



Universidad Central de Venezuela
Facultad de Ciencias
Licenciatura en Matemática

Estudio de la Temperatura y la Pluviosidad en el Lago de Valencia, utilizando Series de Tiempo

Autor: Br. Carla Susana Morillo Blandín.

Tutores: Dr. José Benito Hernández Chaudary.

Trabajo Especial de Grado
Presentada ante la ilustre
Universidad Central de Venezuela
Para optar al título de
Licenciado Matemática.

Caracas, Octubre 2014.

Nosotros, los abajo firmantes, designados por la Universidad Central de Venezuela como integrantes del Jurado Examinador del Trabajo Especial de Grado titulado "El estudio de la Temperatura y la Pluviosidad en el Lago de Valencia, utilizando Series de Tiempo", presentado por la Br. Carla Susana Morillo Blandín, titular de la Cédula de Identidad 18.249.898, certificamos que este trabajo cumple con los requisitos exigidos por nuestra Magna Casa de Estudios para optar al título de Licenciado en Matemática.

Dr. José Benito Hernández
Tutor

Dra. Mairene Colina
Jurado

Dr. José Rafael León
Jurado

Índice general

Contenido	I
Introducción	4
1. Conceptos básicos de meteorología	7
1.1. Temperatura	11
1.1.1. Clases de temperaturas usadas en meteorología y climatología	12
1.1.2. Datos de temperatura más usados	13
1.2. Pluviosidad	14
1.2.1. Teoría de Bergeron-Findeisen	14
1.2.2. Procesos de colisión y coalescencia	14
1.2.3. Tipos de precipitación	15
1.2.4. Clases de Precipitación	15
1.2.5. Instrumentos de medición de la pluviosidad	16
2. Series de Tiempo	17
2.1. Series de Tiempo Estacionarias	20
2.2. Estimación de Correlación	23
2.3. Regresión de Series de Tiempo	24
2.4. Análisis Exploratorio de Datos	27
2.5. Suavizando Series de Tiempo	29
2.6. Autocorrelación y Autocorelación Parcial	32

<i>ÍNDICE GENERAL</i>	1
2.7. Predicciones	34
2.8. Modelos ARIMA	37
2.8.1. Construcción de un modelo ARIMA	37
3. Análisis de Temperatura y Pluviosidad en el Lago de Valencia	39
3.1. Descripción	39
3.2. Comportamiento General de la Temperatura y de la Lluvia	41
3.2.1. Estadística Descriptiva	41
3.2.2. Comparaciones de Temperatura y Pluviosidad	45
3.3. Función de Correlación y Autocorrelación	48
3.4. Comparaciones entre a Temperatura y la Pluviosidad, entre las Estaciones Sucre y Tacarigua . .	51
3.5. Modelo Arima	54
Conclusión	70
Bibliografía	72

A Dios Creador, quien me dio la vida
"Todo lo puedo en Cristo que me fortalece. "
Filipenses 4:13

*

Este trabajo en primer lugar, lo dedico a Dios todopoderoso por darme la oportunidad de llegar hasta donde estoy, por darme la fortaleza para pasar todas las dificultades y por no dejar de bendecirme cada día de mi vida.

A mi madre, por ser tan especial y apoyarme en todo momento, por todos sus esfuerzos y sacrificios para que yo pueda alcanzar mis sueños. Gracias por ser como eres.

A mi padre, por siempre sentirse orgulloso de mi, por apoyarme en todos los momentos difíciles y siempre tener los consejos indicados.

A mi hermana, por ser mi ÚNICA HERMANA EN LA VIDA, gracias por siempre saber escucharme y hacerme reír cuando más lo necesito.

A mis sobrinos, por ser lo más hermoso que veo todos los días al despertarme. Este logro es un ejemplo para ustedes.

A Toda mi familia, por estar siempre conmigo apoyandome cuando más lo he necesitado. Gracias por estar ahí.

A mi tutor Dr. José B., por tener tanta paciencia conmigo y contestar siempre mis dudas a cualquier hora, con la mejor disposición y comprensión. Gracias por creer en mi.

A Dr. Armando Ramirez por ser la primera persona en creer en mi y animarme a conseguir mi meta.

A mis amigos, gracias por todos los buenos y malos momentos que pasamos. Han sido las mejores experiencias de vida.

A la Universidad Central de Venezuela, por educarme y brindarme la oportunidad de culminar mis estudios de Pregrado.

Introducción

El Lago de Valencia o Lago de Tacarigua es el segundo lago en importancia de Venezuela después del Lago de Maracaibo y es el cuerpo de agua dulce sin desagüe al mar más grande de Venezuela. Se encuentra emplazado en una fosa tectónica conocida como Graben de Valencia que se encuentra entre la cordillera de la Costa y la serranía del Interior. La cuenca es de tipo endorreica y cubre 3.150 km^2 , es decir, 0,3 (por ciento) del país. Este lago es el fenómeno hidrográfico más importante del centro del país, tiene una longitud máxima de 30 km, una anchura máxima que alcanza los 20 km, una profundidad de 39 metros y una media de 21 metros. En general, la cuenca del lago de Valencia es también conocida popularmente como los **Valles de Aragua**.

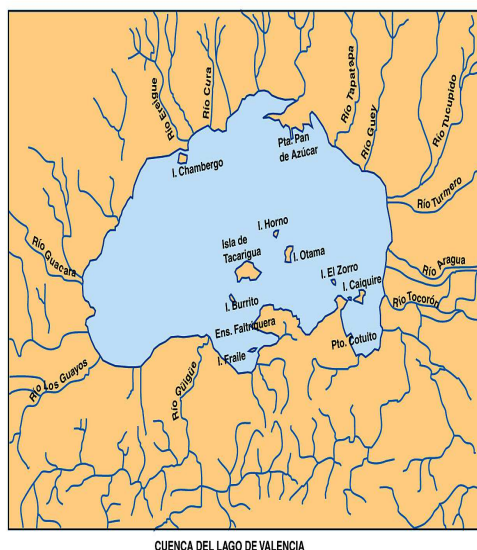


Figura 1: Lago de Valencia

En la actualidad el lago tiene una extensión de 344 km^2 . En sus orillas se levantan dos de las principales ciudades del país como son: Valencia y Maracay y otros centros urbanos importantes como los son Mariara, San Joaquín, Güigüe, Guacara y Palo Negro.

El lago de Valencia se sitúa entre Aragua y Carabobo. El estado Aragua comprende la porción oriental de

la depresión que rodea al lago, se extiende al este formando un valle de estrecha cuña. El estado Carabobo comprende la porción central y occidental de la depresión y está rodeado por los campos de Carabobo.

Geológicamente, el lago es bastante joven, aproximadamente de una edad pre-Pleistocena o Pleistoceno medio y su origen se debe a un hundimiento de bloques de la corteza terrestre. A lo largo de su historia el nivel de sus aguas ha fluctuado significativamente. La desaparición del lago de Valencia, al igual que la de todos los lagos del mundo, es un hecho geológico, pero el proceso debió ser casi imperceptible, lo grave fue que el propio hombre lo aceleró durante el siglo XIX y gran parte del siglo XX. La disminución de sus aguas en aquellos años se debió a múltiples causas entre ellas:

- La evaporación excesiva por las altas temperaturas locales
- La deforestación
- El fenómeno de Piratería Fluvial, captura del río Tuy (que antes drenaba hacia el lago de Valencia) por el cauce actual de dicho río que drena ahora hacia el Caribe, lo que dio origen al desecamiento progresivo de sus aguas y a la formación de extensas planicies de sedimentos lacustres que convierten a sus riberas en importantes centros agrícolas. Ya Alexander von Humboldt mencionaba la creencia de que el lago podía evaporarse.

Para la década de 1970 el lago resistió prolongados períodos de sequías. Esto unido a la utilización de los afluentes en labores agrícolas y la creciente actividad industrial en la zona ocasionaron un descenso progresivo en el nivel de las aguas del lago, que para el año 1980 su cota llegó a tocar los 401 m.s.n.m. Es de hacer notar que el lago de Valencia antes de sufrir todos estos cambios causados tanto por la naturaleza como por el hombre, vertía el exceso de sus aguas de la época de lluvias a la cuenca del río Orinoco a través de un río emisario que hoy en día es el río Paito, afluente del río Pao que a su vez, es un afluente del Orinoco. El río Paito podría considerarse, en realidad, como la parte superior o cabeceras del río Pao.

Sin embargo, en la actualidad se está agravando el problema del aumento del nivel del lago, debido al transvase de las aguas de la cuenca del Orinoco (embalses de Pao Cachinche y Pao La Balsa, en el río Pao) que se hace hacia la cuenca endorreica del Lago de Valencia, sin haber hecho los canales de drenaje para las aguas sobrantes con el fin de sacarlas de la cuenca antes de que lleguen a un nivel inferior al de salida de la cuenca por gravedad, por el antiguo curso del río Cabriales. Como consecuencia de ello, los niveles de contaminación de sus aguas también han venido aumentando debido a la importante industrialización de las ciudades de Maracay y Valencia, cuyas aguas servidas son vertidas directamente al lago, en gran parte, sin un tratamiento previo, lo que hace poco atractivo el desarrollo de actividades turísticas o recreacionales. Para paliar la situa-

ción hace pocos años el río Cabriales que fue desviado hacia el lago en 1979, fue desviado nuevamente, con esto se espera que descendan el nivel de las aguas. Actualmente están en ejecución los proyectos que permitirán el saneamiento de la cuenca del Lago mediante una serie de obras como la adecuación de las plantas de tratamiento de Los Guayos y otras. En el capítulo I, se va a estudiar la teoría de la Meteorología para poder conocer más a fondo la terminología propia de este tema, haciendo enfoque en la temperatura y la pluviosidad, conociendo sus propiedades, instrumentos y mecanismos de medición e información general que ayudará a manejar información básica sobre la meteorología y climatología.

En el capítulo II, se conocerán definiciones, propiedades, proposiciones y ejemplos de las Series de Tiempo, donde se identifica los procedimientos que se llevarán a cabo en el estudio de la información recolectada en el Lago de Valencia, para poder ajustar un modelo adecuado para poder predecir los valores futuros de una serie temporal y poder detectar la presencia o ausencia de correlaciones de largo alcance en una serie. Por último, el tercer capítulo será empleado para mostrar los resultados y procesos de la aplicación del modelo, de series temporales que servirá para poder brindar conclusiones sobre el comportamiento de la pluviosidad y la temperatura en el Lago de Valencia. También, se realizará una estadística descriptiva con los datos de cada mes, y se realizarán comparaciones entre las observaciones por Estación.

Conceptos básicos de meteorología

Para poder entender, los conceptos propios de meteorología y climatología, se debe conocer cierta información. Debe destacarse que parte de la investigación, que está llevando a cabo es en base a la pluviosidad y temperatura en el Lago de Valencia. Se va a empezar por indicar la siguiente definición formal de meteorología.

Definición 1.0.1 *La **meteorología** es la ciencia que estudia el tiempo y el clima. Tiene como objetivo general el entendimiento de los procesos físicos y químicos que determinaran el estado dinámico de la atmósfera a escalas espacial y temporal.*

Tanto la investigación en el campo teórico como en el aplicado de la meteorología se ha demostrado la necesidad de conocer mejor la atmósfera y sus relaciones con la superficie de la tierra para impulsar los principales campos del desarrollo económico. El conocimiento del tiempo en cada país o áreas dentro de un país depende directamente de sus servicios meteorológicos eficientes, basados en una red de estaciones con instrumentos precisos para obtener los datos que forman el material esencial de la meteorología. La meteorología como ciencia de la atmósfera, es sinónimo de ciencia del tiempo. Siendo el mismo, la característica o estado de la atmósfera en cualquier lugar y momento determinados.

Definición 1.0.2 *El **tiempo** es el estado de la atmósfera en un determinado momento, para esto tomaremos en cuenta la humedad absoluta y humedad relativa, la temperatura y la presión en un lugar y momento determinado.*

El término tiempo no se refiere al cronológico sino al estado meteorológico de la atmósfera en un determinado momento. La variación en un mismo lugar y de un lugar a otro es el rasgo más resaltante del tiempo meteorológico y cada uno de sus elementos como: la radiación solar, temperatura, presión, viento, humedad y

precipitación, varía a consecuencia del movimiento incesante de la atmósfera con su traslado de energía de un lugar a otro.

Definición 1.0.3 *La definición de **clima** viene dada por la combinación de una serie de fenómenos atmosféricos, es decir, es una sucesión periódica de tipos de tiempo. Los principales elementos que definen el clima son: temperatura, humedad, pluviosidad, nubosidad y viento.*

La distinción entre el tiempo y clima está dada por sus respectivos conceptos, pero además de esto si, por ejemplo, en una estación meteorológica se lleva registro de humedad para cada hora del día, se puede determinar la media diaria al promediar esos 24 valores. Si se tienen las medias diarias durante un mes, se determina la media mensual y si es para un año la media anual, la cual ya representa cierta idea de la variación de la humedad en este período. Pero la humedad de ese año será distinta a la del próximo y la de un año a otro, necesitándose un período de observación suficientemente largo para que la media de los valores observados varíe en un pequeño grado o no varíe significativamente y así tener una expresión de la humedad media del lugar. Al hacer lo mismo con la temperatura, precipitación, viento y demás elementos de estudio, se logra una idea más completa de las condiciones medias de la atmósfera para el lugar en consideración y es lo que constituye el clima de ese lugar.

Los términos expresados anteriormente están unidos por sus efectos sobre la corteza terrestre. Un área puede tener distintos estados meteorológicos pero un sólo clima, al cambiar de espacio estos integrantes presentarán un cambio según sea dado el caso.

En la meteorología y la climatología se puede distinguir entre elementos y factores. Los elementos del tiempo y clima son las variaciones y características físicas de la atmósfera. Ahora bien, los factores que definen al clima y al tiempo son los fenómenos meteorológicos o geográficos que actuando sobre los elementos producen la variación del tiempo o del clima. Como elementos se pueden considerar los siguientes: radiación, temperatura, presión, viento, humedad, nubosidad, insolación, pluviosidad y evaporación. Por otro lado, los factores son: la latitud, la distribución de tierra y aguas, la proximidad o lejanía del mar, el relieve, las corrientes oceánicas, las masas de aire y frentes, los centros semipermanentes de altas y bajas presiones, las tormentas, la vegetación.

En este trabajo de investigación se realizara especial enfoque en los efectos de los elementos (específicamente temperatura y pluviosidad) sobre los factores del clima y tiempo.

Elementos del clima

- La energía transferida por el Sol a la Tierra es lo que actualmente se conoce como **radiación**. La energía radiante viaja a través del espacio en forma de ondas con una cantidad específica de energía. Dependiendo de la energía de estas ondas obtienen la clasificación de espectro electromagnético.
- Se llama **temperatura** a una magnitud relacionada con la rapidez del movimiento de las partículas que constituyen la materia, lo que indica que a mayor agitación de las partículas mayor será la temperatura.
- La fuerza que ejerce el aire sobre todos los cuerpos es a lo que se denomina **presión**. Esta depende directamente de la altitud y además está íntimamente ligada a otras variables como la situación geográfica, la temperatura, la humedad y las condiciones meteorológicas.
- El movimiento de aire desde una zona a otra, en sentido horizontal, es lo que se denomina como **viento**, mientras que el movimiento en sentido vertical se denomina corriente aérea .

Más adelante se profundizará en los conceptos de temperatura y pluviosidad

- La **humedad** es la cantidad de vapor de agua que contiene el aire. Esta cantidad depende de distintos factores ambientales, como por ejemplo las lluvias recientes, la cercanía al mar, entre otros factores.
- La fracción de cielo cubierto de nubes de un cierto género, de una cierta especie, de una cierta variedad, de una cierta capa o de una combinación de nubes se conoce como **nubosidad**.
- La caída y llegada al suelo de gotas de agua o de partículas de hielo, es lo que se llama **precipitación**. La precipitación caída en un determinado lugar se puede expresar tanto en milímetros (mm) de altura, que es la altura de la capa de agua que cubriría un suelo horizontal sin que el agua se filtrase ni evaporase, como en $l./m^2$ que es la cantidad de agua recogida en una superficie horizontal de un metro cuadrado. Ambas unidades son equivalentes
- La **evaporación** es un proceso físico donde el agua se transforma en vapor de agua desde una superficie líquida o la humedad de los suelos.

Definición 1.0.4 *La atmósfera es la capa gaseosa que envuelve a la Tierra, que la acompaña en sus movimientos de rotación y traslación y que además se adhiere a la Tierra por la acción de la gravedad. Está formada por una mezcla de gases tales como: nitrógeno (78,8 %), oxígeno(20,95 %), argón (0.93 %), dióxido de carbono (0.03 %) y vapor de agua (1 %).*

Su espesor no es uniforme debido a la mayor atracción en los polos y mayor fuerza centrífuga en el Ecuador, resultando mayor espesor en el Ecuador y menor en los polos y forma elipsoidal en vez de completamente esférica. La atmósfera actúa como una pantalla protectora interceptando los rayos ultravioleta y cósmicos, también protege parcialmente de los meteoritos y polvo meteórico. Esta inmensa masa de aire es la responsable de la distribución mas bien moderada de la temperatura en la mayor parte de la superficie de la tierra.

