

UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE LOS VALLES



“SERIES TEMPORALES INVESTIGANDO VARIACIONES
ALEATORIAS DE RADIACIÓN SOLAR”

Tesis que para obtener el título de Doctorado en Ciencias Físico Matemáticas
con orientación en Matemáticas

Presenta:

Gilberto Rubén Aguirre Salas

Director:

Dr. Rubén Sánchez Gómez

Co-director:

Dr. Jaime Alcalá Gutiérrez

Ameca, Jalisco, enero de 2018

UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE LOS VALLES



“SERIES TEMPORALES INVESTIGANDO VARIACIONES
ALEATORIAS DE RADIACIÓN SOLAR”

Tesis que para obtener el título de Doctorado en Ciencias Físico Matemáticas
con orientación en Matemáticas

Presenta:

Gilberto Rubén Aguirre Salas

Director:

Dr. Rubén Sánchez Gómez

Co-director:

Dr. Jaime Alcalá Gutiérrez

Ameca, Jalisco, enero de 2018

Este espacio es para agregar el oficio director

Este espacio es para agregar el oficio de firma de lectores

Agradecimientos

Mi más profundo agradecimiento al *Centro Universitario de los Valles* por esta oportunidad.

Al Dr. *Rubén Sánchez Gómez*, por su paciencia infinita y apoyo incondicional sin el cual no hubiera sido posible la conclusión de este proyecto.

A los *Investigadores* que revisaron este trabajo por sus comentarios y sugerencias que abonaron en su mejora.

A las *Autoridades* del Centro Universitario por su apoyo incondicional en todos los trámites y requisitos necesarios para culminar este trabajo.

Al personal del IAM y del Observatorio Colomos, Servicio Meteorológico Nacional, en particular a la Srta. *Viridiana López* y al Ing. *Rubén Pantoja* por facilitarnos los datos que fueron la base de este trabajo.

A los Compañeros, Amigos y Familia cuyo apoyo moral fue decisivo.

Dedicatorias

A mi Madre (†)

A mi familia

Índice general

Resumen	1
1. Introducción	2
2. Marco teórico	7
2.1. Análisis de series de tiempo	8
2.2. Tendencia temporal	11
2.3. Análisis Espectral	13
2.4. Magnitudes aleatorias	16
2.4.1. Familias de soporte no negativo	17
2.4.2. Métodos de estimación puntual	27
3. Estudio de simulación	34
3.1. Efectos periódicos y componente aleatoria	35
3.2. Factor de tendencia y componente aleatoria	44
3.3. Efectos periódicos y tendencia	46
4. Series de tiempo de radiación solar	49
4.1. Descripción general de la zona de estudio	50
4.1.1. Condiciones geográficas	51
4.1.2. Condiciones meteorológicas	52

4.2. Series de datos	53
4.2.1. Análisis descriptivo de las series de tiempo	56
4.3. Discusión de resultados	58
5. Conclusiones	64
A. Programas utilizados	67
Índice de figuras	71
Índice de tablas	73
Bibliografía	74

Resumen

En este trabajo se presenta un algoritmo confiable para analizar cualquier magnitud impredecible en series temporales observadas en la naturaleza, identificando las componentes típicas como efectos de tendencia temporal, componentes periódicas y asociación entre lecturas consecutivas; así como el impacto que se representa de las interacciones de estos factores.

Se muestran los resultados de un estudio de simulación para verificar la confiabilidad de los algoritmos propuestos y se aplica la metodología a dos conjuntos de datos de radiación solar observada en la Zona Metropolitana de Guadalajara: uno de registros levantados en las instalaciones del Instituto de Astronomía y Meteorología de la Universidad de Guadalajara y el segundo con datos tomados en el Observatorio Colomos.

Las bases de datos utilizadas corresponden a once años de lecturas cada diez minutos en dos estaciones automáticas: la primera en un área urbana, al interior de la ciudad y la segunda en una área forestal. En ambos casos se revisaron por separado efectos periódicos, dependencia temporal en lecturas consecutivas, las variaciones a corto y largo plazo.

Como se puede observar en el contenido de este trabajo, de los resultados obtenidos sobresale que las variaciones a largo plazo muestran evidencia de oscurecimiento en el área urbana y ocurre lo contrario en la zona semi-urbanizada. Además, las variaciones a corto plazo dan evidencia para afirmar que la interacción de factores como efectos periódicos o de tendencia a periodos largos de tiempo deforman el comportamiento de su distribución de probabilidad, encontrando que el modelo Weibull muestra el mejor ajuste a los datos observados en ambas zonas de estudio.

Capítulo 1

Introducción

La radiación solar es una de las fuentes energéticas de origen renovable más importante, esencialmente porque representa una de las mejores alternativas para reducir el impacto ambiental ante el uso indiscriminado de recursos convencionales, tales como petróleo, carbón natural, carbón mineral, gasolina o diesel (SENER, 2013 [1]; REN21, 2010 [2]; REN21, 2014 [3]). Además de que se ha convertido en una fuente importante y rentable con el desarrollo de colectores fotovoltaicos y calentadores térmicos, cada vez más innovadores y eficientes.

Ya sea por factores económicos, por demanda energética o por el desempeño de los colectores fotovoltaicos; cualquier inversión en energía solar requiere de una estimación confiable del recurso solar en el punto de interés (Gueymard y Wilcox, 2011, [4]; Vignola, Grover, Lemon y McMahan, 2012, [5]), principalmente porque algunas moléculas del aire, como aerosoles y nubes, pueden absorber o dispersar una parte importante de la radiación que se recibe en la tierra, provocando que se pierda un porcentaje considerable y a su vez, reduciendo el ingreso de los rayos solares a los colectores. Esto se traduce en un comportamiento impredecible en la cantidad de radiación recibida; entendiendo por impredecible al hecho de que no se puede conocer con certeza la cantidad de radiación solar en un momento específico de tiempo.

Conocer y comprender esta variabilidad impredecible de la radiación solar es de vital importancia para disponer de elementos suficientes que permitan estimar el potencial de cualquier sistema que convierte la radiación solar en energía voltaica. Por ejemplo, Soubdhan, Emilion y Calif (2008, [6]) demuestran que las discrepancias en intervalos

de algunos segundos o minutos, son decisivas en las diferencias observadas en estudios de simulación. Comprobaron de forma experimental en un laboratorio electrónico, que estas variaciones pueden alcanzar hasta $700W/m^2$ a causa de condiciones adversas que interfieren al flujo de radiación solar. En otras palabras, estas variaciones repercuten negativamente en el desempeño de los colectores y en el caso particular de climas tropicales, la complejidad de estos sistemas que interactúan en la atmósfera se incrementa (McKnight y Hess, 2000, [7]), enfrentando escenarios impredecibles en las variaciones de radiación solar por los cambios bruscos en las condiciones meteorológicas locales.

Entender esta componente impredecible de la radiación solar ha sido de interés en diversos trabajos, en los que se destaca el uso de una función de densidad de probabilidad con soporte no negativo o mezclas de dos o más funciones. Por ejemplo, Assunção, Escobedo y Oliveira (2003, [8]) modifican los datos de radiación solar en un índice de difusión y le ajustan una distribución Beta a registros observados en el área de Botucatu, Brasil; presentan además un análisis detallado del índice para diferentes rangos en el intervalo $[0,1]$ y en un segundo ejemplo, Soubdhan, Emilion y Calif (2008, [6]) combinan distribuciones Dirichlet investigando el comportamiento diario de un índice de claridad. En ambos trabajos argumentan un desempeño bimodal de la distribución de las observaciones de radiación solar, es decir, encuentran dos o más picos de frecuencia bien definidos; y los autores señalan que tal comportamiento se puede atribuir a lecturas por bloques de registros en condiciones de cielo despejado, contra escenarios con atmósfera difusa. Asimismo, Jurado, Caridad y Ruiz (1995, [9]) observan en su trabajo de investigación que los registros de radiación solar presentan un comportamiento bimodal en la distribución de los datos para intervalos menores a 60 minutos.

Otra publicación interesante es el trabajo de Chang (2010, [10]) quien publica un análisis similar de lecturas observadas en seis estaciones meteorológicas instaladas en Taiwan, buscando encontrar un modelo de probabilidad que tenga el “mejor” ajuste a los datos. En su análisis comparativo investiga el desempeño de las distribuciones Weibull, logística, normal y lognormal, pero no detecta ese comportamiento bimodal en la distribución de los datos y así lo reporta.

Esta diferencia en los resultados publicados, en donde unos investigadores detectan

distribuciones bimodales y otros no; abre una incógnita sobre las metodologías utilizadas, ya que es contradictorio que conjuntos de datos de una misma magnitud, sigan comportamientos radicalmente diferentes. Más aún, estos resultados discordantes dan la posibilidad de considerar posibles errores en los análisis previos o bien, suponer que se necesita implementar una metodología diferente para entender el comportamiento de estas magnitudes.

Por tanto, qué opciones se tienen para generar una propuesta general que atienda este problema:

- i) Por un lado, se sabe que cualquier conjunto de datos observados en la naturaleza pueden presentar características tales como: una componente *regular* o *determinística*, haciendo referencia a una parte de los datos que no muestra cambios impredecibles. La interacción de efectos *estacionales* o *periódicos* que obedecen ciclos temporales. A manera de ejemplo, trabajos como los de Hayashi (1981, [11]), Zhao y Qu (1995, [12]), Wilks (2002, [13]), Sánchez *et al.* (2010, [14]), Bendat y Pearsol (2010, [15]), muestran evidencia del efecto de componentes *periódicas* o *estacionales* en conjuntos de datos observados en la naturaleza, ya sea por influencia de factores sistemáticos o externos y obedeciendo ciclos temporales semidiurnos, diurnos, semianuales o anuales, entre otros. Un factor de *tendencia* en espacio o tiempo, observable por cambios en el comportamiento de la variable a largo plazo; incluso la integración de los efectos de tendencia en espacio y tiempo de forma simultánea. Trabajos como el de Bhattacharya *et al.* (1983, [16]), Sánchez (2001, [17]), Martínez y Fernández (2004, [18]), Mudelsee (2010, [19]), Gluhovsky (2011, [20]) entre otros, muestran evidencia de un comportamiento de tendencia en series de datos de variables atmosféricas. La influencia entre lecturas consecutivas en espacio o tiempo, que corresponden a un nivel de asociación entre lecturas, en donde sobresalen términos como persistencia de un fenómeno natural. Además de una componente *impredecible* que se observa por cambios bruscos en periodos cortos de tiempo (de segundos a minutos), que puede ser interpretados como un comportamiento de tipo estocástico o aleatorio. En donde, además de los trabajos señalados previamente, estudios como el de Wilks y Wilby (1999, [21]), Karaki *et al.* (1999, [22]), Wilks (2002, [13]), Assunção *et al.* (2003, [8]), Villanueva *et al.* (2013, [23]) por mencionar algunos de fácil acceso.

De este modo, estas variables observadas en la naturaleza pueden incorporar uno o todos estos factores, así como resultados de la interacción de dos o más y este comportamiento puede deformar la componente impredecible de radiación solar. Esto se conecta directamente con el problema que se busca analizar, ya que los conjuntos de datos de radiación solar pueden acumular la interacción de una o más de estas componentes y a su vez, se tiene un argumento probable que provoque efectos bimodales en donde realmente no existen,

Por tanto, analizar o modelar conjuntos de registros observados de una magnitud física, implica la necesidad de conocer e interpretar sus componentes de forma consiente e ignorar cualquiera de ellas provoca conclusiones basadas en resultados limitados o justificaciones no confiables (ver capítulo 3. *Estudio de simulación* de este trabajo).

ii) Por otro lado, revisando la teoría disponible en la literatura (Smith, 1999, [25]; Bloomfield, 2000, [26]; Chatfield, 2003, [27]), el análisis de conjuntos de datos observados en distintos puntos de tiempo, corresponde al concepto de *Series de Tiempo* y algunos de estos registros son: lecturas de radiación solar, volumen precipitado, humedad relativa, nubosidad, presión atmosférica o dirección y velocidad del viento, entre muchas más. De modo que, el reto principal al analizar estas series de tiempo se puede reducir a identificar las componentes que la forman, así como la interacción de cada uno de estos efectos, estableciendo el fundamento teórico que permita sustentar el análisis de estos datos y tomar en cuenta una medición del error que se comete si se ignora la influencia de alguna de estas componentes.

De esta forma, en el presente trabajo se pretende investigar el efecto de los factores de tendencia temporal, efectos periódicos y asociación entre lecturas consecutivas en la componente aleatoria de series de tiempo de radiación solar, suponiendo que la interacción de estas factores pueden alterar y deformar la distribución de probabilidad de la componente aleatoria.

El objetivo general de este estudio es identificar un algoritmo confiable que permita encontrar un modelo de probabilidad adecuado para el factor aleatorio de series temporales bajo el efecto de componentes periódicas y de tendencia temporal; así como el error que comete un usuario si ignora alguna de estas componentes de la serie.

Los objetivos particulares para lograrlo son los siguientes:

1. Detectar e identificar efectos de tendencia temporal.
2. Investigar y analizar periodicidades ocultas en series temporales.
3. Verificar, mediante un estudio de simulación, el efecto de las componentes de tendencia y de efectos periódicos en el factor impredecible de la serie.
4. Identificar un modelo de probabilidad que represente la componente impredecible en series de datos.

Finalmente, este trabajo se estructura de la siguiente forma: en el capítulo 2 se presenta una revisión de las herramientas matemáticas propuestas en la literatura, así como del modelo sugerido en este trabajo, incorporando tanto componentes de tendencia a largo plazo, efectos periódicos con distintas frecuencias, efectos de asociación entre lecturas consecutivas y se analizan las variaciones impredecibles a corto plazo; el capítulo 3 contiene los resultados de un estudio de simulación, para identificar los errores que se cometen al ignorar alguno de los efectos típicos en series temporales; en el capítulo 4 se revisa el comportamiento de las series de tiempo de radiación solar observada en dos puntos de la Zona Metropolitana de Guadalajara, que por sus condiciones geográficas y meteorológicas, presenta fuertes variaciones atribuibles a la dinámica de la atmósfera en climas tropicales; en el capítulo 5 se muestran los resultados obtenidos al aplicar el análisis propuesto, y finalmente, el capítulo 6 contiene las conclusiones y trabajos futuros en esta línea de investigación.

Capítulo 2

Marco teórico

El concepto fundamental en este trabajo se conoce como *series de tiempo* y con él se refiere a conjuntos de observaciones recolectadas en forma secuencial para distintos puntos de tiempo, que pueden o no estar igualmente espaciados (Smith, 1999, [25]; Brockwell y Davis, 2002, [28]). Comúnmente se representa mediante el conjunto de observaciones

$$\{x_{t_1}, x_{t_2}, \dots, x_{t_n}\}$$

registradas en puntos discretos de tiempo $\{t_1, t_2, \dots, t_n\}$, para $t_1 < t_2 \leq \dots \leq t_n$.

En donde una lectura observada x_{t_k} en un momento del tiempo t_k , puede tomar un valor de un conjunto de todos los resultados posibles. Si denotamos con ω al posible resultado y a Ω al conjunto de todos los resultados posibles; revisando el concepto de *variable aleatoria* en cualquier texto de teoría de probabilidad o probabilidad y estadística (*v. gr.* Stuart & Ord, 1994 [29]; Casella & Berger, 2002 [31]; Rohatgi, 2005 [30]), se puede ver que X_{t_k} está sujeto a una regla que asigna a cada salida ω del experimento un número $X_{t_k}(\omega)$.

Por otro lado, revisando la definición de Papoulis (1991 [32], pág. 285) un *proceso estocástico* es una regla que asigna a cada $\omega \in \Omega$ una función $X(t, \omega)$. Así, “un proceso estocástico es una familia de funciones que dependen de t y ω , en donde el dominio de t es el conjunto de los números reales y el dominio de ω es el conjunto de todos los resultados posibles de un experimento aleatorio” (tomado de: Papoulis, 1991 [32], pág. 285).

2.1. Análisis de series de tiempo

De manera formal, para un conjunto no vacío $R \subset \mathbb{R}$ y un espacio de probabilidad (Ω, \mathcal{A}, P) , en donde \mathcal{A} es una σ -álgebra de $\Omega \neq \emptyset$ y P es una medida de probabilidad, la función

$$X : R \times \Omega \rightarrow \mathbb{R}$$

es un proceso estocástico.

Si $R = \{t : a < t < b\}$ para $a, b \in \mathbb{R}$, entonces X_{t_k} es un *proceso en tiempo continuo* y si $R = \{t_n\}_{n \in \mathbb{Z}}$ entonces X_{t_k} es un *proceso en tiempo discreto*. Para un t_k fijo, $X(t_k, \omega) : \Omega \rightarrow \mathbb{R}$ es una variable aleatoria para cada $t_k \in R$ (ver Bailey, 1964 [33]; Kannan, 1979 [34]; Papoulis, 1991 [32]; Tijms, 2003 [35];). Para un ω_p fijo, $X(t, \omega_p) : R \rightarrow \mathbb{R}$ es una realización o una ruta muestral del procesos estocástico para cada $\omega_p \in \Omega$ (ver Bailey, 1964 [33]; Kannan, 1979 [34]; Papoulis, 1991 [32]; Tijms, 2003 [35];). Para t_k y ω_p fijos, $X(t_k, \omega_p)$ es un número real.

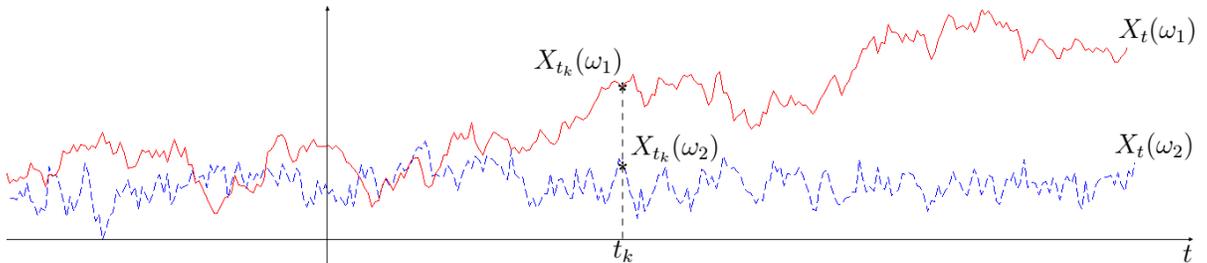


Figura 2.1: Una representación gráfica de dos rutas muestrales de un proceso estocástico.

