como

$$\lim_{n \rightarrow \infty} \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \Rightarrow N(0, 1). \quad (\text{II.3})$$

La prueba de este teorema puede consultarse en la página 271 de Miller (2000).

Una conclusión que se puede extraer del teorema del límite central es que conforme el tamaño de muestra aumenta, la desviación estándar del promedio muestral se reduce; por tanto, la precisión y la exactitud de este mejoran, en torno a la verdadera media de la población.

Definición A.12 *Estimación es la acción de tasar el valor de un parámetro a través de estadísticos, con un nivel de incertidumbre determinable.*

Usualmente el término estimación suele utilizarse para denotar el valor de un estimador. Para un parámetro determinado por estimar, se denota genéricamente por θ , se pueden elegir muchos posibles estimadores. Todos los posibles candidatos a estimadores se denotan por T . Particularmente T es una función de las variables de la muestra aleatoria X_1, X_2, \dots, X_n .

Definición A.13 *Un estimador es un estadístico utilizado para estimar un parámetro.*

En este sentido es una función de variables aleatorias de la muestra.

Definición A.14 *Se define el error cuadrático medio de un estadístico T frente a su parámetro θ , denotado por $ECM(T)$ como $ECM(T) = E[(T - \theta)^2] = \sigma_T^2 + (\theta - E(T))^2$*

Definición A.15 *T es un estimador insesgado de θ , si y sólo si $E[T] = \theta$.*

Por ejemplo, \bar{X} es un estimador insesgado de μ .

Definición A.16 *Sea T el estimador de un parámetro, y defínase T_1, T_2, \dots, T_n como una secuencia de estimadores basados en muestras de tamaño 1, 2, 3, \dots , n respectivamente. T es un estimador consistente de θ si $\lim_{n \rightarrow \infty} P(|T_n - \theta| \leq \epsilon) = 1$ para todos los valores de θ y todo $\epsilon > 0$.*

En particular se tiene que \bar{X} es un estimador consistente de μ .

Definición A.17 *El estimador más eficiente T de un parámetro θ es aquel estimador insesgado que, comparado con todos los estimadores insesgados posibles de θ , sea el que tenga una varianza mínima.*

El principal reto que entraña la definición anterior es el de encontrar todos los estimadores insesgados posibles de θ . Sin embargo, el teorema de la cota inferior de Cramér-Rao simplifica su complejidad, dado que impone un mínimo o cota inferior a la varianza de cualquier estimador; por lo que al encontrar el estimador insesgado cuya varianza sea igual a la cota inferior de Cramér-Rao, ya no es necesario buscar más, porque ese es el estimador más eficiente. Formalmente considere X_1, X_2, \dots, X_n una muestra aleatoria de una población con función de probabilidad $f(x; \theta)$. Si T es un estimador insesgado de θ entonces

$$\sigma_T^2 \geq \frac{1}{nE\left[\left(\frac{\partial \ln f(X; \theta)}{\partial \theta}\right)^2\right]} \quad (\text{II.4})$$

El mejor estimador T posible de un parámetro θ es su estimador más eficiente. La consistencia de un estimador es una ventaja adicional, la cual es importante tener en cuenta para seleccionar los tamaños de muestra en una estimación.

Definición A.18 *La estimación puntual es una técnica estadística que busca determinar el mejor estadístico que, al calcularse por única vez, se aproxime al valor exacto del parámetro.*

Definición A.19 *Sean X_1, X_2, \dots, X_n una muestra aleatoria de una población con función de probabilidad $f(x; \theta)$, entonces la función de verosimilitud de la muestra está dada por:*

$$L(x_1, x_2, \dots, x_n; \theta) = f(X_1; \theta) \cdots f(X_n; \theta) \quad (\text{II.5})$$

El método para determinar el estimador de máxima verosimilitud de un parámetro θ se sigue de la siguiente manera: sea X_1, X_2, \dots, X_n una muestra aleatoria de una población con función de probabilidad $f(x; \theta)$. Si t es el valor del estadístico obtenido experimentalmente que maximiza la función de verosimilitud respecto de θ , entonces el estadístico T es el estimador de máxima verosimilitud de θ y la estadística t es la estimación de máxima verosimilitud. Usualmente resulta más sencillo maximizar el logaritmo natural de la función de verosimilitud, y aprovechar que el máximo de una función es el mismo del de su logaritmo natural, pues el logaritmo es una función siempre creciente.

B. Procesos Estocásticos

Definición B.1 *Proceso estocástico.* Un proceso estocástico es una colección de variables aleatorias $\{X_t : t \in T\}$ parametrizada por un conjunto T , llamado espacio parametral, y con valores en un conjunto S llamado espacio de estados.

En particular se toman como espacio parametral el conjunto discreto $T = \{0, 1, 2, \dots\}$, o bien el conjunto continuo $T = [0, \infty)$, y estos números se interpretan como tiempos. En el primer caso se dice que el proceso es a tiempo discreto, y se denota por $\{X_n : n = 0, 1, \dots\}$, mientras que en el segundo caso el proceso es a tiempo continuo, y se denota por $\{X_t : t \geq 0\}$.

Definición B.2 *Función de media.* Se denomina función de media de un proceso estocástico $\{X_t : t \in T\}$ a la aplicación definida por:

$$m(t) = E(X_t), \forall t \in T. \quad (\text{II.6})$$

Definición B.3 *Función de autocovarianza.* La función de autocovarianza de un proceso estocástico $\{X_t : t \in T\}$ es una aplicación, de $T \times T$ en \mathcal{R} dada por:

$$c(s, t) = \text{Cov}(X_s, X_t) = E[(X_s - m(s))(X_t - m(t))], \forall s, t \in T. \quad (\text{II.7})$$

En la literatura avanzada, un proceso estocástico puede considerarse como una función de dos variables $X : T \times \Omega \rightarrow S$ tal que a la pareja (t, ω) se le asocia el estado $X(t, \omega)$, que puede escribirse como $X_t(\omega)$.

Para cada valor de t en T , el mapeo $\omega \rightarrow X_t(\omega)$ es una variable aleatoria, mientras que para cada ω en Ω fijo, la función $t \rightarrow X_t(\omega)$ es llamada una trayectoria o realización del proceso.

Los diferentes tipos de procesos estocásticos se obtienen al considerar las distintas posibilidades para: el espacio parametral, el espacio de estados, las características de las trayectorias, y principalmente las relaciones de dependencia entre las variables aleatorias que conforman el proceso.

Se presentan algunos de los ejemplos clásicos de los procesos estocásticos.

Ejemplo B.4 *Proceso de ensayos independientes.*

El proceso a tiempo discreto $\{X_n : n = 0, 1, \dots\}$ puede estar constituido por variables aleatorias independientes. Este modelo representa una sucesión de ensayos independientes de un mismo experimento aleatorio, por ejemplo, lanzar un dado o una moneda repetidas veces. El resultado u observación del proceso en un momento cualquiera es, por lo tanto, independiente de cualquier otra observación pasada o futura del proceso.

Ejemplo B.5 *Procesos de Markov.*

Estos tipos de procesos son importantes y son modelos en donde, suponiendo conocido el estado presente del sistema, los estados anteriores no tienen influencia en los estados futuros del sistema. Esta condición se llama propiedad de Markov y puede expresarse de la siguiente forma: Para cualesquiera estados $x_0, x_1, \dots, x_{n-1}, x_n, x_{n+1}$, se cumple la igualdad

$$P(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n) = P(X_{n+1} = x_{n+1} | X_n = x_n). \quad (\text{II.8})$$

De esta forma la probabilidad del evento $X_n = x_n$ sólo depende el evento $X_{n-1} = x_{n-1}$, mientras que la información correspondiente al evento $X_0 = x_0, \dots, X_{n-2} = x_{n-2}$ es irrelevante. En particular, los sistemas dinámicos deterministas dados por una ecuación diferencial pueden considerarse procesos de Markov pues su evolución futura queda determinada por la posición inicial del sistema y una ley de movimiento especificada.

Ejemplo B.6 *Procesos con incrementos independientes.*

Se dice que un proceso $\{X_t : t \geq 0\}$ tiene incrementos independientes si para cualesquiera tiempos $0 \leq t_1 < t_2 < \dots < t_n$, las variables $X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$ son independientes.

Ejemplo B.7 *Procesos estacionarios.*

Se dice que un proceso $\{X_t : t \geq 0\}$ es estacionario (en el sentido estricto) si para cualesquiera tiempos t_1, \dots, t_n , la distribución del vector $(X_{t_1}, \dots, X_{t_n})$ es la misma que la del vector $(X_{t_1+h}, \dots, X_{t_n+h})$ para cualquier valor de $h > 0$. En particular, la distribución de X_t es la misma que la de X_{t+h} para cualquier $h > 0$, y entonces esta distribución es la misma para cualquier valor de t .

Ejemplo B.8 *Martingalas.*

Una martingala a tiempo discreto es, en términos generales, un proceso $\{X_n : n = 0, 1, \dots\}$ que cumple la condición

$$E(X_{n+1} | X_0 = x_0, \dots, X_n = x_n) = x_n \quad (\text{II.9})$$

En palabras, esta igualdad significa que el estado promedio del proceso al tiempo futuro $n + 1$ es el valor del proceso en su último momento observado, es decir, x_n . Esto es, se trata de una ley de movimiento aleatorio que es equilibrada pues en promedio el sistema no se mueve del último momento observado. A estos procesos también se les conoce como procesos de juegos justos pues si se considera una sucesión infinita de apuestas sucesivas y si X_n denota el capital de uno de los jugadores al tiempo n , entonces la propiedad de martingala establece que el juego es justo pues en promedio el jugador no pierde ni gana en cada apuesta.

Ejemplo B.9 *Proceso de Lévy.*

Se dice que un proceso a tiempo continuo $\{X_t : t \geq 0\}$ es un proceso de Lévy si sus incrementos son independientes y estacionarios. De hecho el proceso de Poisson como el movimiento Browniano son ejemplos de este tipo de procesos.

Ejemplo B.10 *Proceso Gaussiano.*

Se dice que un proceso a tiempo continuo $\{X_t : t \geq 0\}$ es un proceso Gaussiano si para cualesquiera colección finita de tiempos t_1, \dots, t_n , el vector $(X_{t_1}, \dots, X_{t_n})$ tiene distribución normal o Gaussiana. Nuevamente, el movimiento Browniano es un ejemplo de este tipo de procesos.

C. Optimización

Definición C.1 *Un conjunto $A \subseteq \mathcal{R}^n$ es convexo si para cada par de elementos $\mathbf{x}, \mathbf{y} \in A$ con $\mathbf{x} \neq \mathbf{y}$ y $\lambda \in (0,1)$, $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in A$.*

Un conjunto $A \subseteq \mathcal{R}^n$ es estrictamente convexo si para cada par de elementos $\mathbf{x}, \mathbf{y} \in A$ con $\mathbf{x} \neq \mathbf{y}$ y $\lambda \in (0,1)$, $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in \text{int}(A)$.

Definición C.2 *Sean $A \subseteq \mathcal{R}^n$ un conjunto convexo, $f : A \rightarrow \mathcal{R}$, \mathbf{x} y \mathbf{y} elementos distintos de A y $\lambda \in (0,1)$. f es convexa si y sólo si*

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y})$$

y f es cóncava si y sólo si

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \geq \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y}).$$

Definición C.3 *Una forma cuadrática $q(\mathbf{x}) = \mathbf{x}A\mathbf{x}^T$ es:*

1. *Definida positiva si $q(\mathbf{x}) > 0$ para todo $\mathbf{x} \neq \mathbf{0}$,*
2. *Semidefinida positiva si $q(\mathbf{x}) \geq 0$ para todo $\mathbf{x} \neq \mathbf{0}$,*
3. *Definida negativa si $q(\mathbf{x}) < 0$ para todo $\mathbf{x} \neq \mathbf{0}$ y*
4. *Semidefinida negativa si $q(\mathbf{x}) \leq 0$ para todo \mathbf{x} .*

Si una forma cuadrática no es semidefinida positiva ni semidefinida negativa, se llama no definida.

Teorema C.4 *La forma cuadrática $q(\mathbf{x}) = \mathbf{x}A\mathbf{x}^T$ es:*

1. *Estrictamente convexa si y sólo si es definida positiva.*
2. *Convexa si y sólo si es semidefinida positiva.*
3. *Estrictamente cóncava si y sólo si es definida negativa.*
4. *Cóncava si y sólo si es semidefinida negativa.*

La prueba de éste teorema se puede consultar en las páginas 97 y 98 de Pecha (2012).

Teorema C.5 *Si la matriz hessiana $H_f(\mathbf{x})$ de f es semidefinida positiva (negativa) para todo \mathbf{x} en un conjunto abierto y convexo A , entonces f es convexa (cóncava) en A .*

Definición C.6 *Sean $A \subseteq \mathcal{R}^n$ un conjunto convexo, $f : A \rightarrow \mathcal{R}$, \mathbf{x} y \mathbf{y} elementos distintos de A y $\lambda \in (0,1)$. f es cuasiconvexa si y sólo si*

$$f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \max\{f(\mathbf{x}), f(\mathbf{y})\}$$

y f es cuasicóncava si y sólo si

$$f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) \geq \min\{f(\mathbf{x}), f(\mathbf{y})\}.$$

Definición C.7 *Argumento maximizador o minimizador. Sean f una función con dominio sobre el conjunto $A \subset \mathcal{R}^n$ y valores en los números reales, $D \subseteq A$ el conjunto de restricciones. Los problemas de optimización buscan valores donde la función f , llamada función objetivo, alcanzan sus valores máximo y mínimo, esto es:*

$$\max\{f(\mathbf{x})|\mathbf{x} \in D\} \quad \text{y} \quad \min\{f(\mathbf{x})|\mathbf{x} \in D\}.$$

El conjunto de puntos de D donde la función alcanza su máximo o mínimo se conoce como conjunto de maximizadores o minimizadores de f sobre D y se denota respectivamente por:

$$\operatorname{argmax}\{f(\mathbf{x})|\mathbf{x} \in D\} = \{\mathbf{x}^* \in D | f(\mathbf{x}^*) \geq f(\mathbf{x}) \quad \forall \mathbf{x} \in D\}$$

y

$$\operatorname{argmin}\{f(\mathbf{x})|\mathbf{x} \in D\} = \{\mathbf{x}_* \in D | f(\mathbf{x}_*) \leq f(\mathbf{x}) \quad \forall \mathbf{x} \in D\}.$$

Definición C.8 *Considere el conjunto de restricciones de la forma:*

$$A \cap \{\mathbf{x} | g_i(\mathbf{x}) = 0 \quad \text{para} \quad i = 1, 2, \dots, m; h_k(\mathbf{x}) \leq 0 \quad \text{para} \quad k = 1, 2, \dots, p\}, \quad (\text{II.10})$$

donde $A \subset \mathcal{R}^n$ es el dominio de f , h_k y g_i son funciones reales con dominio en \mathcal{R}^n . Una solución factible al problema de optimización es un conjunto que satisfaga las restricciones (II.10), esto es, un elemento del conjunto de restricciones.

Se busca en ese sentido el valor óptimo de la función f sobre un conjunto de soluciones factibles.

Para el caso particular de igualdad se encuentran las condiciones que satisfacen los óptimos de una función $f(\mathbf{x})$ sobre un conjunto de la forma:

$$A \cap \{\mathbf{x} | g_i(\mathbf{x}) = 0 \quad \text{para} \quad i = 1, 2, \dots, m\}. \quad (\text{II.11})$$

Teorema C.9 *Lagrange. Condiciones necesarias.* Si f tiene un extremo local en \mathbf{a} sobre el conjunto (II.11), donde A es un subconjunto abierto no vacío de \mathcal{R}^n y las funciones f y g_i para $i = 1, 2, \dots, m$ son diferenciables en A , y si la matriz

$$\left(\frac{\partial g_i}{\partial x_j}(\mathbf{a}) \right)_{m \times n}$$

tiene rango m , entonces existen $\lambda_1, \lambda_2, \dots, \lambda_m$ tales que

$$\nabla f(\mathbf{a}) = \sum_{j=1}^n \lambda_j \nabla g_j(\mathbf{a}).$$

La condición del teorema anterior se puede transformar en

$$\nabla f(\mathbf{a}) - \sum_{j=1}^n \lambda_j \nabla g_j(\mathbf{a}) = 0$$

que equivale a

$$\frac{\partial f}{\partial x_k}(\mathbf{a}) - \sum_{j=1}^n \lambda_j \frac{\partial g_j}{\partial x_k}(\mathbf{a}) = 0, \quad \text{para} \quad k = 1, 2, \dots, n.$$

Teorema C.10 *Condiciones suficientes.* Si los menores principales orlados de orden k evaluados en un punto que cumple las condiciones necesarias tienen signo $(-1)^{k-m}$ para $k = 2m+1, 2m+2, \dots, n+m$, la función tiene un máximo en ese punto. Y si los menores principales orlados de orden k evaluados en un punto que cumple las condiciones necesarias tienen signo $(-1)^m$ para $k = 2m+1, 2m+2, \dots, n+m$, la función tiene un mínimo en ese punto.

Y para el caso de restricciones de desigualdad se tiene el siguiente teorema.

Teorema C.11 *Karush-Kuhn-Tucker.* Sean f y g_k , para $k = 1, 2, \dots, m$, funciones diferenciables en un conjunto abierto no vacío $A \subseteq \mathcal{R}^n$. Si \mathbf{a} es el máximo de la función f sobre el conjunto

$$A \cap \{\mathbf{x} | h_k(\mathbf{x}) \leq 0 \quad \text{para} \quad k = 1, 2, \dots, m\}, \quad (\text{II.12})$$

$I \subseteq 1, 2, \dots, m$ es el conjunto de índices de las restricciones activas en \mathbf{a} y la matriz $\left(\frac{\partial g_i}{\partial x_j} \right)$ para

$i \in I$ de tamaño $\#(I) \times n$, tiene rango $\#(I)$. Entonces existen números reales $\mu_1, \mu_2, \dots, \mu_m$ tales que

$$\nabla f(\mathbf{a}) = \sum_{i=1}^m \mu_i \nabla g_i(\mathbf{a}) \quad (\text{II.13})$$

y

$$\mu_k \leq 0 \quad \text{y} \quad \mu_k g_k(\mathbf{a}) = 0, \quad \text{para} \quad k = 1, 2, \dots, m. \quad (\text{II.14})$$

La matriz $\left(\frac{\partial g_i}{\partial x_j} \right)$ está formada por los gradientes de las funciones g_i para $i \in I$.

Definición C.12 *Problema de programación cuadrática.* Un problema de optimización de n variables y m restricciones, con objetivo de encontrar un vector \mathbf{x} , n -dimensional que:

$$\text{minimice} \quad \frac{1}{2} \mathbf{x}^T Q \mathbf{x} + \mathbf{c}^T \mathbf{x}$$

$$\text{sujeto a} \quad A \mathbf{x} \leq \mathbf{b}$$

donde \mathbf{c} es un vector n -dimensional de valores reales, Q es una matriz de $n \times n$ simétrica de valores reales, A es una matriz $m \times n$ de valores reales y \mathbf{b} es un vector m -dimensional de valores reales se le clasificará como un problema de programación cuadrática.

III Motivaciones de la teoría y al problema

A. Regresión lineal múltiple

Sea $\mathbf{x} \in \mathbb{R}^p$ una variable aleatoria de entrada, de números reales, mientras que $Y \in \mathbb{R}$ una variable aleatoria de salida con la distribución conjunta dada por $Pr(\mathbf{x}, y)$. Se desea encontrar una función $f(\mathbf{x})$ para predecir y dada los valores de \mathbf{x} . En esta teoría se requiere de una función de pérdida, dígase $L(y, f(\mathbf{x}))$, para penalizar el error en la predicción; considere dicha función de pérdida como el error de predicción al cuadrado, dada por

$$L(y, f(\mathbf{x})) = (y - f(\mathbf{x}))^2 \quad (\text{III.1})$$

que facilita un criterio para elegir f ; considere el valor esperado del error de predicción al cuadrado

$$E[(y - f(\mathbf{x}))^2] = \int [y - f(x)]^2 Pr(dx, dy) \quad (\text{III.2})$$

dada la probabilidad condicional en \mathbf{x} , se puede escribir la ecuación anterior como

$$E[(y - f(\mathbf{x}))^2] = E_{\mathbf{x}} E_{\mathbf{x}|Y} [(y - f(\mathbf{x}))^2 | \mathbf{x}] \quad (\text{III.3})$$

por lo que es suficiente minimizar (III.3) puntualmente, considere la siguiente ecuación

$$f(x) = \operatorname{argmin}_c E_{y|\mathbf{x}} [(y - c)^2 | \mathbf{x} = x] \quad (\text{III.4})$$

donde *argmin* es el argumento de valor mínimo, resulta ser que la solución a $f(x)$ es tal que $f(x) = E(y|\mathbf{x} = x)$ que se conoce como la esperanza condicional de y dado \mathbf{x} , en la literaturas avanzadas es conocida como la función de regresión. Entonces el mejor predictor de y en el punto $\mathbf{x} = x$ es la media condicional. Es evidente pensar ya, sobre qué condiciones $f(x)$ es la solución al problema de minimizar el error de predicción al cuadrado.

Sea $SRC(f)$ la suma residual de cuadrados dada por

$$SRC(f) = \sum_{i=1}^N (y_i - f_i)^2 \quad (\text{III.5})$$

con $f = \hat{y}$ una función arbitraria, tal que $f_i = \hat{y}_i$.

Minimizar (III.5) conduce a infinitas soluciones. Cualquier función \hat{f} pasando a través de los puntos de entrenamiento (x_i, y_i) es una solución. Y cualquier solución particular elegida podría ser un predictor deficiente. Si hay múltiples observaciones pares x_i, y_{il} con $l = 1, 2, \dots, N_i$ en cada valor de x_i , el riesgo es limitado, en este caso la solución pasa por el promedio de los valores de y_{il} en cada x_i .

Con el fin de obtener resultados útiles para un N finito, se tiene que restringir las soluciones elegible a un conjunto más pequeño de funciones. Cómo decidir sobre la naturaleza de las restricciones dependerá de la complejidad de las consideraciones tales como la deducción de una representación paramétrica de f o su incorporación en los métodos Bayesianos, Kernel y Regresión Local ya sea implícita o explícitamente.

La motivación al estudio de la regresión lineal múltiple parte del entender las condiciones para la solución del problema de minimizar (III.5) con el método de mínimos cuadrados, además de utilizar las consecuencias del desarrollo de la teoría de regresión lineal y las pruebas estadísticas para encontrar un criterio para la selección de las mejores variables independientes en cuestión que explican mejor estadísticamente la variable dependiente.

B. Cobb-Douglas

La motivación de estudiar la función de Cobb-Douglas es debida a la implicación que tienen las constantes de elasticidad de este modelo con los factores de estudio. Principalmente la deducción en su forma log-lineal, para su cálculo, debido a que permite explicar cuáles de las variables son influyentes en la transparencia del agua del Lago de Atitlán. Además de ver sus implicaciones prácticas al imponerle condiciones, y obtener de ella un problema de programación cuadrática; que gracias a herramientas computacionales implementadas en el software R se pueden determinar las elasticidades de éste enfoque, que facilitará explicar cuáles de los factores físico-químicos son determinantes para la transparencia.

C. Series de Tiempo

Uno de los objetivos es predecir el valor de una serie de tiempo, x_{n+m} , $m = 1, 2, \dots$, basado en un conjunto de observaciones o puntos de entrenamiento desde un punto histórico hasta el presente, $x_n, x_{n-1}, x_{n-2}, \dots, x_1$.

Para el efecto se desea encontrar el predictor mínimo de error cuadrático medio para x_{n+m} ,

que está dada por

$$x_{n+m} = E(x_{n+m} | x_n, x_{n-1}, x_{n-2}, \dots, x_1) \quad (\text{III.6})$$

dado que el valor esperado condicional minimiza el error cuadrático medio,

$$E[x_{n+m} - g(\mathbf{x})]^2 \quad (\text{III.7})$$

donde $g(\mathbf{x})$ es una función de valores observados. Con la teoría que se desarrollará en el tercer capítulo, lo discutido con anterioridad posee argumentos similares. Por lo que el interés de estudiar las series de tiempo surge con algunos conceptos que se delimitarán a continuación.

Una consideración importante que se hace en el desarrollo de la teoría de series de tiempo es restringir a los predictores como funciones lineales construidas a partir de los datos de estudio, en particular tienen la forma:

$$x_{n+m}^n = \alpha_0 + \sum_{k=1}^n \alpha_k x_k \quad (\text{III.8})$$

donde $\alpha_1, \dots, \alpha_n$ son números reales.

Los predictores de la forma (III.8) que minimizan el error cuadrático medio, son llamados los mejores predictores lineales.

Existen componentes fundamentales que se deben considerar en el desarrollo de modelos de predicción con las series de tiempo; particularmente siendo de naturaleza física, química, biológica o económica, que dependen del concepto de un proceso estocástico que puede ser estacionario o no estacionario, autoregresivo o de media móvil.

Los modelos simples de series de tiempo como los *AR* y *MA* sostienen algunos de estos conceptos; pero, la composición de estos modelos extienden la noción de los modelos *ARMA* que son en sí bases para integrar una extensión más robusta de modelos de las series de tiempo, éstos modelos son *ARIMA*, que integran y completan en su estructura los conceptos discutidos en el párrafo anterior. Los modelos *ARIMA* pueden ser ordinarios y generales (*SARIMA*).

La motivación parte de entender las condiciones que hacen que dichos modelos tengan esa robusta estructura y cuál es la importancia de la causalidad e invertibilidad de los modelos *ARMA*.

D. Problema

Lo que motiva principalmente este trabajo de graduación, es establecer una visión sobre la interacción de las variables físico-químicas del Lago de Atitlán. Además de describir la dinámica global del Lago de Atitlán a partir de los sitios de estudios seleccionados. Y explicar el comportamiento (en media) de la transparencia para el año 2018.

IV Regresión lineal múltiple

Se desarrollará en este capítulo la teoría del modelo de regresión lineal múltiple (RLM), exponiendo el método de mínimos cuadrados para la estimación de los coeficientes de modelación para la regresión múltiple. Se exponen los elementos que influyen en el ajuste de datos y se demostrarán los resultados más importantes.

A. Modelo lineal y mínimos cuadrados

Dada una matriz con entradas $\mathbf{X} = (x_{ij})$, $1 \leq i, j \leq n, p$ se puede predecir \mathbf{y} , una variable de interés basado en x_1, x_2, \dots, x_p , con el modelo

$$\hat{y}_i = \hat{\beta}_0 + \sum_{j=1}^p x_{ij} \hat{\beta}_j \quad (\text{IV.1})$$

en particular $\hat{\beta}_0$ es una entrada del vector $\hat{\boldsymbol{\beta}}_0$ de $p \times 1$ de intercepto de $\hat{\mathbf{y}}$, mientras que $\hat{\beta}_j$ es una entrada del vector de coeficientes a estimar del modelo de regresión múltiple $\hat{\boldsymbol{\beta}}$ de $p \times 1$. Una forma conveniente es introducir el vector unidad $\mathbf{1}$ de $n \times 1$ en la matriz \mathbf{X} , siendo ahora $\mathbf{X} = (x_{ij})$ para $1, 0 \leq i, j \leq n, p$, tal que $(x_{i0}) = 1$ para $1 \leq i \leq n$ y la constante $\hat{\beta}_0$ en el vector de coeficientes $\hat{\boldsymbol{\beta}}$, ahora se tiene que $\hat{\boldsymbol{\beta}}$ cumple con $\hat{\beta}_j$ es tal que $0 \leq j \leq p$ por lo que se puede escribir $\hat{\mathbf{y}}$ como un producto interno

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}. \quad (\text{IV.2})$$

Nótese que se está modelando $\hat{\mathbf{y}}$ de valores reales, en general $\hat{\mathbf{y}}$ es un vector de $n \times 1$, \mathbf{X} es una matriz de $n \times (p + 1)$, $\hat{\boldsymbol{\beta}}$ es una matriz de $(p + 1) \times 1$ de coeficientes. En este caso en el espacio $(p + 1)$ -dimensional, $(\mathbf{X}, \hat{\mathbf{y}})$ representa un hiperplano. Si el vector unidad $\mathbf{1}$ está incluida en \mathbf{X} entonces el hiperplano incluye al origen y es un subespacio, de lo contrario es un conjunto afín que corta el eje y en el punto $(0, \dots, \beta_0)$. En este capítulo se asume que β_0 está incluida en $\hat{\boldsymbol{\beta}}$.

Se darán algunas definiciones para la terminología de la teoría.

Definición A.1 *La variable respuesta/dependiente, \mathbf{y} , es la variable que se desea predecir. Las variables predictoras o explicativas/independientes, x_1, x_2, \dots, x_p , son las variables que se utilizarán para realizar la predicción.*

Definición A.2 *La regresión es el estudio de la distribución condicional $y|X$ de la variable respuesta \mathbf{y} dada la matriz de predicción $\mathbf{X} = (x_{ij})$, $1 \leq i, j \leq n, p$.*

Definición A.3 Suponga que \mathbf{y} es la variable respuesta y al menos una variable predictora x_i es cuantitativa. Entonces el modelo de regresión lineal múltiple (RLM) es

$$y_i = x_{i0}\beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{ip}\beta_p + e_i = \sum_{j=0}^p x_{ij}\beta_j + e_i \quad (\text{IV.3})$$

para $i = 1, \dots, n$ donde n es el tamaño de la muestra (observaciones), e_i es una variable aleatoria del i -ésimo error. Y que $(x_{i0}) = 1$, para $i = 1, \dots, n$.

En notación matricial se tiene que $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ donde \mathbf{y} es un vector de $n \times 1$, \mathbf{X} es una matriz de $n \times q$ de variables predictoras con $q = p + 1$, $\boldsymbol{\beta}$ es un vector de coeficientes desconocidos de $q \times 1$, \mathbf{e} el vector de error de $n \times 1$.