Tradicionalmente se consideraron cuatro capas mas o menos concéntricas, desde la superficie hacia arriba: la tropósfera, estratosfera, ionosfera y la exosfera. Estos nombres se establecieron por los rasgos mas importantes de las capas respectivas. Es corriente hablar de baja, media y alta atmósfera. Estos términos, diferentes a los anteriores, son menos precisos, así, la troposfera sería equivalente a la baja, estratosfera y mesosfera a la media y la ionosfera y exosfera a la alta.

Estructura de la atmósfera

En la atmósfera, además de la densidad y la composición del aire, también la temperatura varía con la altura, a partir de esta variación térmica la atmósfera puede dividirse en capas, de la siguiente forma:

- **Troposfera** es la capa que está en contacto con la superficie terrestre donde se desarrolla la vida de la mayoría de los fenómenos meteorológicos. Esta, es la capa de menor espesor con 10 Km en los polos y entre 18 y 20 Km en el Ecuador. La tropopausa es conocida como la sección superior de la troposfera, cuya altura depende directamente de la temperatura y la presión.
 - La **Estratosfera** en esta capa la temperatura se puede incrementar hasta alcanzar aproximadamente los -10°C a unos 50 Km de altitud, en esta capa se localiza la mayor concentración de ozono, el tope de esta capa se denomina estratopausa.
 - La **Mesosfera** es la capa de la atmósfera que abarca desde los 50 Km a 80 Km, al final de los cuales se encuentra la mesopausa. En esta capa la temperatura desciende con la altura y puede llegar hasta los -140°C .
 - La **Ionosfera** es la capa de la atmósfera, se extiende entre 80 y 110 Km; la temperatura se incrementa con la altura y logra alcanzar alrededor de 1500°C . El aumento térmico se debe a la absorción de la radiación solar por los gases del aire, aquí los gases presentan una densidad muy baja y se encuentran ionizados.
 - La **Exosfera** es la capa más externa de la atmósfera. En esta capa no se cumplen las leyes de los gases pero de igual forma se realizan interacciones con sus átomos y moléculas. Esta aproximadamente a 500 Km de la superficie terrestre.
-

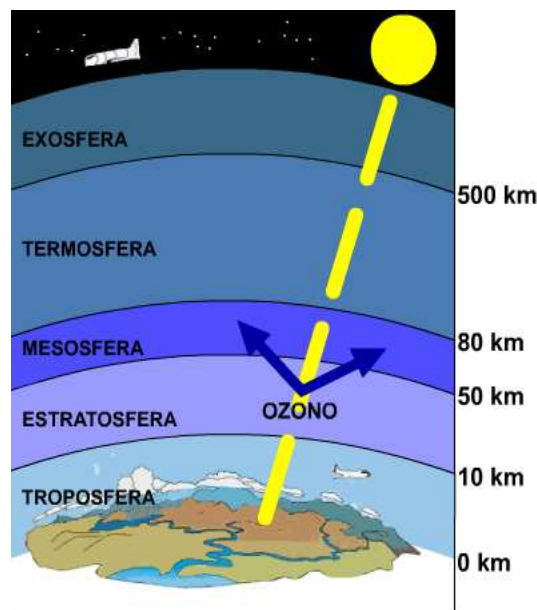


Figura 1.1: Capas de la Atmósfera

1.1. Temperatura

La **temperatura** es un elemento del tiempo y clima, que es utilizada para clasificar o diferenciar las zonas térmicas, es decir, tórrida, templada o fría. Este elemento es de suma importancia para la meteorología pues es una de las magnitudes más utilizadas para definir el estado de la atmósfera. Como se explicó anteriormente la temperatura es una magnitud relacionada con la rapidez del movimiento de las partículas que constituyen la materia, a mayor agitación de las partículas mayor será la temperatura.

Para medir la temperatura, se deben verificar las propiedades de la materia que se ven alteradas cuando ésta cambia. Las escalas de medición de la temperatura más comunes son: Celcius, Fahrenheit, Kelvin y Rankin. Estas escalas tienen el mismo principio pero tienen distintos valores numéricos de medición que dependen directamente de las necesidades para las que fueron creadas.

Todo cuerpo tiene una energía en forma de calor y éste es un estado latente de la energía inherente a la materia. Esa energía calórica o, simplemente calor, se expresa en calorías o kilocalorías. La incomodidad de estas unidades en meteorología y climatología hizo un índice más práctico para medir ese estado de energía calórica con mayor facilidad. Los términos calor y temperatura por estar relacionados, pudieran causar confusión. La diferencia entre ambos, se encuentra si se profundiza un poco en la constitución de los cuerpos para llegar al

estado de movimiento de sus moléculas. Es la energía cinética de las moléculas la que produce el calor; por consiguiente, el calor de un cuerpo será la suma total de las energías cinéticas de sus moléculas, lo que significa que un cuerpo tendrá mayor calor, cuanto más energía cinética tengan sus moléculas y cuanto mayor sea la masa.

1.1.1. Clases de temperaturas usadas en meteorología y climatología

Temperatura del aire es la temperatura actual o la de un momento dado, prescindiendo de la radiación calorífica de los objetos que rodean el ambiente y de los efectos de la humedad relativa y movimientos del aire. Esta temperatura se subdivide en:

- **Temperatura Actual:** es la lectura en un instante cualquiera obtenida mediante un termómetro. Estas lecturas tienen horas fijas, los termómetros deben estar en buen estado y correctamente instalados, es decir, deben estar a la sombra sin ningún contacto directo al sol u otro cuerpo generador de calor, en una garita meteorológica a una altura de 1,50 m sobre el suelo.
- **Temperatura del Termómetro Húmedo:** para este tipo de temperatura del aire es medida por un termómetro normal donde el bulbo ha sido cubierto en una muselina humedecida y se coloca bajo una corriente de aire. Esto se utiliza para hallar la humedad relativa.
- **Temperatura Virtual:** es la temperatura que tiene aire húmedo cuando se desatura y mantiene la densidad y la presión.
- **Temperatura Potencial:** θ es la temperatura que tiene una porción de aire cuando es llevada adiabáticamente¹ al nivel de 1.000 milibares.
- **Temperatura a Punto de Rocío:** esta temperatura es la que se obtiene al momento de empezar la condensación.
- **Temperatura del Nivel de Condensación:** es la temperatura que se toma en el momento que empieza la condensación, la temperatura en la bases de las nubes es lo que indica la información.
- **Temperatura del Suelo:** este tipo de temperatura es utilizado para verificar información en áreas de estudio del suelo, donde se utilizan termómetros especiales llamados **geotermómetros** y se ubican en el suelo con profundidades que varían dependiendo de la necesidad del estudio que se realice.

¹El término adiabático se refiere a la curva de variación del volumen con la presión de un gas, efectuada sin intercambio de calor con el exterior.

- **Temperatura del Agua:** esta temperatura es medida con un termómetro tradicional. En este ambiente la temperatura es mayormente constante. En general, se mide la temperatura en la superficie del agua y en diferentes profundidades, puesto que dependiendo de la profundidad varía.

1.1.2. Datos de temperatura más usados

- **Máximas y mínimas.** Tienen por objeto indicar los valores térmicos extremos del aire ocurridos en un período dado. Su medición se obtiene por los termómetros de máxima y mínima, respectivamente. En un período largo, la máxima y mínima absoluta son los valores más altos o más bajos registrados.
- **Media diaria.** La temperatura media diaria en su más estricto sentido, es el promedio de las 24 lecturas practicadas de hora en hora durante el día. Cuando no es posible efectuar 24 observaciones las lecturas se hacen a ciertas horas fijas, cuyo promedio difiera lo menos posible de la media verdadera de las 24 observaciones horarias.
- **Media mensual.** Se obtiene del valor sumando las temperaturas medias diarias y dividiendo por los días del mes.
- **Anual.** Se obtiene sumando las temperaturas medias mensuales y dividiéndolo entre los doce meses del año.

Para poder medir la temperatura, existe una variedad de termómetros, su uso depende de que tipo de estudio se quiera llevar a cabo. Se tiene la existencia del **Termómetro Común** que es utilizado para verificar la temperatura del aire en un momento dado. También existe el **Termómetro de Máxima y de Mínima** que son utilizados para registrar las temperaturas máximas y mínimas en un lugar establecido. Los termómetros más usados en meteorología utilizan líquido como elemento sensible, especialmente el mercurio y el alcohol. El mercurio tiene ciertas ventajas como: responder rápidamente a los cambios de calor, se distingue fácilmente, no mancha y soporta altas temperaturas. El alcohol se usa cuando las temperaturas a medir son bajas y requiere un líquido que se congele tan bajo como sea posible, el alcohol lo hace a -114°C . La temperatura en cualquier lugar depende directamente de factores externos meteorológicos como la radiación solar, la nubosidad, la altitud.

- **Termómetro común.** La medición de la temperatura del aire en un momento dado, se usa este aparato, con mercurio. Es un dispositivo simple construido por un depósito de mercurio el cual es conocido por bulbo, continuando con un tubo capilar por donde asciende y desciende el mercurio a dilatarse o contraerse en el bulbo por efecto de los cambios del calor.
-

- **Termómetros de usos especiales.** Además de los instrumentos tradicionales de uso en las estaciones meteorológicas, se han fabricado termómetros con diferentes propósitos como:
 - **El globo termómetro.** Consiste en una esfera de cobre pintada de negro en cuyo interior se coloca el bulbo de un termómetro normal.
 - **Termómetros de resistencia de platino.** se utilizan en las estaciones automáticas, conjuntamente con potenciómetros, pero el costo y el consumo eléctrico limita su uso más ampliamente.
 - **Indicador de diferencia de temperatura mediante sensor remoto.** este instrumento permite determinar la diferencia térmica entre dos puntos en el aire, agua o suelo, muy útil para investigaciones ambientales.

1.2. Pluviosidad

Se conoce por pluviosidad, como la precipitación de gotas de agua (nieve, granizo, pedrisco, llovizna) al suelo, que se encontraban en las nubes. Para que pueda haber una precipitación de agua, deben haberse dado ciertas condiciones en la atmósfera, como el vapor de agua condensado en las nubes y tener el suficiente vapor o partículas de agua para que por efecto del peso, esta se precipite. Existen dos procesos que explican el aumento de tamaño de las gotas.

1.2.1. Teoría de Bergeron-Findeisen

La presión de vapor de saturación es menor en el hielo que en el agua. Si en una nube aparecen gotas de hielo el vapor de agua tiende a depositarse sobre el hielo. Una vez formados los minúsculos cristales de hielo, estos crecen rápidamente por condensación. Los cristales de hielo tienden a astillarse por efecto de las corrientes de aire lo que incrementa su número. Los cristales de hielo pueden unirse debido a su forma dendrítica². Cuando la velocidad de la caída del cielo sobrepasa las corrientes de aire ascendente el copo de nieve cae y se transforma en lluvia si atraviesa una capa de aire con temperatura mayor de 0°C suficientemente espesa.

1.2.2. Procesos de colisión y coalescencia

El movimiento de las gotas de agua en las nubes origina colisiones entre las mismas que contribuyen a unir las y por tanto a aumentar su tamaño medio. Por otra parte la velocidad de caída de las gotas de agua

²En el ámbito de la metalurgia, la geología y la cristalografía una dendrita es una estructura con ramificaciones repetitivas características de procesos de crecimiento de cristales.

tiene una velocidad límite proporcional a su diámetro, de este modo las gotas mayores caen más rápido que las pequeñas arrastrándolas. Las gotas pequeñas son apartadas a un lado: la turbulencia y la electrificación en las nubes cumuliformes tiende a incrementar el proceso de coalescencia.

1.2.3. Tipos de precipitación

Dependiendo del tipo de proceso que desencadena el ascenso del aire para enfriarse, condensar y provocar la precipitación, estas se pueden clasificar de la siguiente forma:

- **Lluvias ciclónicas o frontales.** Se genera por el ascenso de masas de aire por convergencia, de distintas temperaturas en un sistema de bajas presiones. La masa de aire frío penetra por debajo de la masa de aire cálido y la eleva con lo que enfría adiabáticamente generando las precipitaciones.
- **Lluvias por convergencia.** Se originan en el acceso del aire por convergencia de dos masas de aire en la Zona de Convergencia Intertropical. Las lluvias también se originan por convergencia de vientos al hacer ascender en su encuentro gran cantidad de aire húmedo.
- **Lluvias Conectivas.** Este tipo de lluvia se genera al ascender el aire que ha sido fuertemente calentado, generalmente en áreas de gran insolación y humedad. Este aire asciende formando nubes de tipo cúmulo
- **Lluvias Orográficas.** En zonas de relieve montañoso es frecuente este tipo de lluvias, por la acción de vientos húmedos. La cantidad de pluviosidad generada por efectos orográficos es menor que la originada por convecciones y frentes pero es de mayor constancia sobre el mismo lugar.

1.2.4. Clases de Precipitación

La forma en que se produce la precipitación, junto a la estructura térmica de la capa de aire situada bajo la atmósfera condicionará la forma de la precipitación. Las formas de precipitación son conocidas como hidrométoros y se dividen de la siguiente forma:

1. Precipitaciones Verticales

- **Lluvia** es una precipitación de partículas de agua con un diámetro mayor de 0,5mm.
 - **Llovizna.** Es una precipitación de partículas de agua con un diámetro menor de 0,5mm. Sus gotas son uniformes y unidas entre sí.
-

- **Nieve.** es una precipitación de partículas de agua solidificadas, son granos blancos y opacos de hielo. Dependiendo de la longitud de la nieve se tienen las siguientes clasificaciones: granos de nieve de 2 a 5 mm y la nieve granulada menor a 1 mm.
- **Granizo.** Esta precipitación está constituida por bolas de nieve de forma irregular y su tamaño varía entre 5 a 50 mm.

Dentro de las lluvias se incluye chaparrón, chubasco o popularmente conocido como **palo de agua** que se caracteriza por tener gotas grandes, corta duración y alta intensidad. Las nubes del trópico, como es el caso de Venezuela tienen más altas temperaturas y mayor contenido de agua que las nubes de latitudes medias.

2. Precipitaciones Horizontales

- **Rocío.** Este fenómeno se origina por la humedad en el aire se condensa en forma de gotas de agua por la disminución brusca de la temperatura.
- **Escarcha.** Es la capa de hielo cristalizado que se forma sobre superficies expuestas a la intemperie y se enfrían lo suficiente para provocar la deposición directa del vapor de agua contenido en aire.

1.2.5. Instrumentos de medición de la pluviosidad

La cantidad de agua caída en una precipitación es expresada en milímetros y se mide con la ayuda de pluviómetros para una medición directa y pluviógrafos para una medición continua.

- **Pluviómetro de Cántaro.** Es un depósito cilíndrico cuyo embudo recibe directamente la lluvia en una superficie de 200cm^2 y termina en su parte inferior en un embudo que lleva el agua hasta el depósito donde se conserva para luego ser medida.
 - **Pluviógrafo de Balancín.** Su embudo receptor es de 1000cm^2 , por donde la lluvia pasa a un depósito dividido en dos partes y funciona como un balancín, al llenarse una de las partes con 0,2m de lluvia se vuelca en un depósito mientras se llena el resto del aparato.
 - **Pluviógrafo de Sifón.** El aparato receptor es de una superficie de 200cm^2 , la lluvia caída pasa un a través de un recipiente cilíndrico mediante un tubo, con un flotador que asciende a medida que va cayendo la lluvia .
-

Capítulo 2

Series de Tiempo

En este capítulo, se estudiarán los conceptos que se manejarán durante el análisis de los datos de temperatura y pluviosidad en el Lago de Valencia, necesarios para poder tener un conocimiento más amplio acerca de lo que significan las Series Temporales.

Una serie de tiempo es una secuencia de datos, medidos en determinado momento y ordenados cronológicamente. Los datos pueden tener registros con intervalos de tiempo iguales como por ejemplo las mediciones de las llamadas que ingresan en un centro telefónico en 6 meses o las mediciones de peso de un paciente, en un consultorio médico. A continuación presentaremos una definición formal.

Definición 2.0.1 Una *serie de tiempo* es un conjunto de observaciones x_t , registradas en un tiempo específico t .

Definición 2.0.2 Un *modelo de series de tiempo* para los datos observados x_t , es una especificación de una distribución conjunta de una sucesión de variables aleatorias X_t de las cuales x_t es una realización.

El principal objetivo de las series de tiempo es hacer proyecciones o pronósticos sobre una actividad futura, suponiendo estables las condiciones y variaciones registradas hasta la fecha, lo cual permite planear y tomar decisiones a corto o largo plazo. Después, con base en esa situación ideal, que supone que los factores que influyeron en la serie en el pasado lo continuarán haciendo en el futuro, se analizan las tendencias pasadas y el comportamiento de las actividades bajo la influencia de ellas.