Además, X_t es un proceso de espacio de estados discreto si Ω es un conjunto contables o finito contable y en caso contrario, será un proceso de espacio continuo de estados.

Por otro lado, un proceso estocástico se puede clasificar con base en la familia de probabilidad que sigue la variable ω ; de modo que, X_t será Gaussiano, si para cualquier conjunto de puntos t_1, \dots, t_n la distribución conjunta de X_{t_1}, \dots, X_{t_n} es una distribución Gaussiana multivariada.

Para fines prácticos, en este trabajo t representa “tiempo”, por lo que no toma valores negativos, con la posibilidad de iniciar en $t_1 = 0$ y además, t es una variable

discreta dado que la serie consiste de lecturas en puntos específicos del tiempo y por comodidad, se denotará con X_t al proceso $X_t(\omega)$.

Así, las lecturas temporales de magnitudes atmosféricas como: humedad relativa, presión, *radiación solar* y velocidad - dirección del viento, entre otras, son rutas muestrales de procesos estocásticos de espacio de estados continuo, en tiempo discreto.

Un proceso arbitrario $\{X_t, t = 1, 2, \dots\}$ se puede clasificar como *fuertemente estacionario* o en forma equivalente, como un proceso *estrictamente estacionario*, si los registros $(X_{t_1}, X_{t_2}, \dots, X_{t_k})$ tienen la misma distribución de probabilidad que

$$(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_k+h}),$$

para cualquier incremento de tiempo $h \geq 0$, en otras palabras, se tiene que las propiedades estadísticas del proceso son invariantes ante desplazamientos temporales, o bien

$$P(X_{t_1+h}, \dots, X_{t_n+h}) = P(X_{t_1}, \dots, X_{t_n}),$$

en donde $P(\cdot)$ representa la medida de probabilidad de (Ω, \mathcal{A}, P) .

Haciendo un pequeño paréntesis, se sabe que la *esperanza matemática* \mathbb{E} de la función $g(X_t)$ para la variable aleatoria X_t , se define como

$$\mathbb{E}[g(X_t)] = \int g(X_t) dP(X_t) \quad (2.1)$$

en donde $P(X_t)$ es la función de distribución de la variable aleatoria X_t .

- Si $g(X_t) = X_t$, $\mathbb{E}X_t$ es el *valor esperado* (la media o el valor promedio) de X_t , $g(X_t) = (X_t - \mathbb{E}X_t)^2$, $\mathbb{E}[g(X_t)]$ es la *varianza* de X_t .
- Para dos variables aleatorias X_t y Y_t , $\mathbb{E}(X_t - \mathbb{E}X_t)(Y_t - \mathbb{E}Y_t)$ es la *covarianza* de las variables (X_t, Y_t) y se denota por $Cov(X_t, Y_t)$, mientras que, para dos momentos de tiempo t y $t + h$, $\mathbb{E}(X_t - \mathbb{E}X_t)(X_{t+h} - \mathbb{E}X_{t+h})$ es la *autocovarianza* y se denota por $\gamma(h)$ o $Cov(h)$.

De esta forma, un proceso arbitrario X_t será *débilmente estacionario* (*estacionario de segundo orden* o *estacionario en sentido amplio*), si para cualquier incremento en el

tiempo $h \geq 0$, se cumple que

$$\begin{aligned}\mu &= \mathbb{E}X_t, \\ \gamma(h) &= \mathbb{E}\{(X_t - \mu)(X_{t+h} - \mu)\} = Cov(h),\end{aligned}$$

en otras palabras, para procesos débilmente estacionarios $\mathbb{E}X_t$ es constante en el tiempo y $\gamma(h)$ depende sólo del incremento temporal.

De este modo, si el valor promedio de un proceso estocástico X_t no es constante para distintos momentos del tiempo o bien, si su autocovarianza no depende únicamente del incremento temporal h , entonces, se dice que X_t no es estacionario.

Por tanto, para garantizar que un proceso es estacionario, el problema se reduce a comprobar si el promedio de las lecturas es igual para distintos periodos de tiempo y que su autocovarianza depende únicamente del incremento temporal en distintos periodos. No obstante, estas condiciones no se cumplen en general. Por ejemplo, las lecturas de velocidad del viento en un punto geográfico para una serie anual muestra diferencias en sus promedios mensuales, afectando directamente al valor esperado que presentaría cambios mensuales y por tanto, los promedios serán diferentes. Por otro lado, una lectura horaria de radiación solar puede mostrar fuerte asociación con sus lecturas previas por el comportamiento natural de rotación terrestre, afectando su autocovarianza y por tanto, se tendrían desempeños diferentes para distintos incrementos temporales.

Además, es evidente que si una serie de tiempo presenta el efecto de tendencia temporal, será monótona creciente (o decreciente) conforme avanza el tiempo y entonces, al calcular promedios μ_t para diferentes momentos de tiempo, se tendrá que

$$\mu_{t_1} \leq \mu_{t_2} \leq \dots \leq \mu_{t_n} \leq \dots \quad \left(\mu_{t_1} \geq \mu_{t_2} \geq \dots \geq \mu_{t_n} \geq \dots \right)$$

y por tal, de forma similar al caso de series con efectos periódicos o de asociación entre lecturas consecutivas, la serie será no estacionaria.

2.2. Tendencia temporal

Cualquier serie temporal de un proceso estocástico no estacionario, presentará cambios en su promedio μ , diferencias en su variabilidad γ_0 o bien en ambas magnitudes, para distintos valores de t . Así, en caso de que la media tiende a crecer - decrecer a largo plazo, es posible que la serie esté bajo la influencia de una componente de tendencia temporal y se debe implementar un análisis que permita identificar si existe evidencia de tendencia en la serie.

Para estimar la tendencia en una serie temporal $\{X_t, t = 1, 2, \dots, n\}$, sin pérdida de generalidad se puede suponer que se tienen observaciones de una trayectoria de un proceso estocástico no estacionario bajo la influencia de una *componente de tendencia* T_t y un residual aleatorio E_t (o error aleatorio) que representa la diferencia de ajuste entre las lecturas disponibles X_t y la componente de tendencia T_t , bajo un modelo de la forma

$$X_t = T_t + E_t. \quad (2.2)$$

Así, para estimar el efecto de tendencia temporal en la serie, se puede considerar la *Tendencia determinística* (Brockwell y Davis, 2002, [28]) dada por una función T_t que puede ser algebraica (lineal, no lineal o racional) o trascendental (exponencial, logarítmica). En el caso más simple, el modelo de tendencia tiene la forma

$$T_t \equiv \alpha + \beta t, \quad (2.3)$$

en donde α y β son los parámetros de la recta de ajuste de la función de tendencia T_t y representan la ordenada y pendiente respectivamente, desde un punto de vista geométrico y además β representa la tasa de crecimiento de la magnitud T_t por unidad de tiempo.

De modo que, para una serie temporal $\{X_t, t = 1, 2, \dots, n\}$, el problema se reduce a resolver el sistema dado por

$$\begin{array}{l} X_1 \approx \alpha + \beta t_1 \\ X_2 \approx \alpha + \beta t_2 \\ \vdots \\ X_n \approx \alpha + \beta t_n \end{array} \equiv \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \approx \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

en donde \mathbf{X} y \mathbf{M} son un vector columna y una matriz, respectivamente.

Así, para estimar el modelo de ajuste de tendencia, en forma matricial se puede implementar el método de mínimos cuadrados (Rohatgi, 2003, [30]; Sánchez *et al.*, 2010, [14]) y denotando con \mathbf{M}' a la matriz transpuesta de \mathbf{M} , se tiene que

$$\mathbf{X} = \mathbf{M} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \iff \mathbf{M}'\mathbf{X} = \mathbf{M}'\mathbf{M} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}.$$

Multiplicando en ambos lados por la matriz inversa $(\mathbf{M}'\mathbf{M})^{-1}$, resulta

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = (\mathbf{M}'\mathbf{M})^{-1}(\mathbf{M}'\mathbf{X}) \quad (2.4)$$

en donde $\hat{\alpha}$ y $\hat{\beta}$ son los estimadores de los parámetros α y β respectivamente.

De la ecuación (2.2), el error de ajuste del modelo de tendencia quedaría expresado como

$$E_t = X_t - (\hat{\alpha} + \hat{\beta}t).$$

Efectos autoregresivos

En consistencia con los modelos lineales, en el análisis de series de tiempo es común suponer alguna medida de asociación entre la lectura tomada en un tiempo X_t y las lecturas previas $X_{t-1}, X_{t-2} \dots, X_{t-k}$ que corresponden a las lecturas con 1, 2 y hasta k retrasos, bajo el supuesto de que el valor de la lectura actual depende en cierta medida de los valores registrados en las lecturas previas de la misma magnitud. Así, lo usual es reportar representaciones gráficas del coeficiente de autocorrelación γ_h , para $h = 1, 2, \dots, k$ retrasos, y mostrar evidencia de efectos de asociación entre la lecturas actual y sus registros previos.

Evidentemente, si una serie está influenciada por una componente periódica, la gráfica de autocorrelación mostrará dicho efecto y posiblemente se tenga una idea errónea, suponiendo que existe evidencia de autocorrelación cuando el análisis muestra la influencia de efectos periódicos en la serie, es por ello que la definición básica de función de autocorrelación se establece sobre series de tiempo estacionarias (Smith,

1999, [25]; Brockwell y Davis, 2002, [28]; Chatfield, 2003, [27]).

Un modelo que incorpora esta asociación temporal de las observaciones (Smith, 1999, [25]) es el *modelo autoregresivo de orden k* , dado por

$$X_t = \sum_{j=1}^k a_j X_{t-j} + E_t \quad (2.5)$$

en donde E_t es el error de ajuste del modelo con parámetros a_1, \dots, a_k ; de modo que si se tiene evidencia de autocorrelación entre las lecturas temporales de radiación solar, una componente autoregresiva asegura un modelo adecuado para representar la dinámica de la serie de datos.

2.3. Análisis Espectral

El *análisis espectral*, *análisis armónico* o *análisis de Fourier* de una serie de tiempo es una herramienta estadística que permite descomponer la información disponible en una serie temporal, en sumas de curvas cosenoidales, *i.e.*, se aplica para medir fluctuaciones periódicas en el tiempo. Por ejemplo, en el caso más simple se puede disponer de datos periódicos que coinciden con una onda coseno dada por

$$X_t = R \cos(2\pi ft + \phi) + E_t,$$

en donde R es la amplitud, ϕ es la fase, f es la frecuencia medida en ciclos por unidad de tiempo y todas estas magnitudes son desconocidas en una serie temporal de valores observados.

Revisando el problema y suponiendo que la serie de tiempo $\{X_t\}_{t=0}^n$ tiene el efecto de una sola frecuencia f , resulta que

$$\begin{aligned} X_t &= \mu + R \cos 2\pi(ft + \phi) + E_t \\ &= \mu + R(\cos 2\pi\phi \cos 2\pi ft - \sin 2\pi\phi \sin 2\pi ft) + E_t, \end{aligned}$$

para un error de ajuste E_t .

Definiendo $A = R \cos 2\pi\phi$ y $B = -R \sin 2\pi\phi$, la serie X_t se puede expresar como

$$X_t = \mu + A \cos 2\pi ft + B \sin 2\pi ft + E_t$$

en donde $R = \sqrt{A^2 + B^2}$ es la amplitud de la onda y $\tan 2\pi\phi = -B/A$.

La descripción de una solución completa para diferentes valores de A y B , considerando que el dominio de la función arcotangente oscila entre $[-\pi/2, \pi/2]$ se puede describir como:

$$2\pi\phi = \begin{cases} \arctan(-B/A), & \text{si } A > 0, \\ \arctan(-B/A) - \pi, & \text{si } A < 0, B > 0, \\ \arctan(-B/A) + \pi, & \text{si } A < 0, B \leq 0, \\ -\pi/2 & \text{si } A = 0, B > 0, \\ \pi/2 & \text{si } A = 0, B < 0. \end{cases}$$

Además, si la frecuencia f es conocida, el problema se reduce al caso del modelo lineal clásico, en donde se busca minimizar la suma del cuadrado de los errores, dada por

$$s(\mu, A, B) = \sum_{t=0}^n E_t^2 = \sum_{t=0}^n (X_t - \mu - A \cos 2\pi ft - B \sin 2\pi ft)^2,$$

cuya solución, por mínimos cuadrados y aprovechando la ortogonalidad de las funciones senoidales (Bloomfield, 2000, [26]; Chatfield, 2003, [27]; Sánchez *et al.*, 2010, [14]), corresponde a

$$\begin{aligned} \hat{\mu} &= \bar{X}_t = \frac{1}{n} \sum_{t=0}^n X_t \\ \hat{A} &= \frac{2}{n} \sum_{t=0}^n X_t \cos 2\pi ft \\ \hat{B} &= \frac{2}{n} \sum_{t=0}^n X_t \sin 2\pi ft \end{aligned}$$

en donde $\hat{\mu}$, \hat{A} y \hat{B} son los estimadores de los parámetros μ , A y B respectivamente.

En el caso general para m frecuencias ($m > 1$), se tiene que

$$\begin{aligned}\widehat{\mu} &= \bar{X}_t \\ \widehat{A}_j &= \frac{2}{n} \sum_{t=0}^n X_t \cos 2\pi f_j t \quad \text{para } j = 1, 2, \dots, m \\ \widehat{B}_j &= \frac{2}{n} \sum_{t=0}^n X_t \sen 2\pi f_j t\end{aligned}\tag{2.6}$$

proporciona una estimación de los parámetros μ y (A_j, B_j) con $j = 1, 2, \dots, m$.

Por otro lado, en una serie temporal de tamaño n , con registros puntuales de un fenómeno arbitrario sujeto a perturbaciones aleatorias, si la amplitud de la componente aleatoria es mayor que la amplitud del efecto periódico, entonces las frecuencias que participan en la onda pueden pasar desapercibidas en el conjunto de datos; por ello, usualmente se dice que las periodicidades están “ocultas” y en tal caso, se suponen frecuencias enteras, múltiplos de $1/n$ que corresponden a periodos n/j y se repiten en toda la extensión de los datos (Bloomfield, 2000). Estos armónicos respecto a la extensión de los datos $f_j = j/n$, para j entero, se conocen como frecuencias de Fourier.

Si los efectos periódicos son desconocidos, se implementan las frecuencias de Fourier y con ellas se pueden determinar aquellas que son significativas bajo cierto nivel de confianza. Para esto, suponiendo que los errores son independientes y distribuidos de forma Gaussiana con media cero y varianza σ^2 y entonces, las estimaciones de los parámetros son incorrelacionadas con varianzas $2\sigma^2/n$, $2\sigma^2/n$, σ^2/n para \widehat{A}_j , \widehat{B}_j y $\widehat{\mu}$, respectivamente.

Además, dado que la amplitud $R_j = \sqrt{A_j^2 + B_j^2}$ mide la intensidad de la oscilación correspondiente a cada frecuencia f_j a partir de los datos, una representación gráfica de R_j contra f_j muestra el comportamiento de las amplitudes para las frecuencias; con la posibilidad de identificar aquellas que son significativas y asegurando magnitudes no negativas lo usual es utilizar

$$I(f_j) = nR^2(f_j)\tag{2.7}$$

en donde $I(f_j)$ se conoce como el *periodograma* de la serie.

Bajo el supuesto de que los errores se ajustan a un modelo Gaussiano, se tiene que las estimaciones $(\widehat{A}_j, \widehat{B}_j)$ también siguen una distribución Gaussiana y por teoría estadística, se tiene que $\frac{I(f_j)}{2\sigma^2}$ se distribuye χ^2 con dos grados de libertad (Stuart y Ord,

1994,[29]). Por lo tanto, para establecer las frecuencias significativas, se implementa un prueba de hipótesis verificando si hay evidencia para suponer que $R_j^2 = 0$ contra la alternativa $R_j^2 > 0$. Una excelente referencia para revisar la teoría de estos argumentos se puede encontrar en [26] y [25].

2.4. Magnitudes aleatorias

Cuando se mide intensidad o concentración de una magnitud física, lo común es encontrar series de tiempo con valores no negativos y en el caso de magnitudes observables en la naturaleza, *v. gr.*, precipitación, presión atmosférica, radiación solar y velocidad del viento; todas presentan valores de magnitudes mayores o iguales a cero y por lo tanto, los modelos de variable aleatoria con soporte no negativo deben ser el tema de estudio.

A continuación, se muestra una lista con algunos trabajos que sugieren modelos de probabilidad con soporte no negativo, en el entendido de que en la siguiente sección se incluye una descripción básica de los modelos que se mencionan en seguida.

- Karaki *et al.* (1999, [22]) muestra un estudio probabilístico del desempeño de sistemas de conversión de energía solar suponiendo un modelo *beta* de soporte $[0, 1]$ para los datos y considera que la función de densidad Weibull representa adecuadamente los datos de velocidad del viento;
- Assunção *et al.* (2003, [8]) señala que la densidad *beta* es un modelo adecuado para analizar la componente irregular de datos de radiación solar registrados en Botucatu, Brasil;
- Chang (2010, [10]) en cambio, un poco más precavido, muestra un análisis en el que incorpora la densidad Weibull, normal y logística y log-normal en su estudio, concluyendo que la densidad log-normal muestra mejor desempeño en términos de error cuadrado medio y de la prueba Kolmogorov-Smirnov;
- mientras que Villanueva *et al.* (2013, [23]) evalúan el desempeño de la distribución Weibull multivariada para representar de manera conjunta la dirección y la velocidad del viento.