Se abordarán las ideas expuestas en la motivación de manera formal, presentando la teoría que sustenta la regresión lineal múltiple.

Definición A.4 Condiciones de Gauss-Markov del modelo de regresión lineal múltiple. Sean los errores e_1, e_2, \dots, e_n del modelo de regresión lineal múltiple, entonces estos cumplen con ser independientes e idénticamente distribuidos con

$$E(e_i) = 0 \quad \text{y} \quad \text{Var}(e_i) = \sigma^2 < \infty$$

Se asume también que los errores son independientes de las variables predictoras $x_i, 1 \leq i \leq p$.

Si las variables predictoras son variables aleatorias, entonces el modelo RLM está condicionada a los valores observados de los x_i .

Definición A.5 Sea $\hat{\boldsymbol{\beta}}$ una estimación de $\boldsymbol{\beta}$, el vector correspondiente a los valores ajustados está dada por $\hat{\mathbf{y}} = \hat{\mathbf{y}}(\hat{\boldsymbol{\beta}}) = \mathbf{X}\hat{\boldsymbol{\beta}}$ por lo que el i -ésimo valor ajustado es

$$\hat{y}_i = \hat{y}_i(\hat{\boldsymbol{\beta}}) = x_{i0}\hat{\beta}_0 + x_{i1}\hat{\beta}_1 + x_{i2}\hat{\beta}_2 + \cdots + x_{ip}\hat{\beta}_p = \sum_{j=0}^p x_{ij}\hat{\beta}_j$$

El vector de residuos es $\mathbf{e} = \mathbf{e}(\hat{\boldsymbol{\beta}}) = \mathbf{y} - \hat{\mathbf{y}}(\hat{\boldsymbol{\beta}})$, entonces el i -ésimo valor residual es

$$e_i = e_i(\hat{\boldsymbol{\beta}}) = y_i - \hat{y}_i(\hat{\boldsymbol{\beta}}) = y_i - x_{i0}\hat{\beta}_0 - x_{i1}\hat{\beta}_1 - x_{i2}\hat{\beta}_2 - \cdots - x_{ip}\hat{\beta}_p$$

con $(x_{i0}) = 1$, para $1 \leq i \leq n$.

El fundamento de predicción con regresión lineal múltiple se centra en encontrar una estimación $\hat{\boldsymbol{\beta}}$ de $\boldsymbol{\beta}$ el cual minimice la suma residual de cuadrados (SRC).

Teorema A.6 El estimador de mínimos cuadrados $\hat{\boldsymbol{\beta}}$ que minimiza

$$SRC(\beta) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (\text{IV.4})$$

es tal que

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X} \mathbf{y} \quad (\text{IV.5})$$

Demostración.

Se empleará el método de mínimos cuadrados para encontrar $\hat{\beta}$. Sea la suma residual de cuadrados

$$SRC(\beta) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n e_i^2 = \mathbf{e}^T \mathbf{e}$$

pero

$$\begin{aligned} \mathbf{e}^T \mathbf{e} &= (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) = (\mathbf{y}^T - \beta^T \mathbf{X}^T) (\mathbf{y} - \mathbf{X}\beta) = \mathbf{y}^T \mathbf{y} - \beta^T \mathbf{X}^T \mathbf{y} - \mathbf{y}^T \mathbf{X}\beta + \beta^T \mathbf{X}^T \mathbf{X}\beta \\ &= \mathbf{y}^T \mathbf{y} - 2\beta^T \mathbf{X}^T \mathbf{y} + \beta^T \mathbf{X}^T \mathbf{X}\beta \end{aligned}$$

dado que

$$\beta^T \mathbf{X}^T \mathbf{y} = (\beta^T \mathbf{X}^T \mathbf{y})^T = \mathbf{y}^T \mathbf{X}\beta$$

al diferenciar $\mathbf{e}^T \mathbf{e}$ respecto de β , se tiene

$$\frac{\partial \mathbf{e}^T \mathbf{e}}{\partial \beta} = \frac{\partial (\mathbf{y}^T \mathbf{y} - 2\beta^T \mathbf{X}^T \mathbf{y} + \beta^T \mathbf{X}^T \mathbf{X}\beta)}{\partial \beta} = -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X}\beta = 0 \quad (\text{IV.6})$$

como

$$\frac{\partial (\beta^T \mathbf{X}^T \mathbf{X}\beta)}{\partial \beta} = 2\mathbf{X}^T \mathbf{X}\beta$$

luego se tiene que

$$-2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X}\beta = 0 \quad \Rightarrow \quad 2\mathbf{X}^T \mathbf{y} = 2\mathbf{X}^T \mathbf{X}\beta$$

por lo que

$$\mathbf{X}^T \mathbf{y} = \mathbf{X}^T \mathbf{X}\beta \quad (\text{IV.7})$$

la última expresión es conocida como la ecuación normal, al cumplir β la condición de ecuación normal se tiene que β es una solución de mínimos cuadrados, por el teorema (XII.A.1.1). Por otro lado \mathbf{X} tiene rango completo, entonces $(\mathbf{X}^T \mathbf{X})^{-1}$ existe (es no singular), por el teorema (XII.A.1.2.), entonces el estimador de mínimos cuadrados de β es único y está dado por

$$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X}\beta = \hat{\beta}$$

por lo tanto

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad \blacksquare$$

Del resultado anterior se puede escribir

$$\hat{\mathbf{y}} = \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

naturalmente la matriz $\mathbf{H} = \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ es conocida como la matriz del sombrero, en la literatura se le describe así debido a que coloca el símbolo de sombrero (circunflejo) sobre \mathbf{y} .

Propiedad A.7 Sea $E[\mathbf{e}] = \mathbf{0}$, los errores son insesgados, y las columnas de \mathbf{X} son linealmente independientes, entonces $E[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$.

Demostración.

$$E[\hat{\boldsymbol{\beta}}] = E[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}] = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E[\mathbf{y}] = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = \boldsymbol{\beta} \quad \blacksquare$$

Por lo que $\hat{\boldsymbol{\beta}}$ es un estimador insesgado de $\boldsymbol{\beta}$. Si se asume que los e_i no están correlacionados y tienen la misma varianza, esto es $Cov[e_i, e_j] = \delta_{ij} \sigma^2$, entonces $Var[\mathbf{e}] = \sigma^2 \mathbf{I}_n$, luego se tiene que $Var[\mathbf{y}] = Var[\mathbf{y} - \mathbf{X} \boldsymbol{\beta}] = Var[\mathbf{e}]$.

Propiedad A.8 Suponga que los y_i no están correlacionados, tal que poseen una varianza constante σ^2 y los x_i son valores fijos, entonces $Var[\hat{\boldsymbol{\beta}}] = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}$

Demostración.

Sea

$$\begin{aligned} Var[\hat{\boldsymbol{\beta}}] &= Var[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}] = Cov[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}, (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}] \\ &= ((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T) Cov[\mathbf{y}, \mathbf{y}] ((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T)^T = ((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T) Var[\mathbf{y}] ((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T)^T \\ &= Var[\mathbf{y}] (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} ((\mathbf{X}^T \mathbf{X})^{-1})^T = Var[\mathbf{y}] (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{X}) ((\mathbf{X}^T \mathbf{X})^T)^{-1} \\ &= Var[\mathbf{y}] (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{X})^T ((\mathbf{X}^T \mathbf{X})^T)^{-1} = Var[\mathbf{y}] (\mathbf{X}^T \mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \quad \blacksquare \end{aligned}$$

Propiedad A.9 Si $E[\mathbf{y}] = \mathbf{X} \boldsymbol{\beta}$, \mathbf{X} es una matriz de $n \times p$ de rango r , ($r < p$), y $Var[\mathbf{y}] = \sigma^2 \mathbf{I}_n$ entonces

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \hat{\mathbf{y}})^T (\mathbf{y} - \hat{\mathbf{y}})}{n - r} = \frac{RSC}{n - r}$$

es un estimador insesgado de σ^2 .

Demostración.

Considere $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$ donde \mathbf{X} es una matriz de $n \times r$ de rango r , entonces

$$\mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I}_n - \mathbf{H})\mathbf{y}$$

con $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$, por (XII.D.1.1) se tiene que

$$\begin{aligned} (n-r)\hat{\sigma}^2 &= (\mathbf{y} - \hat{\mathbf{y}})^T(\mathbf{y} - \hat{\mathbf{y}}) = ((\mathbf{I}_n - \mathbf{H})\mathbf{y})^T((\mathbf{I}_n - \mathbf{H})\mathbf{y}) \\ &= \mathbf{y}^T(\mathbf{I}_n - \mathbf{H})^T(\mathbf{I}_n - \mathbf{H})\mathbf{y} \end{aligned}$$

por (XII.D.1.3)

$$(\mathbf{I}_n - \mathbf{H})^T(\mathbf{I}_n - \mathbf{H}) = (\mathbf{I}_n - \mathbf{H})^2 = (\mathbf{I}_n - \mathbf{H})$$

por lo que

$$(n-r)\hat{\sigma}^2 = \mathbf{y}^T(\mathbf{I}_n - \mathbf{H})\mathbf{y}$$

Como $\mathbf{H}\mathbf{y} = \mathbf{y}$, por (XII.C.2 y XII.D.1.3) aplicado a \mathbf{X} se obtiene

$$E[\mathbf{y}^T(\mathbf{I}_n - \mathbf{H})\mathbf{y}] = \sigma^2 \text{tr}(\mathbf{I}_n - \mathbf{H}) + \hat{\mathbf{y}}(\mathbf{I}_n - \mathbf{H})\mathbf{y} = \sigma^2(n-r)$$

dado que

$$(n-r)E[\hat{\sigma}^2] = E[(n-r)\hat{\sigma}^2] = E[\mathbf{y}^T(\mathbf{I}_n - \mathbf{H})\mathbf{y}] = \sigma^2(n-r)$$

se tiene que

$$E[\hat{\sigma}^2] = \sigma^2 \quad \blacksquare$$

Un resultado importante dentro de la teoría de estimaciones por mínimos cuadrados es el teorema de Gauss-Markov, que presenta argumentos a la inquietud del porqué se elige $\boldsymbol{\beta}$ y no cualquier estimador para $\boldsymbol{\beta}$, para llegar a enunciar dicho teorema se exhibirán algunos hechos que engloba los elementos de Gauss-Markov.

Teorema A.10 *Sea $\hat{\mathbf{y}}$ la estimación por mínimos cuadrados de $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$, donde $\mathbf{y} \in \mathcal{C}(\mathbf{X})$, espacio columna de \mathbf{X} , y \mathbf{X} puede no tener rango completo. Entonces entre la familia de estimaciones lineales insesgadas de $\mathbf{c}^T\mathbf{y}$, $\mathbf{c}^T\hat{\mathbf{y}}$ es la única estimación de varianza mínima. Se dice que $\mathbf{c}^T\hat{\mathbf{y}}$ es la mejor estimación lineal insesgada de $\mathbf{c}^T\mathbf{y}$.*

Demostración.

Se sabe que

$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{y} \Rightarrow \mathbf{H}\mathbf{y} = \mathbf{H}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta} = \mathbf{y}$$

esto es por el teorema (XII.D.1.1), luego considere

$$E[\mathbf{c}^T \hat{\mathbf{y}}] = \mathbf{c}^T \mathbf{H} \mathbf{y} = \mathbf{c}^T \mathbf{y} \quad \forall \mathbf{y} \in \mathcal{C}(\mathbf{X})$$

por lo que $\mathbf{c}^T \hat{\mathbf{y}}$ es una estimación lineal insesgada de $\mathbf{c}^T \mathbf{y}$. Sea $\mathbf{d}^T \mathbf{y}$ otra estimación lineal insesgada de $\mathbf{c}^T \mathbf{y}$. Entonces $\mathbf{c}^T \mathbf{y} = E[\mathbf{d}^T \hat{\mathbf{y}}] = \mathbf{d}^T \mathbf{y}$ por lo que $(\mathbf{c} - \mathbf{d})^T \mathbf{y} = \mathbf{0} \Rightarrow (\mathbf{c} - \mathbf{d}) \perp \mathcal{C}(\mathbf{X})$ de ahí que

$$\mathbf{H}(\mathbf{c} - \mathbf{d}) = \mathbf{0} \Rightarrow \mathbf{H}\mathbf{c} = \mathbf{H}\mathbf{d}$$

Nótese que $\mathbf{c}^T \hat{\mathbf{y}} = \mathbf{c}^T \mathbf{H} \mathbf{y} = (\mathbf{H}\mathbf{c})^T \mathbf{y}$, entonces

$$\begin{aligned} \text{Var}[\mathbf{c}^T \hat{\mathbf{y}}] &= \text{Var}[(\mathbf{H}\mathbf{c})^T \mathbf{y}] = \text{Var}[(\mathbf{H}\mathbf{d})^T \mathbf{y}] \\ &= (\mathbf{H}\mathbf{d})^T \text{Var}[\mathbf{y}] (\mathbf{H}\mathbf{d})^T = \sigma^2 \mathbf{d}^T \mathbf{H}^T \mathbf{H} \mathbf{d} \\ &= \sigma^2 \mathbf{d}^T \mathbf{H}^2 \mathbf{d} = \sigma^2 \mathbf{d}^T \mathbf{H} \mathbf{d} \end{aligned}$$

Por lo que

$$\begin{aligned} \text{Var}[\mathbf{d}^T \mathbf{y}] - \text{Var}[\mathbf{c}^T \hat{\mathbf{y}}] &= \text{Var}[\mathbf{d}^T \mathbf{y}] - \text{Var}[(\mathbf{H}\mathbf{d})^T \mathbf{y}] \\ &= \sigma^2 \mathbf{d}^T \mathbf{d} - \sigma^2 \mathbf{d}^T \mathbf{H} \mathbf{d} = \sigma^2 (\mathbf{d}^T \mathbf{d} - \mathbf{d}^T \mathbf{H} \mathbf{d}) \\ &= \sigma^2 \mathbf{d}^T (\mathbf{d} - \mathbf{H}\mathbf{d}) = \sigma^2 \mathbf{d}^T (\mathbf{I}_n - \mathbf{H}) \mathbf{d} \\ &= \sigma^2 \mathbf{d}^T (\mathbf{I}_n - \mathbf{H})^T (\mathbf{I}_n - \mathbf{H}) \mathbf{d} \\ &= \sigma^2 (\mathbf{d}_1)^T (\mathbf{d}_1) \geq 0 \end{aligned}$$

por lo tanto

$$\text{Var}[\mathbf{d}^T \mathbf{y}] \geq \text{Var}[\mathbf{c}^T \hat{\mathbf{y}}]$$

La igualdad se da cuando $(\mathbf{I}_n - \mathbf{H})\mathbf{d} = \mathbf{0}$ ó $\mathbf{d} = \mathbf{H}\mathbf{d} = \mathbf{H}\mathbf{c}$.

Por lo tanto $\mathbf{c}^T \hat{\mathbf{y}}$ tiene varianza mínima y es única ■

Corolario A.11 Si \mathbf{X} tiene rango completo, entonces $\mathbf{a}^T \hat{\boldsymbol{\beta}}$ es el mejor estimador lineal insesgado de $\mathbf{a}^T \boldsymbol{\beta}$, para todo vector \mathbf{a} .

Demostración. Al ser $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$ se tiene que $\boldsymbol{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$, de igual manera $\mathbf{y} = \mathbf{X} \hat{\boldsymbol{\beta}}$ y $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{y}}$. Por el teorema (IV.A.10) con $\mathbf{c}^T = \mathbf{a}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ se obtiene que $\mathbf{a}^T \hat{\boldsymbol{\beta}} = \mathbf{c}^T \hat{\mathbf{y}}$ es el mejor estimador lineal insesgado de $\mathbf{a}^T \boldsymbol{\beta} = \mathbf{c}^T \mathbf{y} \quad \forall \mathbf{a}$ ■

Teorema A.12 Teorema de Gauss-Markov. Suponga que valen las condiciones de Gauss-Markov, \mathbf{X} es de rango completo. $\mathbf{c}^T \hat{\boldsymbol{\beta}}$, estimador por mínimos cuadrados, es el mejor estimador lineal

insesgado de $\mathbf{c}^T \boldsymbol{\beta}$ que tiene la menor varianza entre la familia de estimadores lineales insesgados de $\mathbf{c}^T \boldsymbol{\beta}$.

Demostración.

Resultado inmediato del teorema (IV.A.10) y corolario (IV.A.11).

De hecho la esencia del teorema de Gauss-Markov responde a la inquietud de estimar los parámetros de $\boldsymbol{\beta}$, por mínimos cuadrados, dado que dicha estimación posee la menor varianza de todas las estimaciones lineales insesgadas.

B. Distribución, correlación y prueba estadística

Se ha supuesto que $E[e] = \mathbf{0}$ y $Var[e] = \sigma^2 \mathbf{I}_n$. Si se asume que los e_i están distribuidos normalmente, entonces $\mathbf{e} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ y por lo tanto $\mathbf{y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$. A partir de los supuestos anteriores se puede establecer el siguiente teorema.

Teorema B.1

Si $\mathbf{y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$, donde \mathbf{X} es una matriz de $n \times p$, con rango p , entonces:

1. $\hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^T \mathbf{X})^{-1})$.
2. $\frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathbf{X}^T \mathbf{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\sigma^2} \sim \chi_p^2$
3. $\hat{\boldsymbol{\beta}}$ es independiente de $\hat{\sigma}$.
4. $\frac{RSC}{\sigma^2} = \frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-p}^2$.

Demostración. 1. Como $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} = \mathbf{C} \mathbf{y}$, \mathbf{C} es una matriz de $p \times n$ tal que

$$rank(\mathbf{C}) = rank(\mathbf{X}^T) = rank(\mathbf{X}) = p$$

$\hat{\boldsymbol{\beta}}$ tiene una distribución normal multivariable (XII.D.2). Por las propiedades (IV.A.7) y (IV.A.8) se cumple que $\hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^T \mathbf{X})^{-1})$.

2. Note que $\frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathbf{X}^T \mathbf{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\sigma^2} = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T (Var[\hat{\boldsymbol{\beta}}])^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$, el cual por la propiedad anterior y el teorema (XII.D.3) se distribuye como χ_p^2 .

3. Sea

$$\begin{aligned} Cov[\hat{\boldsymbol{\beta}}, \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}] &= Cov[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}, (\mathbf{I}_n - \mathbf{H})\mathbf{y}] \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T Cov[(\mathbf{y})] (\mathbf{I}_n - \mathbf{H})^T \\ &= \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I}_n - \mathbf{H}) = \mathbf{0} \end{aligned}$$

esto es por el teorema (XII.D.1), si se hace $\mathbf{U} = \hat{\boldsymbol{\beta}}$ y $\mathbf{V} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ en (XII.D.4), $\hat{\boldsymbol{\beta}}$ es independiente de $\|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2$ y por lo tanto es independiente de $\hat{\sigma}$.

4. Por la propiedad (IV.A.9) de esta sección y (II.D.1.3), se tiene que

$$RSC = \mathbf{y}^T(\mathbf{I}_n - \mathbf{H})\mathbf{y} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{I}_n - \mathbf{H})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \boldsymbol{\epsilon}^T(\mathbf{I}_n - \mathbf{H})\boldsymbol{\epsilon}$$

al ser $(\mathbf{I}_n - \mathbf{H})$ simétrica e idempotente de rango $n - p$. Como $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, se cumple que $RSS/\sigma^2 \sim \chi_{n-p}^2$ ■

Teorema B.2

1. $RSC_H - RSC = \|\hat{\mathbf{y}} - \hat{\mathbf{y}}_H\|^2 = (\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{c})^T[\mathbf{A}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{A}^T]^{-1}(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{c})$
2. $E[RSC_H - RSC] = \sigma^2 q + (\mathbf{A}\boldsymbol{\beta} - \mathbf{c})^T[\mathbf{A}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{A}^T]^{-1}(\mathbf{A}\boldsymbol{\beta} - \mathbf{c}) = \sigma^2 q + (RSS_H - RSS)_y$
3. Cuando se cumple $H : \mathbf{A}\boldsymbol{\beta} = \mathbf{c}$, hipótesis, se tiene que

$$F = \frac{(RSC_H - RSC)/\sigma^2 q}{RSC/\sigma^2(n-p)} = \frac{(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{c})^T[\mathbf{A}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{A}^T]^{-1}(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{c})}{q\hat{\sigma}^2}$$

está distribuida como $F_{q, n-p}$.

4. Cuando $\mathbf{c} = \mathbf{0}$, F puede expresarse de la forma:

$$F = \frac{n-p}{q} \frac{\mathbf{y}^T(\mathbf{H} - \mathbf{H}_H)\mathbf{y}}{\mathbf{y}^T(\mathbf{I}_n - \mathbf{H})\mathbf{y}}$$

donde \mathbf{H}_H es simétrica e idempotente, y $\mathbf{H}_H\mathbf{H} = \mathbf{H}\mathbf{H}_H = \mathbf{H}_H$.

Demostración.

1. Nótese que

$$\|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_H\|^2 = \|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2 + \|\mathbf{X}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_H)\|^2 \quad (\text{IV.8})$$

al hacer $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ y $\hat{\mathbf{y}}_H = \mathbf{X}\hat{\boldsymbol{\beta}}_H$ se tiene

$$\|\mathbf{y} - \hat{\mathbf{y}}_H\|^2 - \|\mathbf{y} - \hat{\mathbf{y}}\|^2 = \|\hat{\mathbf{y}} - \hat{\mathbf{y}}_H\|^2 \quad (\text{IV.9})$$

por lo que

$$RSC = \|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|^2 \quad (\text{IV.10})$$

y

$$RSC_H = \|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_H\|^2 = \|\mathbf{y} - \hat{\mathbf{y}}_H\|^2 \quad (\text{IV.11})$$

de ahí que

$$RSC_H - RSC = \|\hat{\mathbf{y}} - \hat{\mathbf{y}}_H\|^2 = (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_H)^T \mathbf{X}^T \mathbf{X} (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_H) \quad (\text{IV.12})$$

Una forma de expresar $\hat{\beta}_H$ a partir de la solución al problema de minimizar $\epsilon^T \epsilon = SRC(\beta)$ con el método de multiplicadores de Lagrange, con la condición $\mathbf{A}\beta = \mathbf{c}$, dada $\mathbf{y} = \mathbf{X}\beta + \epsilon$ es:

$$\hat{\beta}_H = \hat{\beta} + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{A}^T [\mathbf{A}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{A}^T]^{-1} (\mathbf{c} - \mathbf{A}\hat{\beta}) \quad (\text{IV.13})$$

que al sustituir la expresión anterior en (IV.22) del lado derecho se obtiene

$$\lambda = ((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{A}^T [\mathbf{A}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{A}^T]^{-1} (\mathbf{A}\hat{\beta} - \mathbf{c}))^T \mathbf{X}^T \mathbf{X} ((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{A}^T [\mathbf{A}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{A}^T]^{-1} (\mathbf{A}\hat{\beta} - \mathbf{c}))$$

con $RSC_H - RSC = \lambda$ y al ser $\hat{\beta} - \hat{\beta}_H$ no singular, es fácil notar que λ puede simplificarse como

$$\lambda = (\mathbf{A}\hat{\beta} - \mathbf{c})^T [\mathbf{A}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{A}^T]^{-1} (\mathbf{A}\hat{\beta} - \mathbf{c}) = RSC_H - RSC.$$

2. Las filas de \mathbf{A} son linealmente independientes y $\hat{\beta} \sim N_p(\beta, \sigma^2(\mathbf{X}^T \mathbf{X})^{-1})$, por el teorema (XII.D.2) se tiene que

$$\mathbf{A}\hat{\beta} = N_q(\mathbf{A}\beta, \sigma^2 \mathbf{A}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{A}^T)$$

sea $\mathbf{Z} = \mathbf{A}\hat{\beta} - \mathbf{c}$ y $\mathbf{B} = \mathbf{A}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{A}^T$, entonces

$$E[\mathbf{Z}] = \mathbf{A}\beta - \mathbf{c} \quad y \quad Var[\mathbf{Z}] = Var[\mathbf{A}\hat{\beta}] = \sigma^2 \mathbf{B}$$

por el teorema (XII.C.2) a la condición anterior se tiene:

$$\begin{aligned} E[RSC_H - RSC] &= E[\mathbf{Z}^T \mathbf{B}^{-1} \mathbf{Z}] \\ &= tr(\sigma^2 \mathbf{B}^{-1} \mathbf{B}) + (\mathbf{A}\beta - \mathbf{c})^T \mathbf{B}^{-1} (\mathbf{A}\beta - \mathbf{c}) \\ &= tr(\sigma^2 \mathbf{I}_q) + (\mathbf{A}\beta - \mathbf{c})^T \mathbf{B}^{-1} (\mathbf{A}\beta - \mathbf{c}) \\ &= \sigma^2 q + (\mathbf{A}\beta - \mathbf{c})^T \mathbf{B}^{-1} (\mathbf{A}\beta - \mathbf{c}). \end{aligned}$$

3. Del primer inciso de este teorema se tiene que $RSC_H - RSC$ es una función continua de $\hat{\beta}$ y es por lo tanto independiente de RSC esto es por el teorema (IV.B.1.3), ahora si la hipótesis H es verdadera, $\mathbf{A}\hat{\beta} \sim N_q(\mathbf{c}, \sigma^2 \mathbf{A}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{A}^T)$, entonces por el teorema (XII.D.3)

$$\frac{RSC_H - RSC}{\sigma^2} = (\mathbf{A}\hat{\beta} - \mathbf{c})^T (Var[\mathbf{A}\hat{\beta}])^{-1} (\mathbf{A}\hat{\beta} - \mathbf{c})$$

es χ_q^2 . Como $\frac{RSC}{\sigma^2} \sim \chi_{n-p}^2$, por el teorema (IV.D.1.4), se tiene que

$$F = \frac{(RSC_H - RSC)/\sigma^2 q}{RSC/\sigma^2(n-p)}$$

es de la forma

$$\frac{[\chi_q^2/q]}{[\chi_{n-p}^2/(n-p)]}$$

cuando la hipótesis H es verdadera, por tanto $F \sim F_{q,n-p}$ cuando H es verdadera.

4. Por (IV.13) con $\mathbf{c} = \mathbf{0}$, se tiene que

$$\begin{aligned}\hat{\mathbf{y}}_H &= \mathbf{X}\hat{\boldsymbol{\beta}}_H = \mathbf{X}[\hat{\boldsymbol{\beta}} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{A}^T[\mathbf{A}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{A}^T]^{-1}(-\mathbf{A}\hat{\boldsymbol{\beta}})] \\ \hat{\mathbf{y}}_H &= \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{A}^T[\mathbf{A}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{A}^T]^{-1}(\mathbf{A}\hat{\boldsymbol{\beta}})\end{aligned}\quad (\text{IV.14})$$

de ahí que

$$\begin{aligned}\hat{\mathbf{y}}_H &= \mathbf{X}[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}] - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{A}^T[\mathbf{A}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{A}^T]^{-1}(\mathbf{A})[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}] \\ &= [\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{A}^T[\mathbf{A}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{A}^T]^{-1}(\mathbf{A})(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T]\mathbf{y} \\ &= [\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{A}^T(\mathbf{A}^T)^{-1}((\mathbf{X}^T\mathbf{X})^{-1})^{-1}(\mathbf{A}^{-1})(\mathbf{A})(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T]\mathbf{y}\end{aligned}$$

por lo que

$$\hat{\mathbf{y}}_H = [\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T - (\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)\mathbf{1}]\mathbf{y} \quad (\text{IV.15})$$

$$\hat{\mathbf{y}}_H = [\mathbf{H} - \mathbf{H}_1]\mathbf{y} = \mathbf{H}_H\mathbf{y} \quad (\text{IV.16})$$

dado que \mathbf{H}_H es simétrica. Siguiendo un procedimiento similar al empleado en la demostración de (XII.D.1), se demuestra que $\mathbf{H}_1\mathbf{H} = \mathbf{H}\mathbf{H}_1 = \mathbf{H}_1$, luego

$$\begin{aligned}\mathbf{H}_H^2 &= \mathbf{H}^2 - \mathbf{H}_1\mathbf{H} - \mathbf{H}\mathbf{H}_1 + \mathbf{H}_1^2 \\ &= \mathbf{H} - 2\mathbf{H}_1 + \mathbf{H}_1 = \mathbf{H} - \mathbf{H}_1\end{aligned}$$

por lo que

$$\mathbf{H}_H^2 = \mathbf{H}_H \quad (\text{IV.17})$$

Por otro lado

$$\mathbf{H}_H\mathbf{H} = (\mathbf{H} - \mathbf{H}_1)\mathbf{H} = \mathbf{H} - \mathbf{H}_1 = \mathbf{H}_H \quad (\text{IV.18})$$

nótese que al tomar la transpuesta de (IV.18) se obtiene

$$(\mathbf{H}_H\mathbf{H})^T = \mathbf{H}^T(\mathbf{H}_H^T) = \mathbf{H}\mathbf{H}_H = \mathbf{H}_H$$

por lo que

$$\mathbf{H}\mathbf{H}_H = \mathbf{H}_H \quad (\text{IV.19})$$

Sabemos que $RSC = \mathbf{y}^T(\mathbf{I}_n - \mathbf{H})\mathbf{y}$ y usando un procedimiento similar al (XII.B.1) se obtiene que

$$RSC_H = \|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_H\|^2 = \mathbf{y}^T(\mathbf{I}_n - \mathbf{H})^2\mathbf{y} = \mathbf{y}^T(\mathbf{I}_n - \mathbf{H})\mathbf{y} \quad (\text{IV.20})$$

por lo que

$$RSC_H - RSC = \mathbf{y}^T(\mathbf{H} - \mathbf{H}_H)\mathbf{y}$$

por lo tanto

$$F = \frac{(RSC_H - RSC)/q}{RSC/(n-p)} = \frac{n-p}{q} \frac{\mathbf{y}^T(\mathbf{H} - \mathbf{H}_H)\mathbf{y}}{\mathbf{y}^T(\mathbf{I}_n - \mathbf{H})\mathbf{y}} \quad \blacksquare$$

Para culminar la exposición de la teoría de regresión lineal múltiple, se abordarán algunos conceptos para el enfoque de prueba y criterio de selección de modelo.