Definición 2.0.3 Un *proceso estocástico*, es una familia de variables aleatorias indexadas $x(\omega, t)$ o $x_t(\omega)$, donde t pertenece a un conjunto de índices T y ω pertenece a un espacio muestral Ω . Si $t = t^*$ fijo, $x(\omega, t^*)$ es una variable

aleatoria. Si $\omega = \omega^*$ fijo, $x(\omega^*, t)$ es una función de t y se llama una realización del proceso. Una **serie de tiempo** es la realización de un proceso estocástico.

Un ejemplo es la temperatura en Madrid, aumenta durante el día y disminuye en la noche, aumenta en el verano y desciende mucho en el invierno; su variación es parcialmente determinística y parcialmente aleatoria.

Definición 2.0.4 La función de media se define como

$$\mu_{xt} = \mathbb{E}(x_t) = \int_{-\infty}^{\infty} x f_t(x), dx,$$

donde f_t es la función de densidad de la variable X_t . En caso de que exista, donde $\mathbb{E}$ denota el operador de esperanza.

Ejemplo 2.0.1 *Función de media de una señal con ruido.* Se va a suponer que los datos observados, han sido generados por una señal fija ondulatoria superpuesta por un proceso de ruido de media 0, obteniendo modelos de señal aditivo como se muestra en el siguiente modelo

$$x_t = 2\cos(2\pi t/50 + 0,6\pi) + \omega_t, \text{ para } t = 1, 2, \dots, 500.$$

La señal en la ecuación anterior es una función fija de tiempo entonces tenemos

$$\begin{aligned} \mu_{xt} &= \mathbb{E}(x_t) = \mathbb{E}[2\cos(2\pi t/50 + 0,6\pi) + \omega_t] \\ &= 2\cos(2\pi t/50 + 0,6\pi) + E(\omega_t) \\ &= 2\cos(2\pi t/50 + 0,6\pi) \end{aligned}$$

donde la función de media es la onda del coseno.

Definición 2.0.5 La función de autocovarianza es definida como producto del segundo momento

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

para todo t y s . Cuando no haya confusión en la existencia sobre a la que serie nos referimos, escribiremos $\gamma_x(s, t) = \gamma(s, t)$. Note que $\gamma_x(s, t) = \gamma_x(t, s)$ para todos los puntos s y t . La función de autocovarianza mide la dependencia lineal entre dos puntos de la misma serie en diferentes tiempos. La autocovarianza es el promedio de los productos relacionados con la densidad conjunta $f(x_s, x_t)$. Es claro que, para $s=t$, la autocovarianza se reduce a la varianza (en el caso finito), dado que

$$\gamma_x(t, t) = \mathbb{E}[(x_t - \mu_t)^2]$$

Ejemplo 2.0.2 Autocovarianza de un ruido blanco.

La serie de ruido blanco w_t , tiene $\mathbb{E}(w_t) = 0$ y

$$\gamma_w(s, t) = \mathbb{E}(w_s, w_t) = \begin{cases} \sigma_w^2, & s = t \\ 0, & s \neq t \end{cases}$$

donde, en este ejemplo, $\sigma_w^2 = 1$. Note que w_s y w_t son no-correlacionados para $s \neq t$, por lo que se tiene $\mathbb{E}(w_s w_t) = \mathbb{E}(w_s)\mathbb{E}(w_t) = 0$ porque los valores medios de las variables de ruido blanco son cero.

Definición 2.0.6 La función de autocorrelación (ACF) se define como

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

La ACF mide la predictibilidad lineal de una serie de tiempo en un tiempo t . Se puede predecir x_t exactamente de x_s a través de la relación lineal $x_t = \beta_0 + \beta_1 x_s$ entonces la correlación será 1 cuando $\beta_1 > 0$ y -1 cuando $\beta_1 < 0$.

En síntesis la función de autocorrelación son valores que están relacionados en momentos diferentes en el tiempo. Un valor positivo (o negativo) de la AFC, genera una sucesión de valores positivos (o negativos), esto sería una corelación positiva. Ahora bien, si se manifiesta con alternancia de signos en la sucesión entonces, se está en presencia de un correlación negativa. La correlación cruzada, es conceptualmente similar a la autocorrelación, con la diferencia de que involucra dos series de tiempo que se desplazan una con respecto a la otra con un intervalo o rezago t . Por ejemplo, se puede examinar dinero y precios, dinero y tasa de interés, gasto público y desempleo, si no hay relación entre estas dos series entonces la función de correlación será cero, indicando que no vale la pena incluir una de las series para explicar la otra en una ecuación de regresión.

Definición 2.0.7 La Función de covarianza cruzada entre dos series x_t y y_t se define como

$$\gamma_{xy} = \mathbb{E}[(x_s - \mu_{xs})(y_t - \mu_{yt})]$$

Definición 2.0.8 La Función de Correlación Cruzada (CCF) es definida como

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

2.1. Series de Tiempo Estacionarias

Las series de tiempo se pueden clasificar en estacionarias y no estacionarias. Una serie de tiempo es estacionaria, cuando es estable a lo largo del tiempo, es decir, cuando la media y varianza son constantes en el tiempo. Esto se refleja gráficamente en que los valores de la serie tienden a oscilar alrededor de una media constante y la variabilidad con respecto a esa media también permanece constante en el tiempo. Por otra parte las series de tiempo no estacionarias son series donde la tendencia y/o variabilidad cambian en el tiempo. Los cambios en la media determinan una tendencia, crecer o decrecer a largo plazo, por lo que la serie no oscila alrededor de un valor constante. Definiendo formalmente estos términos, se tiene lo siguiente:

Definición 2.1.1 Una serie de tiempo *estrictamente estacionaria* es una serie para la cual el comportamiento probabilístico de cada sucesión de valores

$$x_{t1}, x_{t2}, \dots, x_{tk}$$

Es idéntico a la serie de trasladada en el tiempo

$$x_{t1+h}, x_{t2+h}, \dots, x_{tk+h}$$

Es decir,

$$Px_{t1} \leq c_1, \dots, x_{tk} \leq c_k = Px_{t1+h} \leq c_1, \dots, x_{tk+h} \leq c_k \quad (2.1)$$

Para todo $k = 1, 2, \dots$ todos los puntos de tiempo $t = t_1, t_2, \dots, t_k$ y números $c_1, c_2, \dots, c_k$ y todo salto $h = 0, \pm 1, \pm 2, \dots$

Observaciones:

1. Si una serie de tiempo es estrictamente estacionaria, entonces todas las funciones de distribución multivariadas para subconjuntos de variables deben coincidir con sus contrapartes en el conjunto trasladado, para todos los valores del parámetro h . Por ejemplo para $k = 1$ la ecuación (2.1) implica que

$$P\{x_s \leq c\} = P\{x_t \leq c\} \quad (2.2)$$

para cada par de puntos s y t .

2. Cuando $k = 2$, podemos escribir la ecuación (2.2) como

$$P\{x_s \leq c_1, x_t \leq c_2\} = P\{x_{s+h} \leq c_1, x_{t+h} \leq c_2\} \quad (2.3)$$

para cada par de puntos s y t y salto h . Entonces, si la función de varianza del proceso existe, (2.3) implica que la función de autocovarianza de la serie x_t satisface $\gamma(s, t) = \gamma(s + h, t + h)$ para todos s y t y salto h .

Podemos interpretar este resultado diciendo que la función de autocovarianza del proceso depende sólo de las diferencias de tiempo entre s y t , y no del tiempo actual.

Definición 2.1.2 Una serie de tiempo *débilmente estacionaria* x_t , es un proceso de varianza finita tal que:

1. La función de media μ_t , es constante y no dependiente del tiempo t .
2. La función de covarianza $\gamma(s, t)$ depende sólo de las diferencias de s y t , $|s - t|$. Se utilizará el término **estacionariedad** para referirnos a estacionariedad débil, si un proceso es estacionario en el sentido estricto usaremos el término **estrictamente estacionario**.

Definición 2.1.3 La función de autocovarianza de una serie de tiempo estacionaria se escribirá como

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

Definición 2.1.4 La función de autocorrelación (ACF) de una serie de tiempo estacionaria será escrita como

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

Definición 2.1.5 Dos series de tiempo x_t y y_t se dice que son **conjuntamente estacionarias** si cada una de ellas es estacionaria y la función de correlación cruzada

$$\gamma_{xy}(h) = \mathbb{E}[(x_{t+h} - \mu_x)(y_t - \mu_y)]$$

es una función sólo del salto h .

Definición 2.1.6 La función de **correlación cruzada (CCF)** de dos series conjuntamente estacionarias x_t y y_t se define como

$$\rho_{xy}(h) = \frac{\gamma_{xy}(h)}{\sqrt{\gamma_x(0)\gamma_y(0)}}. \quad (2.4)$$

La función autocorrelación satisface

$$\rho_{xy}(h) = \rho_{yx}(-h). \quad (2.5)$$

Ejemplo 2.1.1 Estacionaridad conjunta. Considere las series x_t y y_t formadas por las sumas y diferencias de dos valores sucesivos de un ruido blanco respectivamente, esto es

$$x_t = w_t + w_{t-1}$$

y

$$y_t = w_t - w_{t-1}$$

donde w_t son variables aleatorias independientes con media cero y varianza σ_w^2 . También

$$\begin{aligned}\gamma_{xy}(1) &= \mathbb{E}[(x_{t+1} - 0)(y_t - 0)] \\ &= \mathbb{E}[(w_{t+1} + w_t)(w_t - w_{t-1})] \\ &= \sigma_w^2\end{aligned}$$

porque solo uno de los productos es distinto de cero.

Similarmemente, $\gamma_{xy}(0) = 0$, $\gamma_{xy}(-1) = -\sigma_w^2$. Usando (2.4), obtenemos

$$\rho_{xy}(h) = \begin{cases} 0, & h = 0 \\ 1/2, & h = 1 \\ -1/2, & h = -1 \\ 0, & |h| \geq 2. \end{cases}$$

Claramente, las funciones de autocovarianza y correlación cruzada dependen solo del salto h , por lo tanto las series son conjuntamente estacionarias.

Definición 2.1.7 Un **proceso lineal** x_t , con $t = 1, \dots, n$, para un tiempo específico t se define como una combinación lineal de variables aleatorias de ruido blanco ω_t , y está dado por

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

donde los coeficientes satisfacen

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

Definición 2.1.8 Un proceso x_t , con $t = 1, \dots, n$ se dice que es un **proceso gaussiano** si el k -simo vector dimensional $\hat{x} = (x_{t_1}, x_{t_2}, \dots, x_{t_k})'$, para cada conjunto de puntos $t_1, t_2, \dots, t_k$ y cada entero positivo k tiene distribución normal multivariada.

2.2. Estimación de Correlación

A pesar de que las funciones de autocorrelación y correlación cruzada son útiles para describir propiedades de modelos hipotéticos, la mayoría de los análisis se efectúan utilizando datos muestrales. Esta limitación indica que los datos muestrales $x_1, \dots, x_n$ solo pueden ser empleados para estimar las funciones de media, autocorrelación y correlación cruzada. Desde el punto de vista estadístico, esto puede ser una dificultad puesto que no tiene copias idénticamente distribuidas de x_t , que sean posibles de estimar con las funciones de autocorrelación y correlación.

Definición 2.2.1 Sea $x_1, x_2, \dots, x_n$ una muestra de una serie de tiempo. La **media muestral** de $x_1, x_2, \dots, x_n$ es $\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t$. La función de **autocovarianza muestral** se define como

$$\hat{\gamma}(h) = n^{-1} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(x_t - \bar{x})$$

Con $-\hat{\gamma}(h) = \hat{\gamma}(\hat{h})$ para $h = 0, 1, \dots, n-1$. La **función de autocorrelación muestral** se define como

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

Propiedades 2.2.1

Distribución de AFC para muestras grandes. Bajo condiciones generales, si x_t es **Ruido Blanco** para n grande, la muestra de la AFC, $\hat{\rho}_x(h)$, para $h = 1, 2, \dots, H$ donde H es fijo y arbitrario, es aproximadamente una distribución normal, con media cero y desviación estándar dada por

$$\sigma_{\hat{\rho}_x(h)} = 1/\sqrt{n}.$$

Al verificar los picos que están fuera del intervalo $\pm 2/\sqrt{n}$, se puede determinar si los picos en $\hat{\rho}(h)$ son significativos.

Definición 2.2.2 El estimador para la función de covarianza cruzada $\gamma_{xy}(h)$ es la **función de covarianza cruzada muestral** definida como

$$\hat{\gamma}_{xy}(h) = n^{-1} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(y_t - \bar{y})$$

Donde $\hat{\gamma}_{xy}(-h) = \hat{\gamma}_{yx}(h)$ determina la función para saltos negativos, y la **función de correlación cruzada muestral** es

$$\hat{\rho}_{xy}(h) = \frac{\hat{\gamma}_{xy}(h)}{\sqrt{\hat{\gamma}_x(0)\hat{\gamma}_y(0)}}$$

Propiedades 2.2.2

Distribución de la correlación cruzada bajo independencia para muestras grandes. La distribución de muestras grandes de $\hat{\rho}_{xy}$ es normal con media cero y

$$\sigma_{\hat{\rho}_{xy}} = \frac{1}{\sqrt{n}}$$

si al menos uno de los procesos, es ruido blanco independiente.

2.3. Regresión de Series de Tiempo

La regresión lineal en el contexto de una **serie de tiempo**, supongamos que para x_t , con $t = 1, \dots, n$ influenciada por una colección de series independientes $Z_{t1}, Z_{t2}, \dots, Z_{tq}$ donde las entradas son conocidas y fijas. Esta suposición es necesaria para aplicar la regresión lineal clásica. Consideremos el modelo de regresión lineal

$$X_t = \beta_1 Z_{1t} + \beta_2 Z_{2t} + \dots + \beta_k Z_{kt} + \omega_t.$$

Donde $\beta_1, \dots, \beta_k$, son coeficientes de regresión desconocidos y fijos y ω_t es un error cualquiera o un ruido blanco.

Ejemplo 2.3.1 Estimación de una Tendencia Los datos corresponden a 98 años de mediciones de temperaturas, la gráfica presenta la desviación promedio entre las temperaturas en tierra y aire medidos en grados centígrados.

Se utilizará una regresión simple de la forma

$$X_t = \beta_1 Z_{1t} + \beta_2 Z_{2t} + \omega_t, \text{ con } t = 1900, 1901, \dots, 1997.$$

Utilizando los comandos correspondientes a cualquier programa con funciones estadísticas, se obtienen los valores de los coeficientes, siendo los siguientes: $\hat{\beta}_1 = -12,18$, $\hat{\beta}_2 = 0,006$ con un incremento cada 100 años de $0,6^\circ\text{C}$. En la figura se observa la serie de tiempo con la recta de tendencia

$$\hat{x}_t = -12,18 + 0,006t$$

El modelo lineal descrito anteriormente, puede ser convenientemente escrito, en una notación más general, definido en vectores columna

$$Z_t = Z_{t1} + Z_{t2}, \dots, Z_{tq}' \text{ y } \beta = (\beta_1, \beta_2, \dots, \beta_q),$$

esto puede ser escrito de una manera alternativa como

$$X_t = \beta' z_t + \omega_t.$$

donde $\omega_t \sim \text{iid}N(0, \sigma_\omega^2)$. Es habitual considerar el coeficiente del vector desconocido β minimizando el valor de la suma de los cuadrados $S = \sum_{t=1}^n (x_t - \beta' z_t)^2$ con respecto $\beta_1, \beta_2, \dots, \beta_q$. Minimizando S se obtiene el estimador común de mínimos cuadrados. Si se realiza la diferenciación de la ecuación anterior respecto al vector β o usando las propiedades de proyecciones, en la notación anterior se tiene

$$\left(\sum_{t=1}^n z_t z_t'\right) \hat{\beta} = \sum_{t=1}^n z_t x_t, \text{ para } t = 1, \dots, n.$$

La notación puede ser simplificada definiendo $Z = [z_1 | z_2 | \dots | z_n]'$ una matriz de $n \times q$ compuesta de n entradas y el vector observado $x = (x_1, x_2, \dots, x_n)$ y el vector de errores $\omega = (\omega_1, \omega_2, \dots, \omega_n)'$ en este caso el modelo se escribe como

$$X = Z\beta + \omega,$$

escribiéndose la ecuación normal como

$$(Z'Z)\hat{\beta} = Z'x$$

y la solución

$$\hat{\beta} = (Z'Z)^{-1}Z'x.$$

La matriz $Z'Z$ no es singular. El residuo minimizado de la suma de cuadrados tiene la siguiente forma matricial

$$\begin{aligned} S &= (x - Z\hat{\beta})'(x - Z\hat{\beta}) \\ &= x'x - \hat{\beta}'Z'x \\ &= x'x - x'Z(Z'Z)^{-1}Z'x. \end{aligned}$$

$\mathbb{E}(\hat{\beta})$ es el estimador común de mínimos cuadrados de β es insesgado y tiene la menor varianza de todos los valores insesgados lineales. Si los vectores ω_t están normalmente distribuidos, $\hat{\beta}$ es un estimador de máxima verosimilitud para β distribuido, como se indica en lo sucesivo

$$\text{Cov}(\hat{\beta}) = \sigma_\omega^2 \left(\sum_{t=1}^n Z_t Z_t'\right)^{-1} = \sigma_\omega^2 (Z'Z)^{-1} = \sigma_\omega^2 C.$$

donde

$$C = (Z'Z)^{-1}.$$

Un estimador insesgado para una varianza σ_ω^2 sería

$$S_{\omega}^2 = EMC = S/n - q,$$

donde EMC se denota por error medio cuadrado, el cual es contrastado con el estimador de máxima verosimilitud $\hat{\sigma}_{\omega}^2 = S/n - q$. Asumiendo que S_{ω}^2 es distribuido proporcionalmente a una variable aleatoria **chi- cuadrado** con $n-q$ grados de libertad, denotado por χ_{n-q}^2 e independiente de $\hat{\beta}$. Esto muestra que

$$T_{n-q} = \frac{\hat{\beta}_i - \beta_i}{s_{\omega} \sqrt{C_{ii}}}.$$

Tiene una distribución *t* de Student con $n-q$ grados de libertad. C_{ii} denota el *i-ésimo* elemento de la diagonal de la matriz C . Existen varias técnicas para hacer la comparación de modelos regresivos, como por ejemplo la prueba de ANOVA (Analysis of Variance), básicamente es un procedimiento que permite dividir la varianza de la variable dependiente en dos o más componentes, cada uno de los cuales puede ser atribuido a una fuente (variable o factor) identificable. Supongamos que consideramos un modelo de regresión normal con k y denotamos el estimador de máxima verosimilitud para la varianza como

$$\sigma_k^2 = \frac{S_k}{n}. \quad (2.6)$$

S_k denota la suma residual de cuadrados con k coeficientes de regresión.