Ninguno de estos autores da evidencia de un análisis o de incorporar en su modelo efectos de tendencia o periodicidad en las series temporales analizadas. Más aún, los histogramas que muestran en sus trabajos dan evidencia de datos bimodales o de colas pesadas (en ambas direcciones), lo que deja evidencia de efectos de componentes cíclicas en los datos.

De este modo, para indagar el comportamiento bimodal en las magnitudes de radiación solar, así como la función de probabilidad adecuada para la componente aleatoria de la serie, a continuación se muestra una descripción general de los modelos antes mencionados.

2.4.1. Familias de soporte no negativo

A manera de ejemplo, ajustar una función de densidad de soporte real a registros de una magnitud no negativa, suponiendo que puede tener un buen ajuste a los datos, lleva al usuario a errores como el que se representa en la figura 2.2. En la figura se puede observar el ajuste de un modelo Gaussiano a lecturas no negativas, en donde sobresale a simple vista una porción de la población sobre el lado negativo del eje horizontal, sin lecturas disponibles en ese cuadrante.

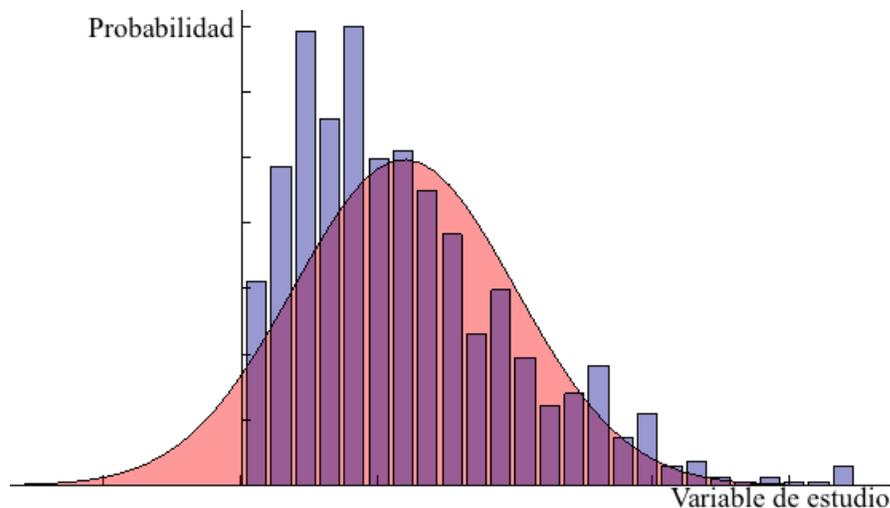


Figura 2.2: Representación gráfica de un esquema que muestra el error que se comete al suponer incorrectamente un modelo.

En este caso, el diagrama de barras se obtiene a partir de datos simulados de una densidad Weibull y el ajuste del modelo Gaussiano se obtiene a partir de los datos simulados. Además, de la gráfica se tiene que aproximadamente el 15 % de la población Gaussiana está del lado negativo en el eje x y en el lado positivo, gran parte de la muestra queda fuera del modelo.

Más aún, las características del modelo, tales como parámetros de localidad (promedios, máximos, mínimos, ...) quedan definidos a partir de los datos, por lo que cualquier intento de “mover” la figura no corresponde al modelo estimado con los datos, en otras palabras, no se tiene la posibilidad de reducir la fracción de la población observada en el lado negativo.

De este modo, cualquier análisis de datos que incorpore una componente impredecible o aleatoria, implica la necesidad de seleccionar un modelo adecuado, entendiéndose por “adecuado” al simple hecho de respetar los supuestos del modelo. Por ejemplo, utilizar un modelo Gaussiano para investigar magnitudes no negativas representa un error que puede llevar a conclusiones incorrectas.

Entonces, es evidente que los modelos paramétricos de soporte no negativo son la mejor alternativa para representar variables aleatorias que midan intensidad, concentración o magnitud de una lectura observable en la naturaleza. En particular, los modelos referidos son las familias *beta* y *beta generalizada*, *exponencial*, *gamma* y *gamma generalizada*, *Rayleigh*, *Weibull*, la *log-normal* y la χ^2 , entre otras de menor uso.

Describiendo algunas propiedades de cada una de estas distribuciones, se tiene por ejemplo:

- La *familia Beta* o $\beta(\cdot)$: la función de densidad de probabilidad beta se puede expresar como

$$\beta(x; a, b) = \frac{x^{a-1}(1-x)^{b-1}}{B(a, b)}$$

en donde $a > 0, b > 0$ son parámetros de forma y $0 \leq x \leq 1$ es el rango de la función. $B(a, b)$ se conoce como la *función beta* y está dada por la expresión

$$B(a, b) = \int_0^1 u^{a-1}(1-u)^{b-1} du = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

para la función gamma $\Gamma(\cdot)$ definida con la integral

$$\Gamma(c) = \int_0^{\infty} e^{-u} u^{c-1} du. \quad (2.8)$$

En las gráficas de la figura siguiente (fig. 2.3) se pueden observar representaciones gráficas para distintos valores de a y b .

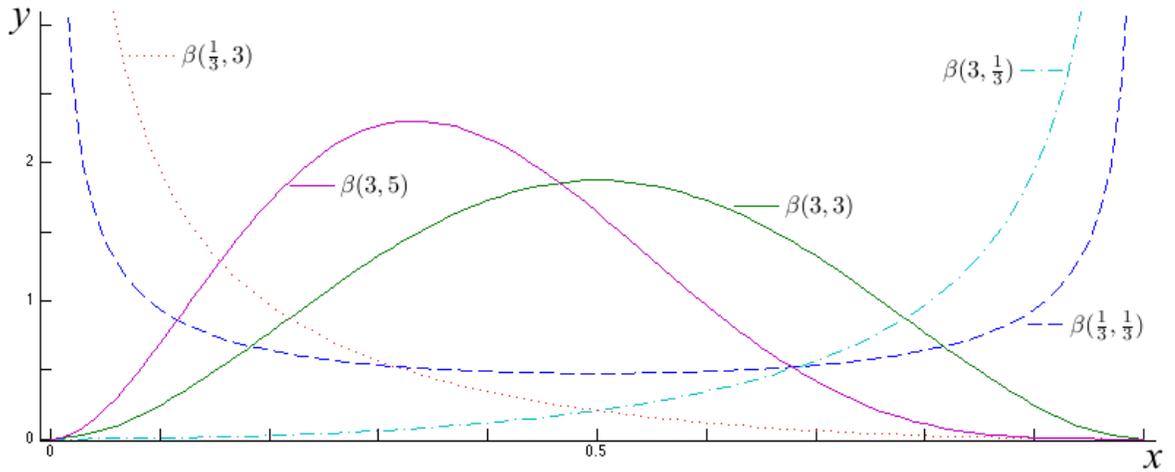


Figura 2.3: Representaciones gráficas para $\beta(\frac{1}{3}, \frac{1}{3})$, $\beta(\frac{1}{3}, 3)$, $\beta(3, 5)$, $\beta(3, \frac{1}{3})$ y $\beta(3, 3)$.

Además, por definición de esperanza matemática (Rohatgi, 2005, [30]) se obtiene que el r -ésimo momento de una variable aleatoria beta está dado por la esperanza matemática (ecuación 2.1)

$$\mathbb{E}X^r = \frac{1}{B(a, b)} \int_0^1 u^r (u^{a-1} (1-u)^{b-1}) du = \frac{1}{B(a, b)} \int_0^1 u^{(a+r)-1} (1-u)^{b-1} du$$

para $r = 1, 2, \dots$ y resolviendo la integral resulta que

$$\begin{aligned} \mathbb{E}X^r &= \frac{B(a+r, b)}{B(a, b)} = \frac{\Gamma(a+r)\Gamma(a+b)}{\Gamma(a)\Gamma(a+b+r)} \\ &= \frac{a(a+1)\cdots(a+r-1)}{(a+b)(a+b+1)\cdots(a+b+r-1)} \end{aligned}$$

y así, dado que la varianza de una variable aleatoria X se obtiene mediante la esperanza (Casella y Berger, 2002, [31]; Rohatgi, 2005, [30])

$$\text{Var}(X) = \mathbb{E}\left\{ (X - \mathbb{E}X)^2 \right\} = \mathbb{E}X^2 - (\mathbb{E}X)^2,$$

el valor esperado y la varianza de una variable beta están dados por

$$\mathbb{E}X = \frac{a}{a+b} \quad \text{y} \quad \text{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)},$$

respectivamente.

Así, para una muestra de tamaño n , $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$ tomada de una función de densidad de probabilidad $\beta(a, b)$, los *estimadores de momentos* de a y b se obtienen resolviendo el sistema no lineal (Casella y Berger, 2002, [31])

$$\begin{aligned} m_1 &= \frac{a}{a+b} \\ m_2 &= \frac{a(a+1)}{(a+b)(a+b+1)}, \end{aligned}$$

de modo que,

$$\hat{a}_{MOM} = \frac{m_1(m_1 - m_2)}{m_2 - m_1^2}, \quad \hat{b}_{MOM} = \frac{(m_1 - 1)(m_2 - m_1)}{m_2 - m_1^2},$$

en donde \hat{a}_{MOM} y \hat{b}_{MOM} son los estimadores de momentos de los parámetros a y b respectivamente, siempre que $m_2 > m_1^2$.

Por otro lado, la función de verosimilitud de la función de densidad de probabilidad beta se escribe como (Casella y Berger, 2002, [31]; Rohatgi, 2005, [30])

$$L(a, b|\mathbf{X}) = \prod_{i=1}^n \frac{x_i^{a-1}(1-x_i)^{b-1}}{B(a, b)} = B(a, b)^{-n} \left(\prod_{i=1}^n x_i \right)^{a-1} \left(\prod_{i=1}^n (1-x_i) \right)^{b-1}$$

o en forma equivalente, la log-verosimilitud está dada por

$$\ell(a, b|\mathbf{X}) = -n \log |B(a, b)| + (a - 1) \sum_{i=1}^n \log |x_i| + (b - 1) \sum_{i=1}^n \log |1 - x_i|.$$

Así, por principio de máxima verosimilitud, diferenciando respecto a los parámetros a y b , e igualando ambas derivadas a cero, se obtiene el sistema de ecuaciones no lineales

$$\begin{aligned} \psi(a) - \psi(a + b) &= \frac{1}{n} \sum_{i=1}^n \log |x_i| \\ \psi(b) - \psi(a + b) &= \frac{1}{n} \sum_{i=1}^n \log |1 - x_i| \end{aligned} \tag{2.9}$$

en donde $\psi(\cdot)$ es la función digamma dada por $\frac{d}{dx} \log(\Gamma(x))$ y dado que se debe resolver el sistema (2.9) en términos de a y b , no es posible obtener una solución analítica, ya que no es posible despejarlas de forma directa en el sistema y por ello, se debe resolver en forma numérica con algún método iterativo propuesto en la literatura, a partir de un valor inicial y el método de Newton-Raphson es el más utilizado en la práctica. Esto puede marcar una pequeña desventaja, en términos de costo computacional, en comparación con los modelos que tienen solución analítica.

En un caso más general, la función de densidad *beta generalizada* (βG) está dada por

$$\beta G(x; a, b) = \frac{x^{a-1}(c-x)^{b-1}}{c^{a+b-1}B(a, b)} \quad \text{para } 0 < x < c,$$

en donde este modelo es más consistente con los datos que se analizan en este trabajo, además de que en el caso particular en que $c = 1$, se tiene la función densidad de probabilidad $\beta(a, b)$.

- La función de densidad de probabilidad *exponencial* con parámetro de forma μ se escribe como

$$f(x) = \frac{1}{\mu} e^{-\frac{x}{\mu}}, \quad \text{para } x > 0,$$

con función de distribución de probabilidad $F(x) = 1 - e^{-\frac{x}{\mu}}$.

En la figura 2.4 se muestran las representaciones gráficas de las funciones de densidad y de distribución del modelo exponencial para $\mu = 1/2, 1$ y 2 .

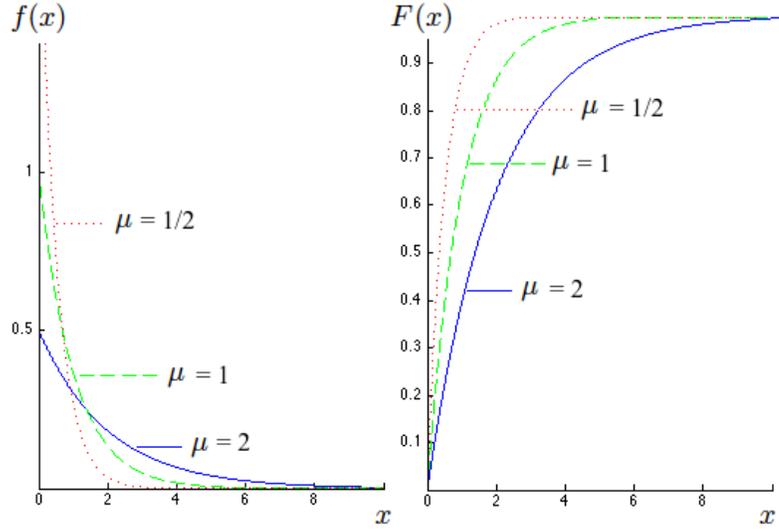


Figura 2.4: Representaciones gráficas de la funciones de densidad y de distribución, familia exponencial, para $\mu = 1/2, \mu = 1$ y $\mu = 2$.

Usando la definición de la función gamma (2.8), el r -ésimo momento de la función de densidad de probabilidad exponencial está dado por

$$\mathbb{E}X^r = \int_0^{\infty} x^r \frac{1}{\mu} e^{-\frac{x}{\mu}} dx = \mu^r \Gamma(r + 1),$$

por lo que, el valor esperado y la varianza de X están dados por $\mathbb{E}X = \mu$ y $Var(X) = \mu^2$.

De esta forma, para una muestra de tamaño n , $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$, tomada de una población exponencial con parámetro μ , la función de verosimilitud se escribe como

$$L(\mu|\mathbf{X}) = \prod_{i=1}^n \frac{1}{\mu} e^{-\frac{x_i}{\mu}} = \mu^{-n} e^{-\frac{n\bar{x}}{\mu}},$$

el estimador del parámetro μ para el modelo exponencial, ya sea por el método de momentos o el método de máxima verosimilitud (Lawless 2002, [36]), está dado

por

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x},$$

lo que representa una gran ventaja, desde el punto de vista computacional, ya que los estimadores se obtienen mediante el cálculo del promedio de los valores observados.

- Otro modelo destacado por sus aplicaciones en distintas áreas como análisis de sobrevivencia, teoría de confiabilidad, análisis de tiempos de falla y control de calidad, por mencionar algunos, es la función de densidad *Weibull*, con parámetro de forma λ y parámetro de escala γ , dado por

$$f(x) = \frac{\gamma}{\lambda} \left(\frac{x}{\lambda}\right)^{\gamma-1} e^{-\left(\frac{x}{\lambda}\right)^\gamma} \quad \text{para } x \geq 0, \text{ con } \lambda, \gamma > 0;$$

con función de distribución $F(x) = 1 - e^{-\left(\frac{x}{\lambda}\right)^\gamma}$. Una representación gráfica de funciones de densidad y de distribución para $\gamma = 2$ y $\lambda = 1/2, 1, 2$ se pueden observar en la figura 2.5

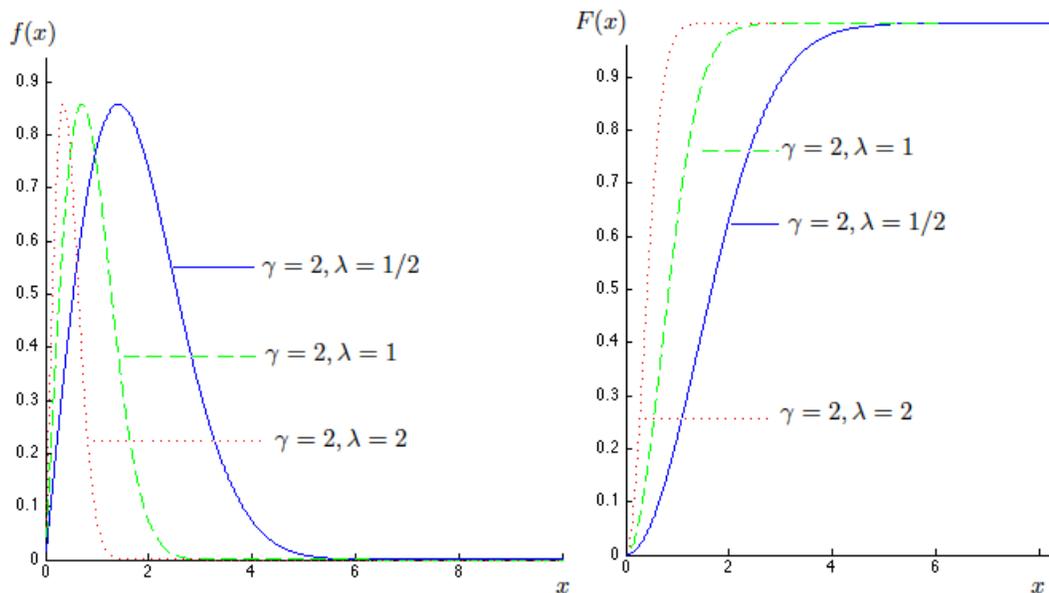


Figura 2.5: Representaciones gráficas de las funciones de densidad y de distribución de la familia Weibull, para $\gamma = 2$ y $\lambda = 1/2, 1, 2$.

En donde su r -ésimo momento está dada por la esperanza matemática

$$\mathbb{E}X^r = \int_0^{\infty} x^r f(x) dx = \int_0^{\infty} x^r \left\{ \frac{\gamma}{\lambda} \left(\frac{x}{\lambda} \right)^{\gamma-1} e^{-\left(\frac{x}{\lambda}\right)^\gamma} \right\} dx$$

y utilizando la definición de la función $\Gamma(u)$, se tiene que

$$\mathbb{E}X^r = \lambda^r \Gamma \left(1 + \frac{r}{\gamma} \right),$$

por tanto, su valor esperado y varianza están dados por

$$\mathbb{E}X = \lambda \Gamma \left(1 + \frac{1}{\gamma} \right) \quad \text{y} \quad \text{Var}(X) = \lambda^2 \left\{ \Gamma \left(1 + \frac{2}{\gamma} \right) - \Gamma \left(1 + \frac{1}{\gamma} \right)^2 \right\},$$

respectivamente.