Definición B.3 Se define el coeficiente de correlación múltiple para una muestra R como la correlación del par (y_i, \hat{y}_i) , y se denota por:

$$R = \frac{\sum(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{[\sum(y_i - \bar{y})^2 \sum(\hat{y}_i - \bar{\hat{y}})^2]^{1/2}}$$

Teorema B.4

1.

$$\sum_i (y_i - \bar{y})^2 = \sum_i (y_i - \hat{y}_i)^2 + \sum_i (\hat{y}_i - \bar{y})^2$$

2.

$$R^2 = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \hat{y}_i)^2} = 1 - \frac{RSC}{\sum_i (y_i - \bar{y})^2}$$

Demostración. Sea $\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$, considere

$$\hat{\mathbf{y}}^T \hat{\mathbf{y}} = \hat{\mathbf{y}}^T \mathbf{H}^2 \hat{\mathbf{y}} = \hat{\mathbf{y}}^T \mathbf{H} \hat{\mathbf{y}} = \mathbf{y}^T \hat{\mathbf{y}} \quad (\text{IV.21})$$

De la ecuación normal se tiene que

$$\sum_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i,1} - \cdots - \hat{\beta}_{p-1} x_{i,p-1}) = 0 \quad (\text{IV.22})$$

esto es al diferenciar la siguiente expresión

$$\sum_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i,1} - \cdots - \hat{\beta}_{p-1} x_{i,p-1})^2 \quad (\text{IV.23})$$

respecto de β_0 . La expresión (IV.22) es una de las ecuaciones normales para $\hat{\boldsymbol{\beta}}$, que puede expresarse

como

$$\sum_i (y_i - \hat{y}_i) = 0 \quad (\text{IV.24})$$

luego

$$\begin{aligned} \sum_i (y_i - \bar{y})^2 &= \sum_i (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 = \sum_i (y_i - \hat{y}_i)^2 + \sum_i (\hat{y}_i - \bar{y})^2 + 2 \sum_i (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) \\ &= \sum_i (y_i - \hat{y}_i)^2 + \sum_i (\hat{y}_i - \bar{y})^2 \end{aligned}$$

esto es al aplicar (IV.22) y (IV.23) a la expresión $\sum_i (y_i - \hat{y}_i)(\hat{y}_i - \bar{y})$:

$$\sum_i (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = \sum_i (y_i - \hat{y}_i)\hat{y}_i = (\mathbf{y} - \hat{\mathbf{y}})^T \hat{\mathbf{y}} = 0$$

2. De la expresión (IV.24), se obtiene que $\bar{\hat{y}} = \bar{y}$, entonces

$$\sum_i (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}}) = \sum_i (y_i - \bar{y})(\hat{y}_i - \bar{y}) = \sum_i (y_i + \hat{y}_i - \hat{y}_i - \bar{y})(\hat{y}_i - \bar{y}) = \sum_i (\hat{y}_i - \bar{y})^2$$

y por la definición (B.3) se tiene que

$$\begin{aligned} R^2 &= \frac{[\sum (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})]^2}{\sum (y_i - \bar{y})^2 \sum (\hat{y}_i - \bar{\hat{y}})^2} = \frac{[\sum (\hat{y}_i - \bar{y})]^2}{\sum (y_i - \bar{y})^2 \sum (\hat{y}_i - \bar{y})^2} = \frac{[\sum (\hat{y}_i - \bar{y})]^2 [\sum (\hat{y}_i - \bar{y})^2]}{\sum (y_i - \bar{y})^2 \sum (\hat{y}_i - \bar{y})^2} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} \\ &= \frac{\sum (y_i - \bar{y})^2 - \sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{RSC}{\sum (y_i - \bar{y})^2} \quad \blacksquare \end{aligned}$$

Para la selección del mejor modelo de regresión lineal múltiple para un fenómeno dado, basta observar los valores de F y R^2 aceptables para el criterio y nivel de estudio. Dado que ello permite identificar cuáles de las variables de estudio pueden ser eliminadas o agregadas al modelo en cuestión. Una alternativa para la selección del modelo de estudio, que no se realiza secuencialmente sino que simplemente evalúa cada modelo propuesto en base a sus propiedades, es el uso del Criterio de Información de Akaike (AIC), la Corrección de Sesgo (AIC_c) o el Criterio de Información Bayesiano, éstos se definen de la siguiente manera.

Definición B.5 *Criterio de Información de Akaike (AIC).*

$$AIC = \ln \hat{\sigma}_k^2 + \frac{n + 2k}{n}$$

con $\hat{\sigma}_k^2 = \frac{RSC_k}{n}$. Siendo $\hat{\sigma}_k^2$ el estimador de máxima verosimilitud para la varianza, k es el número de variables en el modelo, n el número de observaciones y RSC_k es la suma residual de cuadrados del modelo con k variables.

Definición B.6 *AIC, Bias corregido (AIC_c).*

$$AIC_c = \ln \hat{\sigma}_k^2 + \frac{n+k}{n-k-2}$$

con $\hat{\sigma}_k^2 = \frac{RSC_k}{n}$. Siendo $\hat{\sigma}_k^2$ el estimador de máxima verosimilitud para la varianza, k es el número de variables en el modelo, n el número de observaciones y RSC_k es la suma residual de cuadrados del modelo con k variables.

Definición B.7 *Criterio de Información Bayesiano (BIC).*

$$SIC = \ln \hat{\sigma}_k^2 + \frac{k \ln n}{n}$$

con $\hat{\sigma}_k^2 = \frac{RSC_k}{n}$. Siendo $\hat{\sigma}_k^2$ el estimador de máxima verosimilitud para la varianza, k es el número de variables en el modelo y n el número de observaciones y RSC_k es la suma residual de cuadrados del modelo con k variables.

Definición B.8 *El estadístico Durbin-Watson. Se define por el estadístico Durbin-Watson a*

$$d = DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

Definición B.9 *El estadístico Kolmogorov-Smirnov. Se define por el estadístico Kolmogorov-Smirnov a*

$$D = KS = \max |F_n(x) - F_0(x)|$$

siendo $F_n(x)$ la función de distribución normal y $F_0(x)$ es la función teórica.

Definición B.10 *Contraste de Breusch-Pagan.*

Se define el contraste Breusch-Pagan como $H_0 = \text{var}(u_t|X_t) = \sigma^2$ y $H_1 = \text{var}(u_t|X_t) = \sigma_t^2$

V Series de Tiempo

En este capítulo se presentará la teoría de los modelos de series de tiempo, a veces procesos, debido al enfoque se exhibirá la construcción de estos modelos que incorporarán complejidad, durante el desarrollo de esta breve presentación, absorbiendo en su estructura elementos importantes en el modelado.

A. Modelos autoregresivos y de media móvil

Se definirá el operador retroceso y su relación con la diferencia de series, estos conceptos son útiles en el desarrollo de este capítulo.

Definición A.1 *Se define el operador retroceso B como*

$$Bx_t = x_{t-1}$$

que puede extenderse para potencias

$$B^2 = B(Bx_t) = Bx_{t-1} = x_{t-2}, B^3, B^4, \dots$$

y en general para un k entero positivo se tiene que

$$B^k x_t = x_{t-k}$$

Se dirá ∇ es la primera diferencia, y se define como

$$\nabla x_t = x_t - x_{t-1}$$

Por lo que se puede reescribir dicha diferencia como

$$\nabla x_t = (1 - B)x_t$$

Mientras que para la segunda diferencia se tiene que

$$\nabla^2 x_t = \nabla(\nabla x_t) = \nabla(x_t - x_{t-1}) = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2})$$

$$= x_t - 2x_{t-1} - x_{t-2} = (1 - 2B + B^2)x_t = (1 - B)^2 x_t$$

Definición A.2 Se define la diferencia de orden d como

$$\nabla^d = (1 - B)^d$$

Se definirán dos modelos importantes, que son simples y que explican fenómenos de distinta naturaleza.

Definición A.3 Un modelo autoregresivo de orden p , abreviado **AR**(p), es de la forma

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t \quad (\text{V.1})$$

donde x_t es estacionario, y $\phi_1, \phi_2, \dots, \phi_p$ son constantes ($\phi_p \neq 0$). No es necesario aún, pero se asume que w_t es un ruido blanco Gaussiano con media cero y varianza σ_w^2 , a menos que se indique lo contrario.

La media de x_t en (V.1) es cero, pero si la media de x_t es $\mu \neq 0$, se tiene que reemplazar x_t por $x_t - \mu$ en (V.1), esto es $x_t - \mu = \phi_1(x_{t-1} - \mu) + \phi_2(x_{t-2} - \mu) + \cdots + \phi_p(x_{t-p} - \mu) + w_t$ ó de la siguiente manera

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t, \text{ con } \alpha = \mu(1 - \phi_1 - \cdots - \phi_p) \quad (\text{V.2})$$

En terminos del operador retroceso el modelo AR(p) se puede escribir como

$$(1 - \phi B - \phi_2 B^2 - \phi_3 B^3 - \cdots - \phi_p B^p)x_t = w_t$$

o de manera mas compacta como $\phi(B)x_t = w_t$.

Definición A.4 Se define al operador autoregresivo como

$$\phi(B) = 1 - \phi B - \phi_2 B^2 - \phi_3 B^3 - \cdots - \phi_p B^p$$

Definición A.5 El modelo de media movil de orden q , abreviado **MA**(q), se define de la siguiente manera

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q} \quad (\text{V.3})$$

donde hay q diferencias en la media movil y $\theta_1, \theta_2, \dots, \theta_q$ ($\theta_q \neq 0$) son parametros. Aca se asume que w_t es un ruido blanco Gaussiano.

Definición A.6 El operador de media móvil se define como

$$\phi(B) = 1 + \theta_1 B^1 + \theta_2 B^2 + \cdots + \theta_q B^q \quad (\text{V.4})$$

De los modelos anteriores, se pueden extender modelos que absorben los elementos presentados; el estudio principal de estos modelos son las series de tiempo estacionarias.

Definición A.7 Una serie de tiempo $\{x_t : t = 0, \pm 1, \pm 2, \pm 3, \dots\}$ es un modelo autoregresivo de media móvil $ARMA(p, q)$ si es estacionario y

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q} \quad (\text{V.5})$$

con $\phi_p \neq 0$, $\theta_q \neq 0$ y $\sigma_w^2 > 0$. Donde p y q son los ordenes autoregresivo y de media móvil respectivamente, si la media de x_t es $\mu \neq 0$, se tiene que $\alpha = \mu(1 - \phi_1 - \cdots - \phi_p)$ para

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q}$$

A menos que se diga lo contrario, $\{w_t : t = 0, \pm 1, \pm 2, \pm 3, \dots\}$ es una secuencia de ruido blanco Gaussiano.

Nótese que si $q = 0$, el modelo es un $AR(p)$ y si $q = 0$ entonces se tiene que es un modelo $MA(q)$.

Una forma conveniente y concisa de escribir el modelo $ARMA(p, q)$ es:

$$\phi(B)x_t = \theta(B)w_t. \quad (\text{V.6})$$

Este modelo es de mayor extensión en el sentido de abordar fenómenos que poseen componentes autoregresivos y que su media cambia en distintos períodos de estudio. Los modelos $ARMA(p, q)$ pueden no describir con claridad los procesos estacionarios de estudio, para explicar un fenómeno de interés; debido a que presentan aún ciertas debilidades, los posibles problemas que pueden surgir con los modelos $ARMA(p, q)$ y que los hace débiles en su estructura son:

1. Parametros redundantes en el modelo.
2. Modelos estacionarios AR que dependen de valores futuros.
3. Modelos MA que no son únicos.

Por lo que se discutirá y resolverán dichos problemas, a través de condiciones en las cuales el modelo $ARMA(p, q)$ es más robusta; es decir el modelo $ARMA$ es causal e invertible.

Definición A.8 Los modelos **AR** y **MA** polinomiales se definen como

$$\phi(z) = 1 + \phi_1 z + \phi_2 z^2 + \cdots + \phi_p z^p, \quad \text{con} \quad \phi_p \neq 0 \quad (\text{V.7})$$

y

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q, \quad \text{con} \quad \theta_q \neq 0 \quad (\text{V.8})$$

respectivamente, donde z es un número complejo.

Para resolver los problemas mencionados anteriormente, se requiere para el primer problema tomar de aquí en adelante a un modelo $ARMA(p, q)$ cuando está en su forma simple, es decir a parte de las condiciones a la definición de $ARMA$ se requiere que $\phi(z)$ y $\theta(z)$ no tengan factores en común.

Mientras que para el segundo problema, basta introducir en el modelo $ARMA(p, q)$ el concepto de causalidad.

Definición A.9 Un modelo $ARMA(p, q)$,

$$\phi(B)x_t = \theta(B)w_t$$

se dice que es causal si la serie de tiempo $\{x_t : t = 0, \pm 1, \pm 2, \pm 3, \dots\}$ puede ser escrita en un lado como un proceso lineal:

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j} = \psi(B)w_t \quad (\text{V.9})$$

donde $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$, y $\sum_{j=0}^{\infty} |\psi_j| < \infty$; tal que $\psi_0 = 1$.

Propiedad A.10 Causalidad de un proceso $ARMA(p, q)$. Un modelo $ARMA(p, q)$ es causal si y solo si, $\phi(z) \neq 0$ para $|z| \leq 1$. Los coeficientes del proceso lineal que están dadas por la expresión (V.9) pueden determinarse resolviendo

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, \quad |z| \leq 1.$$

Demostración.

(\Rightarrow)

Suponga que x_t es un proceso causal, esto es, posee una representación dada por

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j}, \quad \sum_{j=0}^{\infty} |\psi_j| < \infty$$

se sabe que $x_t = \psi(B)w_t$, que al multiplicarlo por $\phi(B)$ se cumple que

$$\phi(B)x_t = \phi(B)\psi(B)w_t \quad (\text{V.10})$$

además al tratarse de un proceso *ARMA*, se sabe que

$$\phi(B)x_t = \theta(B)w_t \quad (\text{V.11})$$

por (V.10) y (V.11) se obtiene

$$\phi(B)\psi(B)w_t = \theta(B)w_t \quad (\text{V.12})$$

sea ahora $a(z) = \phi(z)\psi(z) = \sum_{j=0}^{\infty} a_j z^j$, $|z| \leq 1$ y por (V.12) se obtiene:

$$\sum_{j=0}^{\infty} a_j w_{t-j} = \sum_{j=0}^q \theta_j w_{t-j} \quad (\text{V.13})$$

luego al multiplicar ambos lados de (V.13) por w_{t-h} con $h = 0, 1, 2, 3, \dots$, y tomando el valor esperado se obtiene que

$$a_h = \theta_h, \quad h = 0, 1, \dots, q$$

y

$$a_h = 0, \quad h > q. \quad (\text{V.14})$$

con lo que

$$\phi(z)\psi(z) = a(z) = \theta(z), \quad |z| \leq 1. \quad (\text{V.15})$$

Por otro lado qué pasa si hay un número complejo en el círculo unitario, dígame z_0 , para el cual $\phi(z_0) = 0$, entonces por (V.15), $\theta(z_0) = 0$. Pero, si ese número complejo es tal como z_0 , entonces $\phi(z)$ y $\theta(z)$ tienen un factor en común, lo cual no es permitido; por lo que se puede escribir

$$\psi(z) = \frac{\theta(z)}{\phi(z)}$$

por hipótesis $|\psi(z)| < \infty$ para $|z| \leq 1$, de ahí que

$$|\psi(z)| = \left| \frac{\theta(z)}{\phi(z)} \right| < \infty, \quad |z| \leq 1. \quad (\text{V.16})$$

por lo que $\phi(z) \neq 0$ para $|z| \leq 1$, esto es consecuencia de (V.16); por lo tanto las raíces de $\phi(z)$ están fuera del círculo unitario.

(\Leftarrow)

Suponga que las raíces de $\phi(z)$, dígame, z_1, z_2, \dots, z_p , se encuentran fuera del círculo unitario.

Considere que las raíces están ordenadas de la siguiente manera

$$1 \leq |z_1| \leq |z_2| \leq \cdots \leq |z_p|$$

aca z_1, z_2, \dots, z_p no necesariamente son únicos; sea $|z_1| = 1 + \epsilon$, para todo $\epsilon > 0$. Entonces $\phi(z) \neq 0$ por lo que $|z| < 1 + \epsilon$ y en consecuencia, $\phi^{-1}(z)$ existe y tiene una expansión en serie de potencias:

$$\frac{1}{\phi(z)} = \sum_{j=0}^{\infty} a_j z^j, \quad |z| < 1 + \epsilon$$

Ahora, consider δ tal que $0 < \delta < \epsilon$, y sea $z = 1 + \delta$, que se encuentra en el radio de convergencia. Luego se tiene que

$$\phi^{-1}(1 + \delta) = \sum_{j=0}^{\infty} a_j (1 + \delta)^j < \infty \quad (\text{V.17})$$

Entonces, se pueden acotar los términos de la suma en (V.17) por una constante M , por lo que $|a_j(1 + \delta)^j| < M$, para $M > 0$. De ahí que $|a_j| < M(1 + \delta)^{-1}$, y por lo que:

$$\sum_{j=0}^{\infty} |a_j| < \infty \quad (\text{V.18})$$

Entonces $\phi^{-1}(B)$ existe y al multiplicarlo en ambos lados de (V.6) se obtiene:

$$x_t = \phi^{-1}(B)\phi(B)x_t = \phi^{-1}(B)\theta(B)w_t$$

haciendo $\psi(B) = \phi^{-1}(B)\theta(B)$, se tiene

$$x_t \psi(B)w_t = \sum_{j=0}^{\infty} \psi_j w_{t-j},$$

donde ψ -pesos son absolutamente sumables, y se pueden evaluar por

$$\psi(z) = \phi^{-1}(z)\theta(z),$$

para $|z| \leq 1$ ■

Una forma de parafrasear la propiedad anterior es decir que un proceso *ARMA* es causal solamente cuando las raíces de $\phi(z)$ se encuentran fuera del círculo unitario, esto es, $\phi(z) = 0$ solamente cuando $|z| > 1$.

Y para atacar el último problema sobre los modelos *ARMA* sobre la unicidad, se introduce el concepto de invertibilidad.

Definición A.11 *Un modelo ARMA(p,q), $\phi(B)x_t = \theta(B)w_t$, se dice que es invertible si la serie*

de tiempo $\{x_t : t = 0, \pm 1, \pm 2, \pm 3, \dots\}$ puede ser escrita como:

$$\pi(B)x_t = \sum_{j=0}^{\infty} \pi_j x_{t-j} = w_t \quad (\text{V.19})$$

donde $\pi(B) = \sum_{j=0}^{\infty} \pi_j B^j$, y $\sum_{j=0}^{\infty} |\pi_j| < \infty$; tal que $\pi_0 = 1$.

Propiedad A.12 *Invertibilidad de un proceso ARMA(p,g). Un modelo ARMA(p,q) es invertible si y solo si, $\theta(z) \neq 0$ para $|z| \leq 1$. Los coeficientes π_j de $\pi(B)$ dadas por la expresión (V.19) pueden determinarse resolviendo*

$$\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}, \quad |z| \leq 1.$$

La demostración de esta propiedad es similar a la propiedad anterior, por lo que no se realizará la prueba.

Una forma de parafrasear la propiedad de invertibilidad es decir que un proceso ARMA es invertible solamente cuando las raíces de $\theta(z)$ están fuera del círculo unitario, esto es, $\theta(z) = 0$ solamente cuando $|z| > 1$.

Las dos propiedades anteriores expuestas, le dan una estructura más robusta a los modelos ARMA.

B. Ecuaciones en diferencia

Esta sección, sólo servirá para ilustrar lo útil que resultan las ecuaciones en diferencias en los métodos de búsqueda de raíces, en particular de los ψ -pesos mencionados en los modelos ARMA.

Las ecuaciones en diferencias son fundamentales en el estudio del dominio de modelos de tiempo y de los procesos estocásticos generales. Por lo que se exhibirán ejemplos nada más.

Ejemplo B.1 *Soluciones a las ecuaciones en diferencias homogéneas.*

Ahora suponga que se tiene una secuencia de números tales que

$$u_n - \alpha u_{n-1} = 0, \quad \alpha \neq 0, \quad n = 1, 2, \dots \quad (\text{V.20})$$

y para dicha secuencia $\rho(h)$, función de autocorrelación de una serie de tiempo estacionario, es tal que

$$\rho(h) - \phi \rho(h-1) = 0, \quad h = 1, 2, \dots$$

de hecho la ecuación (V.20) representa una ecuación en diferencia homogénea de orden 1. Para

resolver dicha ecuación, se escribe

$$\begin{aligned} u_1 &= \alpha u_0 \\ u_2 &= \alpha u_1 = \alpha^2 u_0 \\ &\vdots \\ u_n &= \alpha u_{n-1} = \alpha^n u_0 \end{aligned}$$

dada la condición inicial $u_0 = c$, se desea resolver (V.20), se plantea que $u_n = \alpha^n c$. En términos del operador retroceso se puede escribir (V.20) como $(1 - \alpha B)u_n = 0$ y el polinomio asociado a (V.20) es $\alpha(z) = 1 - \alpha z$, y la raíz, dígase z_0 , de este polinomio es $z_0 = \frac{1}{\alpha}$, esto es $\alpha(z_0) = 0$. Sabemos la solución a (V.20) con la condición inicial, $u_0 = c$, es

$$u_n = \alpha^n c = (z_0^{-1})^n c$$

nótese que la solución a (V.20) depende solamente de la condición inicial y de la inversa de la raíz asociado al polinomio $\alpha(z)$.

Ahora considere una secuencia que satisface

$$u_n - \alpha_1 u_{n-1} - \alpha_2 u_{n-2} = 0, \quad \alpha_2 \neq 0, \quad n = 2, 3, \dots \quad (\text{V.21})$$

esta ecuación en diferencia homogénea es de orden 2, su polinomio correspondiente es

$$\alpha(z) = 1 - \alpha(1)z - \alpha(2)z^2$$

que posee dos raíces, sean éstas z_1 y z_2 ; tal que $\alpha(z_1) = \alpha(z_2) = 0$. Se consideran dos casos, primero suponga que $z_1 \neq z_2$. Luego la solución general a (V.21) es

$$u_n = c_1 z_1^{-n} + c_2 z_2^{-n}$$

esto se puede verificar al tomar la última expresión en (V.21).

Sean las dos condiciones iniciales u_0 y u_1 , se resolverá para c_1 y c_2 :

$$u_0 = c_1 + c_2$$

$$u_1 = c_1 z^{-1} + c_2 z^{-1}$$

en particular z_1 y z_2 pueden resolverse en términos de α_1 y α_2 usando la fórmula cuadrática.

El segundo caso es cuando $z_1 = z_2$, la solución general a (v.21) es

$$u_n = z_0^{-n}(c_1 + c_2n)$$

que se puede verificar al sustituir en (V.21).

Ahora considere las dos condiciones iniciales para u_0 y u_1 , con lo cual se puede resolver para c_1 y c_2 :

$$u_0 = c_1$$

$$u_1 = (c_1 + c_2)z_0^{-1}$$

A manera de resumen, estos resultados, en el caso de raíces distintas, la solución a la ecuación en diferencias homogénea de grado dos fue

$$u_n = z_1^{-n} \times (\text{un polinomio en } n \text{ de grado } m_1 - 1) \\ + z_2^{-n} \times (\text{un polinomio en } n \text{ de grado } m_2 - 1)$$

donde m_1 es la multiplicidad de las raíces de z_1 y m_2 es la multiplicidad de las raíces de z_2 , en el ejemplo que se desarrolló $m_1 = m_2 = 1$, los polinomios fueron de grado 0 tanto c_1 como c_2 .

Mientras que en el caso de raíces repetidas la solución fue

$$u_n = z_0^{-n} \times (\text{un polinomio en } n \text{ de grado } m_0 - 1)$$

donde m_0 es la multiplicidad de la raíz z_0 , en el caso anterior $m_0 = 2$. En este caso se escribió el polinomio de grado 1 como $c_1 + c_2n$. En cualquiera de los casos se resolvió para c_1 y c_2 dada las dos condiciones iniciales u_0 y u_1 .

Ejemplo B.2 *Determinar los ψ -pesos para un modelo ARMA(p, q).*

Para un modelo ARMA(p, q) causal $\phi(B)x_t = \theta(B)w_t$, donde los ceros de $\phi(z)$ están fuera del círculo unitario, nótese que se puede escribir

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j}$$

donde los ψ -pesos pueden determinarse usando la propiedad (IV.A10), sin embargo por otro métodos encontrar dichos pesos resultarán complicados, sin embargo la teoría de ecuaciones en diferencias homogéneas son de gran utilidad para este objetivo.

En general, para resolver este problema sobre ψ -pesos se considera que

$$\psi(z)\phi(z) = \theta(z)$$

$$(1 + \psi_0 + \psi_1 z + \psi_2 z^2 + \cdots)(1 + \phi_0 + \phi_1 z + \phi_2 z^2 + \cdots) = (1 + \theta_0 + \theta_1 z + \theta_2 z^2 + \cdots)$$

de ahí que los primeros valores están dadas de la siguiente manera,

$$\psi_0 = 1$$

$$\psi_1 + \phi_1 \psi_0 = \theta_1$$

$$\psi_2 - \phi_1 \psi_1 - \phi_2 \psi_0 = \theta_2$$

$$\psi_3 - \phi_1 \psi_2 - \phi_2 \psi_1 - \phi_3 \psi_0 = \theta_3$$

\vdots

donde $\phi_j = 0$ para $j > p$, y $\theta_j = 0$ para $j > q$. Los ψ -pesos satisfacen las ecuaciones en diferencias homogéneas dadas por

$$\psi_j - \sum_{k=1}^p \phi_k \psi_{j-k} = 0, \quad j \geq \max(p, q+1), \quad (\text{V.22})$$

dadas las condiciones iniciales

$$\psi_j - \sum_{k=1}^j \phi_k \psi_{j-k} = \theta_j, \quad 0 \leq j \leq \max(p, q+1), \quad (\text{V.23})$$

La solución general depende de las raíces del polinomio AR

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p.$$

Y por su puesto la solución específica depende de las condiciones iniciales. En el caso de las soluciones para ecuaciones en diferencias homogéneas generales de orden:

$$u_n - \alpha_1 u_{n-1} - \cdots - \alpha_p u_{n-p} = 0, \quad \alpha_p \neq 0, n = p, p+1, \cdots. \quad (\text{V.24})$$

el polinomio asociado es

$$\alpha(z) = 1 - \alpha_1 z - \cdots - \alpha_p z^p$$

suponga que $\alpha(z)$ tiene r raíces distintas, sean z_1 con multiplicidad m_1 , z_2 con multiplicidad m_2 , \cdots , z_r con multiplicidad m_r tales que $m_1 + m_2 + \cdots + m_r = p$. La solución general a la ecuación

en diferencia (V.24) es

$$u_n = z_1^{-1}P_1(n) + z_2^{-1}P_2(n) + \cdots + z_r^{-1}P_r(n)$$

donde $P_j(n)$, para $j = 1, 2, \dots, r$, es un polinomio en n , de grado $m_j - 1$. Dada las p condiciones iniciales u_0, u_1, \dots, u_{p-1} , se puede resolver para $P_j(n)$ explícitamente.

C. Función de autocorrelación y autocorrelación parcial

Ya se ha definido la función de autocorrelación en preliminares, en esta sección se ilustrarán con dos ejemplos la obtención de la función ACF para dos procesos en particular, para luego definir la función PACF.