Definición 2.3.1 *Criterio de Información de Akaike (AIC)*

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

Donde $\hat{\sigma}_k^2$ está dado por k y k es el número de parámetros en el modelo.

La idea es que la minimización de $\hat{\sigma}_k^2$ sea un objetivo razonable, exceptuando cuando decrece monótonamente mientras k crece. En consecuencia, se penaliza la variación del error por un término proporcional al número de parámetros.

Definición 2.3.2 *AIC con sesgo corregido (AICc), está dado por la siguiente ecuación*

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

donde $\hat{\sigma}_k^2$ está dado por (2.6), k es el número de parámetros en el modelo y n es el tamaño de la muestra.

Definición 2.3.3 *Criterio de Información de Schwarz(SIC), está definido por*

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

Usando la misma notación de la definición anterior SIC es también llamado Criterio Bayesiano de información (BIC).

Definición 2.3.4 *Definimos el operador de cambio por*

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

y extendemos a la potencia $B^2x_t = B(Bx_t) = Bx_{t-1} = x_{t-2}$ y así sucesivamente. Entonces

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

Definición 2.3.5 *Las diferencias de orden d son definidas como*

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

Donde podemos desarrollar el operador $(1 - B)^d$ algebraicamente para evaluar valores superiores enteros de d . Cuando $d = 1$, podemos prescindir de él en la notación.

2.4. Análisis Exploratorio de Datos

En general, es necesario hacer un promedio de productos en el tiempo, para que los datos de las series de tiempo sean estacionarios. Con estos datos de series de tiempo, es importante medir la dependencia de los valores, para estimar al menos con precisión la autocorrelación. No será fácil medir la dependencia, si la estructura de la misma no es regular o varía cada cierto período de tiempo. En consecuencia, para realizar análisis de datos de series de tiempo, las condiciones de media y varianza deben satisfacer las condiciones de las definiciones vistas anteriormente. A continuación se manejarán ciertos métodos para poder trabajar con datos no estacionarios. La forma más fácil de trabajar con datos no estacionarios, es el modelo de tendencia no estacionaria, donde el proceso tiene un comportamiento estacionario alrededor de una tendencia. Utilizando el siguiente modelo

$$X_t = \mu_t + y_t.$$

donde X_t son observaciones, μ_t la tendencia y y_t es un proceso estacionario. Una tendencia poderosa como μ_t puede traer como consecuencia que entorpezca el proceso estacionario de y_t . De igual forma, existe una ventaja

al remover la tendencia como primer paso en el análisis exploratorio de datos . Los pasos involucrados son para obtener un estimador razonable de la componente de tendencia, siendo esta $\hat{\mu}_t$ y trabajar con el residual

$$\hat{y}_t = x_t - \hat{\mu}_t.$$

Ejemplo 2.4.1 *Suponiendo el modelo*

$$X_t = \mu_t + y_t$$

Un modelo de tendencia para el análisis de temperatura global será lineal, si se cumple lo siguiente

$$\mu_t = \beta_1 + \beta_2 t$$

Estimando la tendencia de mínimos cuadrados

$$\hat{\mu}_t = -12,1864 + 0,0062t.$$

Para deshacerse de la tendencia de la serie, se resta $\hat{\mu}_t$ de las observaciones x_t

$$\hat{y}_t = x_t + 12,1864 - 0,0062t.$$

Se puede modelar la tendencia como una componente estocástica usando camino aleatorio como tendencia

$$\mu_t = \delta + \mu_{t-1} + \omega_t, \text{ donde } \omega_t \text{ es un ruido blanco e independiente de } y_t.$$

Si el modelo es apropiado, es entonces por diferencia de las observaciones se obtiene un modelo estacionario

$$X_t - x_{t-1} = (\mu_t + y_t) - (\mu_{t-1} + y_{t-1}) = \delta + \omega_t + y_t - y_{t-1}$$

Si $z_t = y_t - y_{t-1}$ es estacionario, entonces

$$\gamma_z(h) = \text{cov}(z_{t+h}, z_t) = \text{cov}(y_{t+h} - y_{t+h-1}, y_t - y_{t-1}) = 2\gamma_y(h) - \gamma_y(h+1) - \gamma_y(h-1)$$

es independiente del tiempo. Una ventaja de la diferenciación sobre la eliminación de la tendencia es que no hay que estimar parámetros en la operación de la diferencia. Ahora bien, la dificultad sería que la diferenciación no da un estimador del proceso estacionario y_t . Si un estimador y_t es esencial entonces eliminar la tendencia, sería lo más apropiado. Si el objetivo

es obligar o transformar los datos a estacionarios entonces la difereciación sería lo más viable. La difrenciación es una herramienta viable si la tendencia es ajustada. La primera difrenciación es denotada como

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

La primera diferecniación elimina la tendencia lineal, la segunda puede eliminar la tendencia cuadrática y así sucesivamente.

2.5. Suavizando Series de Tiempo

En general, se debe suavizar una serie de tiempo, para poder descubrir ciertos rasgos en una serie de tiempo, como tendecias de grandes términos o componentes estacionales, por ejemplo un promedio móvil puede suavizar un ruido blanco. En particular si x_t representa las observaciones

$$M_t = \sum_{j=-k}^k a_j x_{t-j}, \text{ donde } a_j = a_{-j} \leq 0 \text{ y } \sum_{j=-k}^k a_j = 1$$

es un promedio móvil simétrico de los datos.

Modelos Autoregresivos de Promedio Móvil

El modelo autoregresivo se define como, el valor actual de la serie x_t , como una función de p valores $x_{t-1}, x_{t-2}, \dots, x_{t-p}$, donde p indica la cantidad de pasos que se debe regresar en el pasado para obtener el valor actual. Los modelos autorregresivos se abrevian con la palabra AR.

Definición 2.5.1 *Un modelo autoregresivo de orden p , abreviado $AR(p)$, es de la forma*

$$X_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \omega_t.$$

Donde x_t es estacionario, $\phi_1, \phi_2, \dots, \phi_p$ son constantes ($\omega_p \neq 0$). A menos que se aclare lo contrario, se asume que ω_t es un ruido gaussiano de media cero y varianza σ_ω^2 . La media de x_t en (1) es cero. Si la media μ de x_t no es cero, reemplazamos x_t por $x_t - \mu$ en (1), es decir

$$x_t - \mu = \psi_1(x_{t-1} - \mu) + \phi_2(x_{t-2} - \mu) + \dots + \phi_p x_{t-p} + \omega_t$$

o se escribe

$$X_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \omega_t$$

donde

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

Definición 2.5.2 El operador autoregresivo se define como

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p.$$

Al estudiar los modelos AR se iniciará el orden del primer modelo, **AR(1)**, dado por $x_t = \phi x_{t-1} + \omega_t$ iterando por el operador de cambio un tiempo k

$$\begin{aligned} x_t &= \phi x_{t-1} + \omega_t = \phi(\phi x_{t-2} + \omega_{t-1}) + \omega_t \\ &\vdots \\ &= \omega^k x_{t-k} + \sum_{j=0}^{k-1} \phi^j \omega_{t-j}. \end{aligned}$$

Este método sugiere que, continuando con el método de iteración y siendo $|\phi| < 1$ y x_t sea estacionario, se puede representar un modelo AR(1) como un proceso lineal dado por

$$x_t = \sum_{j=0}^{\infty} \phi^j \omega_{t-j}.$$

El modelo AR(1), definido en lo anterior es estacionario con media

$$\mathbb{E}(x_t) = \sum_{j=0}^{\infty} \phi^j \omega_{t-1} = 0$$

y función de autocovarianza

$$\begin{aligned} \gamma(h) &= \text{cov}(x_{t+h}, x_t) \\ &= \mathbb{E}[(\sum_{j=0}^{\infty} \phi^j \omega_{t+h-j})(\sum_{k=0}^{\infty} \phi^k \omega_{t-k})] \\ &= \mathbb{E}[(\omega_{h+t} + \dots + \phi^h \omega_t + \phi^{h+1} \omega_{t-1} + \dots)(\omega_t + \phi \omega_{t-1} + \dots)] \\ &= \sigma_{\omega}^2 \sum_{j=0}^{\infty} \phi^{h+j} \phi^j \\ &= \sigma_{\omega}^2 \phi^h \sum_{j=0}^{\infty} \phi^{2j} \\ &= \frac{\sigma_{\omega}^2 \phi^h}{1 - \phi^2}, h \leq 0. \end{aligned}$$

Recordando que $\gamma(h)$ es simétrica, la AFC de un modelo AR(1) es lo siguiente

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \phi^h, h \geq 0$$

Donde $\rho(h)$ satisface

$$\rho(h) = \phi \rho(h-1), \text{ para } h = 1, \dots$$

Definición 2.5.3 El modelo de promedio móvil de orden q o modelo $MA(q)$, se define como

$$X_t = \omega_t + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \cdots + \theta_q \omega_{t-q}.$$

Donde hay q pasos en el promedio móvil y $\theta_1, \theta_2, \dots, \theta_q$ ($\theta_q \neq 0$) son parámetros. El ruido ω_t , donde $t = 1, \dots, n$ se asume como un ruido blanco gaussiano.

Definición 2.5.4 El operador de promedio móvil es

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q.$$

Definición 2.5.5 Una serie de $X_{t \geq 1}$ es **ARMA** (p, q) si es estacionario y

$$X_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \omega_t + \theta_1 \omega_{t-1} + \cdots + \theta_q \omega_{t-q}.$$

Con $\phi_p \neq 0, \theta_q \neq 0$ y $\sigma_\omega^2 > 0$. Los parámetros p y q son llamados ordenes autoregresivos y de promedio móvil respectivamente. Si x_t tiene media μ distinto de cero, se toma $\alpha = \mu(1 - \phi_1 - \cdots - \phi_p)$ y escribimos el modelo como

$$X_t = \alpha + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \omega_t + \theta_1 \omega_{t-1} + \cdots + \theta_q \omega_{t-q}$$

Donde $\omega_t; t = 0, \pm 1, \pm 2, \dots$ será una sucesión de ruido blanco gaussiano.

Definición 2.5.6 Los **Polinomios AR y MA** se definen como

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p, \phi_p \neq 0$$

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q, \theta_q \neq 0$$

respectivamente, donde z es un número complejo.

Definición 2.5.7 Un modelo $ARMA(p, q)$, $\phi(B)x_t = \theta(B)\omega_t$, se dice que es **causal** si la serie de tiempo $x_t; t = 0, \pm 1, \pm 2, \dots$ se puede escribir como un proceso lineal de la siguiente forma

$$X_t = \sum_{j=0}^{\infty} \psi_j \omega_{t-j} = \psi(B)\omega_t \tag{2.7}$$

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

Definición 2.5.8 Un modelo ARMA (p,q) , $\phi(B)x_t = \theta(B)\omega_t$, se dice **invertible** si la serie de tiempo $x_t : t = 0, \pm 1, \pm 2, \dots$ se puede escribir como

$$\pi(B)x_t = \sum_{j=0}^{\infty} \pi_j x_{t-j} = \omega_t$$

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

Propiedades 2.5.1 1. *Causalidad de procesos ARMA (p,q) . Un modelo ARMA (p,q) , es causal si y sólo si $\phi(z) \neq 0$ para $|z| \leq 1$. Los coeficientes del proceso lineal del modelo ARMA (p,q) , pueden ser determinados utilizando la siguiente ecuación*

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

Otra causalidad pudiera ser, sólo cuando las raíces de $\phi(z)$ estén fuera del círculo unitario, esto es $\phi(z) = 0$ sólo cuando $|z| > 1$.

2. *Invertibilidad de un proceso ARMA (p,q) . Un modelo ARMA (p,q) es invertible, si y sólo si $\theta(z) \neq 0$ para $|z| \leq 1$. Los coeficientes π_j de $\pi(B)$ dado por un la definición anterior, pueden ser determinados resolviendo*

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

Otra forma de ver esta propiedad sería sólo cuando las raíces de $\theta(z)$ estén fuera del círculo unitario, esto es, $\theta(z) = 0$ sólo cuando $|z| > 1$.

2.6. Autocorrelación y Autocorelación Parcial

Se empezará analizando la AFC de un proceso MA (q) , $x_t = \theta(B)\omega_t$, donde $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$. Como x_t es un combinación lineal finita de un ruido blanco, el poceso estacionario con media

$$\mathbb{E}(x_t) = \sum_{j=0}^q \theta_j \mathbb{E}(\omega_{t-j}) = 0$$

Donde $\theta_0 = 1$ y con la función de autocovarianza

$$\begin{aligned} \psi(h) &= \text{cov}(x_{t+h}, x_t) \\ &= \text{cov}(\sum_{j=0}^q \theta_j \omega_{t+h-j}, \sum_{k=0}^q \theta_k \omega_{t-k}) \end{aligned}$$

$$= \begin{cases} \sigma_{\omega}^2 \sum_{j=0}^{q-h} \theta_j \theta_{j+h} & \text{si } 0 \leq h \leq q \\ 0 & \text{si } h > q. \end{cases}$$

Recordando que $\psi(h)$ es par; esto es, $\psi(h) = \psi(-h)$. El corte de $\psi(h)$ después de q saltos es el sello del modelo MA(q). Dividiendo lo anterior entre $\psi(0)$ se puede obtener la AFC de un modelo MA(q)

$$\rho(h) = \begin{cases} \frac{\sum_{j=0}^{q-h} \theta_j \theta_{j+h}}{1 + \theta_1^2 + \dots + \theta_q^2} & \text{si } 1 \leq h \leq q \\ 0 & \text{si } h \leq q. \end{cases}$$

Para un modelo ARMA (p, q) causal, $\phi(B)x_t = \theta(B)\omega_t$ donde los ceros de $\phi(z)$ están fuera del círculo unitario

$$X_t = \sum_{j=0}^{\infty} \psi_j \omega_{t-j}.$$

Se sigue inmediatamente que $E(x_t) = 0$. Además, la AFC de x_t puede escribirse como

$$\gamma(h) = \text{cov}(x_{t+h}, x_t) = \sigma_{\omega_{j+j}}^2, \quad h \geq 0.$$

Se puede utilizar posteriormente para resolver el ψ – peso. Paralelamente, se puede resolver para $\gamma(h)$ y la AFC $\rho(h) = \gamma(h)/\gamma(0)$. También es posible obtener una ecuación en diferencias homogéneas directamente en términos de $\gamma(h)$, se escribe

$$\gamma(h) = \text{cov}(x_{t+h}, x_t)$$

$$= \text{cov}(\sum_{j=1}^p \phi_j x_{t+h-j} + \sum_{j=0}^q \theta_j \omega_{t+h-j}, x_t)$$

$$= \sum_{j=1}^p \phi_j \gamma(h-j) + \sigma_{\omega}^2 \sum_{j=h}^q \theta_j \psi_{j-h}, \quad h \geq 0. \quad (2.8)$$

Donde se tiene que para $h \geq 0$,

$$\text{Cov}(\omega_{t+h-j}, x_t) = \text{cov}(\omega_{t+h-j}, \sum_{k=0}^{\infty} \psi_k \omega_{t-k}) = \psi_{j-h} \sigma_{\omega}^2. \quad (2.9)$$

De (2.8) se puede escribir una ecuación general homogénea para la AFC de un proceso ARMA causal:

$$\gamma(h) - \sum_{j=1}^p \phi_j \gamma(h-j) = \sigma_{\omega}^2 \sum_{j=h}^q \theta_j \psi_{j-h}, \quad 0 \leq h < \max(p, q+1). \quad (2.10)$$

Dividiendo (2.9) y (2.10) entre $\psi(0)$ podrá funcionar para resolver la ACF, $\rho(h) = \psi(h)/\psi(0)$.