Así, para una muestra $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$ de una población distribuida Weibull con parámetros (λ, γ) , se tienen dos alternativas para estimar sus parámetros a partir de dicha muestra. El primero es el método de momentos que consiste en resolver el sistema

$$m_1 = \lambda \Gamma \left(\frac{\gamma+1}{\gamma} \right) \tag{2.10}$$

$$m_2 = \lambda^2 \Gamma \left(\frac{\gamma+2}{\gamma} \right) \tag{2.11}$$

en donde $m_1 = \bar{x}$ (primer momento muestral) y $m_2 = \frac{1}{n} \sum_{i=1}^n x_i^2$ (segundo momento muestral), de modo que al despejar λ de (2.10) se tiene que

$$\hat{\lambda} = \frac{m_1}{\Gamma \left(\frac{\gamma+1}{\gamma} \right)} \tag{2.12}$$

por lo que

$$m_2 = \left[\frac{m_1}{\Gamma \left(\frac{\gamma+1}{\gamma} \right)} \right]^2 \Gamma \left(\frac{\gamma+2}{\gamma} \right) \iff m_2 = \frac{m_1^2 \Gamma \left(\frac{\gamma+2}{\gamma} \right)}{\Gamma \left(\frac{\gamma+1}{\gamma} \right)^2} \tag{2.13}$$

reduce el problema a la solución de una ecuación no lineal en una sola variable. Por otro lado, si se prefiere estimar por máxima verosimilitud (Lawless 2002, [36]), se debe resolver el sistema de ecuaciones no lineales dado por

$$\begin{aligned}\hat{\lambda} &= \left[\left(\frac{1}{n} \right) \sum_{i=1}^n x_i^{\hat{\gamma}} \right]^{1/\hat{\gamma}} \\ \hat{\gamma} &= \frac{n}{\left(\frac{1}{\hat{\lambda}} \right)^2 \sum_{i=1}^n x_i^{\hat{\gamma}} \log(x_i) - \sum_{i=1}^n \log(x_i)}\end{aligned}$$

que nuevamente se puede simplificar en una ecuación no lineal con una variable sustituyendo la expresión de $\hat{\lambda}$ en la ecuación dada por $\hat{\gamma}$, obteniéndose

$$\hat{\gamma} = n \left(\frac{\sum_{i=1}^n x_i^{\hat{\gamma}} \log(x_i)}{\left[\left(\frac{1}{n} \right) \sum_{i=1}^n x_i^{\hat{\gamma}} \right]^{2/\hat{\gamma}} - \sum_{i=1}^n \log(x_i)} \right)^{-1} \quad (2.14)$$

En ambos casos se cuenta con funciones en matlab (octave) que permiten obtener estimaciones paramétricas de λ y γ ya sea por método de momentos o por máxima verosimilitud, en donde ambos métodos de estimación representan necesidades similares desde un punto de vista computacional, sin embargo la gran mayoría de autores sugieren implementar el método de máxima verosimilitud, a pesar de que en la práctica ambos métodos muestran un buen desempeño.

- Por último, un modelo que también destaca por sus aplicaciones es la función de *densidad de probabilidad gamma*, que se escribe como

$$f(x) = \frac{x^{k-1} e^{-x/\lambda}}{\Gamma(k) \lambda^k} \quad \text{para } x \geq 0, k > 0, \lambda > 0$$

en donde la función de distribución corresponde a la integral

$$F(x) = \frac{1}{\Gamma(k) \lambda^k} \int_0^x x^{k-1} e^{-x/\lambda} dx.$$

Evidentemente, la distribución de probabilidad no tiene una representación

analítica y comúnmente se deja indicada o bien se representa mediante la función gamma incompleta $I(k, \lambda)$. En la siguiente figura 2.6 se muestran representaciones gráficas de las funciones de densidad y distribución gamma con parámetros $\lambda = 2$ y $k = 1/2, 1, 2$.

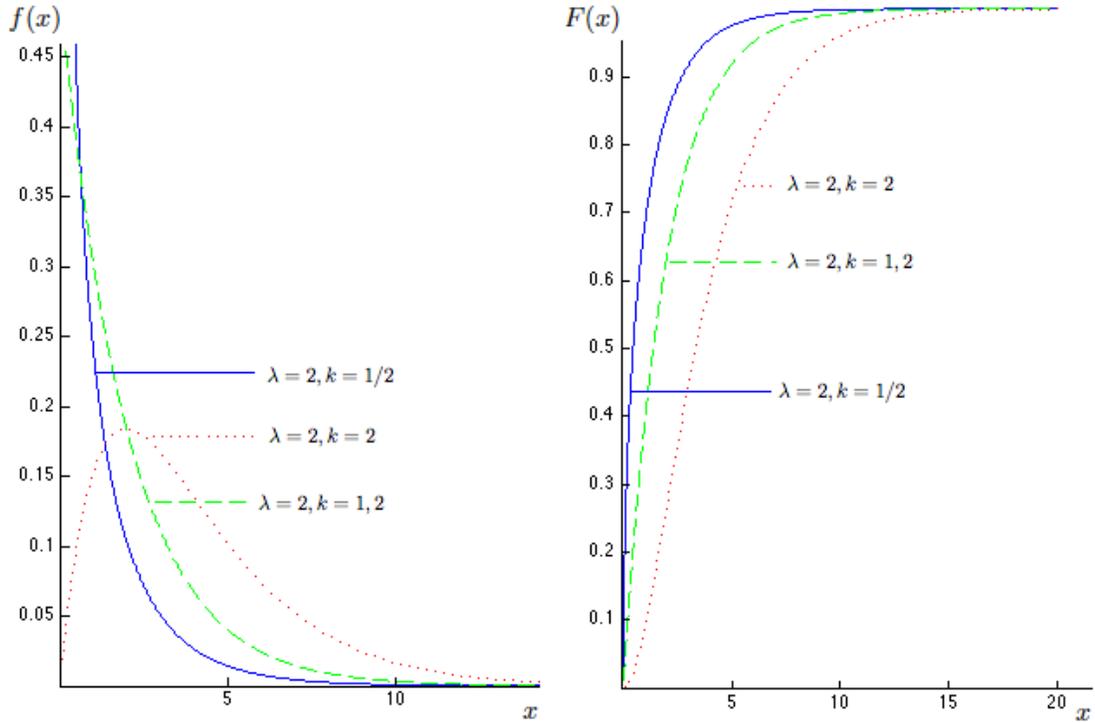


Figura 2.6: Representaciones gráficas de las funciones de densidad y de distribución de la familia gamma, para $\lambda = 2$ y $k = 1/2, 1, 2$.

De forma similar, el r -ésimo momento de una variable distribuida gamma se obtiene mediante la esperanza matemática

$$\mathbb{E}X^r = \int_0^{\infty} x^r \left\{ \frac{x^{k-1} e^{-x/\lambda}}{\Gamma(k)\lambda^k} \right\} dx = \lambda^r \Gamma(r+k)$$

de modo que, para una variable distribuida gamma, el valor esperado y la varianza están dados por

$$\mathbb{E}X = k\lambda \quad \text{y} \quad \text{Var}(X) = k\lambda^2$$

respectivamente.

Además, dada una muestra obtenida de una distribución gamma con parámetros λ, k ; estimando por el método de momentos se tiene que

$$m_1 = k\lambda \quad \text{y} \quad m_2 = (k^2 + k)\lambda^2,$$

y despejando, se tiene que

$$m_2 = m_1^2 + \frac{m_1^2}{k} \quad \Longrightarrow \quad \hat{k}_{MOM} = \frac{m_1^2}{m_2 - m_1^2}, \quad \hat{\lambda}_{MOM} = \frac{m_1}{\hat{k}_{MOM}} = \frac{m_1}{m_2 - m_1^2},$$

son los estimadores de momentos de los parámetros k y λ respectivamente, para $m_2 > m_1^2$.

Por otro lado, los estimadores por máxima verosimilitud de la familia gamma se presentan como ejemplo en la siguiente sección. No obstante, para observar en más detalle cada una de estos modelos de probabilidad, se puede acceder al libro de Lawless (2002, [36]) o bien al de Rohatgi (2005, [30]), el cual es otra excelente opción.

2.4.2. Métodos de estimación puntual

La estadística provee técnicas que permiten obtener conclusiones generales a partir de un conjunto limitado – pero representativo – de datos. Cuando se hace inferencia no se tiene garantía de que la conclusión que obtenemos sea exactamente correcta. Sin embargo, la estadística permite cuantificar el error asociado a la estimación.

La mayoría de los modelos de probabilidad incluyen cierto número de parámetros, generalmente desconocidos y que se deben estimar a partir de una muestra. De esta forma, si de un modelo de probabilidad $f(x; \theta)$ con parámetro θ , se tienen n valores observados x_1, x_2, \dots, x_n entonces, se puede obtener una estimación de θ , digamos $\hat{\theta}$ como función g_f de los valores observados, es decir

$$\hat{\theta} = g_f(x_1, x_2, \dots, x_n)$$

en donde el modelo de probabilidad juega un papel fundamental.

Existen varios métodos para obtener los estimadores de los parámetros de cualquier

modelo paramétrico, tales como el método de momentos, el método de máxima verosimilitud, el método de Bayes y otros. Sin embargo en este capítulo sólo se mencionan los dos primeros, ya que son los únicos utilizados para estimar los parámetros de las distribuciones estudiadas en el presente trabajo.

Método de momentos

El método de momentos que consiste en igualar los momentos muestrales dados por

$$m_1 = \frac{1}{n} \sum_{i=0}^n x_i, \quad m_2 = \frac{1}{n} \sum_{i=0}^n x_i^2, \quad m_3 = \frac{1}{n} \sum_{i=0}^n x_i^3, \dots, \quad m_r = \frac{1}{n} \sum_{i=0}^n x_i^r$$

con los momentos poblacionales, generados a partir del modelo probabilístico

$$\mu_1 = E[x], \quad \mu_2 = E[x^2], \quad \mu_3 = E[x^3], \dots, \quad \mu_r = E[x^r]$$

con base al número de parámetros que tiene la función de probabilidad.

Es una técnica general que consiste en utilizar los momentos de muestra para estimar sus momentos de población correspondientes produce estimadores con las propiedades de mínima varianza e insesgados, en donde se representa a los parámetros en términos de los momentos de la distribución.

Por ejemplo, para el modelo exponencial, dado por

$$f_T(t) = \lambda e^{-\lambda t}$$

solo se tiene un parámetro, en este caso λ , por lo que es suficiente comparar un momento muestral. En este caso, como $\mu_1 = E(t) = \frac{1}{\lambda}$ y $m_1 = \frac{1}{n} \sum_{i=0}^n t_i = \bar{t}$, al igualarlos se tiene que

$$\frac{1}{\lambda} = \bar{t} \quad \implies \quad \hat{\lambda} = \frac{1}{\bar{t}}.$$

En tanto que, para el modelo gamma

$$f_T(t) = \lambda^k e^{-\lambda t} \frac{t^{k-1}}{\Gamma(k)},$$

se tiene dos parámetros k y λ , lo que implica igualar dos momentos.

Dado que $\mu_1 = E(t) = k\lambda$, $\mu_2 = E(t^2) = (k^2 + k)\lambda^2$ y $m_1 = \bar{t}$, $m_2 = \frac{1}{n} \sum_{i=0}^n t_i^2$, al igualar estos momentos respectivamente, se tiene que

$$k\lambda = \bar{t} \quad \text{y} \quad (k^2 + k)\lambda^2 = m_2 \quad \text{luego} \quad \bar{t}^2 + \frac{\bar{t}^2}{k} = m_2 \iff \hat{k} = \frac{\bar{t}^2}{m_2 - \bar{t}^2}$$

y sustituyendo \hat{k} se obtiene el estimador de momentos de λ como $\hat{\lambda} = \frac{\bar{t}}{m_2 - \bar{t}^2}$ y finalmente

$$\hat{k} = \frac{\bar{t}^2}{\frac{1}{n} \sum_{i=0}^n t_i^2 - \bar{t}^2}, \quad \hat{\lambda} = \frac{\bar{t}}{\frac{1}{n} \sum_{i=0}^n t_i^2 - \bar{t}^2}$$

Evidentemente, para un modelo con tres parámetros, se igualan tres momentos y así sucesivamente, de modo que los estimadores de momentos se obtienen simplemente despejando los parámetros en términos de los momentos muestrales.

A pesar de ser un método simple y que generalmente proporciona una solución para cualquier modelo, su uso es muy reducido y por lo regular, la mayoría de usuarios aplican el método de máxima verosimilitud que se presenta en la siguiente sección.

Método de máxima verosimilitud

El método de máxima verosimilitud fue propuesto por Sir Ronald Fisher, a partir de un trabajo desarrollado por Bernoulli y revisado por Euler, fundamentalmente consiste en determinar un valor paramétrico que maximice la posibilidad de que los valores observados en la muestra sean los más probables, es decir, se deben obtener valores paramétricos θ tales que la función de verosimilitud dada por

$$L(\theta) = f(\theta; x_1, x_2, \dots, x_n) \tag{2.15}$$

sea máxima, en donde θ puede ser escalar o vectorial.

El estimador de máxima verosimilitud (o máximo verosímil), como su nombre lo dice, se obtiene calculando el máximo de la función de verosimilitud, por lo que se aplican conceptos elementales del cálculo, buscando maximizar la función $L(\theta)$, utilizando el modelo probabilístico. Por ejemplo, para la densidad exponencial, se tiene

que $f_T(t) = \lambda e^{-\lambda t}$ y entonces, para una muestra t_1, t_2, \dots, t_n se tiene que

$$L(\lambda) = f(\lambda; t_1, t_2, \dots, t_n) = \prod_{i=1}^n \lambda e^{-\lambda t_i} = \lambda^n e^{-\lambda \sum_{i=1}^n t_i}$$

entonces, derivando respecto al parámetro λ e igualando a cero se tiene que

$$n - \lambda \sum_{i=1}^n t_i = 0$$

y despejando λ resulta

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n t_i} = \frac{1}{\frac{1}{n} \sum_{i=1}^n t_i} = \frac{1}{\bar{t}}.$$

En el caso particular de la función de densidad de probabilidad exponencial se puede observar que el estimador por método de momentos es el mismo que el que se obtiene mediante el método de máxima verosimilitud, sin embargo esto no cumple en general. Observando por ejemplo las ecuaciones que resultan para los estimadores de momentos en la expresión (2.13) y el de máxima verosimilitud en (2.14) de la función de densidad de probabilidad Weibull, no se tiene la misma expresión para sus estimadores.

En la mayoría de textos aprovechan el hecho de que el máximo de $f(x)$ es el mismo que el de $\log[f(x)]$, de modo que los cálculos se simplifican de forma significativa al momento de obtener los estimadores por máxima verosimilitud. Como ejemplo, nótese que analizando nuevamente el caso anterior, si

$$L(\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^n t_i}$$

entonces, aplicando el logaritmo natural se tiene que

$$l(\lambda) = \log L(\lambda) = n \log \lambda - \lambda \sum_{i=1}^n t_i$$

así, diferenciando respecto a λ se tiene que

$$\frac{d}{d\lambda} l(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^n t_i$$

que al igualar a cero y despejar se obtiene el mismo resultado ($\hat{\lambda} = 1/\bar{t}$).

Como segundo ejemplo, para el modelo gamma se tiene que

$$L(k, \lambda) = \prod_{i=1}^n \lambda^k e^{-\lambda t_i} \frac{t_i^{k-1}}{\Gamma(k)} = \lambda^{nk} e^{-\lambda \sum_{i=1}^n t_i} \frac{\prod_{i=1}^n t_i^{k-1}}{\Gamma(k)^n}$$

por tanto, aplicando nuevamente el logaritmo natural resulta

$$l(k, \lambda) = (nk) \log \lambda - \lambda \sum_{i=1}^n t_i + (k-1) \sum_{i=1}^n t_i - n \log \Gamma(k)$$

y diferenciando parcialmente respecto a k y λ se obtiene el sistema de dos ecuaciones con dos variables dado por

$$\begin{aligned} \frac{\partial}{\partial k} l(k, \lambda) &= \frac{\partial}{\partial k} \left\{ (nk) \log \lambda - \lambda \sum_{i=1}^n t_i + (k-1) \sum_{i=1}^n t_i - n \log \Gamma(k) \right\} \\ &= n \log \lambda + \sum_{i=1}^n t_i - n \left(\frac{\Gamma'(k)}{\Gamma(k)} \right) \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \lambda} l(k, \lambda) &= \frac{\partial}{\partial \lambda} \left\{ (nk) \log \lambda - \lambda \sum_{i=1}^n t_i + (k-1) \sum_{i=1}^n t_i - n \log \Gamma(k) \right\} \\ &= \frac{nk}{\lambda} - \sum_{i=1}^n t_i \end{aligned}$$

y así, igualando a cero resulta el sistema de ecuaciones no lineales

$$\frac{nk}{\lambda} - \sum_{i=1}^n t_i = 0 \quad (2.16)$$

$$n \log \lambda + \sum_{i=1}^n t_i - n \left(\frac{\Gamma'(k)}{\Gamma(k)} \right) = 0 \quad (2.17)$$

y despejando k de la ecuación (2.16) se tiene que,

$$\widehat{k} = \widehat{\lambda} \left(\frac{1}{n} \sum_{i=1}^n t_i \right) = \widehat{\lambda} \bar{t} \iff \widehat{k} = \widehat{\lambda} \bar{t}$$

$$\log \widehat{\lambda} = \frac{\Gamma'(\widehat{k})}{\Gamma(\widehat{k})} - \bar{t}$$

cuya solución no se puede obtener en forma analítica, por lo que se debe aproximar una solución del sistema mediante algún método numérico o bien, se puede simplificar el problema sustituyendo \widehat{k} para reducir el problema a la solución de una ecuación no lineal en una variable, dada por

$$\widehat{\lambda} = \exp \left(\frac{\Gamma'(\widehat{k})}{\Gamma(\widehat{k})} - \bar{t} \right) \implies \widehat{\lambda} = \exp \left(\frac{\Gamma'(\widehat{\lambda} \bar{t})}{\Gamma(\widehat{\lambda} \bar{t})} - \bar{t} \right),$$

y en este caso, nuevamente se tiene que implementar un método iterativo para aproximar numéricamente los estimadores de máxima verosimilitud.