Ejemplo C.1 *La función de autocorrelación de un modelo MA.*

Considere un proceso $MA(q)$, $x_t = \theta(B)w_t$, donde $\theta(B) = 1 + \theta B + \cdots + \theta_q B^q$. Por que x_t es una combinación lineal finita de términos de ruido blanco, el proceso es estacionario con media

$$E[x_t] = \sum_{j=0}^q \theta_j E[w_{t-j}] = 0$$

aca se tiene que $\theta = 1$, con función de autocovarianza

$$\begin{aligned} \gamma(h) &= cov(x_{t+h}, x_t) = E\left[\left(\sum_{j=0}^q \theta_j w_{t+h-j}\right)\left(\sum_{k=0}^q \theta_k w_{t-k}\right)\right] \\ &= \begin{cases} \sigma_w^2 \sum_{j=0}^{q-h} \theta_j \theta_{j+h}, & 0 \leq h \leq q \\ 0, & h > q \end{cases} \end{aligned}$$

Nótese que $\gamma(h) = \gamma(-h)$, por lo que solo se considerará la parte con $h \geq 0$. Al dividir $\gamma(h)$ entre $\gamma(0)$, se obtiene el ACF de un $MA(q)$:

$$\rho(h) = \begin{cases} \frac{\sigma_w^2 \sum_{j=0}^{q-h} \theta_j \theta_{j+h}}{1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2}, & 1 \leq h \leq q \\ 0, & h > q \end{cases}$$

Ejemplo C.2 *La función de autocorrelación de un modelo ARMA*

Mientras que para un modelo $ARMA(p, q)$, $\phi(B)x_t = \theta(B)w_t$, donde los ceros de $\phi(z)$ están fuera del círculo unitario, se escribe

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j}$$

Se sigue inmediatamente que $E[x_t] = 0$. También, la función de autocorrelación de x_t puede ser

escrito como:

$$\gamma(h) = \text{cov}[x_{t-h}, x_t] = \sigma_w^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}, \quad h \geq 0. \quad (\text{V.25})$$

Se pueden usar las ecuaciones (IV.22) y (IV.23) para resolver los ψ -pesos. Luego se puede resolver para $\gamma(h)$, se obtiene que la función ACF es

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}.$$

Una alternativa para encontrar $\gamma(h)$ es obtener una ecuación en diferencias homogéneas directamente en términos de $\gamma(h)$. Para ello considere

$$\begin{aligned} \gamma(h) &= \text{cov}(x_{t+h}, x_t) = E\left[\left(\sum_{j=1}^p \phi_j x_{t+h-j} + \sum_{j=0}^q \theta_j w_{t+h-j}\right)x_t\right] \\ &= \sum_{j=1}^p \phi_j \gamma(h-j) + \sigma_w^2 \sum_{j=h}^q \theta_j \psi_{j-h}, \quad h \geq 0. \end{aligned} \quad (\text{V.26})$$

Se usará el hecho de que

$$x_t = \sum_{k=0}^{\infty} \psi_k w_{t-k}$$

y para $h \geq 0$ se tiene que

$$E[w_{t+h-j}x_t] = E[w_{t+h-j}\left(\sum_{k=0}^{\infty} \psi_k w_{t-k}\right)] = \psi_{j-h} \sigma_w^2$$

de (IV.26) se puede escribir una ecuación general para el ACF de un modelo causal *ARMA*,

$$\gamma(h) - \phi_1 \gamma(h-1) - \dots - \phi_p \gamma(h-p) = 0, \quad h \geq \max(p, q+1) \quad (\text{V.27})$$

con las condiciones iniciales

$$\gamma(h) - \sum_{j=1}^p \phi_j \gamma(h-j) = \sigma_w^2 \sum_{j=h}^q \theta_j \psi_{j-h}, \quad 0 \leq h < \max(p, q+1) \quad (\text{V.28})$$

que luego al dividir (IV.27) e (IV.28) entre $\gamma(0)$ se tendrá que resolver para el ACF, $\rho(h) = \frac{\gamma(h)}{\gamma(0)}$.

Por notación se dirá que x_h^{h-1} es la regresión de x_h en $\{x_{h-1}, x_{h-2}, x_{h-3}, \dots, x_1\}$, el cual se escribe como:

$$x_h^{h-1} = \beta_1 x_{h-1} + \beta_2 x_{h-2} + \dots + \beta_{h-1} x_1$$

aca no se necesita agregar el término de intercepto pues la media de x_t es cero. Por otro lado sea

x_0^{h-1} la regresión de x_0 en $\{x_1, x_2, x_3, \dots, x_{h-1}\}$, entonces

$$x_0^{h-1} = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{h-1} x_{h-1}.$$

Se definirá la función PACF.

Definición C.3 *La función de autocorrelación parcial (PACF) de un proceso estacionario, x_t , denotado como ϕ_{hh} , para $h = 1, 2, \dots$, es*

$$\phi_{11} = \text{corr}(x_1, x_0) = \rho(1)$$

y

$$\phi_{hh} = \text{corr}(x_h - x_h^{h-1}, x_0 - x_0^{h-1}), \quad h \geq 2.$$

Tanto $(x_h - x_h^{h-1})$ como $(x_0 - x_0^{h-1})$ no están correlacionados con $\{x_1, x_2, x_3, \dots, x_{h-1}\}$. Por la estacionariedad, la función PACF es la correlación de x_t y x_{t-h} con la dependencia lineal de $\{x_{t-1}, \dots, x_{t-(h-1)}\}$, en cada, eliminada.

Ejemplo C.4 *La función PACF de un modelo AR causal.*

Sea $x_t = \sum_{j=1}^p \phi_j x_{t-j} + w_t$ donde las raíces de $\phi(z)$ están fuera del círculo unitario. En particular

$$x_h = \sum_j^p \phi_j x_{h-j} + w_h.$$

Cuando $h > p$, la regresión de x_h en $x_{h-1}, x_{h-2}, \dots, x_1$, es

$$x_h^{h-1} = \sum_j^p \phi_j x_{h-j}.$$

Entonces cuando $h > p$,

$$\phi_{hh} = \text{corr}(x_h - x_h^{h-1}, x_0 - x_0^{h-1}) = \text{corr}(w_h, x_0 - x_0^{h-1}) = 0$$

por la causalidad, $x_0 - x_0^{h-1}$ depende únicamente de $\{w_{h-1}, w_{h-2}, \dots\}$, nótese que $x_0^{h-1} = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{h-1} x_{h-1}$. Por otro lado se sabe que cuando $h \leq p$, se tiene que ϕ_{pp} no es cero y $\phi_{11}, \dots, \phi_{p-1, p-1}$ no necesariamente son cero.

D. Predicción

En esta sección se discutirán las ideas que están inmersas en el desarrollo de la teoría de predicción de las series de tiempo. Debido a la naturaleza de esta teoría, los teoremas fundamentales

desarrollados con la regresión lineal en el capítulo anterior poseen los mismos objetivos que esta sección y son aplicables, por ende sólo se hace una breve exposición.

Para realizar predicciones dada una serie de tiempo, x_{n+m} , $m = 1, 2, \dots$, basado en un conjunto de datos (observaciones) hasta el presente, $x_n, x_{n-1}, x_{n-2}, \dots, x_1$. El predictor de error cuadrático medio mínimo para x_{n+m} es

$$x_{n+m} = E(x_{n+m} | x_n, x_{n-1}, x_{n-2}, \dots, x_1)$$

esto es porque el valor esperado condicional minimiza el error cuadrático medio:

$$E[x_{n+m} - g(\mathbf{x})]^2 \tag{V.29}$$

donde $g(\mathbf{x})$ es una función de valores observados.

Una condición importante aca es, restringir, que los predictores sean funciones lineales de la data empleada, tales que sean de la forma:

$$x_{n+m}^n = \alpha_0 + \sum_{k=1}^n \alpha_k x_k \tag{V.30}$$

donde $\alpha_1, \dots, \alpha_n$ son números reales. Los predictores de la forma (V.30) que minimizan el error cuadrático medio, son llamados los mejores predictores lineales.

Dentro de esta teoría, se establece que a partir del conjunto de datos observados, x_1, x_2, \dots, x_n , el mejor predictor lineal, $x_{n+m}^n = \alpha_0 + \sum_{k=1}^n \alpha_k x_k$, de x_{n+m} , para $m \geq 1$, está dada por solución de

$$E[(x_{n+m} - x_{n+m}^n)x_t] = 0, \quad k = 0, 1, \dots, n. \tag{V.31}$$

donde $x_0 = 1$. La ecuaciones especificadas en (V.31) son llamadas las ecuaciones de predicción y son usadas para encontrar los coeficientes $\{\alpha_0, \dots, \alpha_n\}$.

Si $E(x_t = \mu)$, la primera ecuación ($k = 0$) de (V.30) implica que

$$E(x_{n+m}^n) = E(x_{n+m}) = \mu$$

entonces, tomando el valor esperado en (V. 30), se tiene que

$$\mu = \alpha_0 + \sum_{k=1}^n \alpha_k \mu$$

o

$$\alpha_0 = \mu \left(1 - \sum_{k=1}^n \alpha_k\right).$$

Por lo que, la forma del mejor predictor lineal es

$$x_{n+m}^n = \mu + \sum_{k=1}^n \alpha_k (x_k - \mu) \quad (\text{V.32})$$

Considere ahora la primera predicción, es decir de un paso adelante. Esto es, dada $\{x_1, \dots, x_n\}$, se desea pronosticar el valor de la serie de tiempo al próximo punto, x_{n+1} , el mejor predictor lineal para dicho valor es

$$x_{n+1}^n = \phi_{n1}x_n + \phi_{n2}x_{n-1} + \dots + \phi_{nn}x_1 \quad (\text{V.33})$$

dedido al propósito se escribirá α_k en (V.30) como ϕ_n, ϕ_{n+1-k} en (V.33), para $k = 1, 2, \dots, n$. Usando las condiciones de (V.31), los coeficientes $\phi_{n1}, \phi_{n2}, \dots, \phi_{nn}$ satisfacen

$$E\left[x_{n+1} - \sum_{j=1}^n \phi_{nj}x_{n+1-j}\right]x_{n+1-k} = 0, \quad k = 1, 2, \dots, n$$

o

$$\sum_{j=1}^n \phi_{nj}\gamma(k-j) = \gamma(k) \quad , k = 1, 2, \dots, n. \quad (\text{V.34})$$

Que puede expresarse en notación matricial como

$$\Gamma_n \phi_n = \gamma_n \quad (\text{V.35})$$

donde $\Gamma_n = \{\gamma(k-j)\}_{j,k=1}^n$ es una matriz $n \times n$, $\phi_n = (\phi_{n1}, \phi_{n2}, \dots, \phi_{nn})^T$ es un vector de $n \times 1$ y $\gamma_n = (\gamma(1), \gamma(2), \dots, \gamma(n))^T$ es un vector de $n \times 1$.

La matriz Γ_n es no negativa definida. Si Γ_n es singular entonces existen muchas soluciones a (V.35) pero, por el teorema de unicidad (II.A.1.2), x_{n+m}^n es única. Si Γ_n es no singular, los elementos de ϕ_n son únicos por el teorema de unicidad y están dados por

$$\phi_n = \Gamma_n^{-1} \gamma_n \quad (\text{V.36})$$

En el caso de un modelo ARMA, el hecho de que $\sigma_w^2 > 0$ y $\gamma(h) \rightarrow 0$ cuando $h \rightarrow \infty$ es suficiente asegurar que Γ_n es positiva definida. Es conveniente escribir a veces el pronóstico de la serie de tiempo del próximo valor en la notación vectorial

$$x_{n+1}^n = \phi_n^T \mathbf{x} \quad (\text{V.37})$$

donde $\mathbf{x} = (x_n, x_{n-1}, \dots, x_1)^T$

Propiedad D.1 *El predictor del valor próximo de la serie de tiempo de error cuadrático medio mínimo es:*

$$P_{n+1}^n = E(x_{n+1} - x_{n+1}^n)^2 = \gamma(0) - \boldsymbol{\gamma}_n^T \Gamma_n^{-1} \boldsymbol{\gamma}_n \quad (\text{V.38})$$

Demostración.

Se sabe que

$$\begin{aligned} E(x_{n+1} - x_{n+1}^n)^2 &= E(x_{n+1} - \boldsymbol{\phi}_n^T \mathbf{x})^2 = E(x_{n+1} - \boldsymbol{\gamma}^T \Gamma_n^{-1} \mathbf{x})^2 \\ &= E(x_n^2 - 2\boldsymbol{\gamma}_n^T \Gamma_n^{-1} \mathbf{x} x_{n+1} + \boldsymbol{\gamma}_n^T \Gamma_n^{-1} \mathbf{x} \mathbf{x}^T \Gamma_n^{-1} \boldsymbol{\gamma}_n) \\ &= \gamma(0) - 2\boldsymbol{\gamma}_n^T \Gamma_n^{-1} \boldsymbol{\gamma}_n + \boldsymbol{\gamma}_n^T \Gamma_n^{-1} \Gamma_n \Gamma_n^{-1} \boldsymbol{\gamma}_n \\ &= \gamma(0) - \boldsymbol{\gamma}_n^T \Gamma_n^{-1} \boldsymbol{\gamma}_n \quad \blacksquare \end{aligned}$$

Para los modelos *ARMA* en general, la ecuación de predicción resulta ser no simple pues requiere de inversión de matrices grandes en dimensión, sin embargo en métodos tradicionales se simplifica este problema utilizando soluciones iterativas que no requieren de inversión de matrices. En particular éstas soluciones recursivas se deben a Levinson (1947) y Durbin (1960), quienes establecieron el algoritmo Durbin-Levinson que resuelve de manera iterativa las ecuaciones (V.36) y (V.38) como sigue:

$$\phi_{00} = 0, \quad P_1^0 = \gamma(0)$$

Para $n \geq 1$,

$$\phi_{nn} = \frac{\rho_n - \sum_{k=1}^{n-1} \phi_{n-1,k} \rho(n-k)}{1 - \sum_{k=1}^{n-1} \phi_{n-1,k} \rho(k)}, \quad P_{n+1}^n = P_n^{n-1} (1 - \phi_{nn}^2) \quad (\text{V.39})$$

donde para $n \geq 2$,

$$\phi_{nk} = \phi_{n-1,k} - \phi_{nn} \phi_{n-1,n-k}, \quad k = 1, 2, \dots, n-1 \quad (\text{V.40})$$

como consecuencia del algoritmo anterior, con la ecuación (V.39) de manera iterativa se puede calcular la función PACF de un modelo estacionario x_t , como

$$\phi_{nn}, \quad \text{para } n = 1, 2, \dots$$

E. Modelos ARIMA

Esta sección expone elementos que hacen de los modelos *ARIMA* sean de estructura robusta, es decir absorben la mayoría de fenómenos de modelado como los discutidos en los modelos *ARMA*.

En muchas situaciones una serie de tiempo puede pensarse como una composición de dos componentes; uno de ellos es un proceso no estacionario de tendencia y una componente estacionaria con media cero.

De ilustración veamos algunos ejemplos.

Ejemplo E.1

1. Sea x_t una caminata aleatoria, $x_t = x_{t-1} + w_t$, que por ser una diferencia se tiene que $\nabla x_t = w_t$ por lo que es un proceso estacionario.
2. Considere ahora el modelo $x_t = \mu_t + y_t$, donde $u_t = \beta_0 + \beta_1 t$ y y_t es estacionario. Aplicando una diferencia a dicho proceso se tiene que

$$\begin{aligned}\nabla x_t &= x_t - x_{t-1} = \mu_t + y_t - \mu_{t-1} - y_{t-1} = \beta_0 + \beta_1 t + y_t - \beta_0 - \beta_1 t + \beta_1 - y_{t-1} \\ &= \beta_1 + \nabla y_t\end{aligned}$$

es estacionario.

Ejemplo E.2

Sea x_t como en el caso anterior, con $u_t = u_{t-1} + v_t$ un proceso estocástico, con variaciones pequeñas de acuerdo a una caminata aleatoria, v_t es estacionario.

En este caso

$$\nabla x_t = x_t - x_{t-1} = \mu_t + y_t - \mu_{t-1} - y_{t-1} = v_t + y_t - y_{t-1} = v_t - \nabla y_t$$

el cual es estacionario.

Un modelo *ARMA* integrado o un modelo *ARIMA*, es una expansión de la clase de modelos *ARMA* que incluyen diferencias.

Definición E.3 Un proceso, x_t se le dice **ARIMA**(p, d, q) si

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

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

$$\phi(B)(1 - B)^d x_t = \theta(B)w_t. \tag{V.41}$$

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 $\alpha = \mu(1 - \phi_1 - \dots - \phi_p)$.

El interés ahora será discutir la construcción de los modelos ARIMA.

Existen algunos pasos para el ajuste de modelos ARIMA de las series de tiempo. Estos pasos envuelven graficar los datos respecto del tiempo de los fenómeno de estudio e identificar anomalías, una posible transformación a la data, indentificación de posibles dependencias de los órdenes del modelo y los parámetros de estimación.

Por ejemplo, si la variabilidad en la data crece con el tiempo, será necesario aplicar una transformación adecuada para para estabilizar la varianza, una de esas transformaciones puede ser la de Box-Cox. Para poner un contexto suponga por ejemplo, un proceso $x_t = (1 + p_t)x_{t-1}$ donde x_t son los valores de cierta inversión en el tiempo t y p_t es el porcentaje de cambio de un período $t - 1$ a t , el cual puede tomar valores negativos. Aplíquese el \ln en ambos lados del proceso, se obtiene:

$$\ln(x_t) = \ln(1 + p_t) + \ln(x_{t-1}),$$

que es equivalente a decir

$$\ln(x_t) - \ln(x_{t-1}) = \ln(1 + p_t)$$

Si el porcentaje p_t se mantiene relativamente bajo en magnitud, entonces $\ln(1 + p_t) \approx p_t$ y entonces

$$\nabla[\ln(x_t)] \approx p_t$$

será un proceso relativamente estable, eventualmente se le dice a $\nabla[\ln(x_t)]$ como la razón de crecimiento.

Antes de aplicar alguna transformación a la data de estudio, se requiere identificar antes los valores de órdenes de autoregresividad p , de diferencia d , y de media móvil q . Para el efecto se empieza analizando la gráfica de la data, que mostrará incidencias para sugerir una posible diferencia de orden d , además un criterio alternativo es ver la variabilidad de gráfica de ACF para la variable de estudio.

Si $p = 0$ y $q > 0$, el ACF se corta después de q diferencias, y el PACF se desvanece. Si $q = 0$ y $p > 0$, el PACF se corta después de p , y el ACF se desvanece. Si $p > 0$ y $q > 0$, ambos ACF y PACF desaparecerán porque se está tratando con estimados.

Por último se realiza alguna prueba estadística a los modelos de estudio, como el de Q-estadístico, y en base a un criterio de selección de modelo definir cuál es la que mejor se ajusta al fenómeno de investigación, este tipo de criterio puede ser el de AIC, AICc o SIC.

F. Modelos ARIMA multiplicativos estacionales

Para culminar este capítulo se introducirán algunas modificaciones a los modelos *ARIMA*, que le darán una expansión importante para el objeto de estudio de este texto. Estas modificaciones integrarán en el modelo una cuantificación sobre la estacionalidad y la no estacionariedad de los modelos *ARIMA*.

Muchos fenómenos físicos, biológicos, químicos y económicos poseen tendencias estacionarias y fluctuantes. Por lo que es conveniente para el propósito de este texto introducir la noción de modelos polinomiales autoregresivos y de media móvil que se identifican con las diferencias estacionales, que se dirá por $ARMA(P, Q)_s$ y será de la forma:

$$\Phi_P(B^s)x_t = \Theta_Q(B^s)w_t \quad (\text{V.42})$$

donde s es el período estacional.

Consideremos la siguiente definición.

Definición F.1 *Los operadores*

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps} \quad (\text{V.43})$$

$$\Theta_Q(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \dots + \Theta_Q B^{Qs} \quad (\text{V.44})$$

son los operadores autoregresivo estacional y de media móvil estacional de orden P y Q respectivamente, de período estacional s .

En esta definición son válidas las condiciones de causalidad e invertibilidad de las propiedades (IV.10) e (IV.12) respectivamente.

En general, se pueden combinar los operadores de estacionalidad con la no estacionalidad en un modelo multiplicativo estacional autoregresivo de media móvil, denotado por:

$$ARMA(p, q) \times (P, Q)_s \quad (\text{V.45})$$

y se escribe como

$$\Phi_P(B^s)\phi(B)x_t = \Theta_Q(B^s)\Theta(B)w_t \quad (\text{V.46})$$

como el modelo en general.

El comportamiento de los ACF Y PACF tienden a mostrar patrones de la forma indicada del modelo de estudio.

La tendencia no estacionaria puede ocurrir, por ejemplo considere un proceso cercano a una periodicidad en una estación, sea éste proceso el de la temperatura media mensual a lo largo de los años, cada mes de Enero puede aproximarse dada su estacionalidad, de igual manera cada Febrero y así sucesivamente. Piense que la temperatura media mensual es x_t siguiendo el siguiente modelo:

$$x_t = S_t + w_t$$

donde S_t es la parte estacional que varía lentamente de un año al otro, de acuerdo a una caminata aleatoria: $S_t = S_{t-12} + v_t$ en éste modelo, v_t y w_t son procesos de ruidos blancos que no están correlacionados.

Suponga que la tendencia de la data al graficar muestra que el ACF de la temperatura decae lentamente en diferencias de $h = 12k$ con $k = 1, 2, \dots$. Si se substraen los años de la temperatura registrada, de manera sucesiva de los años correspondientes, se encontrará que

$$(1 - B^{12})x_t = x_t - x_{t-12} = v_t + w_t - w_{t-12} \quad (\text{V.47})$$

el cual resulta ser un modelo $MA(1)_{12}$.

En general, el estudio de la diferencia estacional es un indicador del decaimiento lento de la función ACF como múltiplo de una estación de período s , pero es insignificante entre los períodos.

Por lo que, a una diferencia estacional de orden D se define como

$$\nabla_s^D x_t = (1 - B^s)^D x_t \quad (\text{V.48})$$

donde $D = 1, 2, \dots$ son valores enteros. Particularmente es suficiente para $D = 1$ conseguir la estacionariedad estacional.

Incorporaremos estas ideas para culminar esta pequeña ilustración sobre los modelos ARIMA.

Definición F.2 *El modelo multiplicativo estacional autoregresivo integrado de media móvil o SARIMA, de Box y Jenkins (1970) está dada por*

$$\Phi_P(B^s)\phi(B)\nabla_s^D \nabla^d x_t = \alpha + \Theta_Q(B^s)\theta(B)w_t \quad (\text{V.49})$$

donde w_t es el usual proceso de ruido blanco Gaussiano.

Mientras que al modelo general se le denomina $ARIMA(p, d, q) \times (P, D, Q)_s$.

Los componentes ordinarios autoregresivo y de media móvil son representados por polinomios $\phi(B)$ y $\theta(B)$ de orden p y q respectivamente, los componentes autoregresivo estacional y de media móvil estacional son $\Phi_P(B^s)$ y $\Theta_Q(B^s)$ de orden P y Q , y los componentes en diferencia ordinaria y estacional son ∇^d y ∇_s^D respectivamente.

VI Cobb Douglas

En este capítulo se desarrollará el modelo Cobb-Douglas, un tipo de función de producción, considerado como el modelo empírico de estudios económicos más usado.

Se centrará en los conceptos importantes de la función Cobb-Douglas, dada su naturaleza empírica, el cual posee limitada teoría.

Definición .1 *El espacio de producción.*

Se define por $SP = \{(x_1, x_2, x_3, \dots, x_n) | x_i \geq 0, \text{ para } i = 1, 2, \dots, n\}$ con n fijo.

Para un subconjunto convexo $D_P \subset SP$, se le llamará como dominio de producción.

Definición .2 *Se define por función de producción a una aplicación*

$$Q : D_P \rightarrow \mathcal{R}^+$$

$$(x_1, x_2, x_3, \dots, x_n) \rightarrow Q(x_1, x_2, x_3, \dots, x_n) \in \mathcal{R}^+ \quad \forall (x_1, x_2, x_3, \dots, x_n) \in D_P$$

que cumple con

1. $Q(0, \dots, 0) = 0$
2. Q es de clase C^2 en D_P , esto es admite derivadas parciales de orden 2 y éstas son continuas en D_P .
3. Q es monótona en el sentido de que $\frac{\partial Q}{\partial x_i} \geq 0$ para $i = 1, 2, \dots, n$.
4. Q es cuasi cóncava.

Definición .3 *Sea Q una función de producción, $Q : D_P \rightarrow \mathcal{R}^+$, y $Q_0 \in \mathcal{R}^+$ valor fijo, al conjunto de valores de entrada que generan la producción Q_0 se dirá que es una isocuanta.*

Se dirá que Q , $Q : D_P \rightarrow \mathcal{R}^+$, es una función de producción de retorno constante a escala (inelástica) si

$$Q(\lambda x_1, \dots, \lambda x_n) = \lambda Q(x_1, \dots, x_n),$$

con aumento en el retorno a escala (elástica) si

$$Q(\lambda x_1, \dots, \lambda x_n) > \lambda Q(x_1, \dots, x_n),$$

y con decremento en el retorno a escala (elástica) si

$$Q(\lambda x_1, \dots, \lambda x_n) < \lambda Q(x_1, \dots, x_n) \quad \forall (x_1, x_2, x_3, \dots, x_n) \in \mathcal{D}_P$$

Definición .4 *La función de producción Cobb-Douglas.*

Se define la función Cobb-Douglas como

$$f : \mathcal{D} \subset \mathbf{R}_+^n \rightarrow \mathcal{R}^+$$

$$f(x_1, x_2, x_3, \dots, x_n) = Ax_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n} \quad (\text{VI.1})$$

con $A, \alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n$ constantes positivas.

Las constantes α_i , con $i = 1, 2, \dots, n$ para un n fijo, suelen llamarse elasticidades del i -ésimo factor de entrada, que jugarán un papel importante en la parte práctica de esta teoría. Por lo que cualquier función de producción que adopte la forma de la función Cobb-Douglas se dirá que es una función de producción de Cobb-Douglas.

Se deducirán algunos hechos útiles. Al aplicar el \ln en ambos lados a (VI.1) se tiene:

$$\ln[f(x_1, \dots, x_n)] = \ln(Ax_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n}) \quad (\text{VI.2})$$

$$= \ln(A) + \sum_{i=1}^n \alpha_i \ln(x_i) \quad (\text{VI.3})$$

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

$$y = \alpha_0 + \sum_{i=1}^n z_i \alpha_i \quad (\text{VI.4})$$

que es básicamente una regresión lineal, precisamente un modelo log-lineal. Con (VI.4) se pueden estimar los valores α_i con el método de mínimos cuadrados, presentado en el primer capítulo. Dado que y, x_1, \dots, x_n poseen valores observados.

Si $\sum_{i=1}^n \alpha_i = 1$ entonces se estará hablando de una función de producción inelástica, si $\sum_{i=1}^n \alpha_i < 1$ es una función elástica y si $\sum_{i=1}^n \alpha_i > 1$ es una función elástica.

Considere la derivada parcial de (VI.3) sobre $\ln(x_i)$:

$$\frac{\partial \ln[f]}{\partial \ln x_i} = \frac{x_i}{f} \frac{\partial f}{\partial x_i} = \alpha_i \quad (\text{VI.5})$$

En la literatura $\frac{\partial f}{\partial x_i}$ suele llamarse la productividad marginal.

Una consideración a la teoría es imponer a (VI.1) la condición de elasticidad:

$$\sum_{i=1}^n \alpha_i = 1 \quad (\text{VI.6})$$

que en conjunto forman un problema de programación no lineal; las condiciones suficientes y necesarias para considerar y transformar (VI.1) en un problema de programación no lineal de optimización puede verse en A.3 de la sección de apéndice en Ioan y Ioan (2015), el cual puede resolverse con algoritmos computacionales ya establecidos, como la técnica de programación cuadrática.

Para culminar éste capítulo, considere la derivada de (VI.3) respecto del tiempo:

$$\frac{1}{f} \frac{df}{dt} = \frac{1}{A} \frac{dA}{dt} + \sum_{i=1}^n \alpha_i \frac{1}{x_i} \frac{dx_i}{dt} \quad (\text{VI.7})$$

denotemos a $\frac{1}{f} \frac{df}{dt} = \nabla f$ como la razón de crecimiento medio de f , así mismo para $\frac{1}{A} \frac{dA}{dt} = \nabla A$ y para cada factor $\frac{1}{x_i} \frac{dx_i}{dt} = \nabla x_i$. Por lo que (VI.7) puede expresarse como

$$\nabla f = \nabla A + \sum_{i=1}^n \alpha_i \nabla x_i \quad (\text{VI.8})$$

el cual puede ajustarse por regresión lineal múltiple.

Tanto en (VI.4) y (VI.8) se asume que A es constante, de manera empírica.

Una discusión interesante sobre considerar (VI.1) y (VI.6) de manera generalizada y como un teorema homogéneo puede verse en Cooper *et al.* (1976).

El modelo de Cobb-Douglas a ha sido criticado sobre su formulacion experimental (posee limitada fundamentación), principalmente sobre su constante de tecnología. Existen mejoras a la técnica de la función de Cobb-Douglas, que se presentan en Brown (1970), Arrow *et al.* (1961), Fletcher y Lu (1968), Christensen *et al.* (1973), Ryu y Zellner (1998).

VII Pronóstico de la transparencia del agua del Lago de Atitlán

En éste capítulo se aplicarán las teorías desarrolladas en los capítulos anteriores; centrándose principalmente en la simulación de la data recolectada por la AMSCLAE, se estudiará la data de entrenamiento (simulada), desde el punto de vista de regresión lineal múltiple; se establecerán algunas relaciones entre los parámetros de estudio a partir las correlaciones existentes, se obtendrán modelos tipo Cobb-Douglas; que explicarán las elasticidades de los parámetros importantes de modelado y por último se realizará el pronóstico de la transparencia del agua del Lago de Atitlán.

Dada la información solicitada, se observó que las mediciones de los parámetros de estudio no se realizaron de manera continua (mes a mes) y no se midieron en los mismos días de cada mes, en definitiva ésto afecta la distribución de las muestras de los parámetros de estudio. Pues para el estudio de un parámetro en particular se desea mantener un diseño experimental fijo, en este caso realizar las mediciones mensuales continuamente, fijando un día de muestreo de cada mes y la hora de medición. Conservando la periodicidad de los parámetros de estudio.

Con la data, registrada por la AMSCLAE, se tomaron las observaciones que contenían información completa de los parámetros temperatura (TE), nitratos (NI), fosfatos (FOS), fósforo total (FOST), clorofila a (CLA), sólidos disueltos totales (SDT), oxígeno disuelto (OD), saturación de oxígeno (SO), radiación (PAR), salinidad (SAL), potencial de hidrógeno (PH), conductividad (COND) y transparencia (TR) 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. Aca transparencia hace referencia a la medición del disco Secchi. Además las abreviaciones no son acrónimos de los parámetros en cuestión, es una notación que se adopta en éste trabajo.

Para la parte de aplicación de éste capítulo las simulaciones, cálculos y gráficas fueron realizadas con el software *R*.

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*:

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

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*

$$V_i = M_{V_i} + F_i SD_{V_i} \quad i = 1, 2, \dots, 13 \quad (\text{VII.2})$$

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 obtenidos 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 en éste capítulo es verificar si

$$TR = \beta_0 + \sum_{i=1}^{12} x_i \beta_i \quad (\text{VII.3})$$

es un buen predictor para la transparencia del agua 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 siguiente algoritmos.