Definición 2.6.1 La *Función de Correlación Parcial (PACF)* de un proceso estacionario x_t , denotado por ϕ_{hh} , para $h = 1, 2, \dots$ es

$$\phi_{11} = \text{corr}(x_{t+h} - \hat{x}_{t+h}, x_t - \hat{x}_t), h \geq 2$$

Tanto $(x_{t+h} - \hat{x}_{t+h})$ y $(x_t - \hat{x}_t)$ no están correlacionados con $x_{t+1}, \dots, x_{t+h-1}$.

Ejemplo 2.6.1 Considerando la PACF de un proceso AR(1) dado por $x_t = \phi x_{t-1} + \omega_t$ con $|\phi| < 1$. Por definición, $\phi_{11} = \rho(1) = \phi$. Para calcular ϕ_{22} , se considerará la regresión de x_{t+2} en x_{t+1} , diciendo que $\hat{x}_{t+2} = \beta x_{t+1}$. Se escoge β para minimizar

$$\mathbb{E}(x_t - \hat{x}_t)^2 = \mathbb{E}(x_t - \beta x_{t+1})^2 = \gamma(0) - 2\beta\gamma(1) + \beta^2\psi(0).$$

Esta es la misma ecuación anterior, entonces $\beta = \phi$.

En consecuencia,

$$\phi_{22} = \text{corr}(x_{t+2} - \hat{x}_{t+2}, x_t - \hat{x}_t) = \text{corr}(x_{t+2} - \phi x_{t+1}, x_t - \phi x_{t+1}) = \text{corr}(\omega_{t+2}, x_t - \phi x_{t+1}) = 0.$$

2.7. Predicciones

En predicciones, la meta es poder proyectar el futuro de las series de tiempo x_{n+m} , $m = 1, 2, \dots$, basado en los datos recolectados del presente, $x = x_n, x_{n-1}, \dots, x_1$, se asumirá que x_t es estacionario y los parámetros del modelo son conocidos. El mínimo error cuadrático del predictor x_{n+m} es

$$X_{n+m}^n = \mathbb{E}(x_{n+m}|x)$$

porque la esperanza condicional minimiza el error cuadrático medio

$$\mathbb{E}[x_{n+m} - g(x)]^2$$

donde $g(x)$ es la función de observaciones x . Primero, se restringirá la atención para predecir que hay funciones lineales de la data, esto es, predicciones de la forma

$$X_{n+m}^n = \alpha_0 + \sum_{k=1}^n \alpha_k x_k, \quad (2.11)$$

donde $\alpha_0, \alpha_1, \dots, \alpha_n$ son números reales. Las predicciones lineales de la forma (2.11) que minimiza la predicción del error cuadrático medio son llamadas la mejores proyecciones lineales (BLPs). Los predictores lineales dependen solo del segundo momento del proceso, lo que es fácil de estimar en los datos.

Propiedades 2.7.1 1. *Mejor Predicción Lineal para un Proceso Estacionario.*

Dados los datos $x_1, x_2, \dots, x_n$, el mejor predictor lineal ,

$$x_{n+m}^n = \alpha_0 + \sum_{k=1}^n \alpha_k x_k, \text{ de } x_{n+m} \text{ para } m \neq 1,$$

es encontrado para resolver

$$\mathbb{E}[(x_{n+m} - x_{n+m}^n)x_k] \text{ para } \alpha_0, \alpha_1, \dots, \alpha_n. \quad (2.12)$$

Las ecuaciones mencionadas en la propiedad anterior son llamadas ecuaciones de predicción, y son usadas para resolver los coeficientes $\alpha_0, \alpha_1, \dots, \alpha_n$. Si $\mathbb{E}(x_t) = \mu$, la primera ecuación ($k = 0$) de (2.12) implica que

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

Tomando la esperanza en (2.12)

$$\mu = \alpha_0 + \sum_{k=1}^n \alpha_k \mu \text{ ó } \alpha_0 = \mu(1 - \sum_{k=1}^n \alpha_k).$$

En consecuencia, la forma del BLP es

$$x_{n+m}^n = \mu + \sum_{k=1}^n \alpha_k (x_k - \mu)$$

También, se puede considerar cuando $\mu = 0$, en este caso, $\alpha_0 = 0$. Primero, se debe considerar la predicción de un paso. Esto es, dado $x_1, x_2, \dots, x_n$, se desea predecir el valor de la serie de tiempo en el momento x_{n+1} . El BLP de x_{n+1} es de la forma

$$x_{n+1}^n = \phi_{n1}x_n + \phi_{n2}x_{n-1} + \dots + \phi_{nn}x_1, \quad (2.13)$$

se escribió anteriormente en (2.11), como $\phi_{n,n+1-k}$ en (2.13), para $k = 1, \dots, n$. Usando la propiedad 2.7.1, los coeficientes $\phi_{n1}, \phi_{n2}, \dots, \phi_{nn}$ satisfacen

$$\mathbb{E}[(x_{n+1} - \sum_{j=1}^n \phi_{nj}x_{n+1-j})x_{n+1-k}] = 0, k = 1, \dots, n,$$

o

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

Estas ecuaciones de predicción pueden ser escritas en forma matricial como

$$\Gamma_n \phi_n = \gamma_n. \quad (2.14)$$

Donde $\Gamma_n = \gamma(k-j)_{j,k=1}^n$ es una matriz de $n \times n$, $\phi_n = (\phi_{n1}, \dots, \phi_{nn})'$ es un vector de $n \times 1$ y $\gamma_n = (\gamma(1), \dots, \gamma(n))'$ es un vector de $n \times 1$. La matriz Γ_n es definida no negativa. Si Γ_n es singular, se obtienen muchas soluciones de (2.14), pero por el Teorema de Proyección, x_{n+1}^n es único. Si Γ_n es no singular, los elementos de ϕ_n son únicos y vienen dados por

$$\phi_n = \Gamma_n^{-1} \gamma_n. \quad (2.15)$$

Para un modelo ARMA, el hecho de que $\sigma_\omega^2 > 0$ y $\gamma(h) \rightarrow 0$ cuando $h \rightarrow \infty$ es suficiente para asegurar que Γ_n sea definida positiva.

$$x_{n+1}^n = \phi_n' x. \quad (2.16)$$

Donde $x = (x_n, x_{n-1}, \dots, x_1)'$. El error cuadrático medio de la predicción de un paso es

$$P_{n+1}^n = \mathbb{E}(x_{n+1} - x_{n+1}^n)^2 = \gamma(0) - \gamma_n' \Gamma_n^{-1} \gamma_n.$$

Propiedades 2.7.2 1. El Algoritmo de Durbin-Levinson

Las ecuaciones (2.15) y (2.16) pueden ser resueltas iterativamente como sigue

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

para $n \geq 1$

$$\phi_{nn} = \frac{\rho(n) - \sum_{k=1}^{n-1} \phi_{n-1,k} \rho(n-k)}{1 - \sum_{k=1}^{n-1} \phi_{n-1,k} \rho(k)}, \quad (2.17)$$

$$P_{n+1}^n = P_n^{n-1} (1 - \phi_{nn}^2),$$

para $n \geq 2$,

$$\phi_{nk} = \phi_{n-1,k} - \phi_{nn} \phi_{n-1,n-k}, k = 1, 2, \dots, n-1.$$

2. La solución iterativa para la PACE.

La PACF de un proceso estacionario x_t , puede ser obtenido por vía iteración de (2.17) como ϕ_{nn} para $n = 1, 2, \dots$

2.8. Modelos ARIMA

Si x_t es un camino aleatorio, $x_t = x_{t-1} + \omega_t$, luego diferenciando x_t , se consigue que $\nabla x_t = \omega_t$ es estacionario. En muchas ocasiones, se puede pensar que las series de tiempo están compuestas por dos componentes, una componente con tendencia no estacionaria y otra con media cero estacionaria. Por ejemplo, considerando el siguiente modelo

$$x_t = \mu_t + y_t,$$

donde $\mu_t = \beta_0 + \beta_1 t$ y y_t es estacionario. Diferenciando este proceso, se obtiene otro estacionario

$$\nabla_t = x_t - x_{t-1} = \beta_1 + y_t - y_{t-1} = \beta_1 + \nabla y_t.$$

Otro modelo estadístico para Series Temporales que se puede mencionar es el **ARIMA (p, d, q) (AutoRegressive Integrated Average)**, proveniente de los componentes AR (Autoregresivos), I (Integrado) y MA (Medias Móviles). Este modelo permite describir un valor como una función lineal de datos anteriores y errores recibidos al azar, además debe contener todos los elementos necesarios para describir el fenómeno.

Definición 2.8.1 Un proceso x_t se dice que es ARIMA (p, d, q) si

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

es ARMA (p, q). En general, escribiremos el modelo como

$$\phi(B)(1 - B)^d x_t = \theta(B)\omega_t, \text{ siendo } \omega_t \text{ un ruido blanco.}$$

Si $\mathbb{E}(\nabla^d x_t) = \mu$, escribiremos el modelo como

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

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

2.8.1. Construcción de un modelo ARIMA

Existen ciertos pasos para ajustar un modelo ARIMA a una serie de tiempo. Estos pasos involucran la graficación de los datos, transformación, identificación de la dependencia de orden del modelo, parámetros de estimación, diagnostico y modelo de escogencia. Primero, como cualquier otro análisis de datos, se debe construir una gráfica de tiempo de estos datos y verificar si presenta algún tipo de anomalías. Si por ejemplo, la variabilidad de los datos crece con el tiempo, será necesario transformar los datos para estabilizar la varianza.

En tal caso, se puede utilizar la transformación de potencia de Box-Cox. Además, la aplicación particular puede sugerir una transformación apropiada. Por ejemplo, se supondrá un proceso que evolucione con un porcentaje de cambio pequeño y estable, como una inversión. Se puede tener

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

donde x_t es el valor de la inversión en un tiempo t y p_t es el porcentaje de cambio en el período $t-1$ hasta t , donde este período puede ser negativo. Tomando logaritmos se obtiene

$$\log(x_t) = \log(1 + p_t) + \log(x_{t-1}), \text{ ó } \nabla \log(x_t) = \log(1 + p_t).$$

Si el porcentaje de cambio p_t es relativamente pequeña en magnitud, luego $\log(1 + p_t) \approx p_t$

$$\nabla \log(x_t) \approx p_t$$

será un proceso estable relativo. Frecuentemente, $\nabla \log(x_t)$ es llamado la tasa de crecimiento o el regreso. Luego de la transformación de los datos, el próximo paso es identificar los valores preliminares del orden autoregresivo p , el orden de la diferenciación d y el orden del promedio móvil q . Luego de tener, en parte, direccionado el problema de seleccionar d , se puede realizar una gráfica del tiempo con esos datos, si es necesario introducir la diferenciación. Si en efecto, esta es necesaria, entonces se diferencia una vez, $d=1$ y se procede a analizar la gráfica de ∇x_t . Si se necesita otra diferenciación, se intenta nuevamente y se analiza la gráfica $\nabla^2 x_t$. A pesar de esto, no se debe sobre diferenciar puesto que indicaría dependencias donde no existen.

Análisis de Temperatura y Pluviosidad en el Lago de Valencia

3.1. Descripción

En la climatología el análisis de series de tiempo de elementos meteorológicos, como la temperatura del aire en la superficie y la pluviosidad es útil en análisis sinópticos para pronósticos del tiempo y el análisis de variabilidad climática. Uno de los problemas que intenta resolver las serie de tiempo es la predicción. Esto es dada una serie $x(t_1), x(t_2), \dots, x(t_n)$, los objetivos de interés son describir el comportamiento de la serie, investigar el mecanismo generador de la serie temporal y buscar posibles patrones temporales que permitan sobrepasar la incertidumbre del futuro. En adelante se contruirá un modelo matemático, para explicar la estructura y prever la evolución de una variable que se observa a lo largo del tiempo.

La variable temperatura se expresa en la grados centígrados ($^{\circ}\text{C}$), determinado por la escala Celcius y la lluvia o pluviosidad en milímetros cúbicos (mm). Toda la información estudiada, fue recolectada en el Lago de Valencia en dos estaciones de medición, la Estación Sucre y la Estación Tacarigua, ambas ubicadas en distintos extremos del Lago, utilizando una consola llamada Vantage Pro2, que permite una visualización local de las variables metereológicas como: temperatura, lluvia, presión barométrica, velocidad y dirección del viento.

En el transcurso del análisis de la información, se demuestra que en las estaciones no se maneja la misma información, puesto que hay ciertos vacíos en esos datos, por lo que se puede presumir que los dispositivos de medición tuvieron fallas al momento de procesar esa infomación , también que para la pluviosidad (que

no tuvo altos niveles), los dispositivos no pudieran registrar bajos volúmenes de lluvia o que en efecto, exista para una zona del Lago un mayor índice de pluviosidad que en la otra. De igual forma, a pesar de estos factores se logra recolectar información importante que pueda funcionar para indicar el comportamiento de estas variables en el Lago. Estos datos se obtuvieron en el período 2007-2008, empezando por los meses de Noviembre y Diciembre para el año 2007 y de Enero a Octubre en el año 2008. En Venezuela existen dos estaciones climáticas: temporada de sequía o período seco, que empieza en el mes de Noviembre hasta el mes de Marzo y temporada de lluvia o período de lluvia, que va a partir de Abril hasta Octubre. Por las posibles dificultades de medición mencionadas anteriormente, es evidente que no se manejan la misma cantidad de datos ni de días registrados para cada estación, en consecuencia se estructura la siguiente tabla, para poder conocer cuántos días de observaciones se tienen para cada mes y por estación.

Meses	Días Observados por Estaciones	
	E. Sucre	E. Tacarigua
Nov-07	21	4
Dic-07	3	21
Ene-08	13	23
Feb-08	16	Sin Información
Mar-08	25	
Abr-08	9	19
May-08	Sin información	3
Jun-08	21	12
Jul-08	26	28
Ago-08	31	16
Sep-08	26	Sin Información
Oct-08	23	9

Tabla 3.1: Cantidad de días con observaciones para las Estaciones Sucre y Tacarigua.

3.2. Comportamiento General de la Temperatura y de la Lluvia

3.2.1. Estadística Descriptiva

El primer paso para poder entender el comportamiento de la información otorgada por las variables estudiadas es efectuar un análisis de estadística descriptiva, que ayudará a calcular los parámetros estadísticos que permiten calcular el conjunto de variables estudiadas. Se utilizan histogramas para ver las características descriptivas de la distribución, diagramas de caja (boxplot), pues estos gráficos que sumistran información de la mediana, el primer y tercer cuantil, existencia de datos atípicos y la simetría de la distribución siendo una herramienta visual para comparar la distribución de dos conjuntos de datos, en esta gráfica se representan los cuantiles de una distribución contra los cuantiles correspondiente de la otra. Es importante destacar que en Venezuela, existen dos estaciones en el año; el período de sequía y el período de lluvia, cada uno con duración de seis meses aproximadamente.

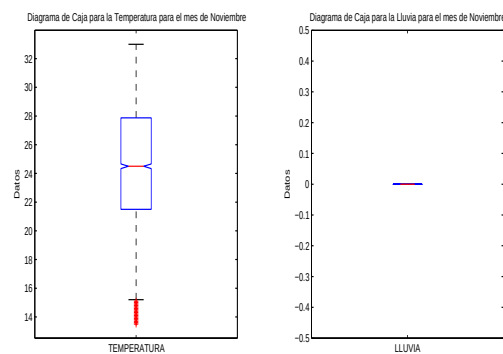


Figura 3.1: Diagrama de Caja para el mes de Noviembre en la Estación Sucre (Temperatura-Pluviosidad).