Finalmente, suponiendo un modelo aditivo para las componentes revisadas en este capítulo, una propuesta general que incorpore cada una de los efectos antes mencionados queda descrito por

$$X_t = P_t + T_t + A_t + E_t$$

en donde E_t es la componente impredecible de la serie de datos y sigue una distribución de probabilidad y

$$P_t = a_0 + \sum_{i=1}^m a_i \cos(2\pi\omega_i t) + b_i \sin(2\pi\omega_i t) \quad (2.18)$$

$$T_t = b_1 + b_2 t \quad (2.19)$$

$$A_t = \sum_{i=1}^k c_i X_{t-i} \quad (2.20)$$

De modo que, el modelo se puede representar como

$$X_t = A_0 + b_2 t + \sum_{i=1}^m a_i \cos(2\pi\omega_i t) + b_i \sin(2\pi\omega_i t) + \sum_{i=1}^k c_i X_{t-i} + E_t \quad (2.21)$$

en donde $A_0 = a_0 + b_1$, m son las frecuencias significativas de la serie y se tienen hasta k retrasos entre lecturas consecutivas.

Capítulo 3

Estudio de simulación

Como ya se ha mencionado en el capítulo previo, cuando se analizan series temporales de magnitudes observables en la naturaleza, se espera encontrar efectos de tipo estacional o periódico P_t , de tendencia temporal a largo plazo T_t , de asociación entre lecturas consecutivas A_t y una componente aleatoria de variable no negativa E_t , atribuible a variaciones impredecibles a corto plazo.

Por tal, la idea principal en este capítulo es construir series temporales de longitud fija n y simular, por ejemplo, una medición por hora (24 horas) durante 365 días, generando $n = 8760$ lecturas y a partir de ellas, se incorporarían los efectos observados en la naturaleza mediante la expresión

$$X_t = P_t + T_t + A_t + E_t, \quad (3.1)$$

con el fin de obtener series de tiempo que reproduzcan la estructura temporal de magnitudes observables en la naturaleza, y con estas magnitudes simuladas, medir el efecto que produce cualquiera de estas componentes en la función de densidad de probabilidad de una variable aleatoria.

De modo que, si E_t corresponde a una densidad de probabilidad exponencial con parámetro λ , se buscará medir el efecto que genera, por ejemplo, una componente periódica en el comportamiento de la magnitud exponencial y repetir el proceso de simulación de forma iterativa, hasta obtener $m = 10,000$ muestreos de series temporales de tamaño 8760, para lograr estimadores confiables.

De esta forma, suponiendo que se conoce la distribución de probabilidad de la componente aleatoria, se pueden obtener los estimadores paramétricos y compararlos con los valores exactos, a partir de los cuales se generan los datos simulados, dando la posibilidad de medir los errores que se cometen cuando existen efectos de tendencia temporal a largo plazo o de asociación entre lecturas consecutivas y que, el usuario no los toma en cuenta para estimar el ajuste del modelo de probabilidad a los datos.

Por lo tanto, se analizan los errores en tres apartados:

- en la interacción entre efectos periódicos y la componente aleatoria,
- en la interacción entre efectos periódicos y el factor de tendencia y
- en la interacción entre la componente aleatoria y el factor de tendencia,

investigando y reportando en cada caso, cualquier evidencia de influencia entre los componentes que se analizan. Las rutinas utilizadas para este estudio están disponibles en el apéndice A de este trabajo.

3.1. Interacción entre componentes aleatorias y efectos periódicos

En un primer paso, si t representa los puntos discretos de muestreo en el tiempo, para $t = 1, 2, \dots, 8760$, entonces, la función de distribución de probabilidad de una variable exponencial x con parámetro μ está dada por

$$F(x) = 1 - e^{-\mu x} \quad (3.2)$$

y así, para cualquier número pseudoaleatorio $0 < r < 1$, se puede igualar $r = F(x)$, obteniendo por despeje la expresión

$$x = -\frac{1}{\mu} \ln(r'), \quad (3.3)$$

con $r' = 1 - r$. Destacando que, para todo r entre cero y uno, es evidente que $0 < 1 - r < 1$, además si r es un número pseudoaleatorio, entonces $1 - r$ también lo es,

con la posibilidad de ahorrar una operación en cada iteración. Así, en lugar de $1 - r$ se utiliza r' , dado que el resultado es el mismo, a con un menor costo computacional.

A manera de ejemplo, en la figura 3.1 se puede ver el histograma de una serie simulada con $\mu = 0.2$, obteniendo la forma típica de una función de densidad exponencial, es decir, de forma decreciente, asimétrica y con una alta concentración de los datos a la izquierda (ver figura 2.4).

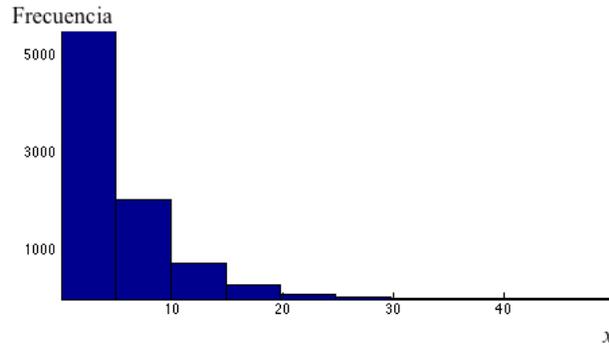


Figura 3.1: Histograma de frecuencias de una serie simulada con $\mu = 0.2$.

La representación gráfica del mismo conjunto de datos como una serie de tiempo se puede observar en la figura 3.2, en donde el eje horizontal representa el tiempo y el eje vertical muestra la amplitud de una variación exponencialmente distribuida

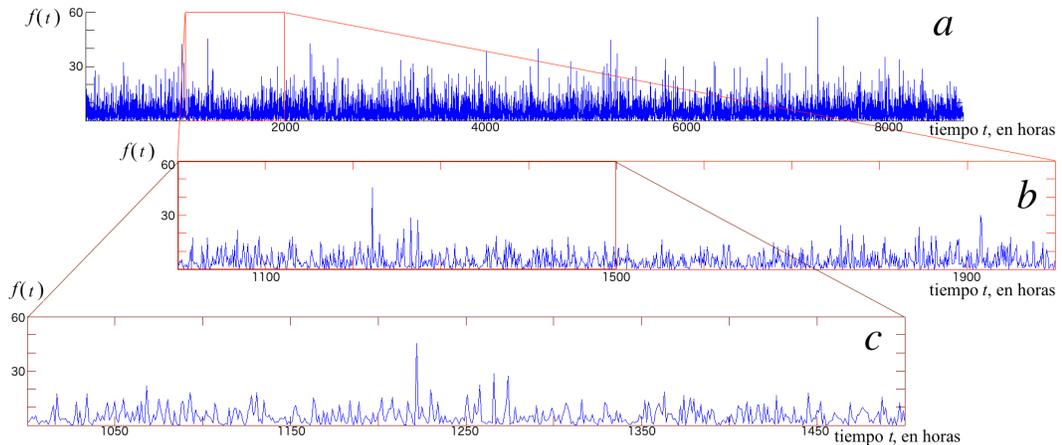


Figura 3.2: Representación gráfica de la componente aleatoria de una serie temporal simulada, variable aleatoria exponencial con $\mu = 0.2$; serie completa (a), haciendo un acercamiento para $1000 \leq t \leq 2000$ (b) y para $1000 \leq t \leq 1500$ (c).

Por otro lado, para construir una serie periódica se parte de la ecuación 2.18, utilizando sumas de funciones seno y coseno con n frecuencias y amplitudes, de la forma

$$P_t = a_0 + \sum_{i=1}^n a_i \cos(2\pi\omega_i t) + b_i \operatorname{sen}(2\pi\omega_i t),$$

y a manera de ejemplo, se simula una serie con tres frecuencias: una diurna que se repite 365 veces en el periodo de un año con $\omega_1 = \frac{1}{24}$, una mensual que se repite 12 veces en el mismo periodo $\omega_2 = \frac{1}{730}$ y una frecuencia anual que se presenta una sola vez en el periodo con $\omega_3 = \frac{1}{8760}$. Además, las amplitudes simuladas para este caso son cero en las funciones coseno ($a_1 = a_2 = a_3 = 0$), se asigna una amplitud de 20 para las funciones seno ($b_1 = b_2 = b_3 = 20$) y se toma $a_0 = 0$; de modo que la serie periódica de este ejemplo tiene la forma

$$P_t = 20 \left\{ \operatorname{sen} \left(\frac{2\pi t}{24} \right) + \left(\frac{2\pi t}{730} \right) + \operatorname{sen} \left(\frac{2\pi t}{365} \right) \right\}$$

ya así, en la figura 3.3 se pueden observar las representaciones gráficas de las frecuencias diurna (c), mensual (b) y anual (a), todas con amplitud 20.

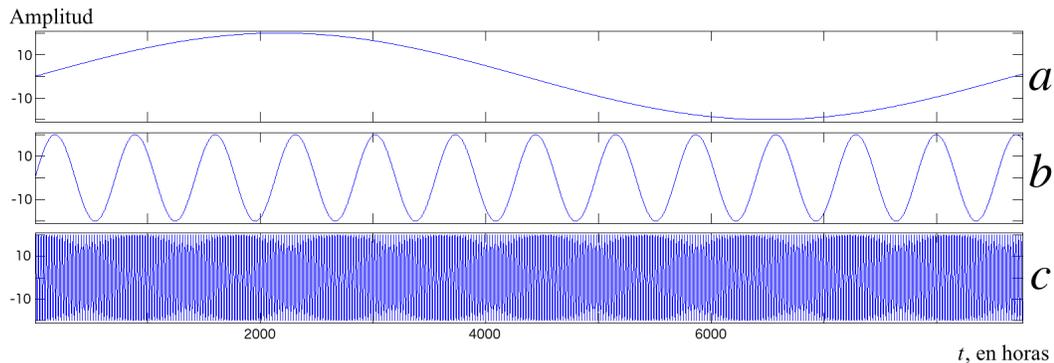


Figura 3.3: Representación gráfica de la componente periódica de la serie simulada para frecuencias diurna (c), mensual (b) y anual (a), todas con amplitud 20.

Por tanto, la serie periódica con las tres frecuencias (diurna, mensual y anual) se puede ver en la figura 3.4, en donde destacan principalmente la frecuencia diurna (la de mayor oscilación) y la mensual con doce picos bien definidos en la gráfica; en tanto que, la serie anual se percibe ligeramente en la serie completa (gráfica 3.4a) por una

caída en la magnitud de los datos en la segunda parte del año.

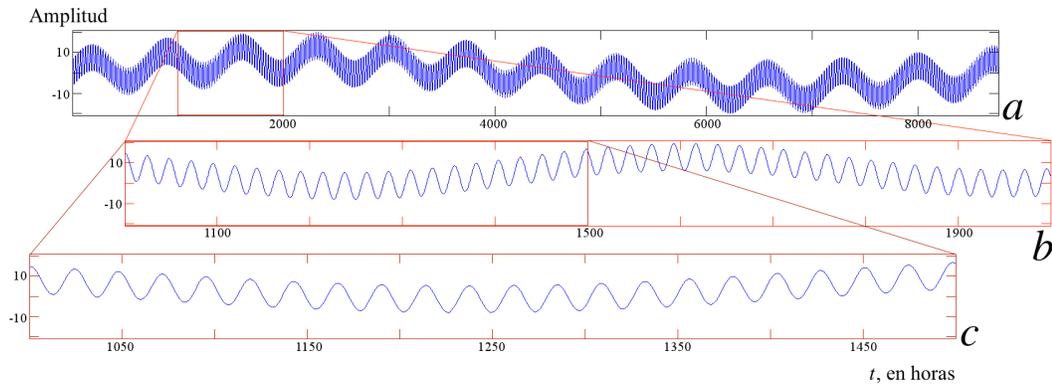


Figura 3.4: Representación gráfica de la componente periódica de la serie temporal simulada, incorporando las frecuencias diurna, mensual y semianual; serie completa (a), haciendo un acercamiento para $1000 \leq t \leq 2000$ (b) y para $1000 \leq t \leq 1500$ (c).

Combinando las dos componentes (el efecto periódico P_t y la componente aleatoria E_t), en la figura 3.5 se puede ver la representación gráfica de la interacción $X_t = P_t + E_t$, en donde se hace evidente el efecto estacional para la frecuencia mensual (doce repeticiones), así como la variación de 365 repeticiones (figura 3.5b) junto con la componente aleatoria que muestra diferencias significativas al comparar con la figura 3.4c.

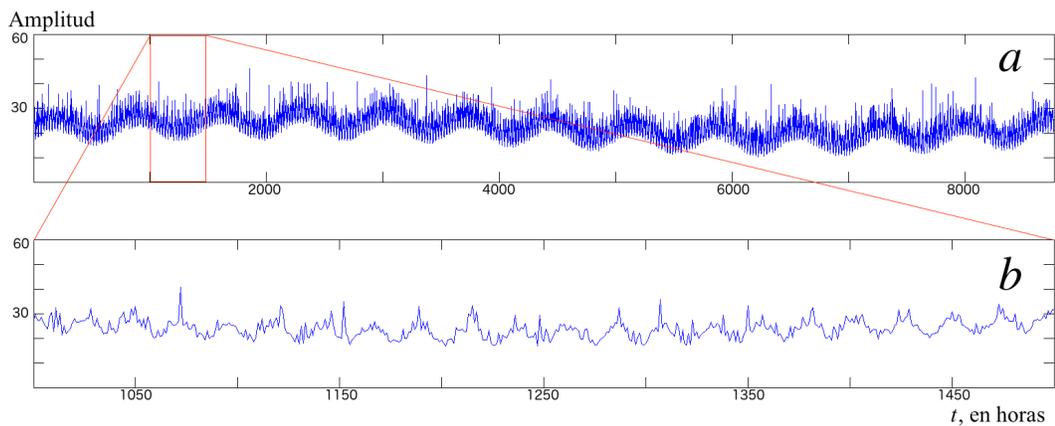


Figura 3.5: Representación gráfica de la serie temporal simulada $X_t = P_t + E_t$ para la serie completa (a) y haciendo un acercamiento para $1000 \leq t \leq 1500$ (b).

Investigando ahora el efecto de la componente aleatoria sobre el factor periódico, en la figura 3.6 se pueden ver las representaciones gráficas de los periodogramas (eje vertical Amplitud R_j , eje horizontal frecuencia $f : j$) de las transformadas de la componente periódica P_t (figura 3.6a) y la interacción de ambas componentes X_t (figura 3.6b).

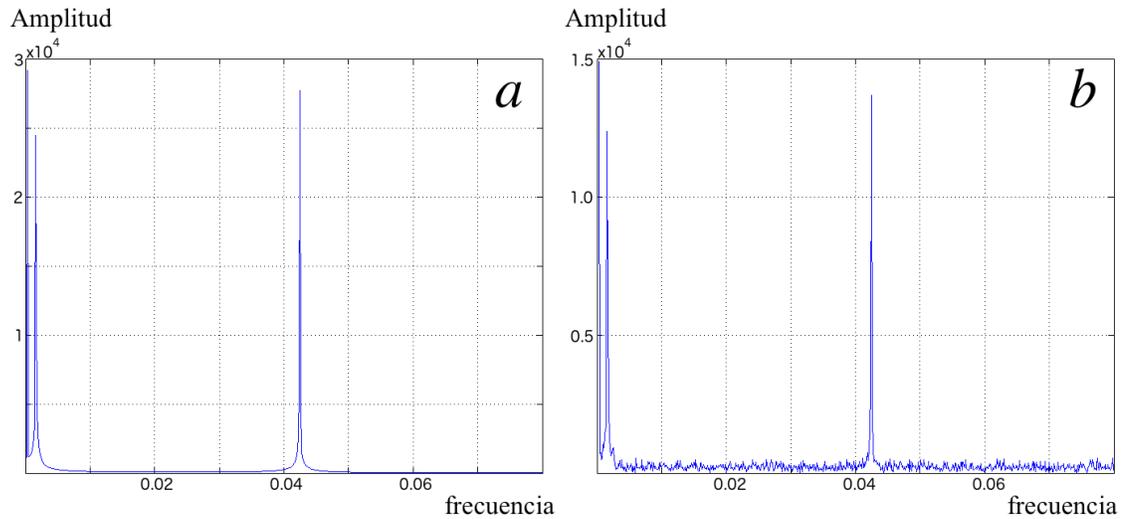


Figura 3.6: Representación gráfica de los periodogramas de la serie temporal simulada: (a) únicamente la componente periódica P_t y (b) la interacción de las componentes aleatoria y periódica $X_t = P_t + E_t$.

En la figura 3.6 se puede ver que las frecuencias significativas se presentan de forma idéntica en ambos casos, *i.e.*, las frecuencias con mayor amplitud se repiten en las dos gráficas. Por tanto, hay evidencia para creer que la magnitud aleatoria no repercute en las periodicidades ocultas en la serie.

Se distingue además una reducción en la amplitud al comparar las representaciones gráficas de ambos periodogramas, ya que en la figura 3.6b la amplitud máxima de la gráfica es aproximadamente la mitad de la amplitud máxima que se observa en la figura 3.6a.

Además, en el periodograma de la serie con la componente aleatoria (figura 3.6b) se pueden observar pequeñas oscilaciones de frecuencias con amplitudes cercanas a cero y por ello, podría afirmarse que no son significativas.

Por otro lado, comparando ahora los histogramas de las series E_t y X_t , para comprobar si la componente periódica influye en el comportamiento de la magnitud aleatoria; en la figura 3.7 se puede ver que el comportamiento de la variable exponencialmente distribuida E_t (figura 3.7a) se deforma completamente en la interacción de las dos componentes X_t (figura 3.7b).