Algoritmo B.1 *Análisis de modelo.*

1. *Plantear el modelo*

$$TR = \beta_0 + \sum_{i=1}^{12} x_i \beta_i \quad (\text{VII.4})$$

donde β_0 es el intercepto, $x_1 = TE$, $x_2 = COND$, $x_3 = SAL$, $x_4 = SDT$, $x_5 = NI$, $x_6 = FOS$, $x_7 = FOST$, $x_8 = CLA$, $x_9 = PAR$, $x_{10} = OD$, $x_{11} = SO$ 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 A.2, se pueden plantear modelos tipo Cobb-Douglas 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 Coob-Douglas planteado.
3. Encontrar los α_i (elasticidades) del modelo log-ineal con el método de mínimos cuadrados.

D. Aplicación de series de tiempo

Mientras que a partir de la data de entrenamiento para la transparencia; se desea saber cómo este parámetro físico del agua 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.

Para ese objetivo se presenta el algoritmo de construcción de modelos ARIMA.

Algoritmo D.1 Modelos ARIMA para transparencia.

1. Graficar los valores de entrenamiento de la transparencia del agua del Lago de Atitlán.
2. Identificar anomalías en la gráfica.
3. Graficar 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 anexos contienen los resultados más importantes de la investigación realizada. El código que se desarrolló e implementó en R de este trabajo de graduación, se basa en las librerías, paquetes y funciones creadas por otros autores.

Para la parte de simulación mensual se utilizaron tidyverse, ggplot2 y GGally desarrollados en Wickham y Grolemund (2016), Wicklam (2009) y en Schloerke *et al.* (2011) respectivamente. Para las gráficas se utilizó plotly desarrollado en Sievert *et al.* (2017). Para la regresión lineal múltiple se emplearon lmtest, car, DAAG, faraway, reshap2 y corrplot desarrollados en Hothorn *et al.* (2018), Fox *et al.* (2017), Maindonald y Braun (2015), Faraway (2016) y Wicklam (2017) respectivamente. Mientras que para las elasticidades del modelo de Cobb-Douglas se empleó quadprog de Turlach (2015). Y para los modelos ARIMA se usaron forecast, xts y hexbin de Hyndman *et al.* (2017), Shumway y Sttofer (2006), Ulrich *et al.* (2017), Koh y Maechler (2018).

VIII Resultados

Se presentan los resultados obtenidos al seguir los algoritmos planteados en el capítulo anterior. Dichos resultados provienen del estudio de la transparencia en distintos sitios; las relaciones que posee con otros parámetros, existen muchas otras relaciones importantes e interesantes (que se pueden deducir con el Algoritmo B.2) en la matriz de correlaciones, debido al enfoque se exponen las relaciones fuertes de los parámetros con la transparencia y luego su pronóstico para el año 2018.

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 - San Pablo - San Marcos (WA). Para apreciar la ubicación de dichos sitios véase la Figura no. 3 en la parte de anexos.

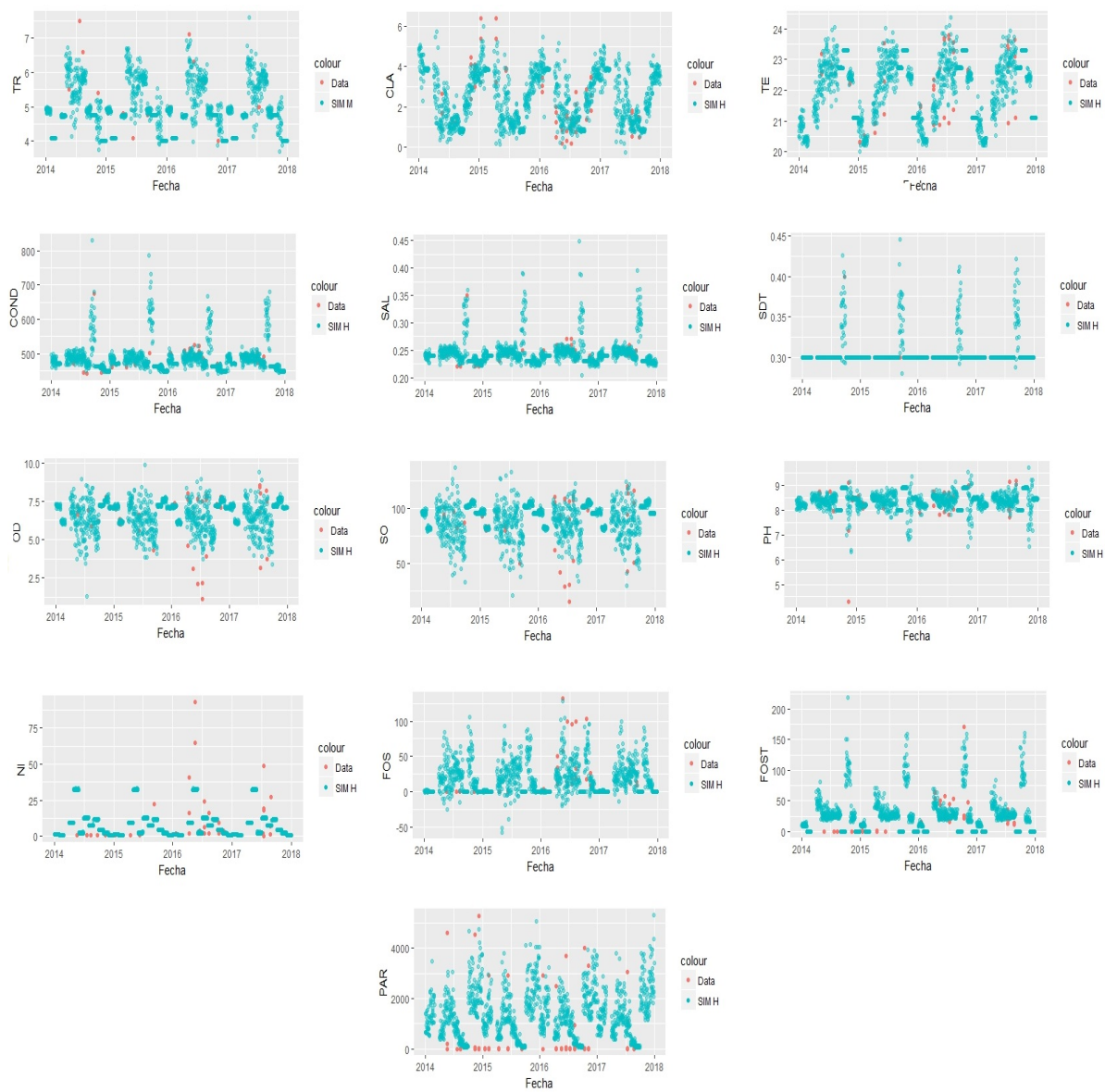
1. Los modelos de regresión lineal múltiple encontrados para los sitios de muestreo en el Lago de Atitlán con el criterio de AIC, describen de manera global la transparencia esto es debido al p-valor de los modelos que se aprecia en el cuadro (XIII. 1) en anexo; sin embargo los modelos encontrados para Bahía Santiago y Bahía San Lucas explican un 68.48 % y 60.61 % de variabilidad de la transparencia, mayores a los otros tres modelos.
2. En los sitios Centro WG y San Pedro - San Pablo - San Marcos todos los parámetros estudiados exceptuando SDT son componentes importantes en el modelado de la transparencia, como se aprecia en el cuadro (XIII. 2) y (XIII.3) en anexo.
3. Los modelos de regresión lineal múltiple encontrados para la transparencia no son buenos predictores para la transparencia del agua del Lago de Atitlán, debido a que no cumplen todas las condiciones de Gauss-Markov, véase (XIII.4) en anexo. Dicho de otra manera, los estimadores lineales insesgados que se encontraron no son de mínima varianza.
4. La transparencia posee relaciones fuertes con temperatura, clorofila a, salinidad y oxígeno disuelt, mientras que relaciones medias con sólidos disueltos totales, nitratos, fosfatos, fósforo total y conductividad como se aprecia en el cuadro (XIII.5) en anexo.
5. En los modelos de Cobb-Douglas encontrados, véase el cuadro (XIII.6) en anexo para modelos abiertos, a partir de las elasticidades se evidencia que en la Bahía de Santiago Atitlán la temperatura es un parámetro influyente pues, un aumento del 10 % de temperatura y clorofila

a, 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 a, 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-San Pablo-San Marcos, 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 (XIII.6) se supone que A representa los otros parámetros físico químicos 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.

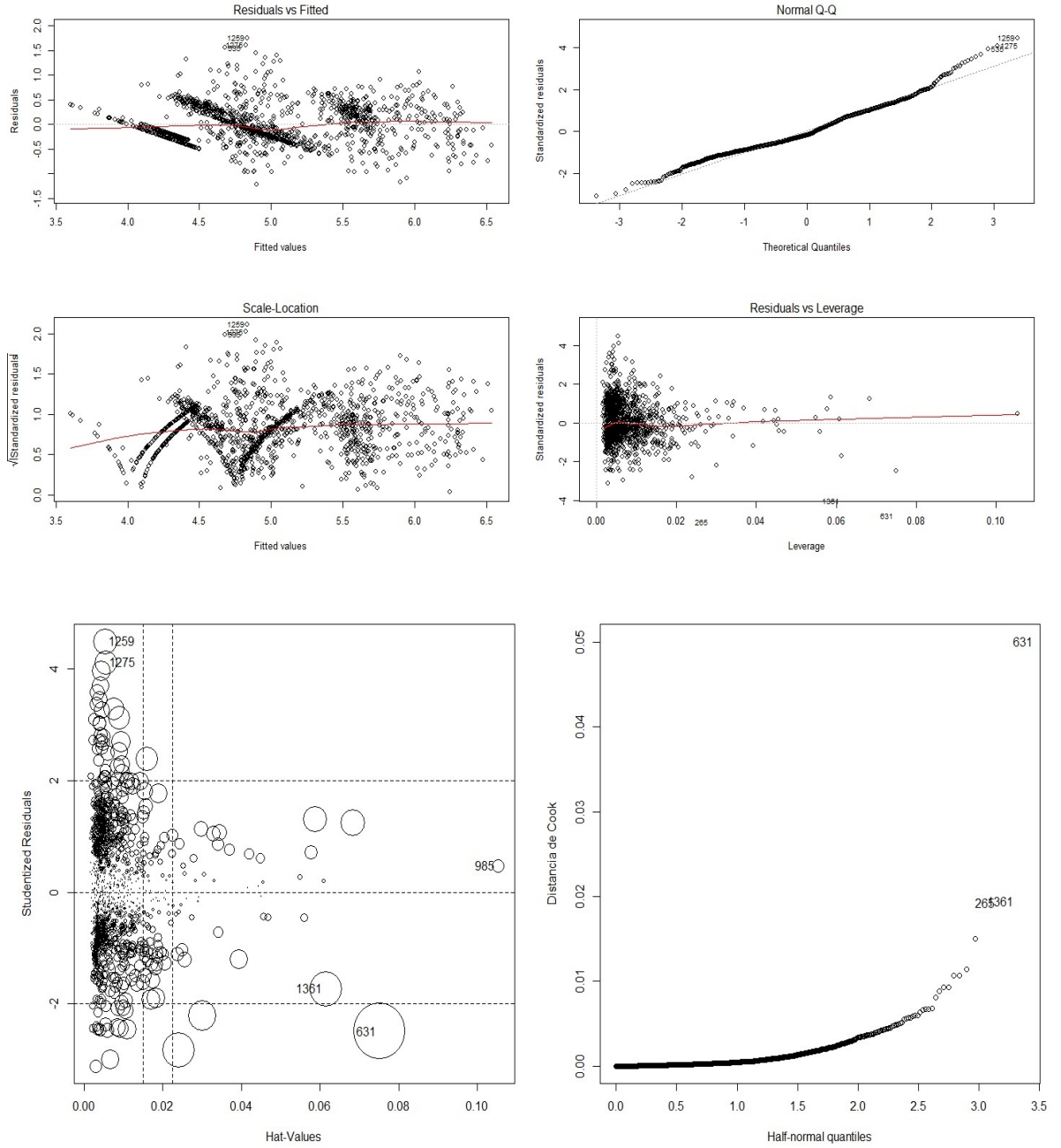
6. Si se considera que los modelos de Cobb-Douglas encontrados en el cuadro (XIII.7), modelos cerrados con la condición de $A = 1$ y $\sum_{i=1}^n \alpha_i = 1$, es decir los parámetros usados para la estimación de estos modelos representan todas las variables que inciden en la transparencia del agua 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-San Pablo-San Marcos la temperatura, los sólidos disueltos totales y salinidad son los factores más influyentes en la transparencia.
7. El modelo ARIMA, para el pronóstico de la transparencia, de la Bahía de Santiago Atitlán posee el menor AIC, AICc y BIC.
8. Los modelos ARIMA encontrados, que se exponen en el cuadro (XIII.9) describen modelos estacionarios; ésto debido a los p-valores de la prueba Box-Pierce.
9. Del pronóstico de transparencia para el año 2018 para los sitios estudiados: Bahía Santiago, Centro WG, Panajachel, Bahía San Lucas y San Pedro-San Pablo-San Marcos tendrán una media en transparencia de 4.984, 6.07, 6.085, 6.642 y 6.787 metros respectivamente. Notando que Bahía Santiago es el más bajo y San Pedro-San Pablo-San Marcos el más alto, véase el cuadro (XIII.9) en anexos.

Figura VIII.1: Simulación vrs real. Bahía Santiago.



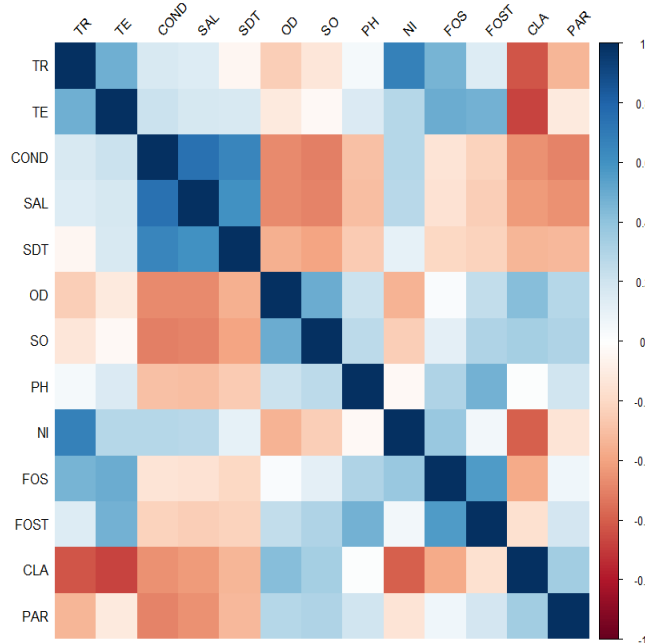
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.2: Análisis de residuos, atípicos e influyentes. Bahía Santiago.



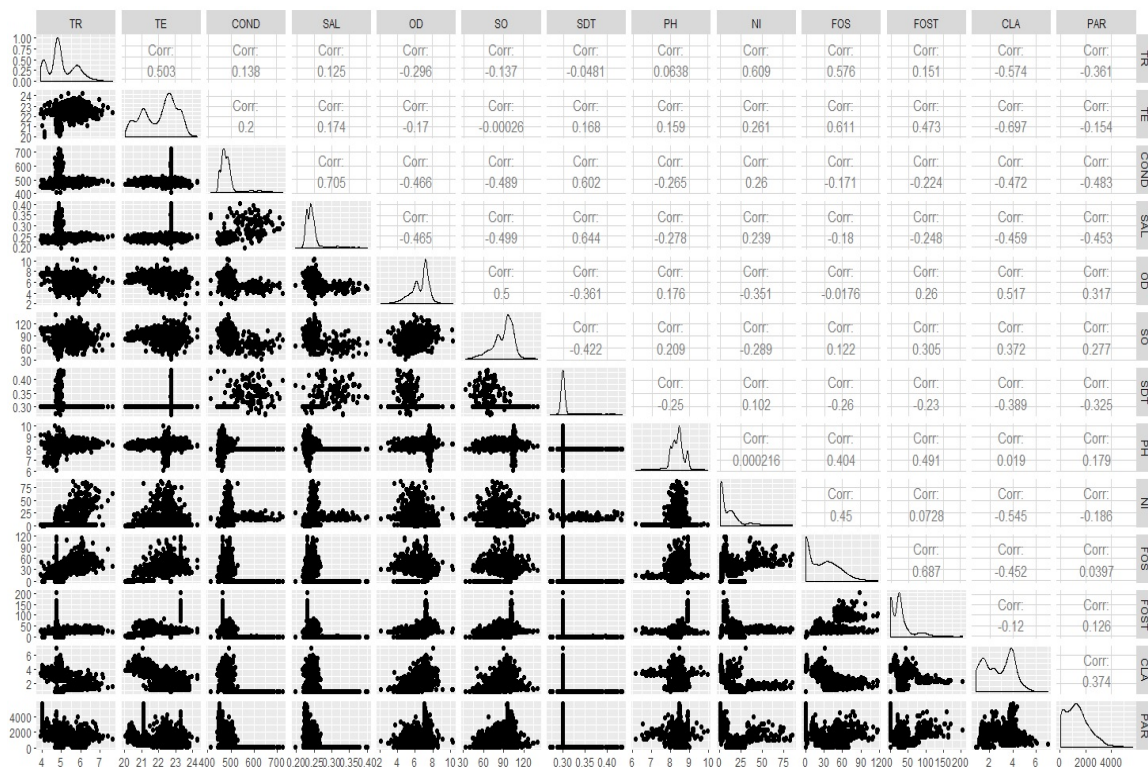
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.3: Matriz de correlaciones. Bahía Santiago.



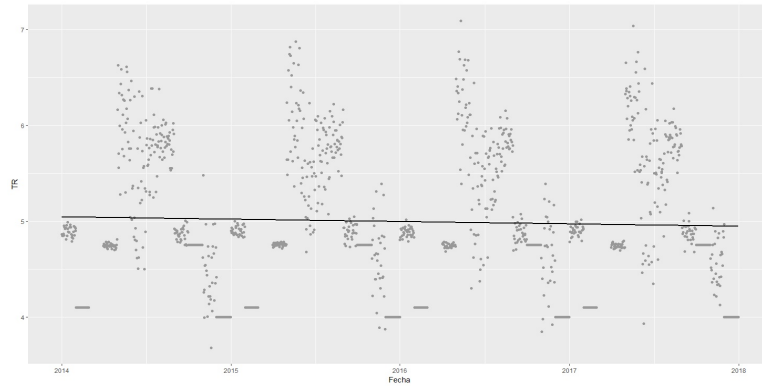
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.4: Matriz de correlaciones. Bahía Santiago.



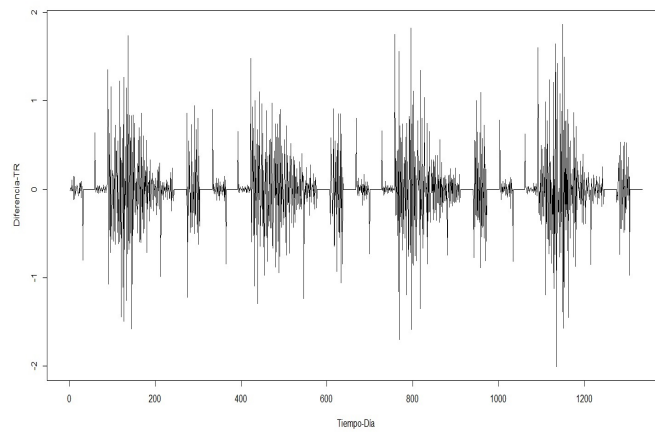
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.5: Transparencia. Bahía Santiago.



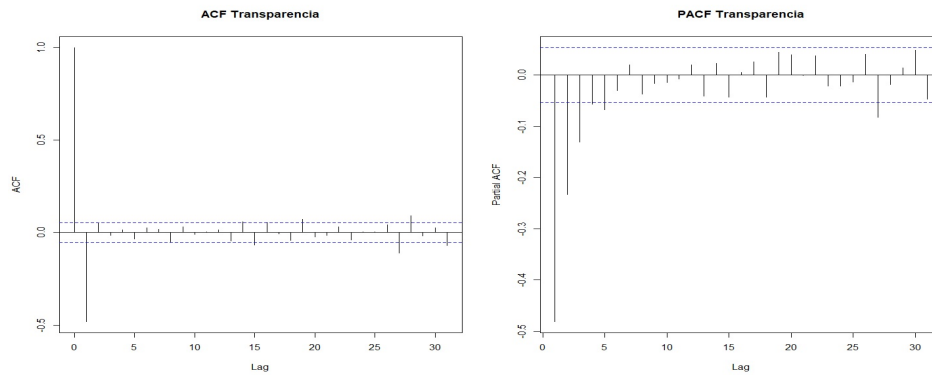
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.6: Diferencia de transparencia. Bahía Santiago.



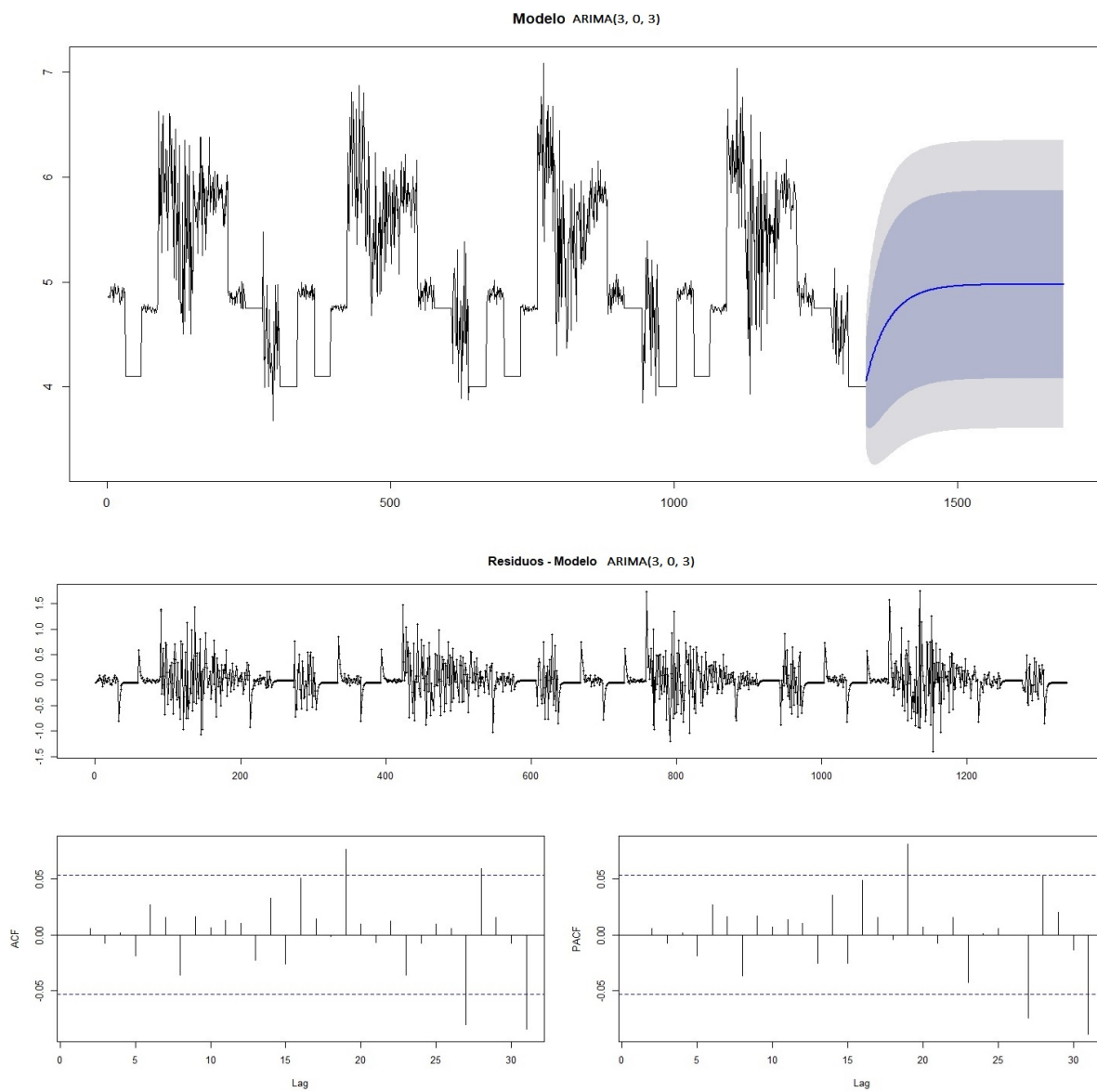
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.7: ACF y PACF de transparencia. Bahía Santiago.



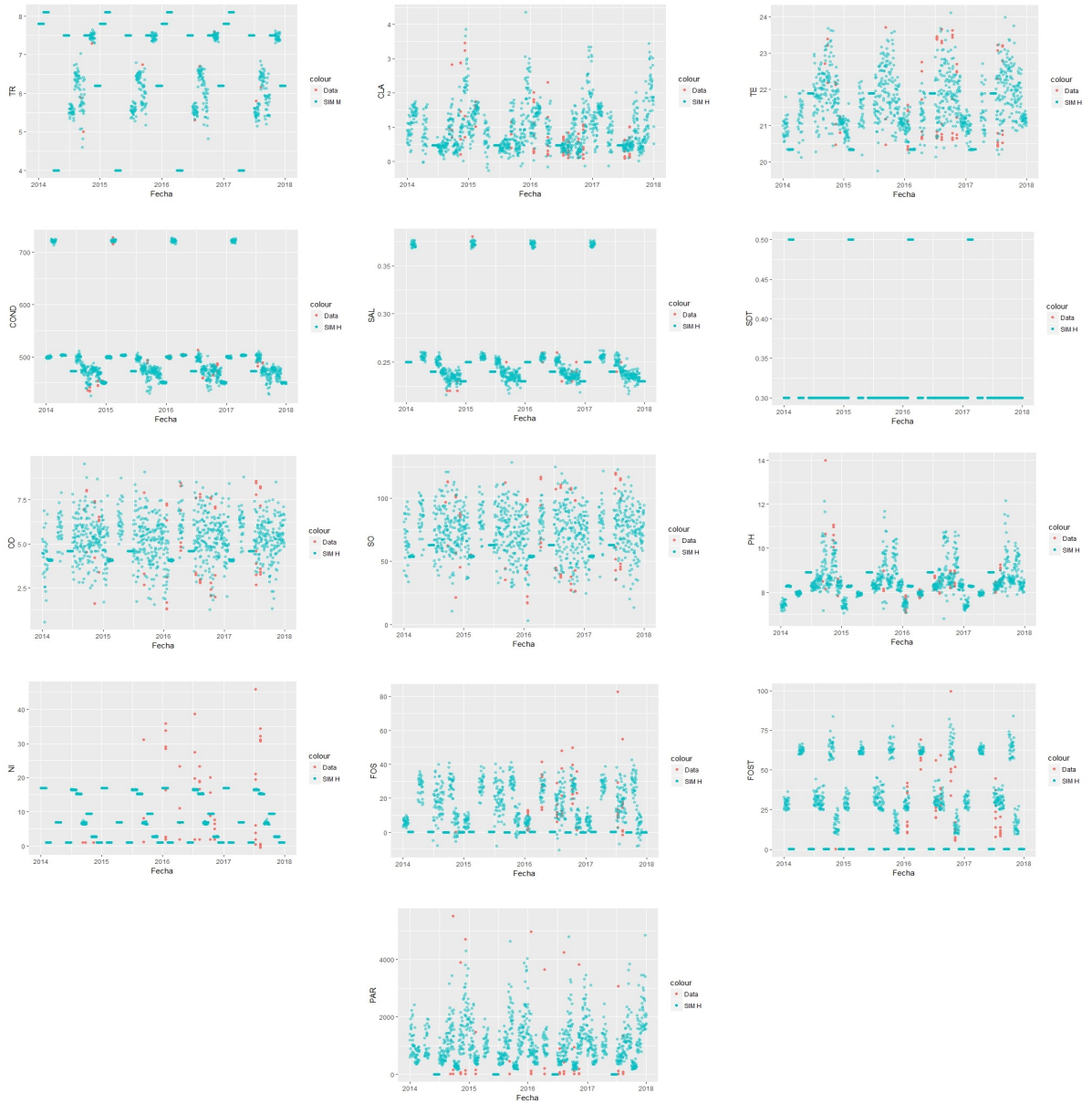
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.8: Pronóstico 2018 de TR y su diagnóstico. Bahía Santiago.