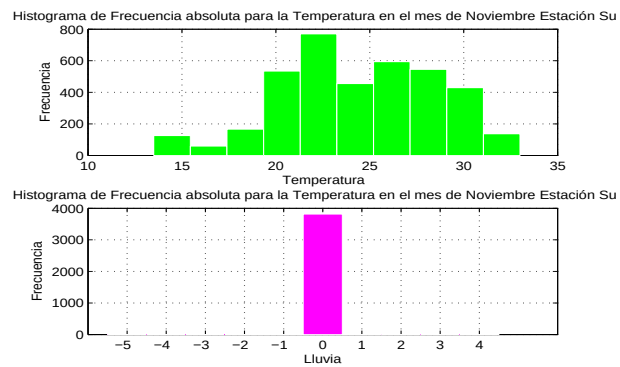


Figura 3.2: Histograma para el mes de Noviembre en la Estación Sucre (Temperatura-Pluviosidad).

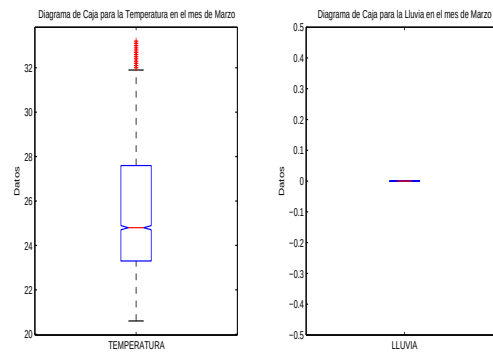


Figura 3.3: Diagrama de Caja para el mes de Marzo en la Estación Tacarigua (Temperatura-Pluviosidad).

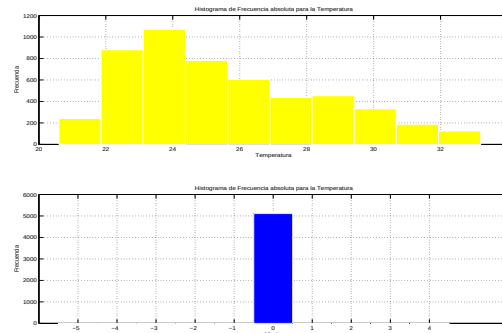


Figura 3.4: Histograma para el mes de Marzo en la Estación Tacarigua (Temperatura-Pluviosidad).

En las gráficas anteriores, se puede observar que los niveles de pluviosidad son bajos. Se debe tomar en cuenta que los meses observados, son del período seco y en consecuencia es de esperar que no sean registrados altos niveles de pluviosidad. Esta información es similar para los meses desde Noviembre hasta Abril, por lo que si se observan los Diagramas de Caja para la lluvia no indica ningún tipo de información significativa. Ahora bien, la otra variable observada es la temperatura, donde en promedio el rango de temperatura para el período seco es de 20°C - 34°C y el 75 % de la información esta por encima de 24°C , lo que indica que la temperatura en gran parte del mes fue elevada, para las Estaciones Sucre y Tacarigua. El Histograma muestra la distribución de los datos durante los meses estudiados. Según la gráfica se observa que no existe una distribución normal, de los datos observados.

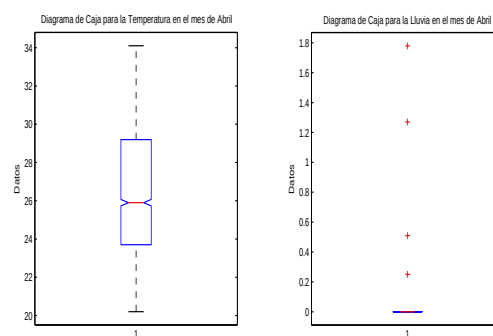


Figura 3.5: Diagrama de Caja para el mes de Abril en la Estación Sucre (Temperatura-Pluviosidad).

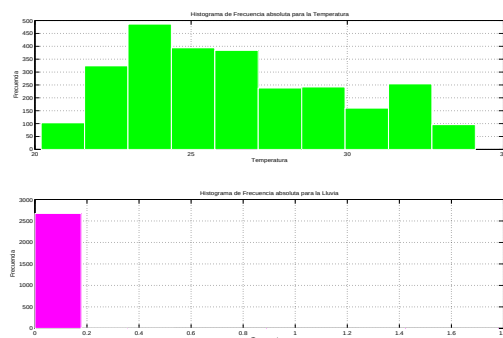


Figura 3.6: Histograma para el mes de Abril en la Estación Sucre (Temperatura-Pluviosidad).

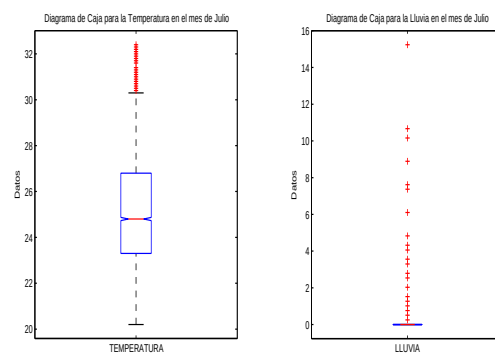


Figura 3.7: Diagrama de Caja para el mes de Julio en la Estación Tacarigua (Temperatura-Pluviosidad).

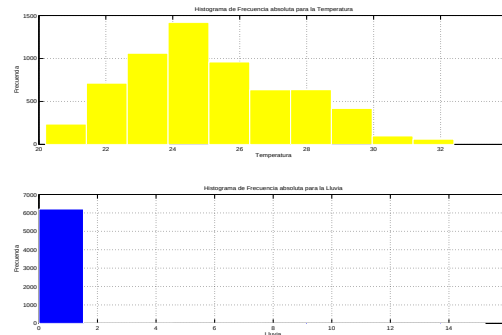


Figura 3.8: Histograma para el mes de Julio en la Estación Tacarigua (Temperatura-Pluviosidad).

Si se observa el diagrama de caja para la lluvia, muestra un incremento en las precipitaciones del mes, marcando el comienzo del período lluvioso, de igual forma en la mayor parte de los días no hubo pluviosidad. La precipitación en la cuenca muestra una distribución unimodal¹. El diagrama de caja de la temperatura, se observa con un comportamiento constante y con cambios progresivos, lo que tendría sentido al considerar que la temperatura no varía bruscamente, de un grado a otro. De igual forma se observa que el 75 % de las observaciones de la temperatura está por encima de los 24°C y que hay un incremento de los niveles máximos de temperatura. Durante este período se observa en el diagrama de caja de la lluvia, denotado con cruces rojas, la cantidad de mediciones sobre los registros de pluviosidad durante el mes, es importante recordar que el mes de Abril marca el final de la temporada seca en el país. Como consecuencia de este hecho, en el transcurso de Noviembre de 2007 y Octubre de 2008 se puede observar el marcado contraste entre ambos períodos del año.

3.2.2. Comparaciones de Temperatura y Pluviosidad

Luego de realizar la estadística descriptiva de los datos, se procede a realizar una gráfica de datos (temperaturas en °C o lluvia en mm) versus tiempo, lo que reflejará el comportamiento de la temperatura y la pluviosidad según cada mes y por estaciones.

¹Una distribución de probabilidad unimodal es una distribución de probabilidad que tiene un único modo.

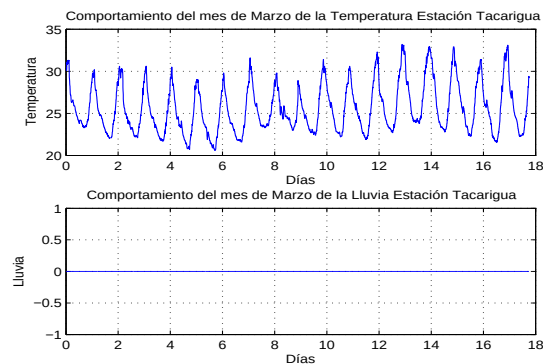


Figura 3.9: Comportamiento de la Temperatura y la Lluvia durante el mes de Marzo, Estación Tacarigua.

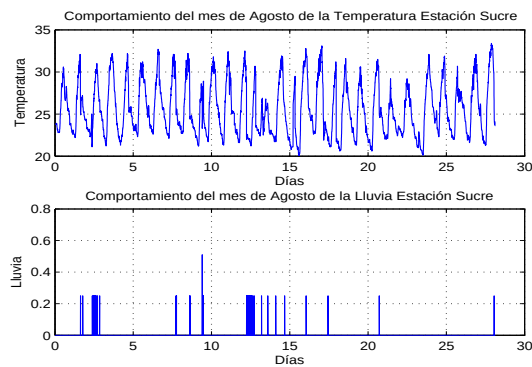


Figura 3.10: Comportamiento de la Temperatura y la Lluvia durante el mes de Agosto, Estación Sucre.

Para los meses Noviembre, Diciembre, Enero, Febrero y Marzo, se observa que hay mínimos registros de lluvia, también se verifica que Marzo es uno de los meses más calurosos del año. En efecto, se tuvo para este mes como máximo valor $33,2^{\circ}\text{C}$, la temperatura mínima se mantiene con $20,6^{\circ}\text{C}$ y la media mensual sería de $25,5^{\circ}\text{C}$. A pesar de ser uno de los meses con mayor temperatura, esta variable se mantiene con un comportamiento constante durante el año, bien sea con la presencia de lluvias o no, como se observa en la segunda gráfica a pesar de que hay precipitaciones la temperatura se mantiene en un rango aproximado de 19°C - 34°C , lo que puede sugerir que estas variables son independientes entre sí.

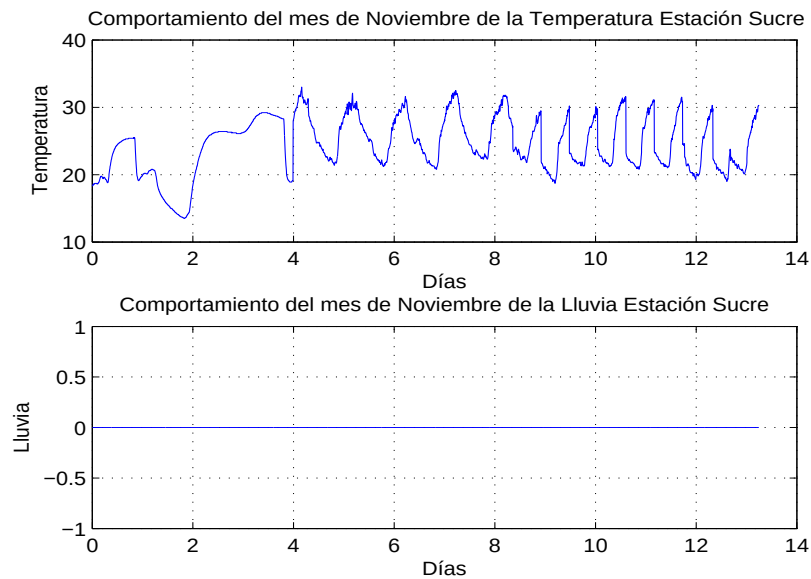


Figura 3.11: Comportamiento de la Temperatura y la Lluvia durante el mes de Noviembre, Estación Sucre

Para el mes de noviembre de 2007, se obtuvo una temperatura máxima de 33°C , mínima de $13,5^{\circ}\text{C}$ y la temperatura promedio durante el mes fue de $24,4^{\circ}\text{C}$. Por otra parte, la pluviosidad para este período no tuvo mayor información puesto que sus valores máximos y mínimos fueron de 0 mm y en consecuencia el promedio fue de 0. Lo que efectivamente para este mes refleja que la pluviosidad no cumple con el período de invierno o con la época de lluvias, sin embargo, el promedio de la temperatura se adapta a la temporada. Si se observa la gráfica de la temperatura con detenimiento, se observa que en las primeras observaciones del mes el comportamiento es diferente al del resto del mes. Este hecho, puede que se manifieste por poca información para esos primeros días del mes. Hay que descartar que los dispositivos de medición pudieron haber tenido interrupciones en la medición de las variables estudiadas.

Durante este mes, sólo se realizó el registro de 3 días, por lo que en efecto se puede observar una diferencia marcada en comparación a el comportamiento general del año. La temperatura se mantuvo con un promedio de $26,7^{\circ}\text{C}$, alcanzando un máximo valor de $32,3^{\circ}\text{C}$ y uno mínimo de $23,3^{\circ}\text{C}$, mostrando un comportamiento similar al mes anterior, a pesar del corto tiempo registrado. Además, se visualiza que la pluviosidad estuvo presente en estos días con un máximo valor de 1,2 mm y en promedio 0,003 mm.

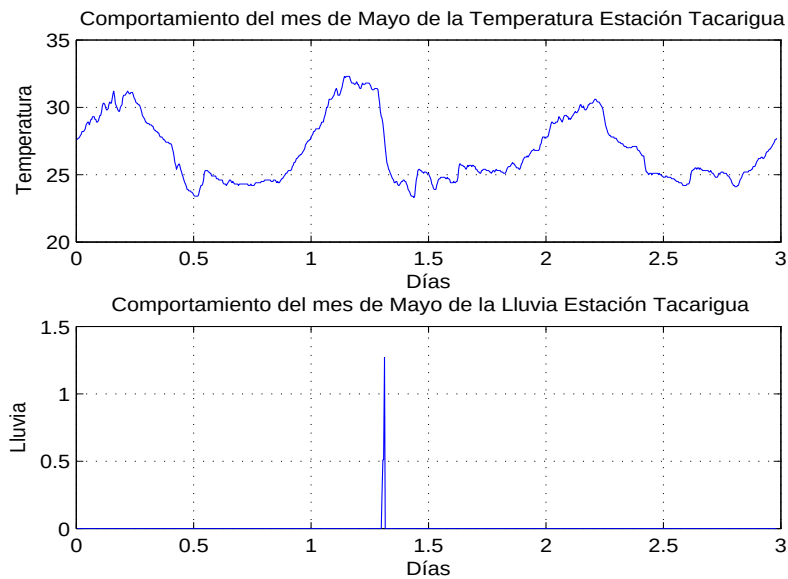


Figura 3.12: Comportamiento de la Temperatura y la Lluvia durante el mes de Mayo, Estación Tacarigua

3.3. Función de Correlación y Autocorrelación

La Función de Correlación Cruzada, es empleada para estudiar la fortaleza y la dirección lineal entre dos procesos estocásticos conjuntamente estacionarios y la Función de autocorrelación mide la relación estadística entre las observaciones de una serie de tiempo. Ahora bien, en esta sección se buscarán estas funciones, para poder identificar si en efecto, la temperatura y la pluviosidad tienen relación en su desarrollo mensual y si una variable afecta a la otra directamente.

Según lo que se observa durante el año 2007-2008, las variables que se están evaluando muestran que no hay dependencia entre ellas, esto puede ser motivado a que en primera instancia, no llovió todos los días en el Lago como era de esperarse y en la temporada de lluvia también, según las observaciones que se manejan no hay altos niveles de pluviosidad bien sea por las condiciones climáticas propias del Lago de Valencia o a fallos en el dispositivo de medición, de igual forma las precipitaciones si bien pueden ocasionar una ligera disminución de la temperatura, esto no se debe considerar como una dependencia entre la temperatura y la lluvia.

Por otra parte, también se observa la relación existente entre las mediciones de la temperatura, es decir;

cada registro depende directamente del anterior, puesto que la temperatura no tiene cambios bruscos en su comportamiento, si bien puede tomar valores muy bajos o muy altos, siempre será un cambio progresivo puesto que no se puede tener una temperatura de 22°C y 5 minutos después estar en 35°C . Como consecuencia, es un indicativo de la alta correlación entre los datos de la temperatura, a diferencia de la pluviosidad que si puede cambiar bruscamente y el hecho de que en un momento del día llueva no implica que ese hecho suceda todos los días posteriores o durante ese mismo día lo que indica que para la pluviosidad la autocorrelación es muy baja. Se tienen los siguientes ejemplos:

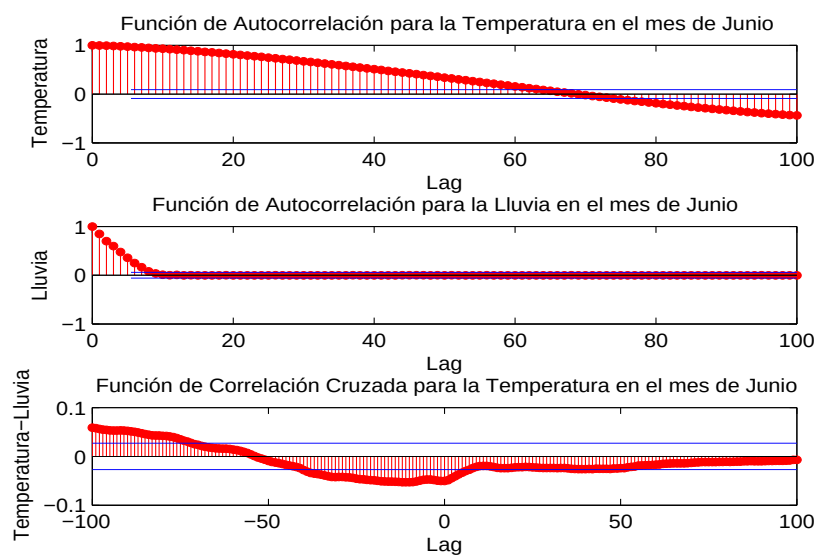


Figura 3.13: Función de AFC y CCF de la Temperatura y de la Pluviosidad durante el mes de Junio.