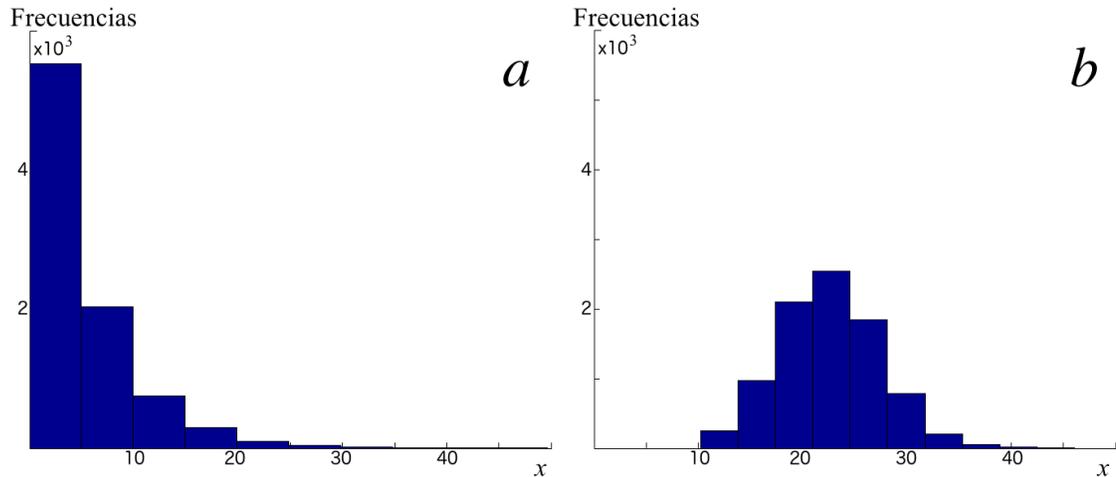


Figura 3.7: Representación gráfica de los histogramas de la serie temporal simulada: (a) únicamente la componente aleatoria E_t y (b) la interacción de las componentes aleatoria y periódica $X_t = P_t + E_t$.

Revisando ambos histogramas en la figura 3.7, se tiene suficiente evidencia para afirmar que la componente periódica genera un efecto contundente sobre la parte aleatoria de la serie. Más aún, observando de forma aislada el histograma de la figura 3.7b, no hay evidencia para creer que los datos se ajustan a un modelo exponencial ya que no parece asimétrica, no es decreciente y no muestra una alta concentración de los datos a la izquierda (ver figura 2.4).

Partiendo de los resultados observados en este caso, se ejecutaron 10,000 simulaciones para cada uno de los modelos de magnitud no negativa (sección 2.4) y se agregó además el modelo Gaussiano, que es un modelo de soporte real (abarcando tanto valores positivos como negativos) y tener la posibilidad de resultados más generales.

Los parámetros de cada modelo se asignaron en forma aleatoria entre $[5, 6, \dots, 100]$ y en el caso de las frecuencias, se tomaron periodos diarios, mensuales y de un año.

De este modo, a partir del modelo de simulación se conocen las frecuencias verdaderas f_j y sus respectivas amplitudes a_0 , a_j y b_j (son las que se utilizaron para obtener la muestra). Por lo que el problema se reduce a verificar si las frecuencias estimadas de la interacción $\hat{f}_{j_{X_t}}$ y amplitudes estimadas $\hat{a}_{0_{X_t}}$, $\hat{a}_{j_{X_t}}$ y $\hat{b}_{j_{X_t}}$ para $X_t = P_t + E_t$, son iguales a las verdaderas, midiendo las diferencias a partir de la *Raíz del Error Cuadrado Medio (RECM)* dado por

$$\begin{aligned} RECM_f &= \sqrt{\frac{1}{m} \sum_{j=1}^m (f_j - \hat{f}_{j_{X_t}})^2} \\ RECM_{a_0} &= \sqrt{\frac{1}{m} \sum_{j=1}^m (a_0 - \hat{a}_{0_{X_t}})^2} \\ RECM_{a_j} &= \sqrt{\frac{1}{m} \sum_{j=1}^m (a_j - \hat{a}_{j_{X_t}})^2} \\ RECM_{b_j} &= \sqrt{\frac{1}{m} \sum_{j=1}^m (b_j - \hat{b}_{j_{X_t}})^2} \end{aligned}$$

con la posibilidad de analizar el comportamiento de los errores para cada uno de los parámetros del modelo.

Además, se ejecutaron en total 5×10^4 simulaciones y de éstas, los resultados apuntan de manera contundente (el 100 % de los casos) a que la función de distribución se modifica completamente por el efecto periódico, es decir, por lo menos en 10^4 corridas de simulación con diferentes combinaciones de parámetros de forma, de escala y de localidad, con base en las características de cada modelo (densidades beta generalizada, exponencial, Weibull, gamma y Gaussiano).

En todos los casos se observa que la componente periódica transforma el comportamiento de la magnitud aleatoria y dado que se busca evidencia de igualdad en la densidad empírica f de la variable aleatoria, se comparan las distribuciones F de ambos casos (E_t y X_t), midiendo la distancia entre ellas mediante la *RECM*, dado por

$$RECM = \sqrt{\frac{1}{m} \sum_{j=1}^m (F_{E_{t_j}} - F_{X_{t_j}})^2} \quad (3.4)$$

De esta forma, analizando los resultados del estudio de simulación, se tiene que la gran mayoría de los casos (el 99.97 %) las frecuencias simuladas y estimadas mostraron

el mismo comportamiento en términos de frecuencias significativamente diferentes de cero (picos que sobresalen en el periodograma), con un error promedio estimado de $\widehat{Error}_{f_j} = 0.00001122$ en todas las simulaciones, todos los modelos y considerando los casos individuales para cada modelo, en la tabla 3.1 se pueden observar los errores estimados con sus respectivos intervalos de confianza del 95 %.

Tabla 3.1: Errores observados al comparar las frecuencias simuladas f_j y las frecuencias estimadas \hat{f}_{jX_t} , en la interacción de componentes periódicas con magnitudes aleatorias.

Modelo	<i>RECM</i>	Límites del 95 % de confianza
Beta Generalizado	0.000113284	[0.000093207, 0.000114088]
Exponencial	0.000112980	[0.000093207, 0.000114094]
Weibull	0.000114155	[0.000114153, 0.000114148]
Gamma	0.000108688	[0.000093207, 0.000114084]
Gaussiano	0.000112084	[0.000093207, 0.000114069]

En los pocos casos en los que no se tenía el mismo comportamiento (el 0.03 %), se observó que la amplitud de la componente aleatoria excedía la de las frecuencias simuladas, es decir, se localizaron las frecuencias pero no fueron significativas debido a que su amplitud no superaba la magnitud aleatoria.

Revisando ahora el comportamiento de los parámetros a_0 , a_j y b_j del modelo de efectos periódicos, en los resultados de simulación se observan RECM muy grandes, de orden de 10^5 en todos los casos, y a manera de ejemplo, en la siguiente tabla (3.2) se muestran las RECM al comparar los a_j de la componente periódica P_t y la interacción X_t , para cada uno de los modelos.

Tabla 3.2: Errores observados al comparar las amplitudes simuladas a_j y las estimadas a_{jX_t} , en la interacción de componentes periódicas con magnitudes aleatorias.

Modelo	<i>RECM</i>	Límites del 95 % de confianza
Beta Generalizado	250396.4399352	[108438.2321104, 367042.6064073]
Exponencial	250423.9867693	[108494.8424091, 367199.2542605]
Weibull	250375.1334927	[108391.7203738, 367104.7230265]
Gamma	252903.6753076	[120115.6357493, 367199.9015274]
Gaussiano	250562.3162148	[109113.7100903, 367153.3007603]

Así, de los resultados de la tabla 3.2 se puede concluir que las amplitudes de la componente periódica se modifican al interactuar con las magnitudes aleatorias de los modelos utilizados en este experimento de simulación.

Por otro lado, analizando el comportamiento de la RECM entre la densidad empírica de la magnitud aleatoria E_t y la que corresponde a la interacción X_t , se tiene que todos los experimentos de simulación dan evidencia para rechazar igualdad entre densidades, por lo que se concluye que la componente periódica deforma la densidad de la variable aleatoria, al menos para los modelos probados.

Tabla 3.3: Errores observados al comparar las densidades empíricas de probabilidad para la magnitud E_t y la interacción de componentes periódicas con magnitudes aleatorias X_t .

Modelo	RECM	Límites del 95 % de confianza
Beta Generalizado	1897.8986628	[1845.8135847, 1971.6198523]
Exponencial	1574.8219873	[1170.3737144, 1867.7804446]
Weibull	1901.1651044	[1852.3683556, 1972.6511015]
Gamma	1265.8976106	[920.3877082, 1676.3972763]
Gaussiano	1757.7729508	[1288.0214913, 1887.7156203]

Por tanto, los resultados de simulación sugieren que lo recomendable es estimar primero las componentes periódicas, ya que el efecto de la componente aleatoria no repercute en la estimación de las frecuencias significativas y en el caso en que la amplitud de las frecuencias ocultas sea menor al efecto de la componente aleatoria, no se tendría evidencia de efectos periódicos y se investiga directamente la componente aleatoria.

De igual forma, se concluye que la amplitud de las frecuencias significativas se afecta con la interacción de la componente aleatoria, ya que al comparar las amplitudes antes y después de la interacción con la componente aleatoria, se observan diferencias muy grandes y por ésto, se recomienda estimar las amplitudes a partir de la serie completa, mediante el método de mínimos cuadrados y de este modo, para investigar la componente aleatoria es necesario retirar el efecto periódico de la serie, antes de buscar un modelo adecuado.

3.2. Interacción entre factores de tendencia y componentes aleatorias

Con un procedimiento similar al anterior, para construir una serie de tiempo con un efecto de tendencia, de la expresión 2.3 se tiene que un modelo adecuado puede tener la forma

$$T_t = \alpha + \beta t$$

y entonces, una representación gráfica de un caso con $\alpha = 5$ y $\beta = 0.002$ se puede observar en la figura 3.8a, la componente aleatoria corresponde a la figura 3.8b y la figura 3.8c muestra la interacción de ambos componentes.

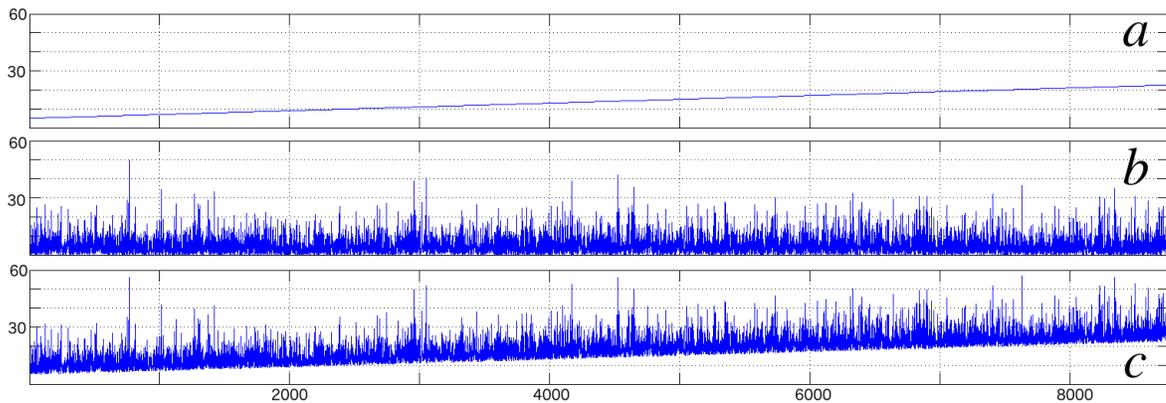


Figura 3.8: Representación gráfica de la serie temporal simulada, (a) componente de tendencia temporal T_t , (b) componente aleatoria E_t y la interacción de las componentes aleatoria y temporal (c).

Por otro lado, implementando el método de mínimos cuadrados (expresión 2.4) para estimar los parámetros de tendencia de la serie de datos, se tiene que

$$\hat{\alpha} = 9.7864 \quad \text{y} \quad \hat{\beta} = 0.0021$$

y así, comparando con los valores exactos ($\alpha = 5$, $\beta = 0.002$), se puede ver una sobre-estimación en el parámetro α con un error de 4.7864, en tanto que la pendiente muestra un error más pequeño (0.0001). Lo que es natural, ya que al incorporar cualquier magnitud monótona no negativa (no decreciente, no creciente) a una función de tendencia, la recta transformada se traslada hacia arriba, en comparación con la

recta original. Por tanto, se debe tomar en cuenta que la componente aleatoria muestra mayor afectación en el termino independiente.

Comparando ahora el desempeño de la magnitud aleatoria, antes y después del efecto de tendencia, en los histogramas se puede ver de forma análoga que la componente aleatoria se deforma al incorporarle el efecto de tendencia anual (figura 3.9). En este caso, en los histogramas de los datos que se muestran en la figura 3.9 sobresale nuevamente que la representación gráfica de la figura 3.9b muestra un desempeño muy diferente a un comportamiento exponencial (figura 3.9a).

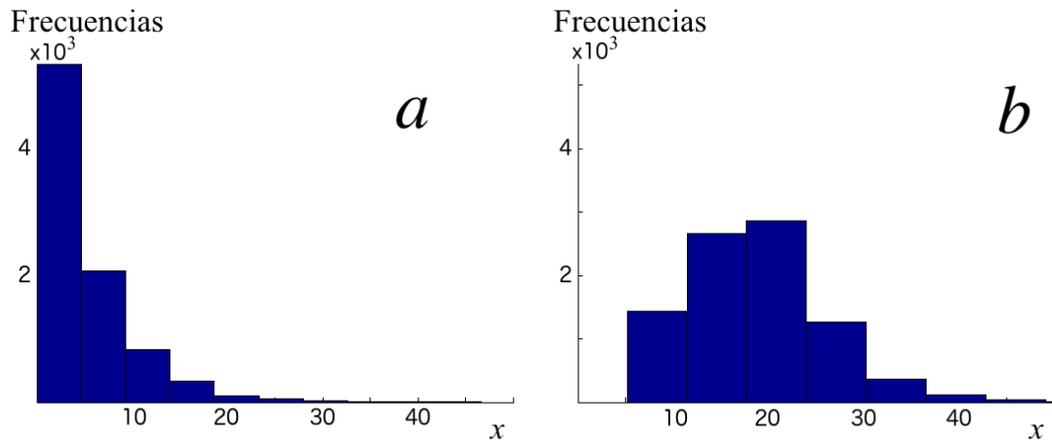


Figura 3.9: Histogramas de frecuencia de una variable aleatoria exponencial (a) y la interacción de las componentes aleatoria y de tendencia temporal (b).

En este caso se implementó un esquema de simulación similar al del caso previo, es decir, se ejecutaron 50,000 simulaciones, evaluando el desempeño de los cinco modelos de densidad de probabilidad, 10,000 repeticiones para cada modelo (densidades beta generalizada, exponencial, Weibull, gamma y Gaussiano) y se incorporó el efecto de tendencia T_t seleccionando de forma aleatoria los parámetros de los modelos de probabilidad y de tendencia.

Así, de los resultados de simulación se observa que el 97.43 % de los casos, la función de distribución se deforma al incorporar el efecto de tendencia temporal, en donde el 2.57 % de los casos en los que no se deformó la densidad de probabilidad fue porque los coeficientes α y β simulados tenían una magnitud muy cercana a cero, es decir, el efecto de tendencia era muy pequeño y no alteraba el comportamiento de los datos.

De este modo, se recomienda nuevamente estimar el ajuste del modelo de probabilidad una vez que se retire el efecto de tendencia en la serie.

3.3. Interacción entre efectos periódicos y factor de tendencia

Analizando ahora la interacción entre componentes periódicos y de tendencia, en la figura 3.10 se muestra las representaciones gráficas de las series de tendencia temporal, de efectos periódicos diurno, mensual y anual y la interacción entre ambos componentes.

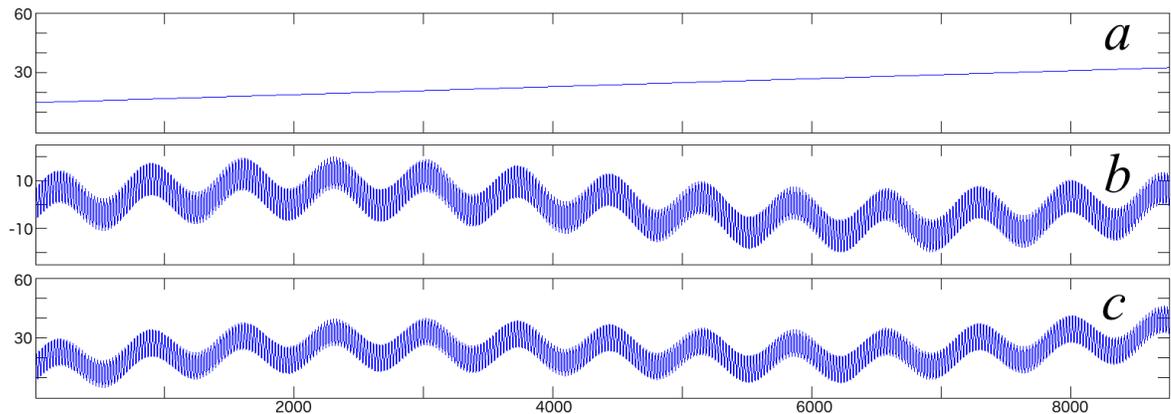


Figura 3.10: Representación gráfica de la serie temporal simulada, (a) componente de tendencia temporal T_t , (b) componente periódica P_t y la interacción de las componentes periódica y temporal $P_t + T_t$ (c).