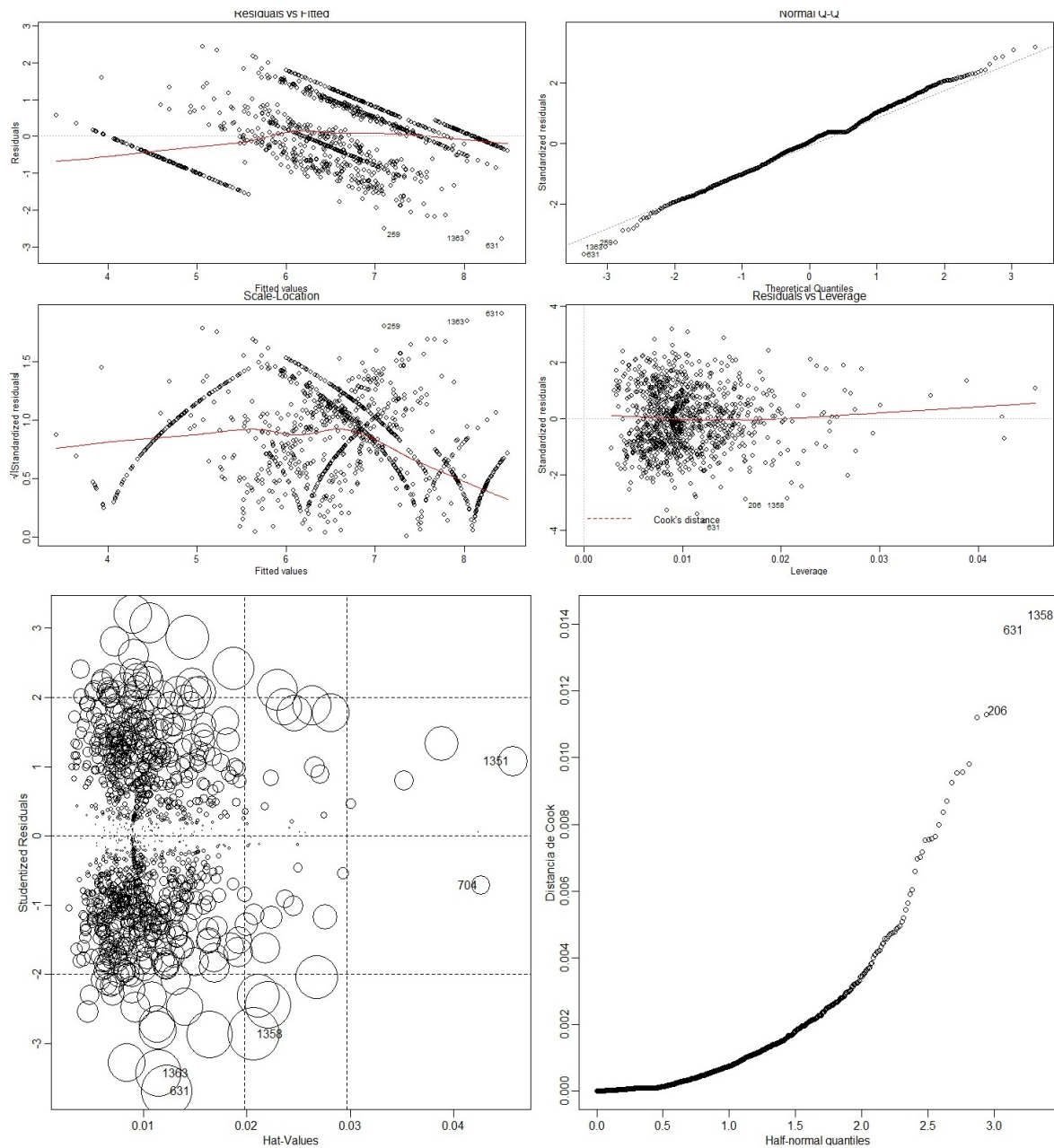
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.9: Simulación vrs real. Bahía San Lucas.



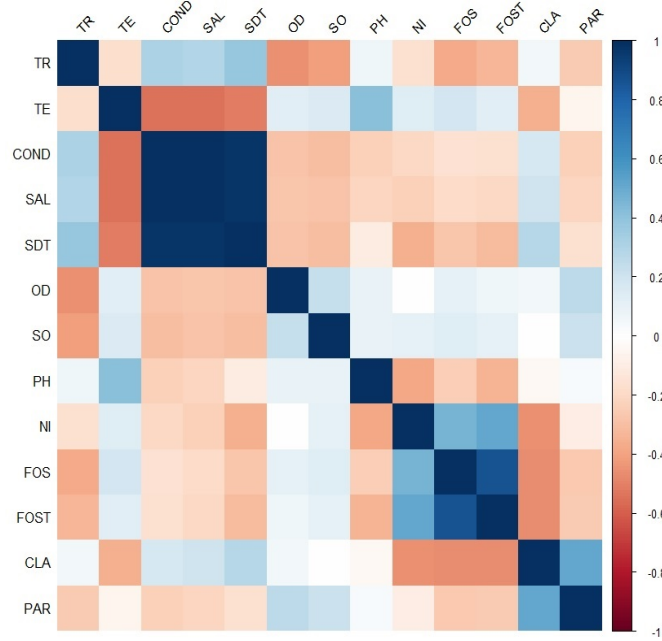
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.10: Análisis de residuos, atípicos e influyentes. Bahía San Lucas.



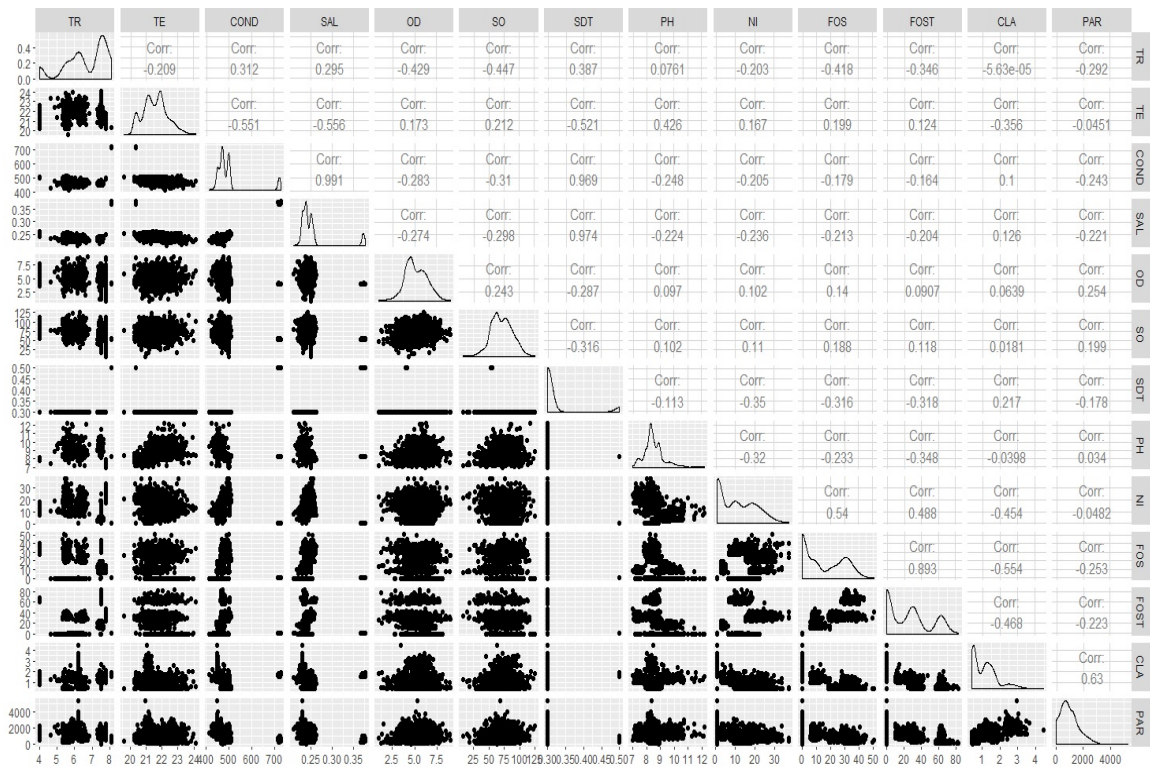
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.11: Matriz de correlaciones. Bahía San Lucas.



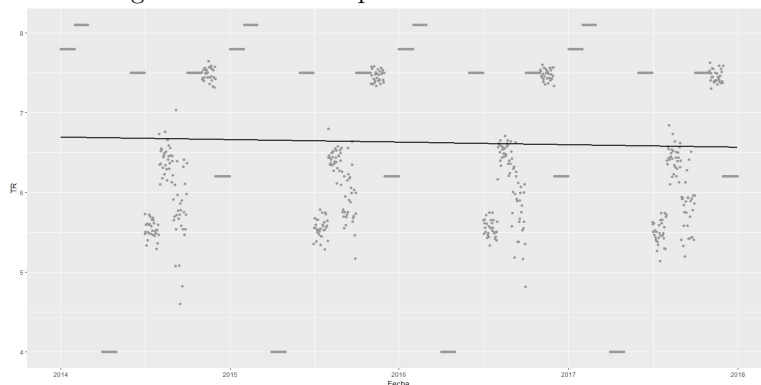
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.12: Matriz de correlaciones. Bahía San Lucas.



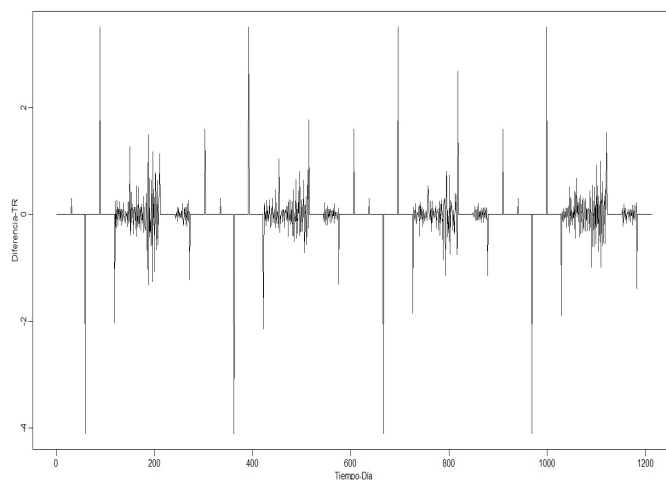
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.13: Transparencia. Bahía San Lucas.



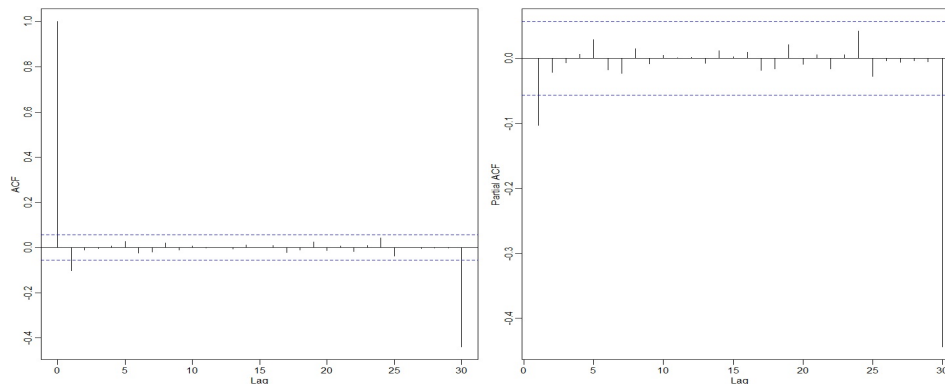
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.14: Diferencia de transparencia. Bahía San Lucas.



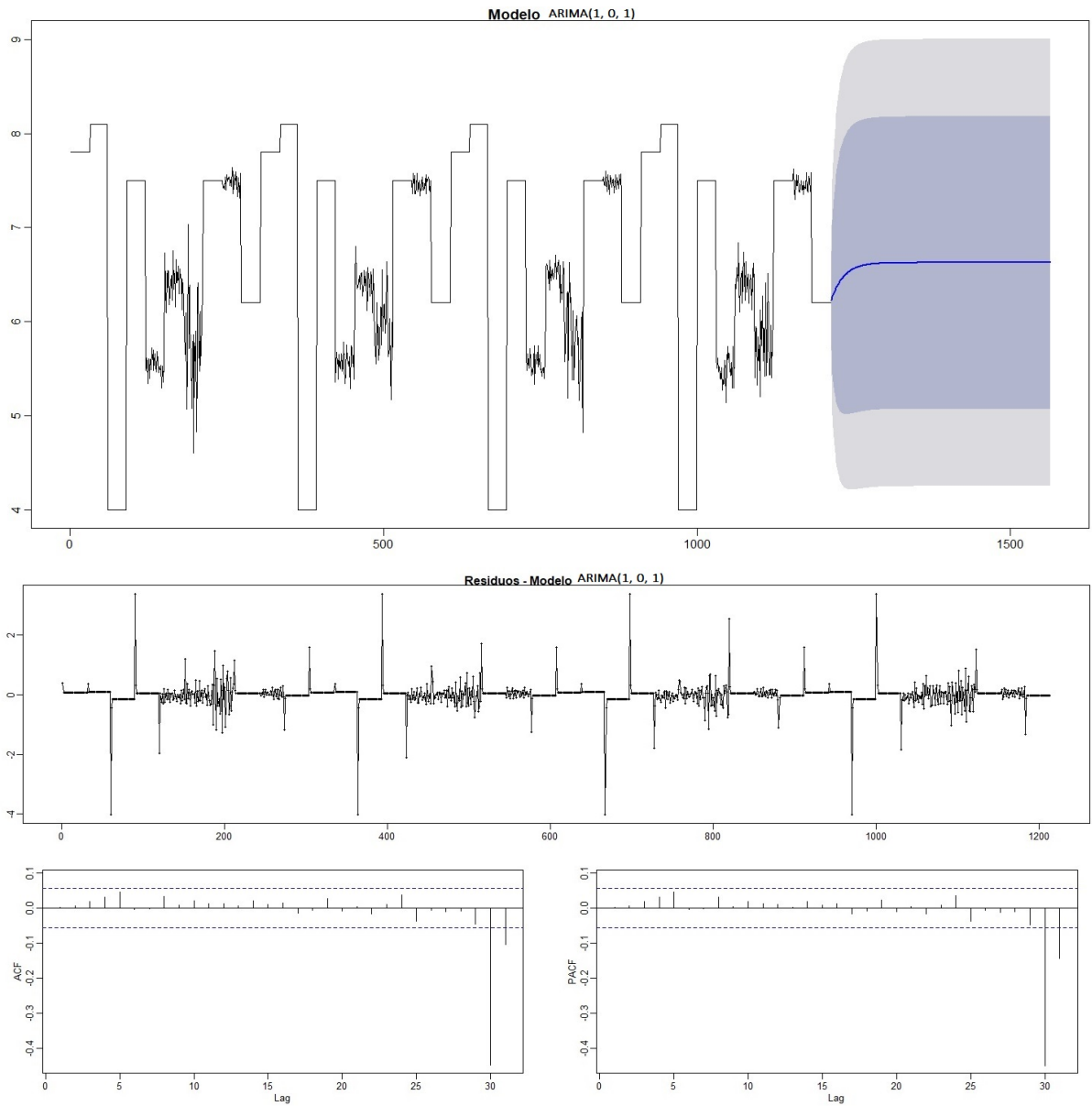
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.15: ACF y PACF de transparencia. Bahía San Lucas.



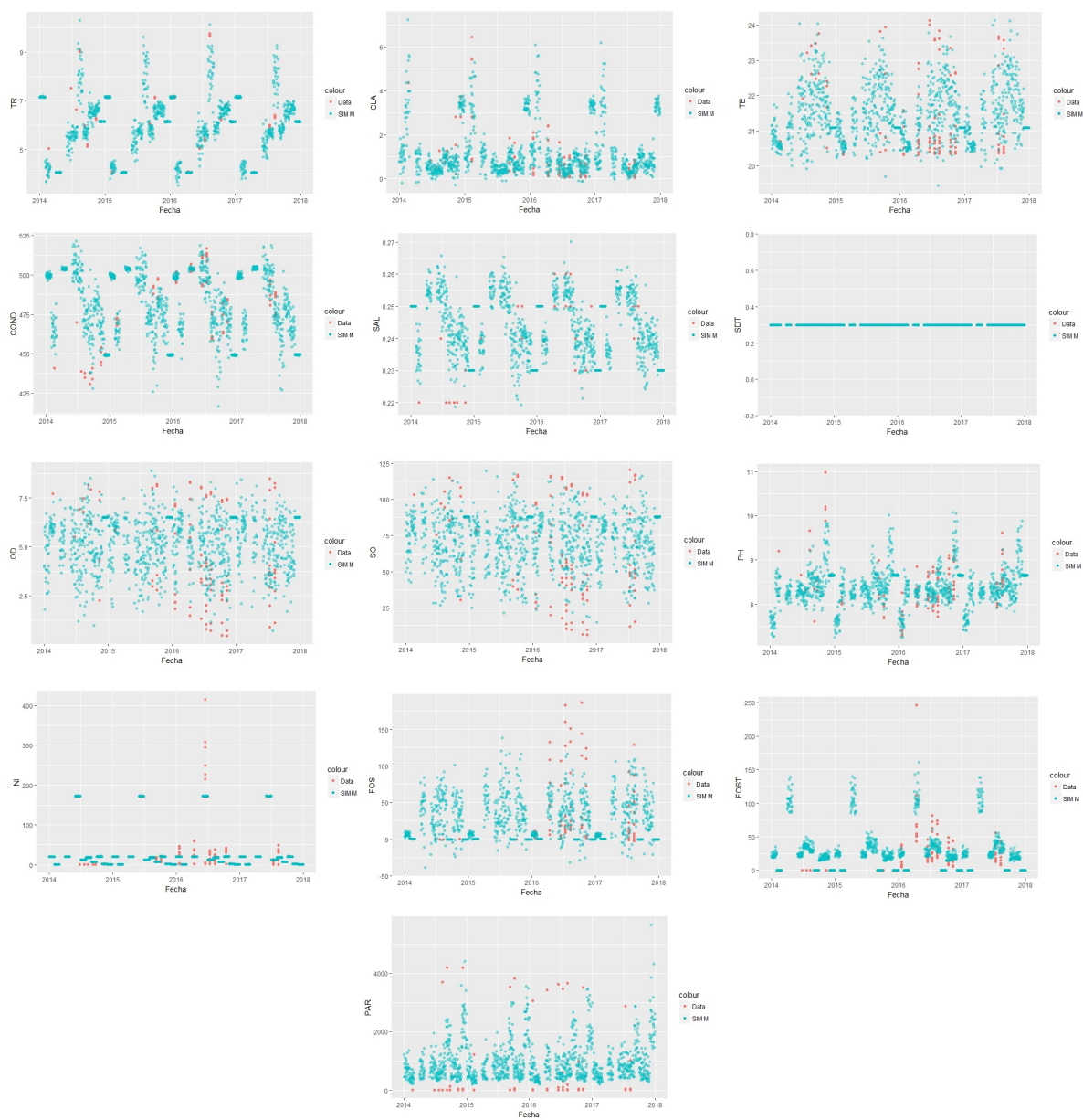
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.16: Pronóstico 2018 y su diagnóstico. Bahía San Lucas.



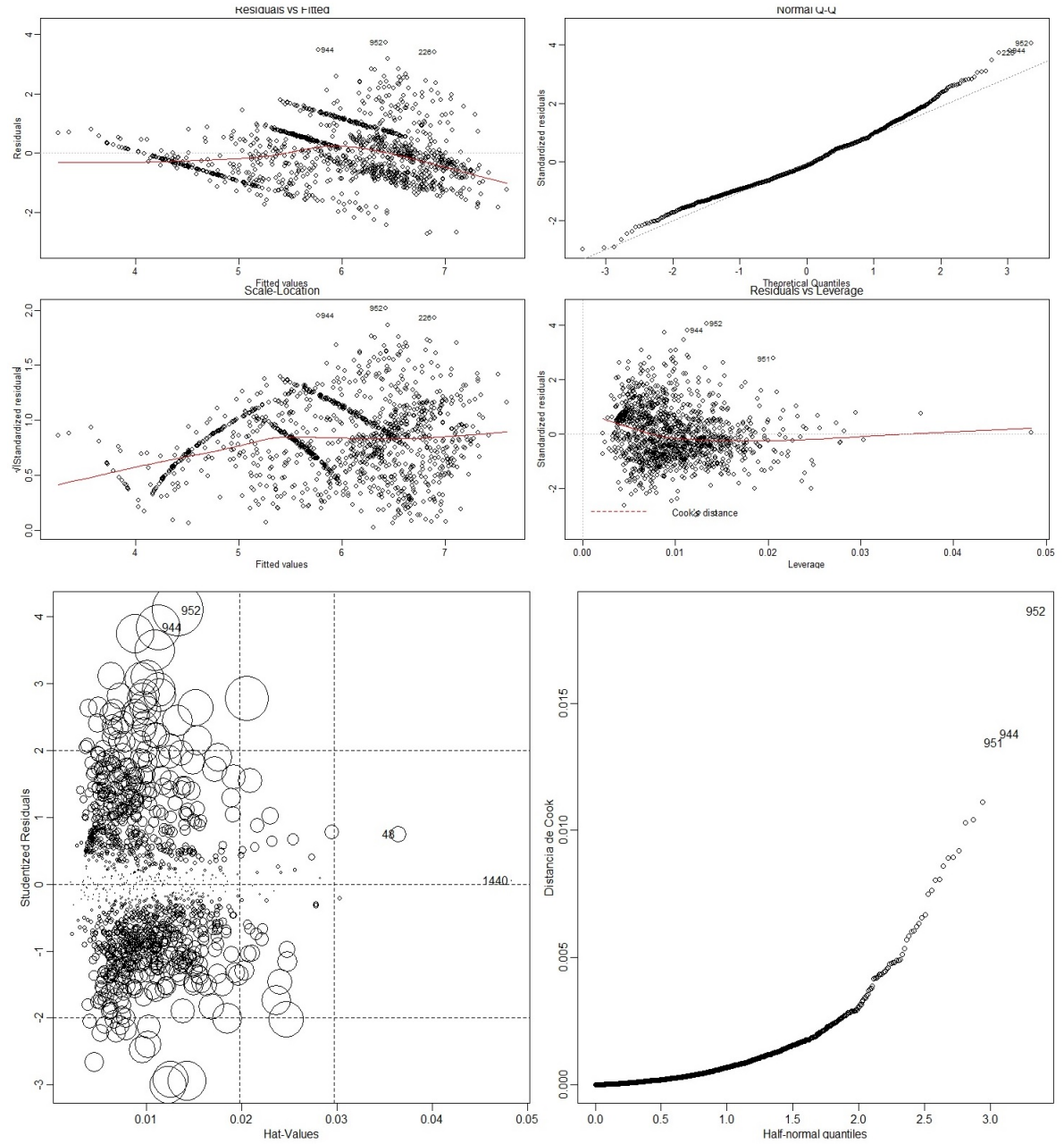
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.17: Simulación vrs real. Centro WG.



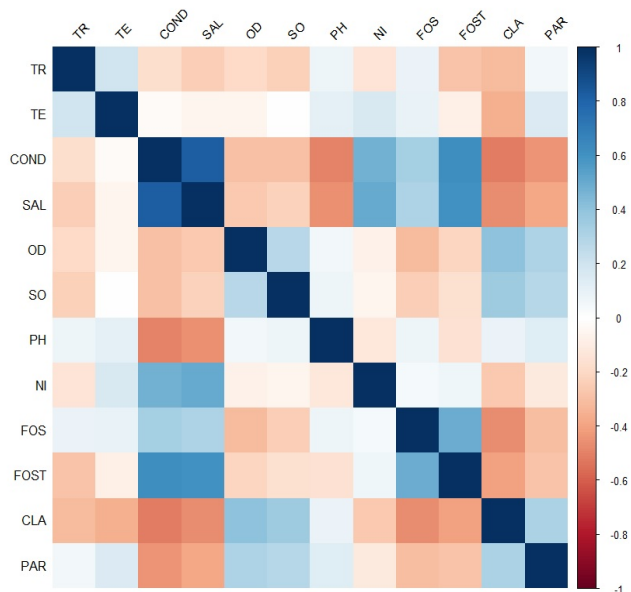
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.18: Análisis de Residuos. Centro WG.



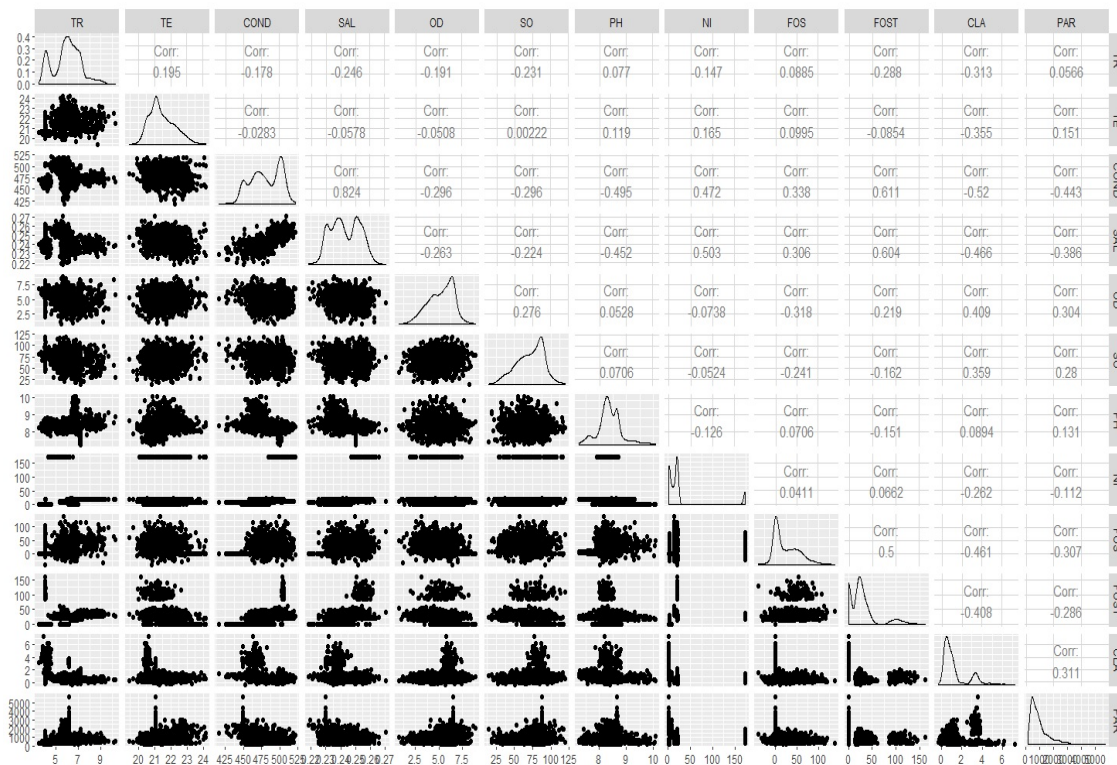
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.19: Matriz de correlaciones. Centro WG.



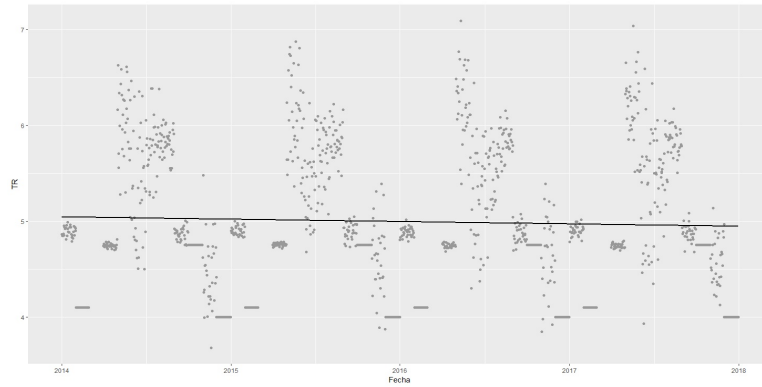
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.20: Matriz de correlaciones. Centro WG.



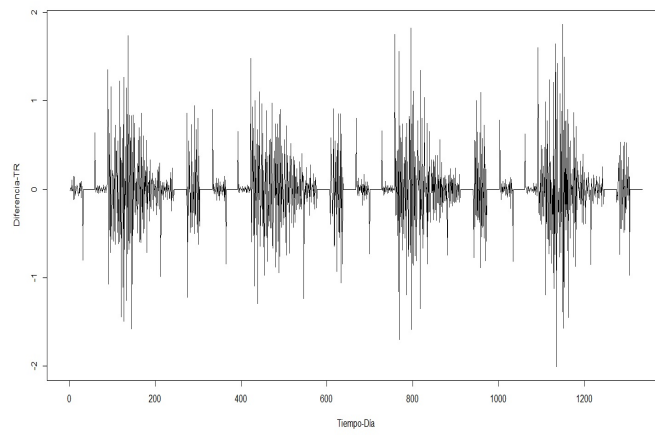
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.21: Transparencia. Centro WG.



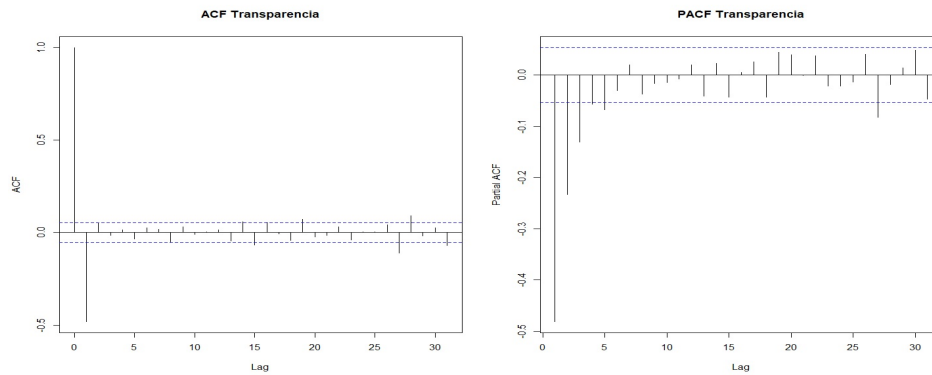
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.22: Diferencia de transparencia. Centro WG.