En Junio se mantiene el coeficiente de correlación negativo de -0.0504 lo que indica que a pesar de que llueva eso no tiene incidencia en el cambio de la temperatura en el Lago de Valencia. A pesar de eso, los cambios de temperatura siguen siendo dependientes entre sí.

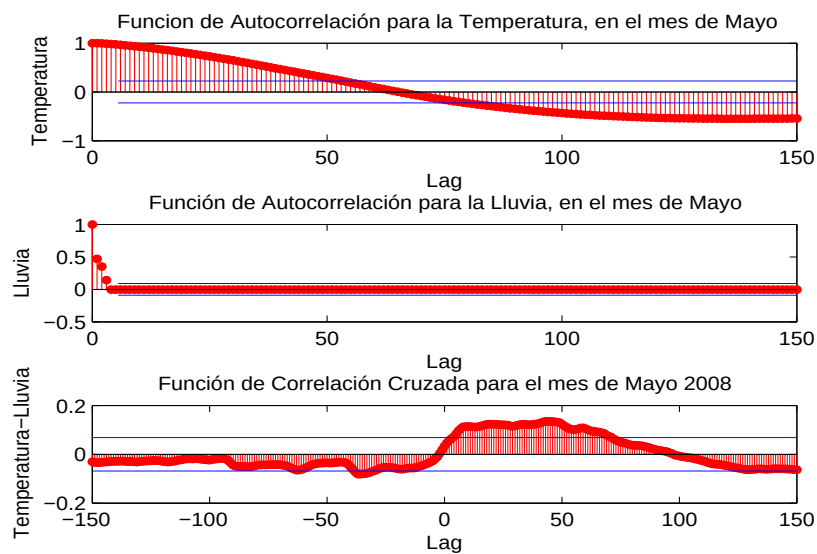


Figura 3.14: Función de AFC y la CCF de la Temperatura y la Pluviosidad durante el mes de Mayo.

Los resultados de este mes muestran que hubo evidencia de lluvia y al igual que los meses anteriores hay una ligera dependencia de estos registros. De igual forma, se visualiza que el coeficiente de correlación positivo 0.0336 , lo que indica que para este mes las variables se vieron afectadas por la incidencia de las precipitaciones a la temperatura. Lo que puede indicar, que la lluvia baja (o aumenta) la temperatura para este mes.

3.4. Comparaciones entre a Temperatura y la Pluviosidad, entre las Estaciones Sucre y Tacarigua

Además de la información recolectada y estudiada anteriormente, se debe establecer si existen diferencias entre las mediciones estadísticas que se realizaron entre las estaciones. Tomando en cuenta que si ambas estaciones están en el Lago, deberían tener la misma información. Sin embargo, por los datos recolectados se nota que a pesar de tener ciertos días la misma medición la información no necesariamente es la misma. En este apartado, se realizaran las comparaciones de Temperatura y Pluviosidad entra cada estación.

La Estación Sucre aportó mayor información durante el año evaluado, de igual forma no hay información del mes Mayo y en la Estación Tacarigua para los meses Febrero y Septiembre. Entonces, para la Estación Sucre durante el mes de Diciembre,(3.15) se tiene como máxima temperatura alcanzada en el año $34,6^{\circ}\text{C}$, una mínima de $13,5^{\circ}\text{C}$ y un promedio de $25,3^{\circ}\text{C}$ y para la lluvia se tiene que el valor mínimo fue 0 mm, el promedio 0 mm y el nivel máximo de pluviosidad 10,2 mm con 48761 datos registrados, mientras que para la Estación Tacarigua se obtuvo como máxima temperatura $34,4^{\circ}\text{C}$, una mínima $13,5^{\circ}\text{C}$ y un promedio $25,7^{\circ}\text{C}$ y para la precipitaciones los valores fueron, como máximo 15,2 mm, mínimo 0 mm y un promedio de 0 mm con 39490 datos observados. Lo que nos indica que la Estación con la mayor temperatura fue la Estación Sucre y con la mayor precipitación del año.

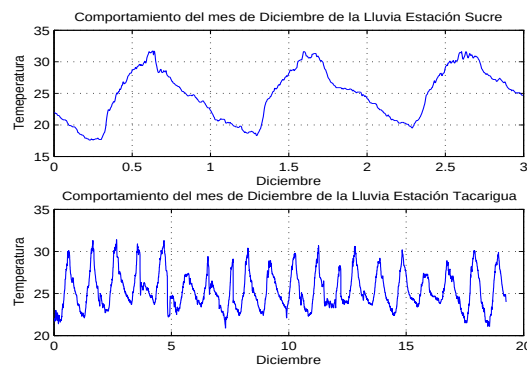


Figura 3.15: Gráfica de comparación de la Temperatura entre las Estaciones Sucre- Tacarigua para el mes de Diciembre.

Estadísticos	Diciembre	
	Sucre	Tacarigua
Promedio	24,80	25,40
Máximo	31,70	31,40
Mínimo	17,6	20,90
Mediana	25,20	24,90
Moda	20,30	23,90
Desviación Estándar	4,02	2,19

Tabla 3.2: Comparación de estadísticos entre las Estaciones Sucre y Tacarigua, para el mes de Diciembre

Las temperaturas promedio registradas fueron de $24,8^{\circ}\text{C}$, en la Estación Sucre y de $25,4^{\circ}\text{C}$ en la Estación Tacarigua. En este período la temperatura máxima no tuvo gran variación entre las estaciones estudiadas, manteniéndose con una diferencia de $31,7^{\circ}\text{C}$ - $31,4^{\circ}\text{C}$ respectivamente. Sin embargo, los valores mínimos presentan una marcada diferencia entre sí ($20,9^{\circ}\text{C}$ Estación Sucre - $17,6^{\circ}\text{C}$ Estación Tacarigua). Para este mes se va notando un ligero cambio de temperatura en las estaciones siendo Tacarigua más calurosa que Sucre. Los valores más repetidos en las estaciones son de $20,3^{\circ}\text{C}$ y $23,9^{\circ}\text{C}$, con medianas de $25,2^{\circ}\text{C}$ y de $24,9^{\circ}\text{C}$ y desviación estándar de 4,02 y 2,19 respectivamente. Gráficamente se puede observar, que la Estación Tacarigua otorga mayor información, por la cantidad de días que reflejan información, con 21 días de observaciones.

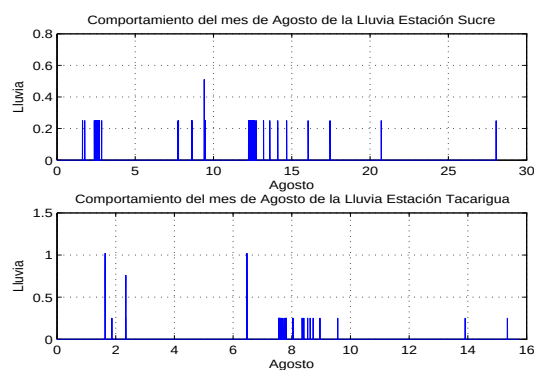


Figura 3.16: Gráfica de comparación de la Pluviosidad entre las Estaciones Sucre-Tacarigua para el mes de Agosto.

En la figura (3.16) se observa que bajan los niveles máximos de precipitación con 0,51 mm de Sucre y 1,02 mm de Tacarigua, disminuyendo considerablemente al mes anterior. De igual forma el promedio mensual disminuye con 0 mm para ambas estaciones. Se observa en la gráfica de comportamiento que no se manejan la misma cantidad de observaciones en ambas estaciones, puesto que se manejan 31 días de observaciones en la Estación Sucre y 16 en la Estación Tacarigua. También se evidencia que no se alcanzan los mismos niveles máximos de lluvia y que además no se registran los mismos datos en las observaciones, lo que en parte puede venir dado por la situación explicada en el mes anterior.

3.5. Modelo Arima

El marco teórico para el estudio de las series de temperatura es la teoría de los procesos estocásticos, en particular el ARIMA, para el análisis univariante de estas series temporales. Este modelo se utiliza para explicar la estructura de la temperatura mensual correspondientes y se puede considerar como una primera aproximación objetiva en la predicción de la temperatura.

- Estación Sucre

Primero se observaran el resumen de los modelos que ajustan cada mes con los los índices del Criterio de Información de Akaike (AIC) de la Estación Sucre.

MESES	DÍAS	MODELO	AIC
nov-07	20	ARIMA(3,1)	-2.8575
dic-07	3	ARIMA(3,1)	-3.9092
ene-08	13	ARIMA(3,1)	-2.8945
feb-08	16	ARIMA(4,2)	-3.54
mar-08	26	ARIMA(4,4)	-3.711
abr-08	10	ARIMA(2,4)	-3.7242
jun-08	22	ARIMA(4,3)	-4.1678
jul-08	27	ARIMA(3,2)	-3.2305
ago-08	31	ARIMA(4,1)	-2.9519
sep-08	27	ARIMA(3,4)	-3.0603
oct-08	23	ARIMA(3,3)	-2.8299

Tabla 3.3: Índices de AIC para la Estación Sucre

Ahora se mostrará las tablas de los coeficientes ϕ y θ de cada modelo en la Estación Sucre.

Meses	Coeficientes de ϕ			
nov-07	-2.088	1.213	-0.1251	
dic-07	-2.338	1.694	-0.3553	
ene-08	-2.145	1.32	-0.1747	
feb-08	-1.231	-0.737	1.213	-0.2446
mar-08	-1.975	0.4553	1.028	-0.5085
abr-08	-1.963	0.9626		
jun-08	-0.6012	-0.549	-0.2072	-0.3665
jul-08	-2.482	1.981	-0.4983	
ago-08	-2.18	1.342	-0.1139	-0.0488
sep-08	-2.515	2.05	-0.5349	
oct-08	-2.575	2.166	-0.5917	

Tabla 3.4: Valores de Coeficientes de ϕ

Meses	Coeficientes de θ			
nov-07	-0.8953			
dic-07	-0.8950			
ene-08	-0.8967			
feb-08	-0.0984	-0.9004		
mar-08	-0.5484	-0.7617	0.2427	0.1373
abr-08	-0.4377	-0.2277	-0.1300	-0.0387
jun-08	1.1450	0.5842	0.0801	
jul-08	-1.1370	0.1840		
ago-08	-0.9136			
sep-08	-1.2280	0.2959	0.0037	-0.0224
oct-08	-1.1900	0.2412	0.0028	

Tabla 3.5: Valores de Coeficientes de θ

Luego de manejar esta información, se mostrará las gráficas correspondientes a cada de mes de los datos estudiados en la Estación Sucre. Donde se podrá observar los vacíos de información o días en que no hubo ningún registro y además se podrá observar en la parte inferior la simulación de cómo pudo haber sido el comportamiento de esos datos faltantes.

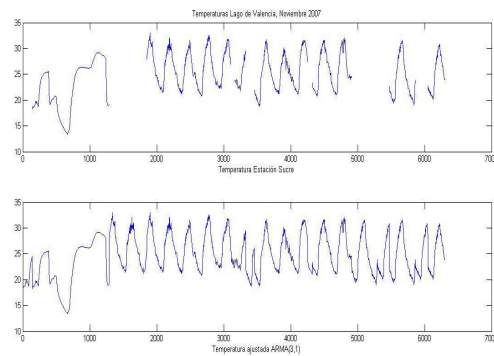


Figura 3.17: Noviembre 2007, Estación Sucre

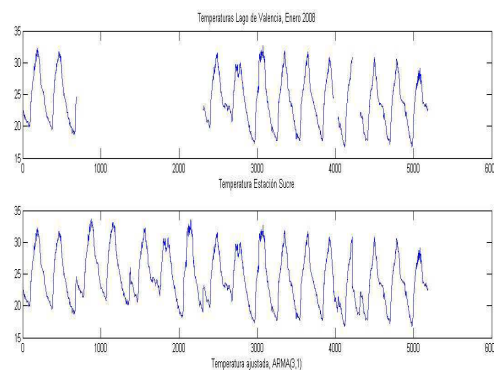


Figura 3.18: Enero 2008, Estación Sucre

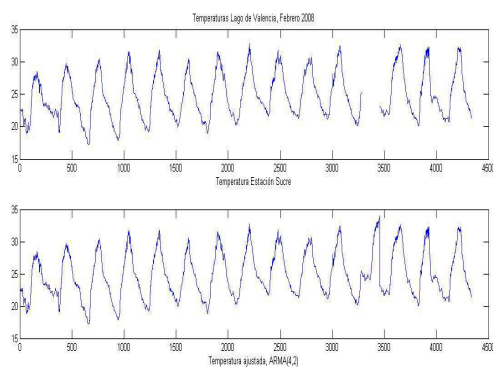


Figura 3.19: Febrero 2008, Estación Sucre

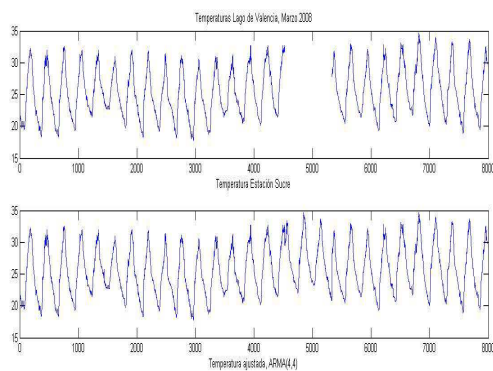


Figura 3.20: Marzo 2008, Estación Sucre

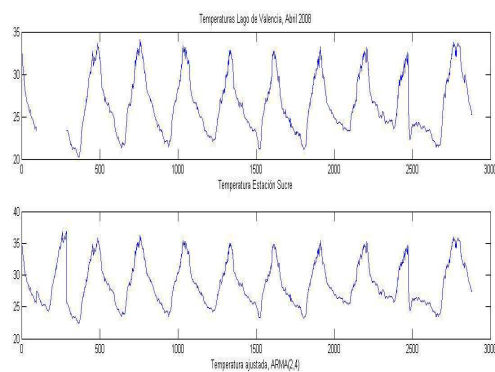


Figura 3.21: Abril 2008, Estación Sucre

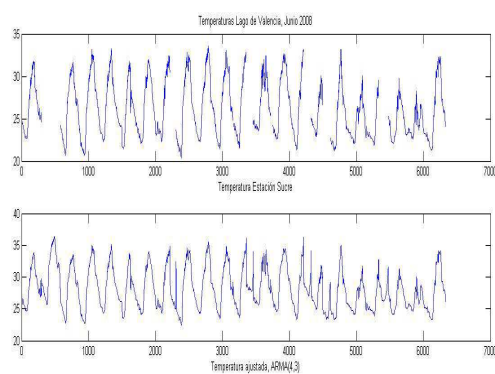


Figura 3.22: Junio 2008, Estación Sucre

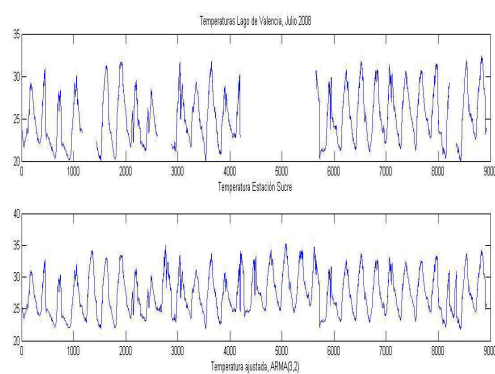


Figura 3.23: Julio 2008, Estación Sucre

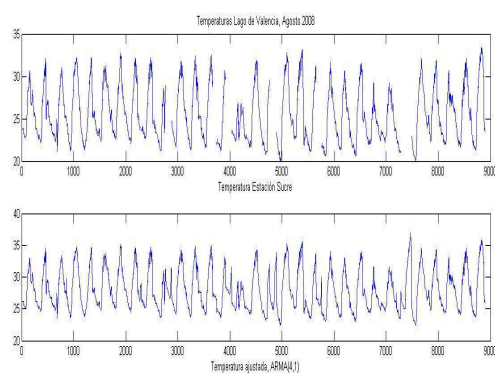


Figura 3.24: Agosto 2008, Estación Sucre

■ Estación Tacarigua

Ahora se va a manejar la información para la Estación Tacarigua de manera análoga a la E. Sucre. Primero se observaran los índices del Criterio de Akaike.

MESES	DÍAS	MODELO	AIC
nov-07	5	ARIMA(2,1)	-2.236
dic-07	22	ARIMA(4,3)	-3.7685
ene-08	23	ARIMA(5,5)	-3.7745
mar-08	20	ARIMA(5,2)	-4.1314
abr-08	17	ARIMA(5,1)	-3.7676
may-08	3	ARIMA(5,4)	-4.3719
jun-08	18	ARIMA(4,1)	-4.0575
jul-08	26	ARIMA(4,3)	-4.2213
ago-08	16	ARIMA(3,5)	-4.1704
oct-08	10	ARIMA(2,5)	-3.6679

Tabla 3.6: Índices de AIC para la Estación Tacarigua

Para cada modelo de la Estación Tacarigua, se deben tener los coeficientes ϕ y θ , que serán mostrados en las tablas mostradas a continuación.