Así, para medir el efecto de la componente periódica en el efecto de tendencia, basta con estimar los parámetros de tendencia en el modelo $T_t = \alpha + \beta t$ y resolviendo mediante el método de mínimos cuadrados se tiene que

$$\hat{\alpha} = 21.5276 \quad \text{y} \quad \hat{\beta} = 0.000539.$$

Por tanto, se concluye que la componente periódica P_t afecta el comportamiento de la componente de tendencia, ya que los estimadores de los parámetros de tendencia

estimados difieren de los valores verdaderos ($\alpha = 5$, $\beta = 0.002$). La distancia entre los valores exactos y estimados es significativamente diferente.

Por otro lado, para comprobar si existe evidencia de algún efecto de la componente de tendencia sobre el efecto periódico, en la figura 3.11 se pueden ver las representaciones gráficas del periodograma de la componente periódica P_t (figura 3.11 *a*) y el de la interacción de ambos componentes $P_t + T_t$ (figura 3.11 *b*), *i.e.*, el periodograma de la componente periódica bajo la influencia de tendencia temporal.

Así, revisando los periodogramas de la figura 3.11 sobresale nuevamente que las frecuencias significativas permanecen invariantes en ambos casos, es decir, nuevamente se tiene evidencia de que las frecuencias no presentan cambios ante efectos de la componente de tendencia y destaca de forma análoga al caso de magnitudes aleatorias que la amplitud muestra pequeñas diferencias y en conclusión, la mejor opción para estimar las amplitudes es el método de mínimos cuadrados.

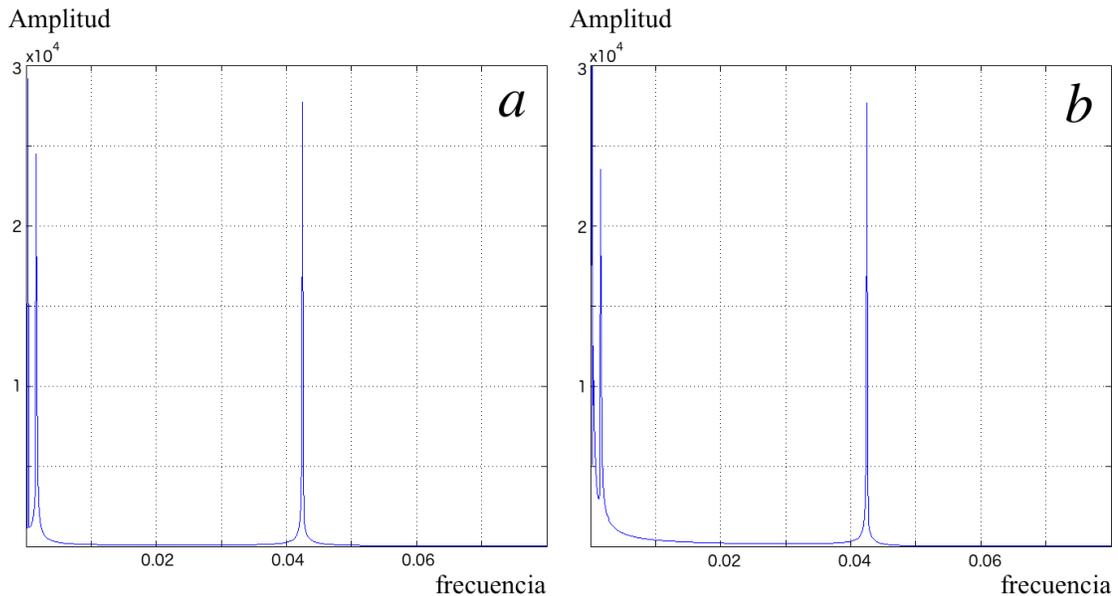


Figura 3.11: Representación gráfica de los periodogramas de la serie temporal simulada: (a) únicamente la componente periódica P_t y la interacción de las componentes de tendencia temporal y la periódica (b).

Más aún, analizando los resultados de una muestra de 10,000 simulaciones, procediendo de forma análoga, es decir, construyendo una serie periódica a partir de

la ecuación 2.18, con sumas de funciones seno y coseno con 3 frecuencias y amplitudes, de la forma

$$P_t = a_0 + \sum_{i=1}^3 a_i \cos(2\pi\omega_i t) + b_i \sin(2\pi\omega_i t),$$

y asignando las amplitudes de forma aleatoria para obtener las frecuencias y amplitudes $a_{0_{X_t}}$, $a_{j_{X_t}}$ y $b_{j_{X_t}}$ para $X_t = P_t + T_t$. En el 100 % de los casos los estimadores de frecuencias a partir del periodograma permanecen invariantes, en tanto que el error promedio que se observa en los estimadores de los parámetros de tendencia es de 17.9789 con una desviación estándar de ± 7.3382 .

Finalmente, con base en los resultados de simulación, se puede plantear un algoritmo para obtener un análisis de la componente aleatoria de una serie temporal en la que se tienen los efectos periódicos y de tendencia temporal de modo que, se puede abreviar en los siguientes pasos:

1. Bajo el supuesto de una componente periódica, primero se debe investigar si existe evidencia de frecuencias significativas ocultas en la serie.
2. Si se tienen efectos periódicos, una vez que se estimaron las frecuencias significativas, un segundo paso es calcular con el método de mínimos cuadrados las amplitudes correspondientes a cada una de las frecuencias estimadas.
3. En un tercer paso, se debe verificar si existe evidencia de tendencia temporal en la serie mediante el método de mínimos cuadrados.
4. Si los parámetros del modelo de tendencia son significativos, se debe retirar el efecto de tendencia y así, el paso final corresponde al análisis del comportamiento de la componente aleatoria.

Capítulo 4

Series de tiempo de radiación solar

Una vez que se establece un algoritmo para analizar la componente aleatoria de una serie de tiempo que incorpora efectos de tendencia, componentes periódicas, asociación de lecturas consecutivas y una componente impredecible, en este capítulo se revisa el comportamiento de series de radiación solar registrada en dos puntos de la Zona Metropolitana de Guadalajara (**ZMG**), que por su ubicación geográfica muestra las características típicas de los climas tropicales, es decir, una dinámica atmosférica y meteorológica que favorece fuertes variaciones en las series de datos de radiación solar, produciendo escenarios impredecibles en las variaciones a corto plazo por los cambios bruscos en sus condiciones locales. Además, es muy importante contar con una estimación confiable del recurso solar en la región, así como de sus variaciones a corto plazo, principalmente porque estas variaciones típicas de climas tropicales repercuten de forma negativa en el desempeño de los colectores fotovoltaicos.

Tener la posibilidad de calcular adecuadamente el desempeño esperado de dichos colectores permite asegurar una inversión que funcione correctamente y a su vez, estimar la recuperación adecuada de la inversión en estos recursos de energías alternativas.

Por tanto, lo primero es describir en términos generales las características de la zona de estudio, así como los conjuntos de datos disponibles para implementar el algoritmo propuesto y encontrar un modelo adecuado a las series temporales de radiación solar.

4.1. Descripción general de la zona de estudio

La ZMG abarca varios municipios, además de Guadalajara, se suman Tlaquepaque, Tonalá y Zapopan, y a partir de 2004 se adicionaron Tlajomulco de Zúñiga, El Salto y posteriormente Juanacatlán e Ixtlahuacán de los Membrillos como municipios exteriores (Figura 4.1); en donde tanto los municipios interiores como exteriores se consideran parte de la ZMG.



Figura 4.1: Localización de la Zona Metropolitana de Guadalajara (obtenida a partir de imágenes google).

Además de las áreas urbanizadas, se cuenta con una amplia zona industrial conformada por numerosas empresas nacionales y extranjeras que durante los años setentas del siglo pasado se establecieron en Guadalajara y entre las que sobresalen la industria fotográfica, la cigarrera, la hulera, de productos químicos, producción de alimentos, bebidas y metalurgia entre otras (Programa para el Mejoramiento de la Calidad del Aire en la ZMG, 1997; [43]).

Un parque vehicular que va en incremento, con automotores relativamente antiguos, destacando que en horas de mayor tráfico se excede la capacidad de tránsito en la ciudad, así como la emisión de contaminantes de los automotores. La reducción en áreas

verdes que ha sido significativa al interior de la metrópoli y la contribución de partículas suspendidas en áreas periféricas de la ciudad con escasa vegetación, rebasando con frecuencia las normas de calidad del aire por el exceso de partículas suspendidas, menor a 10 micras el que ha excedido la norma (Programa para el Mejoramiento de la Calidad del Aire en la ZMG, 1997; [43]). Esto llama la atención porque se configura en el aire una barrera de aerosoles que reducen la cantidad de radiación solar que llega a la superficie.

4.1.1. Condiciones geográficas

La ZMG se ubica en el centro del estado de Jalisco, a una latitud de $20^{\circ} 39' 54''$ Norte, longitud de $103^{\circ} 18' 42''$ Oeste y una altitud promedio de 1,540 metros sobre el nivel medio del mar. Las montañas que circundan la ZMG son: al noroeste la Sierra de San Esteban; al sureste, la Serranía de San Nicolás y los conjuntos montañosos Cerro Escondido-San Martín y El Tapatío-La Reyna; al sur el Cerro del Cuatro-Gachupín Santa María; y al oeste la Sierra de la Primavera (Figura 4.2).

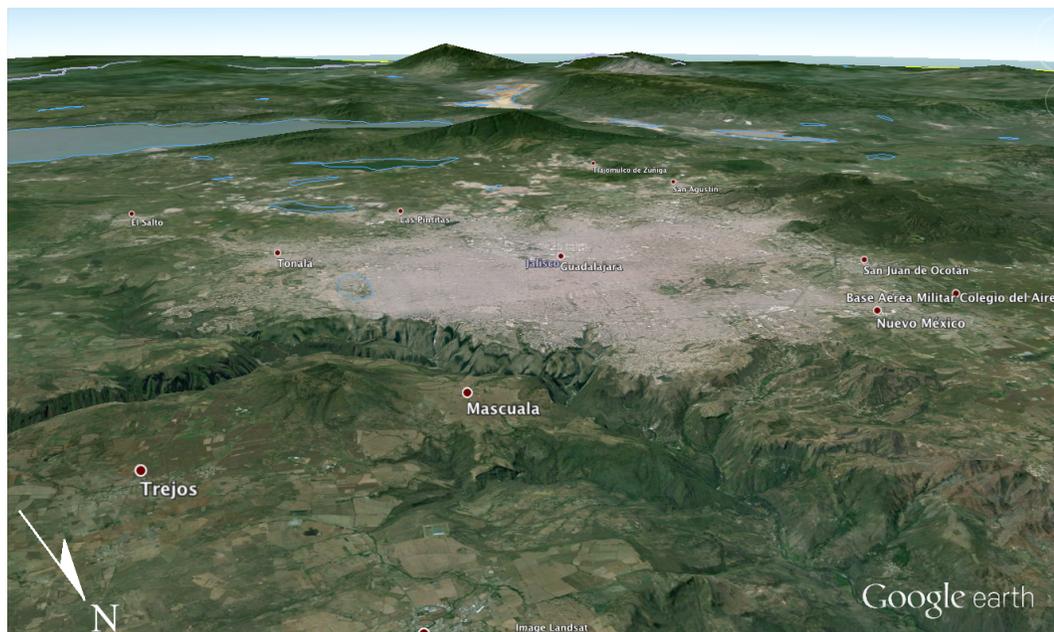


Figura 4.2: Condiciones geográficas de la Zona Metropolitana de Guadalajara (Elaborado con imágenes el software libre *Google Earth Pro* Ver. 7.3.1.4507).

Se sitúa en la cuenca del Valle del Río Grande de Santiago, en los Valles de Atemajac y la Planicie de Tonalá, entre las zonas montañosas de la Sierra Madre Occidental y el Cinturón Volcánico Transmexicano. Estas sierras constituyen parcialmente una barrera física natural para la circulación del viento, impidiendo el desalojo del aire contaminado fuera de la ZMG. Además, el terreno donde se ubica la ZMG tiene pendientes variables con un promedio de 3% (Programa para el Mejoramiento de la Calidad del Aire en la ZMG, 1997; [43]), lo que es un factor importante ya que se tienen alturas bajas (barrancas) y altas (cerros) en el interior de la ZMG, que en conjunto aportan a un escenario complejo en la dinámica atmosférica de la metrópoli.

4.1.2. Condiciones meteorológicas

La ZMG, se encuentra afectada por la afluencia de aire marítimo tropical; sin embargo, en el transcurso del año una gran variedad de fenómenos meteorológicos de escala regional, en superficie y en la atmósfera superior, tienen influencia sobre las condiciones meteorológicas de la zona metropolitana. Al estar en la región central del país, está sujeta a la influencia de sistemas anticiclónicos, generados en el Golfo de México y en el Océano Pacífico, y estos sistemas pueden ocasionar gran estabilidad atmosférica inhibiendo el mezclado vertical del aire.

El valle de Atemajac recibe una abundante radiación solar debido a su latitud de 20°N, lo que hace que la atmósfera sea altamente fotorreactiva; tiene un clima tropical con una temperatura media anual de 19°C y precipitación de 900mm por año. (Jauregui *et al.*, 1992; [68]). En la época de invierno conocida también como estación seca (o de pocas lluvias), el anticiclón del Pacífico (es decir, la zona de alta presión, con movimiento descendente del aire y vientos débiles en superficie), al bajar de latitud genera vientos occidentales que favorecen la entrada de aire marítimo tropical (Corriente en Chorro), con un ligero contenido de humedad, generando lluvias poco abundantes.

Sin embargo, de manera temporal, la presencia de la Corriente en Chorro (o de vientos máximos) en los niveles superiores de la atmósfera, llega a reflejar una intensificación de los vientos occidentales en superficie y un aumento en el aporte de humedad, generando algunas precipitaciones. Durante la misma época se presenta el

ingreso de masas de aire frío y seco que penetran por la región norte del país y avanzan hacia la región central, llegando a extenderse hasta la ZMG, provocando descensos de temperatura, algunas heladas y estratificación de las capas atmosféricas, intensificando con ello el fenómeno de inversión térmica (Programa para el Mejoramiento de la Calidad del Aire en la ZMG, 1997; [43]).

En la época de verano, conocida también como de lluvias, la ZMG se ve afectada por la entrada de aire cálido y húmedo procedente del Océano Pacífico, del Golfo de México y del Mar Caribe, provocando altas temperaturas y favoreciendo el movimiento vertical ascendente del aire, disminuyendo con ello la presencia, intensidad y espesor de las inversiones térmicas.

Por otra parte, la influencia del Golfo de México y Mar Caribe se manifiesta como un aumento de humedad en la ZMG, debido al constante avance de ondas tropicales, provocando importantes cantidades de lluvia y vientos de componente oriental. En el caso del Océano Pacífico, el transporte de humedad es favorecido por el desplazamiento de masas de aire hacia latitudes mayores de la Zona Intertropical de Convergencia (**ZIC**), la que al activarse ocasiona fuertes desprendimientos de humedad hacia los estados del Pacífico central, generando un aumento de cielos nublados e intensas precipitaciones.

Durante la misma época se desarrollan importantes sistemas tropicales sobre el Océano Pacífico oriental, como son las perturbaciones, depresiones, tormentas tropicales y huracanes.

Dichos sistemas llegan a provocar, debido a su circulación ciclónica contrarias a las manecillas del reloj y cercanía a la tierra, la entrada de bandas nubosas y altas cantidades de precipitación con intensificación de los vientos, los cuales alcanzan a llegar hasta la ZMG.

4.2. Series de datos

Al interior de la ZMG se cuenta con dos estaciones meteorológicas automáticas tipo Davis, modelo Vantage Pro2 (figura 4.3). En donde la primera se localiza en el interior del observatorio “*Colomos*”, ubicado en el interior del Parque Colomos, en una zona

boscosa semiurbanizada a una longitud $-103^{\circ}23'33''$ Este, latitud $20^{\circ}42'24''$ Norte y a una altitud de $1559m$ sobre el nivel medio del mar (figura 4.4); a cargo de la Comisión Nacional del Agua (CNA), que provee información al Servicio Meteorológico Nacional (SMN).



Figura 4.3: Estación meteorológica automática Davis Vantage Pro2.

La segunda se encuentra instalada en un ambiente urbanizado, al interior del Instituto de Astronomía y Meteorología (IAM) de la Universidad de Guadalajara, a una longitud de $-103^{\circ}23'03''$ Este, latitud $20^{\circ}40'29''$ Norte y a una altitud de $1592m$ sobre el nivel medio del mar y se muestra de forma georeferenciada en la misma figura (4.4).

Cabe señalar que ambas generan registros cada diez minutos de lecturas de temperatura (promedio, mínima y máxima), humedad relativa, punto de rocío, precipitación, dirección – velocidad del viento y radiación, entre otras; utilizan radio

intercepción de espectro amplio de frecuencia y son capaces de recibir la señal hasta 300m de distancia, no obstante en ambos casos (el observatorio Colomos y el IAM) la antena y el aparato receptor está a menos de 10m.

De forma gráfica, en la figura 4.4 se puede ver una imagen google en donde se localizan de manera georreferenciada las dos estaciones meteorológicas instaladas en la ZMG.

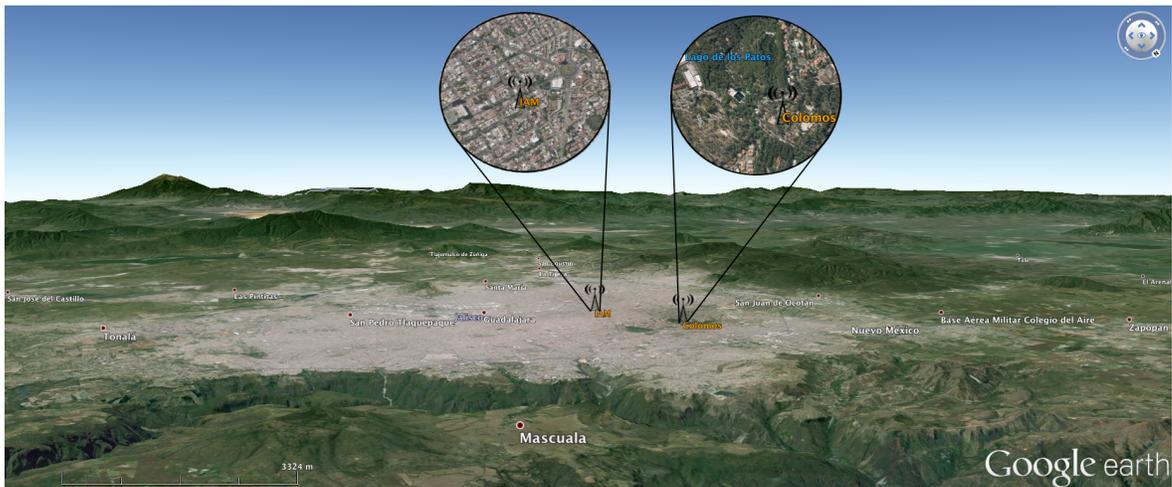


Figura 4.4: Estaciones meteorológicas Davis georreferenciadas en la Zona Metropolitana de Guadalajara (Elaborado con imágenes el software libre *Google Earth Pro* Ver. 7.3.1.4507).