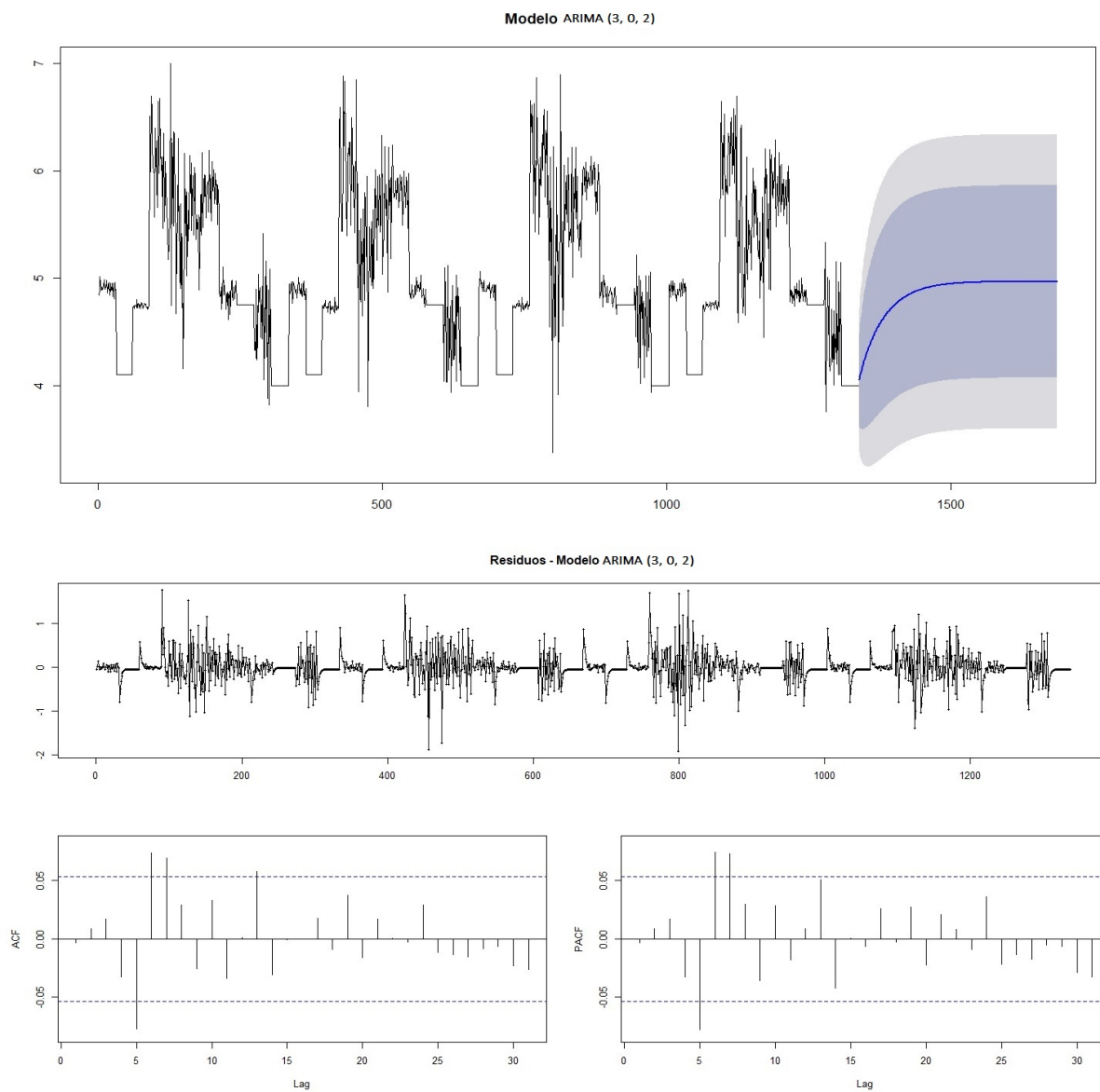
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.23: ACF y PACF de transparencia. Centro WG.



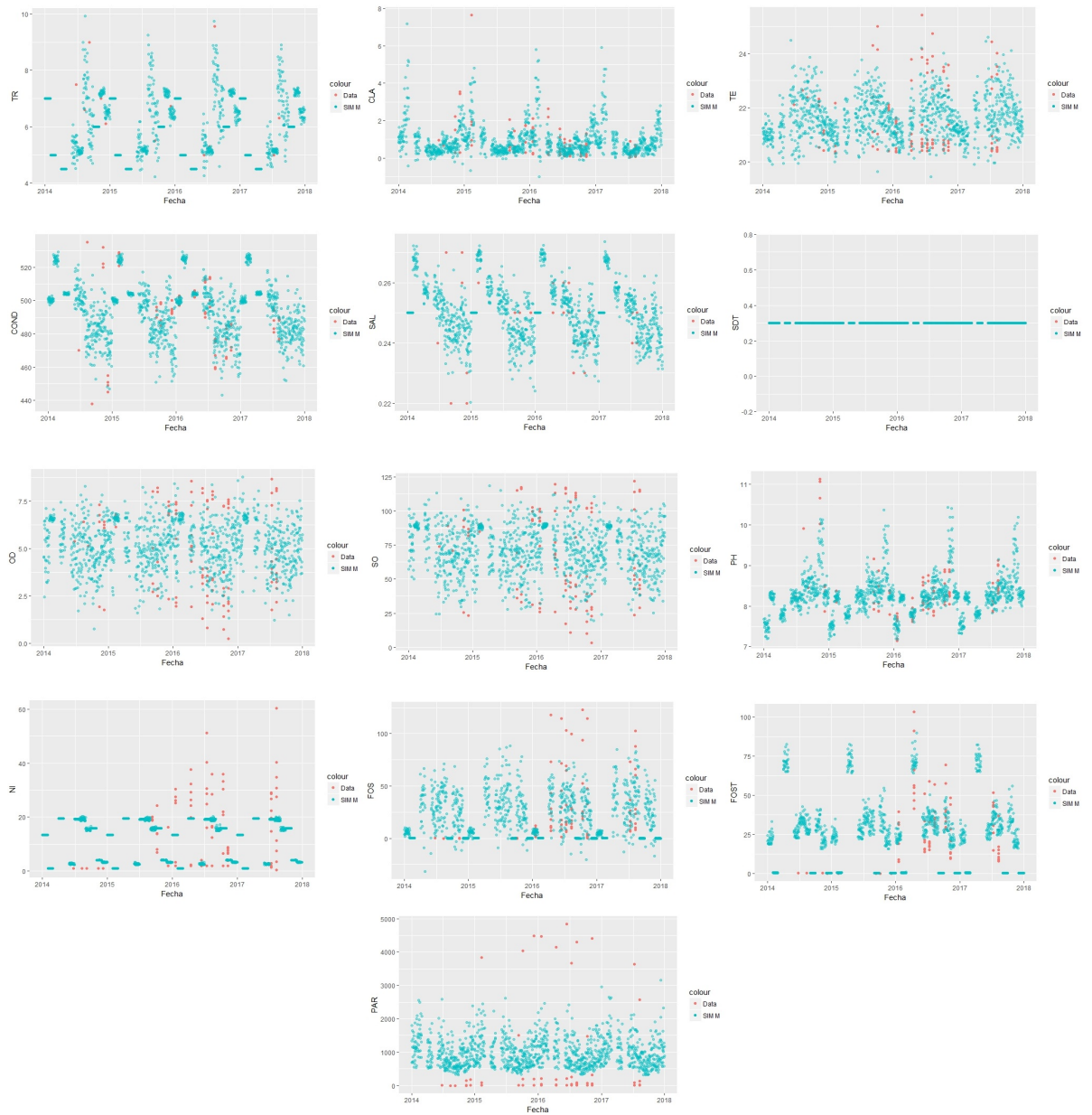
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.24: Pronóstico 2018 y su diagnóstico. Centro WG.



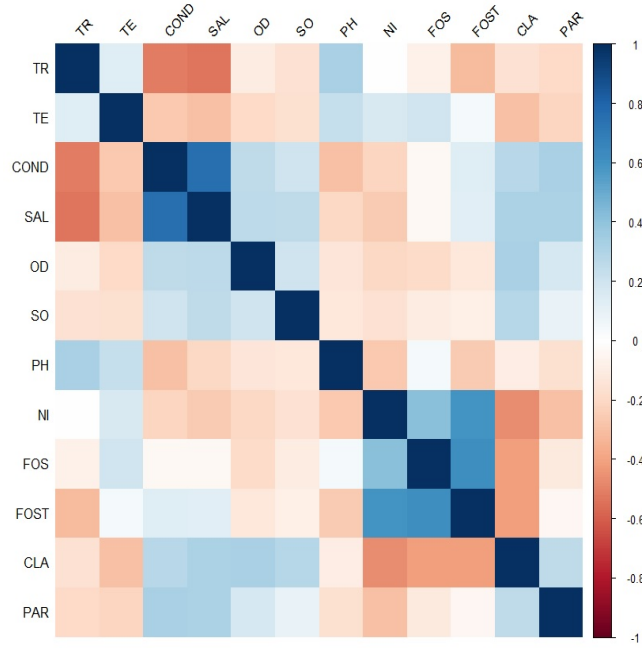
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.25: Simulación vrs real. Panajachel.



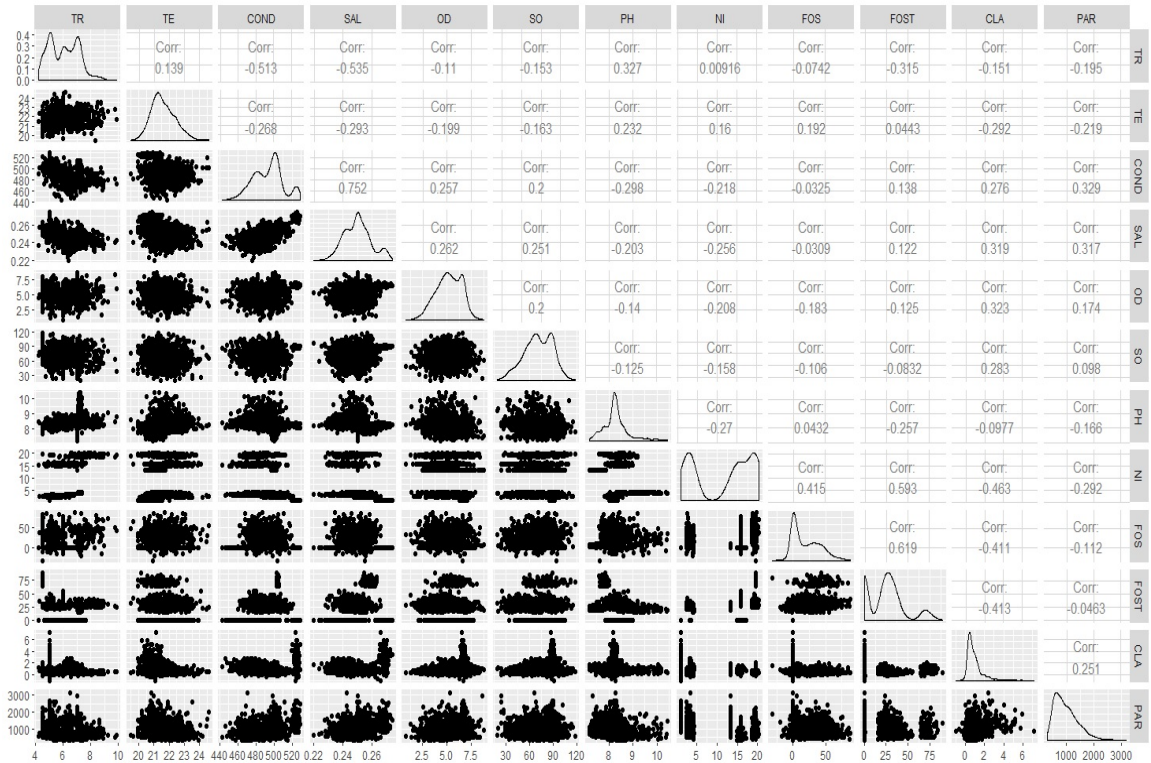
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.27: Matriz de correlaciones. Panajachel.



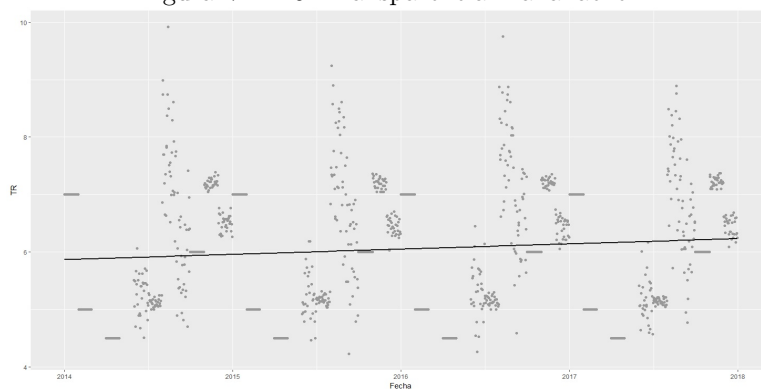
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.28: Matriz de correlaciones. Panahachel.



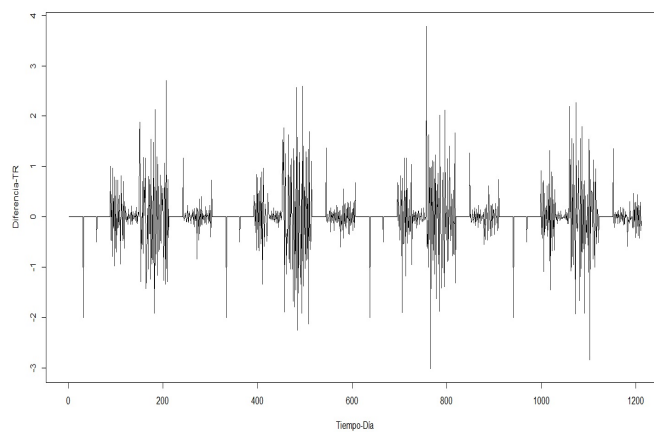
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.29: Transparencia. Panahachel.



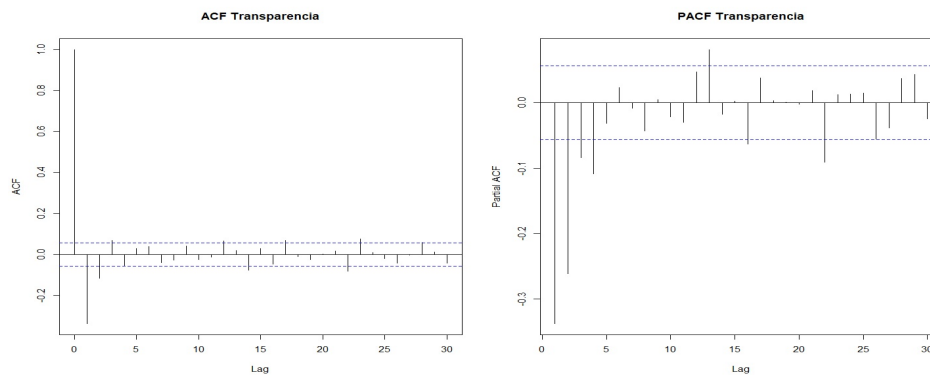
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.30: Diferencia de transparencia. Panajachel.



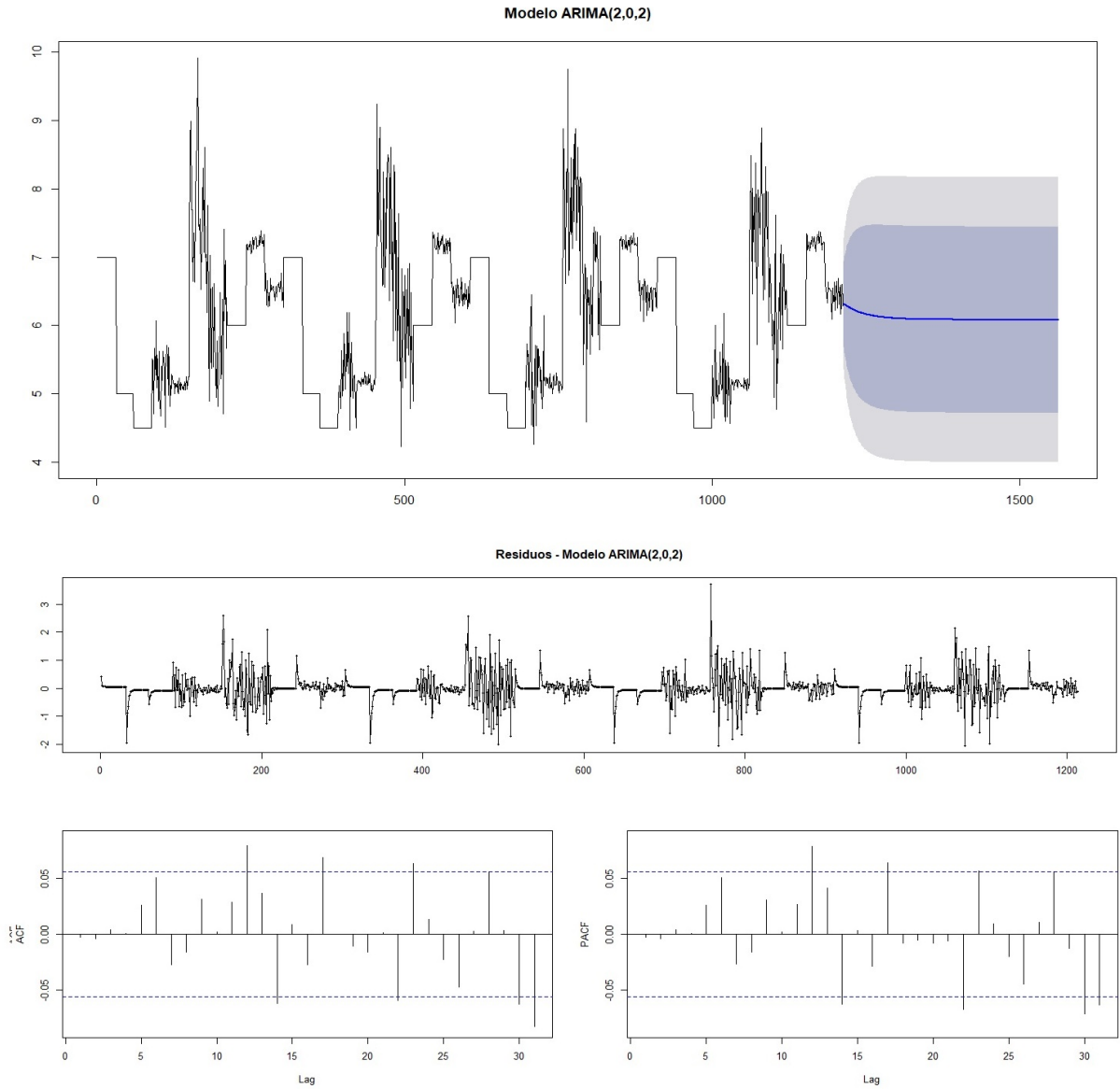
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.31: ACF y PACF de transparencia. Panahachel.



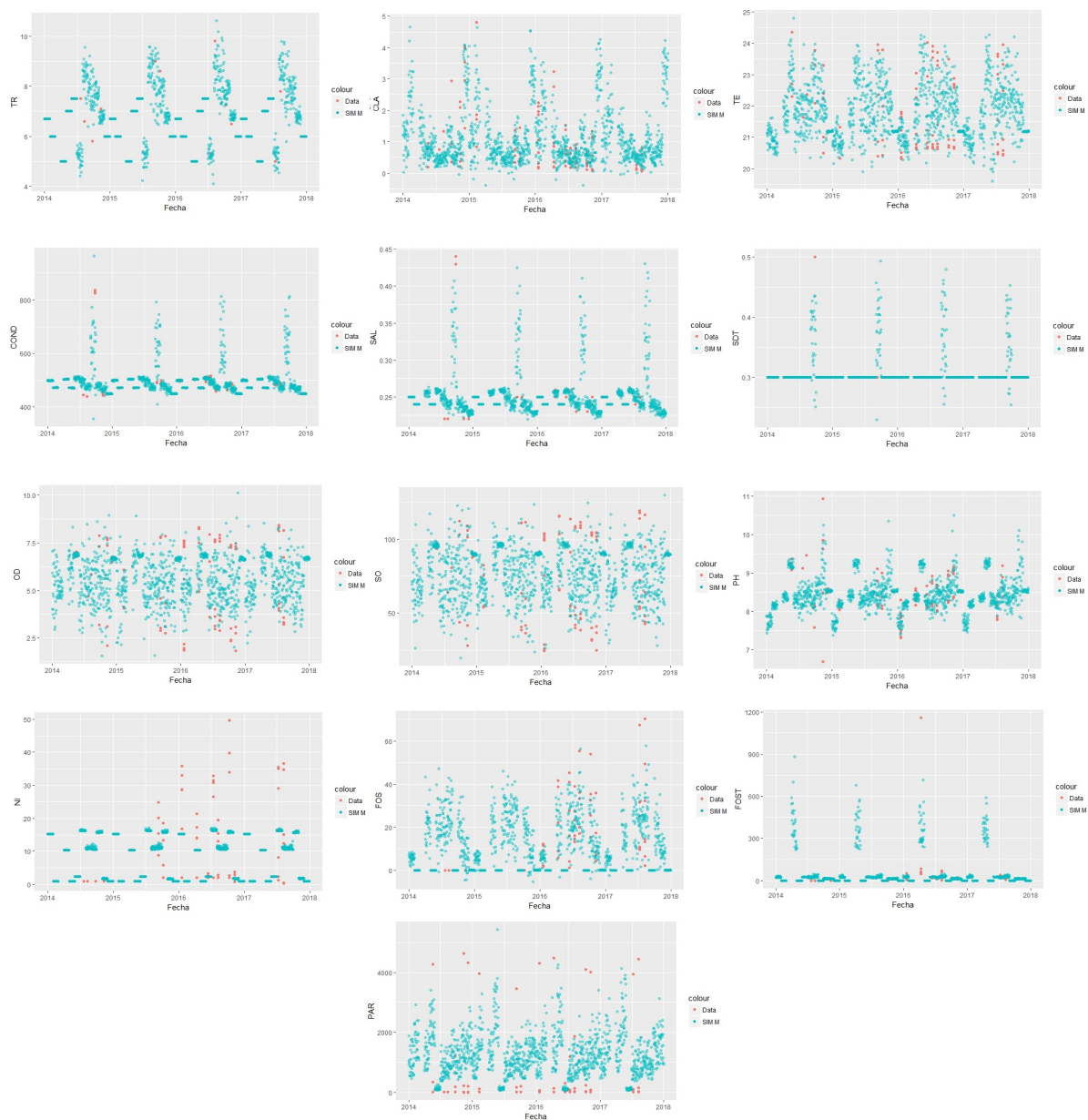
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.32: Pronóstico 2018 y su diagnóstico. Panahachel.



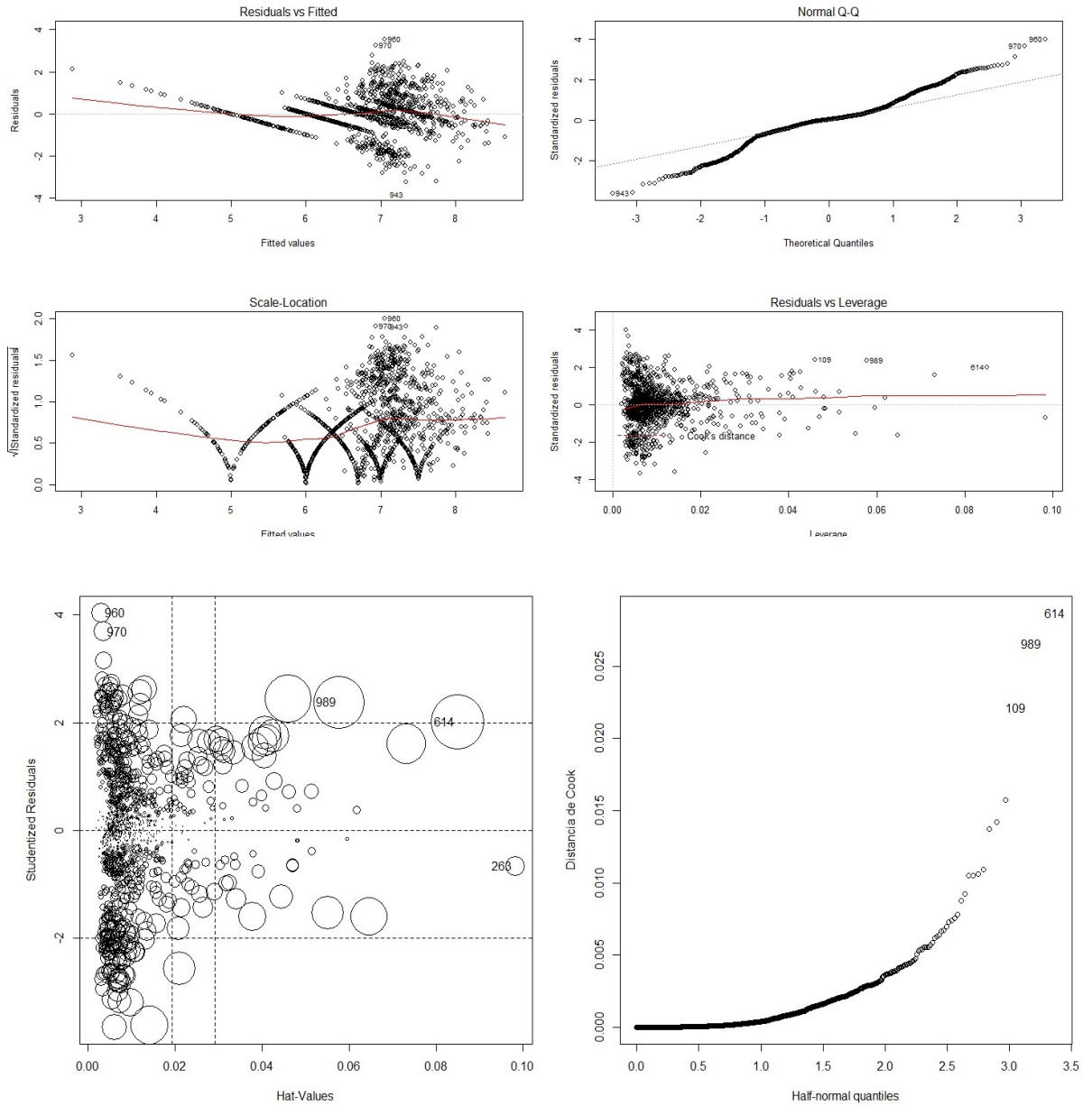
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.33: Simulación vrs real. San Pedro-San Pablo.



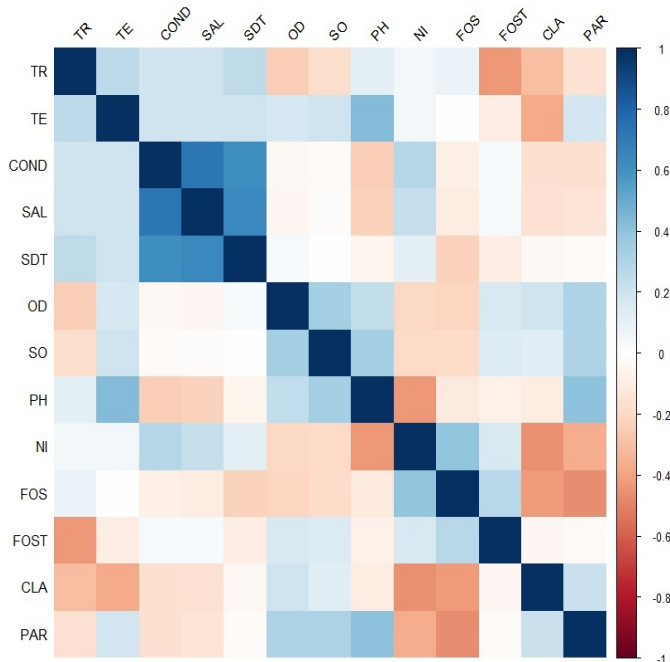
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.34: Análisis de residuos, puntos atípicos e influyentes. San Pedro-San Pablo.



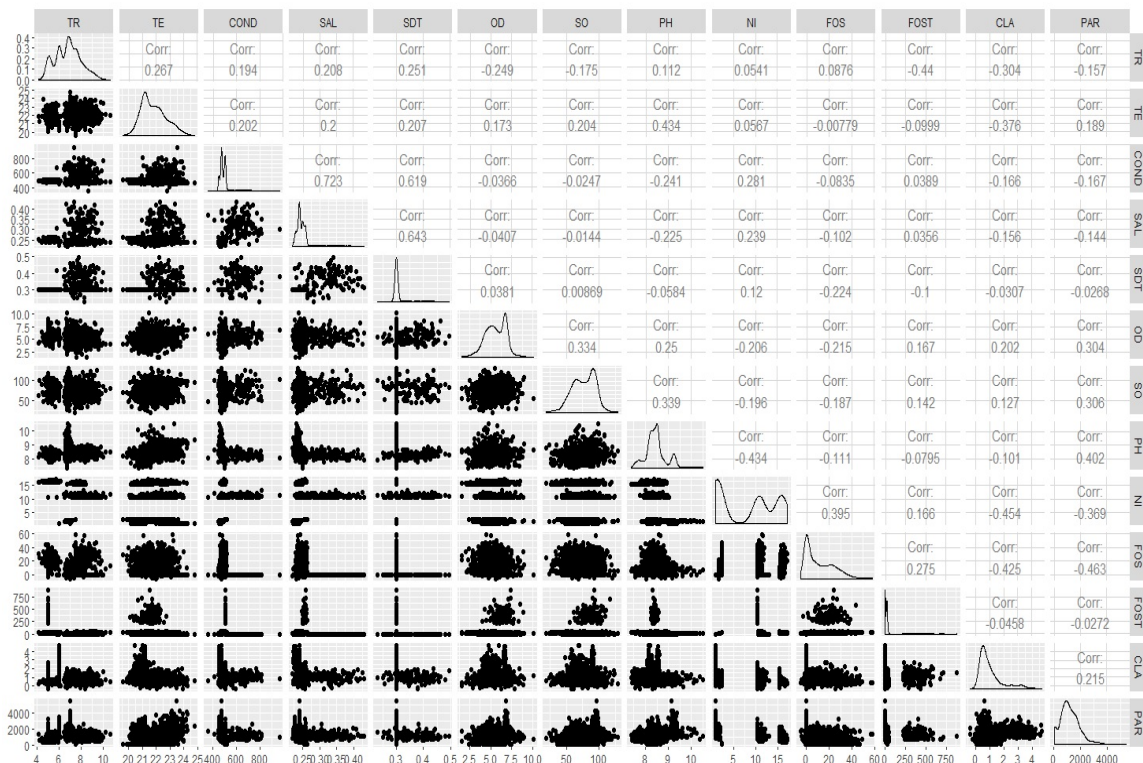
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.35: Matriz de correlaciones. San Pedro-San Pablo.