Meses	Coeficientes de ϕ				
nov-07	-1.9670	0.9672			
dic-07	-1.4170	0.5485	1.3910	0.4256	
ene-08	-1.4680	0.4632	0.3832	-1.1980	0.8203
mar-08	-1.2440	-0.7504	1.2510	0.2351	-0.0219
abr-08	-2.2600	1.4630	-0.1634	0.0009	-0.0401
may-08	-1.0300	0.6963	0.0148	1.2470	-0.5349
jun-08	-2.2259	1.4420	-0.0843	-0.0989	
jul-08	-1.3220	0.6028	0.9803	0.1766	
ago-08	-0.8359	-0.9925	0.8285		
oct-08	-1.9380	0.9383			

Tabla 3.7: Valores de Coeficientes de ϕ

Meses	Coeficientes de θ				
nov-07	-0.9232				
dic-07	-0.1467	0.9332	0.2073		
ene-08	-0.2829	-0.0030	0.4198	-0.7204	0.0971
mar-08	0.0975	-0.9008			
abr-08	-0.9308				
may-08	0.5385	-0.3822	-0.8992	0.0268	-0.5349
jun-08	-0.9246				
jul-08	0.1911	0.7481	-0.2945		
ago-08	0.7534	-0.3467	-0.1799	-0.1341	-0.0559
oct-08	-0.5882	-0.1131	-0.0321	-0.0657	-0.0293

Tabla 3.8: Valores de Coeficientes de θ

Ya con esta información anterior, entonces se procedera a realizar la aplicación del modelo, para que pueda completar los espacios sin información para así poder tener una posible idea del comportamiento de la temperatura en el Estación Tacarigua. Cabe destacar, que no hubo información para los meses de Febrero y Septiembre del año 2008.

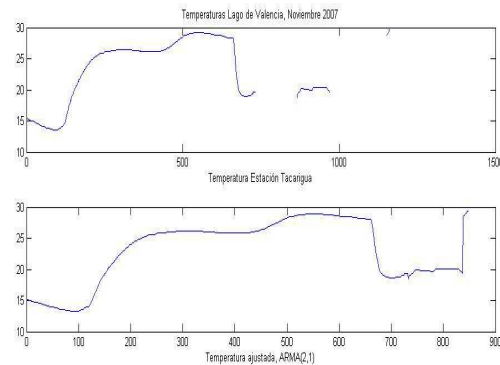


Figura 3.25: Noviembre 2007, Estación Tacarigua

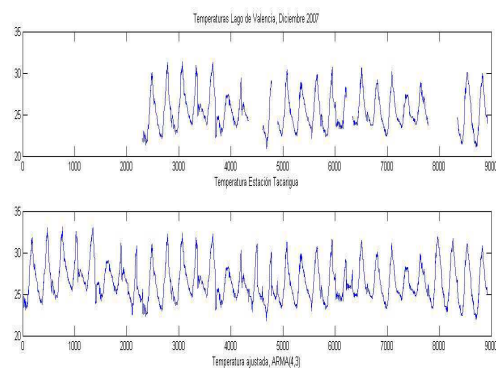


Figura 3.26: Diciembre 2007, Estación Tacarigua

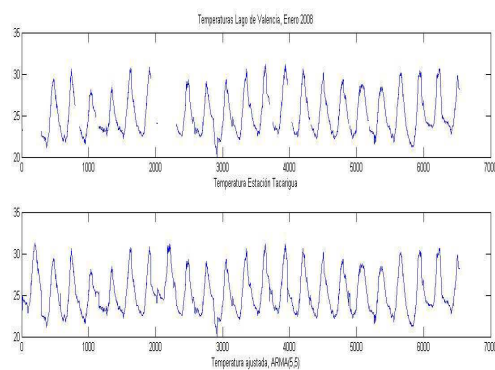


Figura 3.27: Enero 2008, Estación Tacrigua

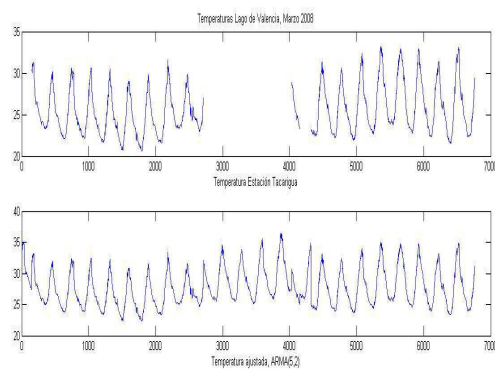


Figura 3.28: Marzo 2008, Estación Tacrigua

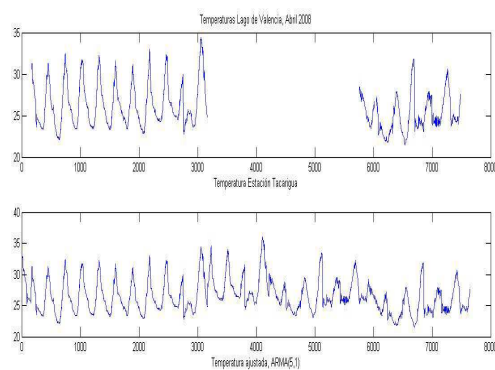


Figura 3.29: Abril 2008, Estación Tacrigua

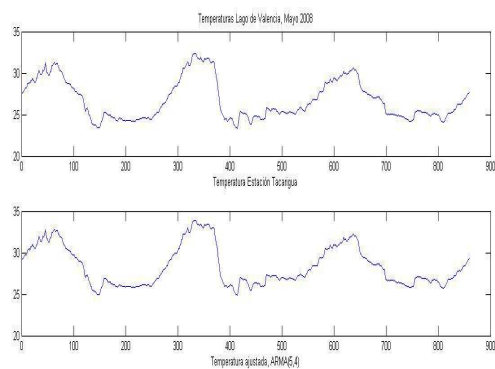


Figura 3.30: Mayo 2008, Estación Tacrigua

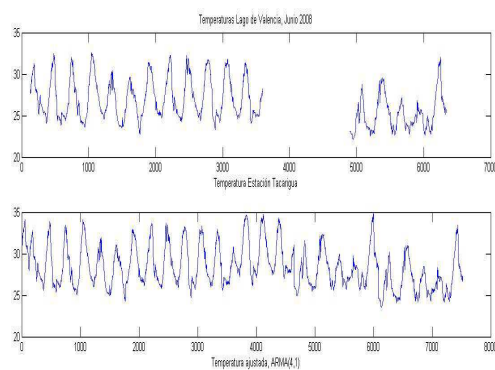


Figura 3.31: Junio 2008, Estación Tacrigua

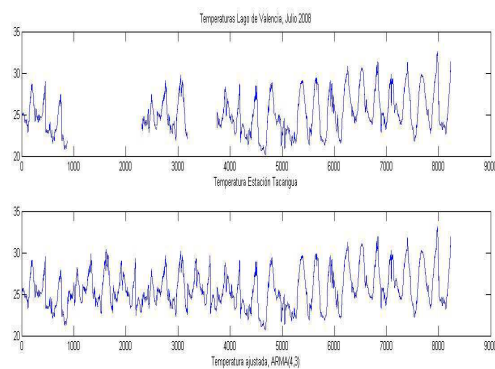


Figura 3.32: Julio 2008, Estación Tacrigua

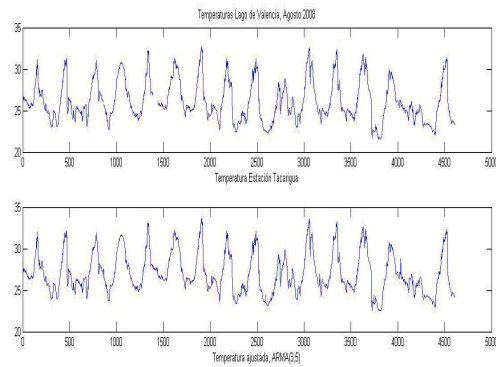


Figura 3.33: Agosto 2008, Estación Tacrigua

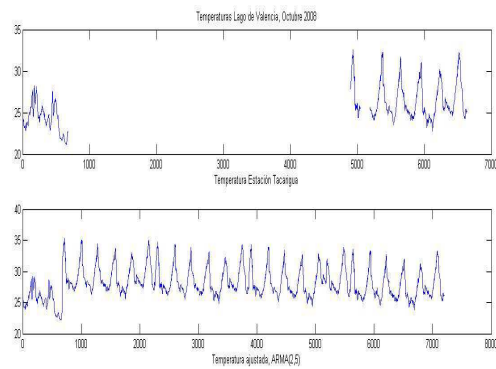


Figura 3.34: Septiembre 2008, Estación Tacrigua

Al tener estos resultados gráficos, se realizará una comparación de promedios, máximos y mínimos entre los datos de pluviosidad y temperatura y la simulación realizada con el modelo ARIMA, para completar la información faltante.

	Resultados de los registros de Temperatura		
Meses	Máximos	Mínimos	Promedio
nov-07	33,00	13,50	24,30
ene-08	32,70	16,80	24,00
feb-08	32,70	17,20	24,70
mar-08	34,60	17,70	25,40
abr-08	34,10	20,20	26,40
jun-08	33,60	20,40	26,10
jul-08	32,40	20,00	25,10
ago-08	33,40	20,10	25,50
sep-08	32,80	20,00	25,40

Tabla 3.9: Resultados de los registros de la Estación Sucre

	Resultados de Simulación de Temperatura		
Meses	Máximos	Mínimos	Promedio
nov-07	33,00	13,50	24,73
ene-08	33,61	16,80	24,36
feb-08	33,91	17,20	24,88
mar-08	34,60	17,70	25,58
abr-08	36,77	22,34	28,62
jun-08	36,37	22,43	28,17
jul-08	35,20	21,82	27,34
ago-08	36,88	22,52	28,01
sep-08	32,80	20,00	27,90

Tabla 3.10: Resultados de simulación Estación Sucre

En las tablas 3.9 3.10 y anteriores, se muestra en primera instancia el promedio, máximo y mínimo mensual de las simulaciones realizadas con el modelo ARIMA para poder entender el comportamiento de la temperatura en las estaciones, donde no se obtiene la información. Por otra parte, en la siguiente tabla se tiene la misma información, es decir, máximos, mínimos y promedios de la información que si se pudo recolectar en el Lago

de Valencia. Se puede observar que el máximo valor se alcanzó para el mes de Marzo con 34,60 °C mientras que en la simulación el mayor valor se encuentra en el mes de Agosto con 36,88 °C, el mínimo valor se puede observar en el pronóstico del mes de Noviembre con 13 °C, siendo el mismo valor real observado.

	Resultados de los registros de Temperatura		
Meses	Máximos	Mínimos	Promedio
nov-07	29,70	13,50	23,20
ene-08	33,10	20,10	25,10
mar-08	33,20	20,60	25,50
abr-08	33,20	20,60	25,50
may-08	32,30	23,30	26,70
jun-08	32,60	22,80	27,20
jul-08	32,40	20,20	25,10
ago-08	32,70	21,60	26,10
oct-08	32,60	21,10	25,80

Tabla 3.11: Resultados de los registros de la Estación Tacarigua

	Resultados de Simulación de Temperatura		
Meses	Máximos	Mínimos	Promedio
nov-07	29,41	13,21	23,01
ene-08	33,61	16,80	24,36
feb-08	33,09	21,80	26,50
mar-08	31,21	20,10	25,25
abr-08	36,43	22,35	27,78
may-08	33,87	24,87	28,34
jun-08	34,87	23,57	28,41
jul-08	33,06	20,66	25,60
ago-08	33,60	22,50	27,06
oct-08	35,34	22,11	28,13

Tabla 3.12: Resultados de simulación Estación Tacarigua

Para la Estación Tacarigua se realiza el procedimiento análogo, al de la Estación Sucre donde se desea comparar la información real con el pronóstico, el promedio de temperatura simulada se encuentra en un rango de 23-28 °C y para la temperatura observada es de entre 23-27°C, lo que muestra un pequeño aumento de las temperaturas con la simulación en ambas estaciones según la información suministrada en los indicadores anteriores.

Conclusión

El estudio de este trabajo, se llevó acabo en el Lago de Valencia durante el período 2007-2008, utilizando información suministrada por dispositivos colocados en dos estaciones de medición, Sucre y Tacarigua, las cuales registran información como radiación solar, temperatura del aire, temperatura del agua, precipitaciones, velocidad del viento, entre otras.

El propósito principal era conocer el comportamiento durante un año de la temperatura y la pluviosidad en el Lago de Valencia para poder realizar simulaciones basadas en ese comportamiento y posteriormente realizar comparaciones entre las temperaturas reales y las simuladas, y así para poder determinar sus diferencias y similitudes.

La Estación Sucre brindó mayor infomación en cuanto a registros, puesto que tuvo mayor cantidad de días con información que la Estación Tacarigua. Además de esto se realizó un análisis de ambas variables para validar si el comportamiento de una afectaba directamente a la otra, para así determinar el nivel de correlación existente entre ellas. También se estudiaron los valores máximos, mínimos, mediana, promedio, moda y desviación estándar de cada una. Después de haber analizado los resultados de la investigación se obtuvo lo siguiente:

- Luego de analizar los resultados, se observa el bajo nivel de precipitaciones en ambas estaciones, notando que durante el período de lluvia a pesar de tener precipitaciones no fueron tan significativas como para poder realizar la aplicación de modelo ARIMA para esta variable.
 - En el estudio de la información, se observa que en las estaciones no se maneja la misma información, puesto que hay ciertos vacíos en esos datos, por lo que se puede presumir que los dispositivos de medición tuvieron fallas al momento de procesar esa información.
 - En la información estudiada durante el año 2007-2008, las variables que se están evaluando denota que no hay correlación entre ellas, según el resultado de las gráficas de AFC y el coeficiente de correlación.
 - El modelo ARIMA se utilizó para realizar simulaciones y poder observar como pudo haber sido el comportamiento en las Estaciones Sucre y Tacarigua para la temperatura, se observa que ésta en ciertos meses tuvo un ligero incremento, sobre todo en los meses del período seco lo que pudo ocasionar una pequeña reducción en el nivel del agua.
-

El Lago de Valencia es el segundo lago de mayor importancia del país, actualmente el Lago vive un mal momento debido a su alto nivel de contaminación que afecta la salubridad de la población, además de sus necesidad de reubicarse. Lo que origina la necesidad de estudiar estas variables para poder determinar el comportamiento del lago y así predecir sus condiciones climáticas.

Bibliografía

- [1] Guevara D., J. M. (2004). **Meteorología**, Caracas: Consejo de Desarrollo Científico y Humanístico, Segunda Edición.
- [2] Guevara E., Guevara J. and García E. (2008). **Régimen climático y patrón espacial de las lluvias en la cuenca del Lago de Valencia, Venezuela**. Revista de Climatología. Vol. 8:29-40. Disponible en: <http://webs.ono.com/reclim3/reclim08c.pdf>.
- [3] Hernández J. (2009) **Serie de Tiempo**. Universidad Central de Venezuela, clases 2-14.
- [4] Martelo M. (2003) **La precipitación en Venezuela y su relación con el sistema climático**. [en línea]. [Consulta: 10 Febrero 2014]. Caracas. Disponible en: http://www.inameh.gob.ve/documentos/precipitacion_venezuela_relacion_sistema%20climatico.pdf.
- [5] Mundaray R. **Análisis de Series Temporales de la Amplitud Térmica para 7 localidades de la zona Centro-Norte-Costera de Venezuela con datos del Modelo de Reanalysis-NCAR** [en línea]. Caracas. [Consulta: 24 Febrero 2014]. Disponible en: http://www.inameh.gob.ve/documentos/Analisis_de_Series_Temporales_Amplitud_Termica.pdf.
- [6] Murray R., Schiller, J. & Srinivasan, A. (2009). **Probabilidad y Estadística**. México: McGraw Hill, Tercera Edición, pag. 265-274.

-
- [7] Pacheco P. (2008) **Análisis Estadístico de Datos de Contaminación Ambiental del Lago de Valencia**. Tesis Msc en Modelos Aleatorios, Univ. Central de Venezuela, Fac. Ciencias.
- [8] Shumway R. & Stoffer, D. (2011). **Time Series Analysis and Its Applications with R examples**. 3rd edition. Springer. pag. 47-121.
- [9] Umpiérrez O. **ANÁLISIS DE IMPACTOS HIDROLÓGICOS DEL "EL NIÑO"-COMPILACIÓN, ESTUDIOS E INVESTIGACIÓN**. [en línea]. [Consulta: 26 Febrero 2014]. Chile. Disponible en: http://dgasatel.mop.cl/ar3/docs/Venezuela/CC_IMPACTOS-EL%20NI%C3%91O-ARIII.pdf. pag. 27-31.
-