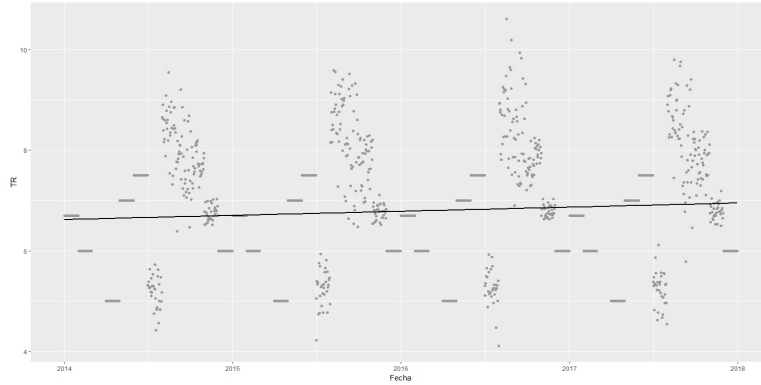
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.36: Matriz de correlaciones. San Pedro-San Pablo.



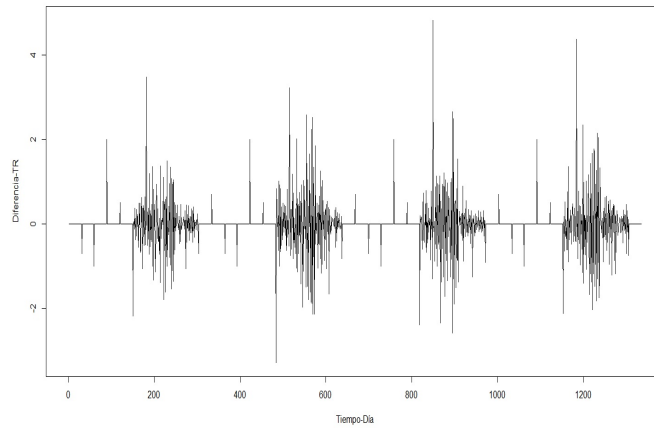
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.37: Transparencia. San Pedro-San Pablo.



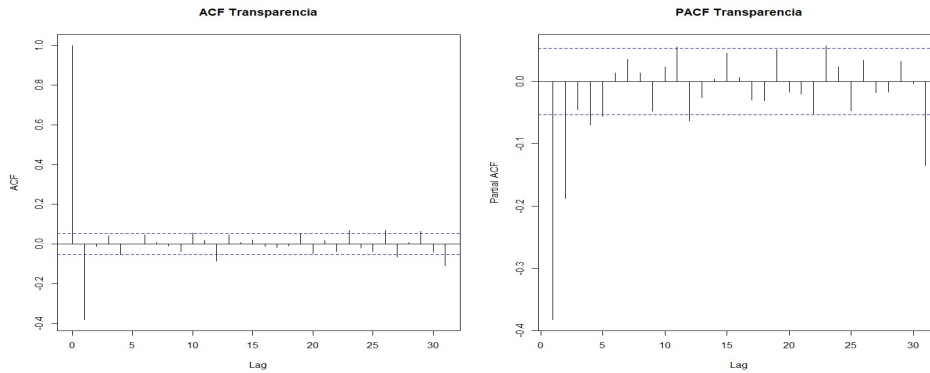
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.38: Diferencia de transparencia. San Pedro-San Pablo.



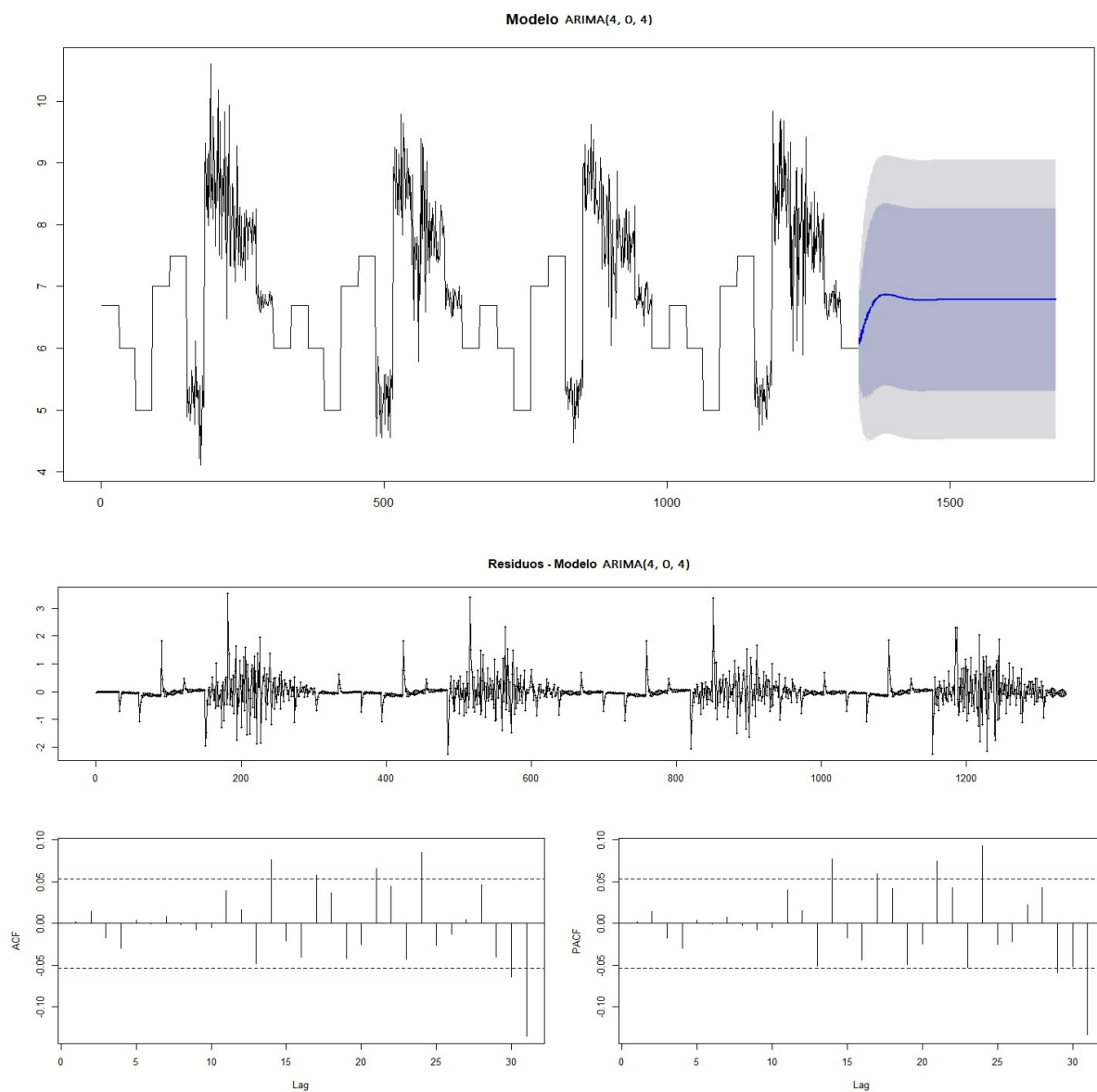
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.39: ACF y PACF de transparencia. San Pedro-San Pablo.



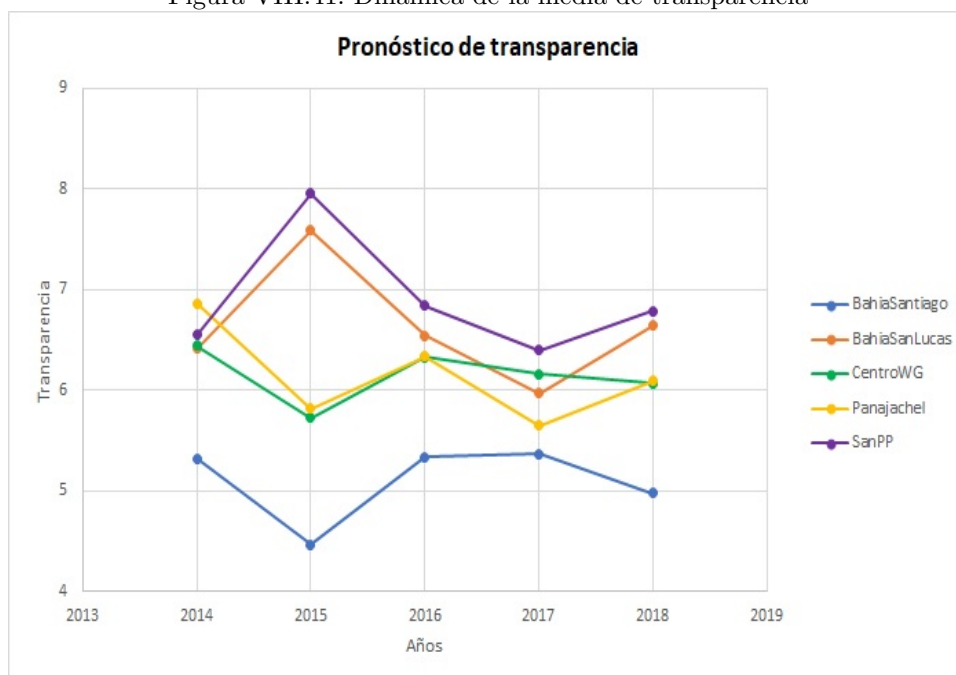
Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.40: Pronóstico 2018 y su diagnóstico. San Pedro-San Pablo.



Fuente: Elaboración propia con base en los datos de AMSCLAE.

Figura VIII.41: Dinámica de la media de transparencia



Fuente: Elaboración propia con base en los datos de AMSCLAE.

IX Conclusiones

1. Se cumplieron con los objetivos principales de describir la interacción de los factores físico-químicos con la transparencia, clasificar las relaciones más importantes de los factores físico-químicos con la transparencia hasta explicar su influencia y pronosticar la transparencia para el año 2018; encontrando que en los sitios Centro WG y San Pedro-San Pablo-San Marcos todos los factores físico-químicos poseen una interacción mayor para el modelado de la transparencia, los parámetros temperatura, clorofila a, salinidad y oxígeno disuelto están fuertemente relacionados con la transparencia en la mayoría de sitios y que el sitio Bahía Santiago tendrá la transparencia media más baja con 4.98 metros, mientras que San-Pedro-San Pablo-San Marcos tendrá la transparencia media más alta con 6.78 metros para el año 2018.
2. Los modelos de regresión lineal múltiple encontrados para la transparencia, poseen estimadores lineales insesgados que no son de mínima varianza.
3. 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.
4. Los nitratos y fosfatos parecen no incidir en la transparencia, como se aprecia en los resultados de éste trabajo. La validez de lo anterior depende principalmente de la toma correcta de las mediciones de dichos factores en los últimos años. Por lo tanto es necesario estudiar dichos parámetros con más atención.
5. El modelo ARIMA de pronóstico del sitio Bahía Santiago posee mayor calidad y eficiencia comparado con los otros modelos encontrados.
6. El estado ecológico del Lago de Atitlán no es un problema homogéneo; como se evidencia con la transparencia (véase la Figura no. 41) no es posible definir, inferir y generalizar la calidad del agua del Lago de Atitlán sólo con un sitio de estudio.

X Recomendaciones

1. Para futuras investigaciones relacionados al tema y desarrollo abordado en este trabajo de graduación, se sugiere estudiar la transparencia del agua del Lago de Atitlán basándose desde un diseño experimental fijo. Es decir partiendo de un diseño experimental tal que mantenga bajo control ciertos parámetros de medición, como la hora de muestreo. Implementando una periodicidad continua en la toma de muestras.
2. La transparencia es una variable importante y práctica para determinar la calidad del agua del Lago de Atitlán; por lo que es un buen candidato a ser un predictor para las variables que requieren de mediciones sofisticadas como clorofila a, fosfatos, nitratos, temperatura y sólidos disueltos totales.
3. La temperatura es un parámetro interesante de estudio con la transparencia, según el análisis de este trabajo es un parámetro que influye en el comportamiento de la transparencia; por lo que sugiere plantearse 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.
4. Se recomienda medir los cambios de transparencia durante el día, en un sitio de estudio.
5. Se recomienda evaluar y aplicar las técnicas desarrolladas en este trabajo para otros lagos de Guatemala.
6. Se recomienda abordar el problema del estudio de la transparencia/ecología como un problema matemático y no como un problema limnológico; desde esa concepción describir y explicar la fenomenología matemática de la transparencia/ecología del agua del Lago de Atitlán (y otros Lagos), y que los resultados obtenidos se le de una interpretación limnológica.
7. Para futuras investigaciones de esta índole, se recomienda delimitar y concentrarse en la matemática del problema de estudio. Debido a que genera una visión pura para la solución del problema.
8. Se recomienda plantear modelos estocásticos de otra índole para la transparencia, para el estudio de los parámetros más importantes establecidos en esta investigación, principalmente de modelos estocásticos que contemplen volatilidad y variabilidad de los parámetros físico-químicos del agua del lago de Atitlán.

9. Se recomienda complementar la parte del análisis por regresión lineal múltiple desarrollado en este trabajo con el análisis por componentes principales y de análisis por modelos robustos generalizados lineales.

XI Bibliografía

- Arrow, K., Chenery, H., Minhas, B. & Solow, R. (1961). Capital-labor substitution and economic efficiency. *The Review of Economics and Statistics*. 43(43).
- Braun, J. & Maindonald, J. H. (2015). *Package DAAG*. Recuperado de <https://cran.r-project.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*. Recuperado de <https://cran.r-project.org/web/packages/hexbin/index.html>
- Casella, G. & Berger, R. L. (2002). *Statistical Inference* (2nd ed.). Pacific Grove: Duxbury, Thomson Learning.
- 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).
- Cowpertwait, P. S. P. & Meltcalfe, A. V. (2008). *Introductory Time Series With R* (2nd ed.). New York: Springer.
- Cryer, J. D. & Chan, Kung-Sik. (2008). *Time Series Analysis, With Applications in R* (2nd ed.). New York: Springer.
- Faraway, J. (2016). *Package faraway*. Recuperado de <https://cran.r-project.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).
- Fox, J., Weisberg, S. & Adler, D. (2017). *Package car*. Recuperado de <https://cran.r-project.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.). New York: Springer.
- Hothorn, T., Zeileis, A., Farebrother, R. W., Cummins, C. & Mitchell, D. (2018). *Package lmtree*. Recuperado de <https://cran.r-project.org/web/packages/lmtree/index.html>
- Hyndman, R., O'Hara, M., Bergmeir, C., Razbashm, S. & Wang, E. (2017). *Package forecast*. Recuperado de <https://cran.r-project.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.
- Ryu, H. & Zellner, A. (1998). Alternative functional forms for production, cost and returns to scale function. *Journal of Applied Econometrics*. 13.
- Schoerke, B., Crowley, J., Cook, D., Hofmann, H. & Wickham, H. (2017). *Package GGally*. Recuperado de <https://cran.r-project.org/web/packages/GGally/index.html>
- Seber, G. & Lee, A. (2003). *Lineal Regression Analysis* (2nd ed.). New Jersey: Jhon Wiley & Sons, Inc.
- Seivert, C., Parmer, C., Hocking, T., Chamberlain, S. & Ram, K. (2017). *Package Plotly*. Recuperado de <https://cran.r-project.org/web/packages/plotly/index.html>
- Shumway, R. & Stoffer, D. (2006). *Time Series Analysis and Its Applications: With R examples* (2nd ed.). New York: Springer.
- Turlach, B. A. (2015). *Package quadprog*. Recuperado de <https://cran.r-project.org/web/packages/quadprog/index.html>
- Ulrich, J., Ryan, J. A. & Bennett, Ross. (2017). *Package xts*. Recuperado de <https://cran.r-project.org/web/packages/xts/index.html>
- Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y. & Zemlea, J. (2017). *Package corrplot*. Recuperado de <https://cran.r-project.org/web/packages/corrplot/index.html>
- Wickham, H. & Grolemund, G. (2016). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Canada: O'Reily Media, Inc.
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.
- Wicklam, H. (2017). *Package reshape2*. Recuperado de <https://cran.r-project.org/web/packages/reshape2/index.html>

XII Anexo

A. Propiedades de mínimos cuadrados

Teorema A.1 *Teorema de existencia y unicidad de mínimos cuadrados.*

1. Sea $\mathbf{X} \in \mathbb{R}^{n \times q}$ ($n \geq q$) y $\mathbf{y} \in \mathbb{R}^n$, un vector $\boldsymbol{\beta} \in \mathbb{R}^q$ es una solución de mínimos cuadrados de $\mathbf{X}\boldsymbol{\beta} = \mathbf{y}$ si y sólo si $\boldsymbol{\beta}$ satisface la ecuación normal

$$\mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = \mathbf{X}^T \mathbf{y}. \quad (\text{XII.1})$$

2. La solución de mínimos cuadrados, cuando existe, es única si y sólo si \mathbf{X} tiene rango completo.

Demostración.

(1) Se probará primero la condición de $\boldsymbol{\beta}$ como solución de mínimos cuadrados bajo la ecuación normal.

Sea $\mathbf{e}(\boldsymbol{\beta}) = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$ el vector residual, luego considere

$$\mathbf{e}(\boldsymbol{\alpha}) = \mathbf{y} - \mathbf{X}\boldsymbol{\alpha} = \mathbf{e}(\boldsymbol{\beta}) + \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\boldsymbol{\alpha} = \mathbf{e}(\boldsymbol{\beta}) + \mathbf{X}(\boldsymbol{\beta} - \boldsymbol{\alpha})$$

entonces

$$\|\mathbf{e}(\boldsymbol{\alpha})\|^2 = \|\mathbf{e}(\boldsymbol{\beta}) + \mathbf{X}(\boldsymbol{\beta} - \boldsymbol{\alpha})\|^2 = \|\mathbf{e}(\boldsymbol{\beta})\|^2 + 2(\boldsymbol{\beta} - \boldsymbol{\alpha})^T \mathbf{X}^T \mathbf{e}(\boldsymbol{\beta}) + \|\mathbf{X}(\boldsymbol{\beta} - \boldsymbol{\alpha})\|^2$$

ahora suponga que $\boldsymbol{\beta}$ satisface la ecuación normal, por lo que $\boldsymbol{\beta}$ es tal que $\mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = \mathbf{X}^T \mathbf{y}$ entonces

$$\mathbf{X}^T \mathbf{e}(\boldsymbol{\beta}) = \mathbf{X}^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{X}^T \mathbf{y} - \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = \mathbf{0}$$

por lo tanto

$$\|\mathbf{e}(\boldsymbol{\alpha})\|^2 = \|\mathbf{e}(\boldsymbol{\beta})\|^2 + \|(\boldsymbol{\beta} - \boldsymbol{\alpha})\|^2 \geq \|\mathbf{e}(\boldsymbol{\beta})\|^2$$

entonces $\boldsymbol{\beta}$ es una solución de mínimos cuadrados.

(2) Se probará por contradicción. Primero, suponga que \mathbf{X} es de rango completo, pero $\mathbf{X}^T \mathbf{X}$ es singular. Entonces $\mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = \mathbf{0}$ para algún $\boldsymbol{\beta}$ diferente de cero, lo que significa que $\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = 0$,

de ahí que $\mathbf{X}\boldsymbol{\beta} = 0$, implicando que \mathbf{X} es de rango deficiente, lo cual es una contradicción. Para el caso contrario, suponga que \mathbf{X} es de rango deiciente, pero $\mathbf{X}^T \mathbf{X}$ es no singular. Como \mathbf{X} es de rango deficiente, existe un $\boldsymbol{\beta}$ diferente de cero, tal que $\mathbf{X}\boldsymbol{\beta} = 0$, mostrando que $\mathbf{X}^T \mathbf{X}\boldsymbol{\beta} = 0$, esto implica que $\mathbf{X}^T \mathbf{X}$ es singular, lo cual es una contradicción. ■

B. Valor esperado y covarianza

Teorema B.1 Sea $\mathbf{x} = (x_i)$ un vector de variables aleatorias de $n \times 1$ y sea \mathbf{A} una matriz simétrica de $n \times n$. Si $E[\mathbf{x}] = \boldsymbol{\mu}$ y $Var[\mathbf{x}] = \boldsymbol{\Sigma} = (\sigma_{ij})$, entonces

$$E[\mathbf{x}^T \mathbf{A} \mathbf{x}] = tr(\mathbf{A} \boldsymbol{\Sigma}) + \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu}$$

La prueba puede consultarse en Seber y Lee (2003), en la página 9.

C. Estimación y distribución

Teorema C.1 Suponga que \mathbf{X} es de $n \times p$ y de rango p , tal que $\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$. Entonces se cumplen las siguientes propiedades.

1. \mathbf{H} y $\mathbf{I}_n - \mathbf{H}$ son simétricas e idempotentes.
2. $rank(\mathbf{I}_n - \mathbf{H}) = tr(\mathbf{I}_n - \mathbf{H}) = n - p$.
3. $\mathbf{H}\mathbf{X} = \mathbf{X}$.

La prueba puede consultarse en Seber y Lee (2003), en la página 40.

Teorema C.2 Sea $\mathbf{y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, \mathbf{C} una matriz de $m \times n$ de rango m , y \mathbf{d} un vector de $m \times 1$. Entonces $\mathbf{C}\mathbf{y} + \mathbf{d} \sim N_m(\mathbf{C}\boldsymbol{\mu} + \mathbf{d}, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T)$.

Véase la prueba en Seber y Lee (2003), en la página 20.

Teorema C.3 Suponga que $\mathbf{y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, donde $\boldsymbol{\Sigma}$ es positiva definida. Entonces

$$\mathbf{Q} = (\mathbf{y} - \boldsymbol{\mu})^T (\boldsymbol{\Sigma})^{-1} (\mathbf{y} - \boldsymbol{\mu}) \sim \chi_n^2$$

Consultar la prueba en Seber y Lee (2003) en la página 30.

Teorema C.4 Sea $\mathbf{y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, y defina $\mathbf{U} = \mathbf{A}\mathbf{y}$, $\mathbf{V} = \mathbf{B}\mathbf{y}$. Entonces \mathbf{U} y \mathbf{V} son independientes si y sólo si $Cov[\mathbf{U}, \mathbf{V}] = \mathbf{A}(\boldsymbol{\Sigma})\mathbf{B}^T = \mathbf{0}$

La prueba puede verse en Seber y Lee (2003) en la página 35.

D. Autocorrelaciones

Definición D.1 La función de autocovarianza se define como el segundo momento producto

$$\gamma_x(s, t) = E[(x_s - \mu_s)(x_t - \mu_t)]$$

para todo s y t . Generalmente se se escribe $\gamma_x(s, t)$ como $\gamma(s, t)$.

Notese que $\gamma(x)(s, t) = \gamma_x(t, s)$

Definición D.2 La función de autocorrelacion (ACF) se define como

$$\rho(s, t) = \frac{\gamma(s, t)}{\sqrt{\gamma(s, s)\gamma(t, t)}}$$

La ACF mide la linealidad de la predecibilidad de la serie al tiempo t, x_t , usando solamente el valor de x_s .

Definición D.3 La función cruzada de autocorrelacion (CCF) se define como

$$\rho_{xy}(s, t) = \frac{\gamma_{xy}(s, t)}{\sqrt{\gamma_x(s, s)\gamma_y(t, t)}}$$

La ACF mide la linealidad de la predecibilidad de la serie al tiempo t, x_t , usando solamente el valor de x_s .

Definición D.4 La función de autocovarianza de una serie de tiempo estacionario se define como

$$\gamma_x(h) = E[(x_{t+h} - \mu)(x_t - \mu)]$$

Definición D.5 La función de autocorrelacion de una serie de tiempo estacionario (ACF) se define como

$$\rho(h) = \frac{\gamma(t+h, t)}{\sqrt{\gamma(t+h, t+h)\gamma(t, t)}} = \frac{\gamma(h)}{\gamma(0)}$$

Definición D.6 Proceso Lineal. Un proceso lineal, x_t , se define como la combinación lineal de ruido blanco con variantes w_t , y se escribe como

$$x_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j}$$

donde los coeficientes ψ_j satisfacen

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty.$$

Cuadro XII.1: Constantes de determinación.

Sitio	R cuadrático	R ajustado	p-valor
Bahía Santiago	0.684	0.682	<2.2E-16
Bahía San Lucas	0.606	0.602	<2.2E-16
Centro WG	0.443	0.438	<2E-16
Panajachel	0.406	0.402	<2E-16
San Pedro - San Pablo	0.408	0.403	<2E-16

Cuadro XII.2: AIC.

Sitio	AIC inicial	AIC final
Bahía Santiago	-2476.04	-2481.12
Bahía San Lucas	-631.46	-632.45
Centro WG	-193.00	-193.00
Panajachel	-451.00	-454.00
San Pedro - San Pablo	-303.00	-303.00

Cuadro XII.3: Coeficientes lineales.

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 Sisueltos 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 (ug/L)	0.038	0.050	-0.003	0.012	-0.016
Fosfatos (ug/L)	0.004	-0.037	0.005	0.002	0.014
Fósforo total (ug/L)	-0.003	0.004	-0.016	-0.016	-0.004
Clorofila a(ug/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 XII.4: Pruebas.

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 XII.5: Relaciones.

Sitio	Relaciones
Bahía Santiago	(Transparencia: temp, nitratos, fosfatos, Chl- α)
Bahía San Lucas	(Transparencia: SDT, OD, SO, fosfatos)
Centro WG	(Transparencia: temp, cond, salinidad, OD, SO, nitratos, fosfatos, Chl- α)
Panajachel	(Transparencia: pH, salinidad, cond, fósforo total)
San Pedro	(Transparencia: Fósforo total, Chl- α , OD, SDT, salinidad, temp)

Cuadro XII.6: Elasticidades, primer enfoque.

Parámetro	Bahia Santiago	Bahia 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 Sisuelos 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 a(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 XII.7: Elasticidades, segundo enfoque.

Parámetro	Bahia Santiago	Bahia 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 Sisuelos 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 a(ug/L)	0	0	0	0	0.00009
PAR ($\mu Em^{-2}s^{-1}$)	0	0	0	0	0

Cuadro XII.8: Criterios.

Sitio	AIC	AICc	BIC
Bahía Santiago	903	904	945
Bahía San Lucas	1275	1277	1297
Centro WG	917.76	917.79	938.55
Panajachel	1712	1712	1743
San Pedro - San Pablo	1916	1916	1968

Cuadro XII.9: Modelos de pronóstico.

Sitio	Modelo	ar1	ar2	ar3	ar4	ma1	ma2	ma3	ma4	Media	Box-Pierce
Bahia Santiago	Arima(3,0,3)	-0.11	0.13	0.92	0	0.45	0.30	-0.54	0	4.98	0.92
Bahia San Lucas	Arima(1,0,1)	0.95	0	0	0	-0.08	0	0	0	6.64	0.97
Panajachel	Arima(3,0,2)	1.42	-0.82	0.36	0	-0.83	0.37	6.00	0	6.07	0.91
Panajachel	Arima(2,0,2)	0.37	0.16	0	0	0.16	-0.38	0	0	6.08	0.9
San Pedro	Arima(4,0,4)	0.09	1.68	0.06	-0.85	0.46	-1.27	-0.48	0.45	6.79	0.9

Figura XII.1: Estado del Lago, 2009.



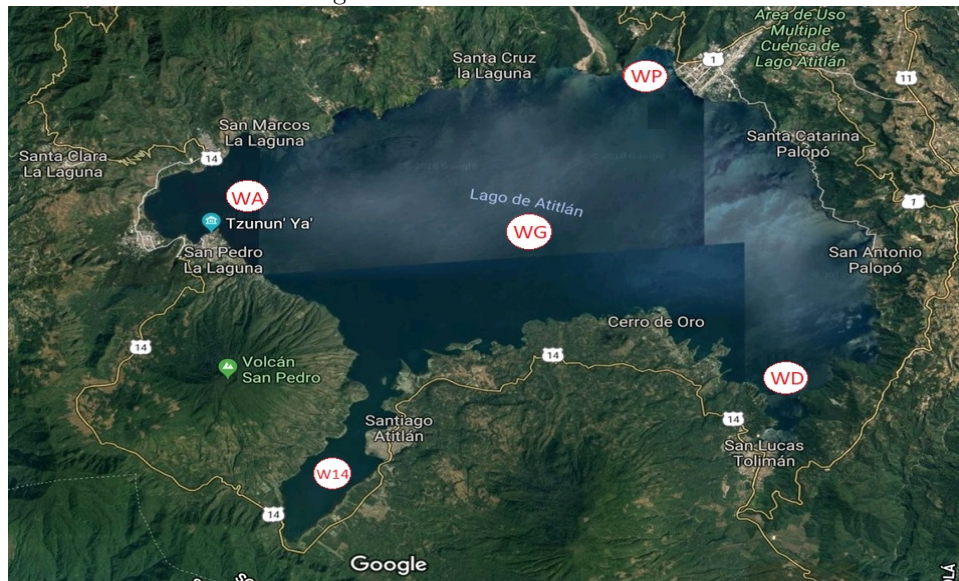
Fuente: ASTER/NASA, 2009.

Figura XII.2: Estado del Lago, 2015.



Fuente: LANDSAT/SERVIR&GLOBAL, 2015.

Figura XII.3: Sitios de estudio.



Fuente: Google imágenes/CNES/AIRBUS/LANDSAT, 2018.