Los sensores de temperatura y radiación solar se ubican dentro de un protector solar por lo que se tiene fiabilidad y precisión en sus lecturas. Se alimenta de energía eléctrica, pero cuentan con un sistema alternativo de tres baterías tipo C de reserva durante cortes breves de energía; no obstante, ambas estaciones tienen problemas de lecturas perdidas desde su instalación.

Las series de datos disponibles van desde el 28 de febrero del 2000 para la estación del observatorio Colomos, mientras que los registros en la estación del IAM inician el 23 de junio del 2003, ambas series muestran periodos largos de datos perdidos; ya sea por fallas de energía, datos sobrepuestos o fallas en los dispositivos de almacenamiento.

A manera de ejemplo, en la figura 4.5 se muestran las representaciones gráficas de los datos, en donde sobresale claramente que a partir de su fecha de inicio, en la estación del observatorio Colomos hay periodos de datos perdidos durante algunos días

en el año 2006; algunos meses de los años 2001, 2008, 2010 y 2011; alrededor de medio año perdido en el 2012 y el año completo en el 2009. Mientras que, para la estación instalada en el IAM, se tienen lecturas erróneas en el 2007 (se detecta por un periodo de lecturas bajas) y periodos mensuales de datos perdidos en los años 2004, 2012 y 2014.

De este modo, en este trabajo se muestran los resultados obtenidos con las bases de datos de ambas estaciones desde sus fechas respectivas de inicio hasta el siete de marzo del 2013, fecha máxima con registros en la estación del observatorio Colomos, con un promedio de once años de información disponible.

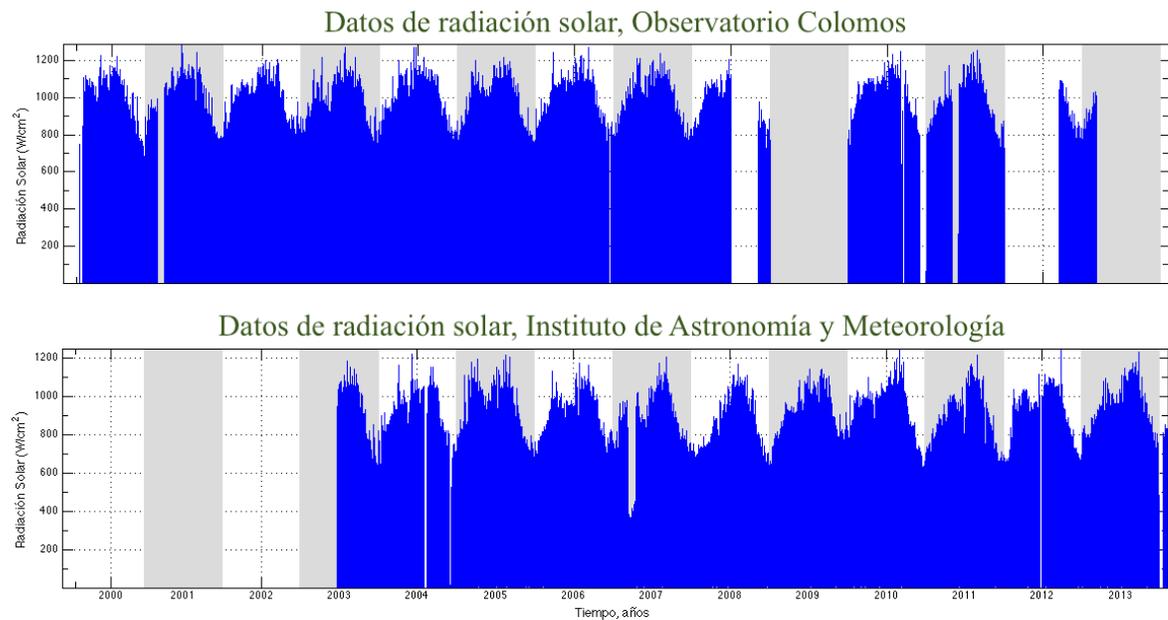


Figura 4.5: Series temporales de datos registrados en dos estaciones meteorológicas instaladas al interior de la ZMG: Observatorio Colomos y en el IAM.

4.2.1. Análisis descriptivo de las series de tiempo

Hasta este punto, es claro que los datos de radiación solar observados en la ZMG contienen una componente aleatoria, es decir, muestran fuertes variaciones en pequeños intervalos de tiempo por la dinámica típica en la atmósfera de climas tropicales. Además, partiendo de los resultados observados en el estudio de simulación, los

histogramas de magnitudes no negativas muestran un comportamiento específico cuando no están bajo la influencia de componentes de tendencia temporal o efectos estacionales.

Así, suponiendo que no existen otros efectos, en la figura 4.6 se muestran las representaciones gráficas de los histogramas de ambas series temporales de radiación solar para verificar si se tiene el escenario típico del comportamiento de una magnitud aleatoria.

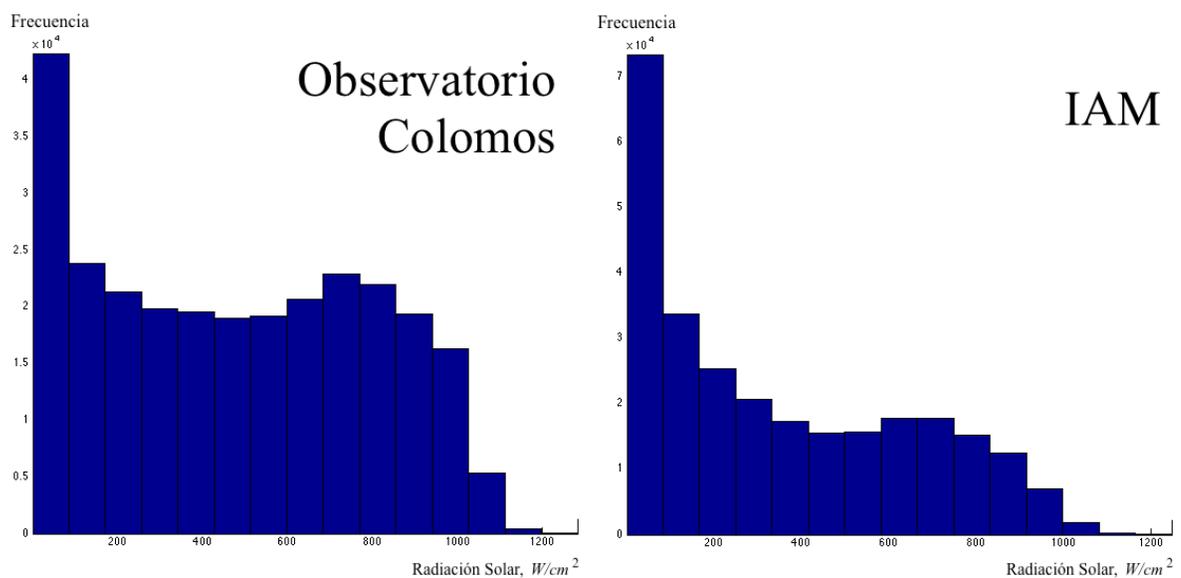


Figura 4.6: Histogramas de frecuencias observadas para las dos series temporales de radiación solar recibida en dos estaciones meteorológicas instaladas en la Zona Metropolitana de Guadalajara.

Sin embargo, en los dos histogramas de la figura 4.6 se puede ver un desempeño *bimodal* lo que puede ser resultado de la interacción de efectos de factores periódicos, de tendencia o de asociación entre lecturas consecutivas.

Así, el siguiente paso es estimar las componentes de la serie para determinar el comportamiento de la componente aleatoria sin influencia de otros factores.

Los resultados de este capítulo se obtuvieron con las rutinas en matlab que se presentan en el apéndice A.

4.3. Discusión de resultados

Analizando las series temporales de radiación solar en la ZMG, una observación general en lo que respecta al efecto de tendencia, los datos no dan evidencia para afirmar algún efecto significativo, es decir, al implementar el método de mínimos cuadrados se obtienen los valores de tendencia dados por

$$\widehat{\beta}_{Colomos} = 0.000010653,$$

para la estación Colomos, mientras que

$$\widehat{\beta}_{IAM} = -0.00001407,$$

para la estación instalada en el IAM, sin embargo en ambos casos son aproximadamente cero, por lo que no hay evidencia para afirmar efectos de tendencia en ambas series.

A pesar de que los estimadores son no significativos (ambos son aproximadamente cero), estos resultados son interesantes porque la estación Colomos está dentro de la zona boscosa o semiurbana y dado que $\beta > 0$, se observa un efecto de brillantés *i.e.*, la cantidad de radiación recibida en la superficie se incrementa en el tiempo; mientras que la estación instalada en el IAM, dentro de una zona totalmente urbanizada, muestra un $\beta < 0$ lo que implica un efecto de obscurecimiento, *i.e.*, la cantidad de radiación recibida en la superficie se reduce en el tiempo. Este resultado coincide, de manera localizada, con reportes de autores como Pinker, Zhang, Dutton (2005, [71]), Ye, Li, Sun y Guo (2009, [72]), quienes señalan un efecto llamado como “obscurecimiento global”, asociado fuertemente al cambio climático reportado en las últimas décadas. Sin embargo, detectar obscurecimiento en la zona urbana y brillantés en la zona boscosa puede reflejar la influencia de niveles elevados en la concentración de contaminantes como el ozono o las partículas menores a $10\mu m$ en la ZMG y sobre el IAM (Sánchez *et al.* 2015, [73]).

Por otro lado, al investigar efectos estacionales, en ambas series se tiene evidencia de efectos estacionales de 12 horas ($f_1 = 0.01389$), 24 horas ($f_2 = 0.006948$), 15 días ($f_3 = 0.000463$), 30 días ($f_4 = 0.000231$) y de aproximadamente 6 meses ($f_5 = 0.000038$). En la figura 4.7 se muestran las representaciones gráficas de los periodogramas estimados

para las dos series disponibles. En ellas se puede observar evidencia de frecuencias significativas para un intervalo de confianza del 95 %, destacando que en las gráficas de la figura solo se resaltan las frecuencias de 12 y 24 horas con los de 15 y 30 días, mientras que la frecuencia correspondiente a seis meses no se pueden ver en las representaciones gráficas porque está muy cercana al origen.

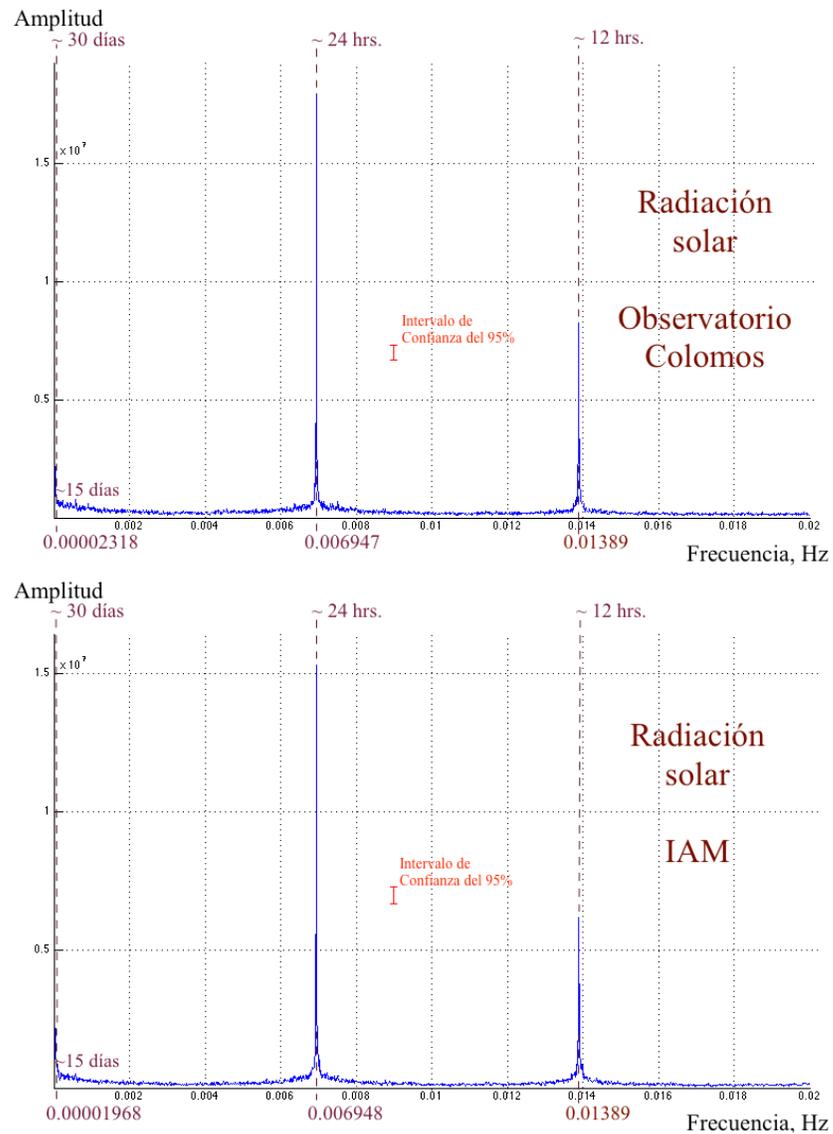


Figura 4.7: Periodogramas estimados con los datos de radiación solar observados en el Observatorio Colomos y en el IAM.

Una vez identificadas las frecuencias significativas, el siguiente paso es aplicar un filtro para eliminarlas. Técnicamente el uso del filtro consiste en eliminar las frecuencias significativas y conservar las frecuencias que corresponden a la componente aleatoria. En este caso se pueden suprimir las frecuencias menores a $0.014Hz$ y con ello, retirar el efecto estacional en las dos series de datos disponibles. Así, se obtiene la variación impredecible una vez que se retiraron los efectos periódicos de la serie.

Para verificar evidencia de efectos de asociación entre lecturas consecutivas, en la figura 4.8 se pueden observar los coeficientes de la autocorrelación estimadas $\hat{\rho}$ en ambas series temporales de radiación solar, dados por (Casella, 2002, [31])

$$\rho(h) = \frac{\mathbb{E}[(X_t - \mu)(X_{t+h} - \mu)]}{\sqrt{\mathbb{E}[(X_t - \mu)^2]}\sqrt{\mathbb{E}[(X_{t+h} - \mu)^2]}}, \quad \text{para } h = 1, 2, \dots, k.$$

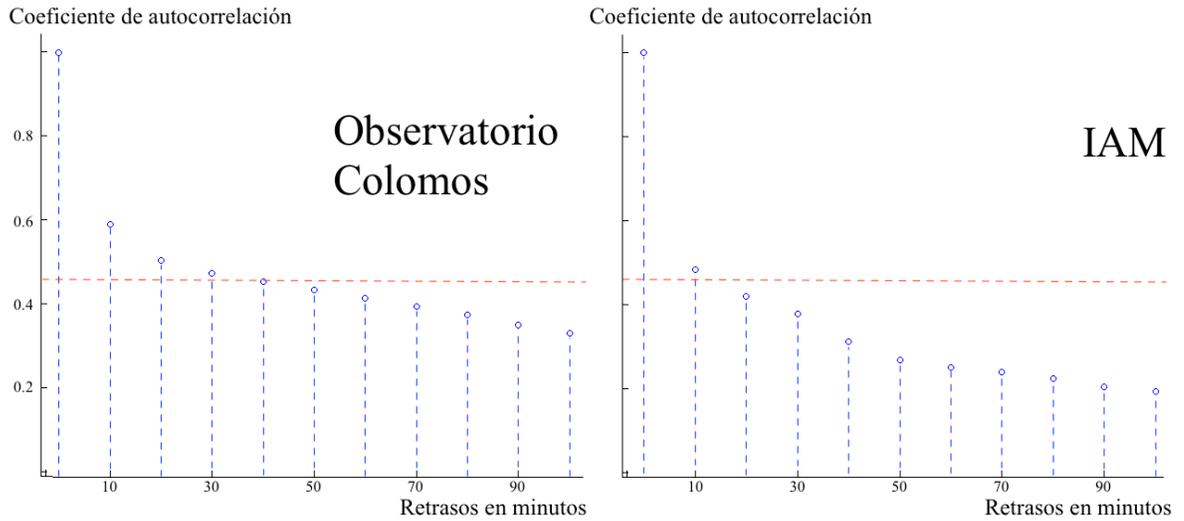


Figura 4.8: Periodogramas estimados con los datos de radiación solar observados en el Observatorio Colomos y en el IAM.

Además, en la figura 4.8 sobresale que para retrasos de diez minutos, la correlación estimada en ambas estaciones es menor a 0.60 y además, que en observaciones horarias (60 minutos) de la autocorrelación estimada es menor a 0.45, lo que hace suponer que la asociación entre lecturas consecutivas no es un factor relevante para las magnitudes impredecibles de las series, ya que una asociación perfecta se presenta para valores

$\rho = \pm 1$ y la ausencia de correlación corresponde a $\rho = 0$ (Casella y Berger, 2002, [31]; Rohatgi, 2005, [30]).

Aplicando los modelos en las variaciones aleatorias de las series una vez que se suprimieron los efectos estacionarios, en la tabla 4.1 se presentan dos medidas de distancia propuestas en la literatura para identificar el modelo que mejor representa los datos; en particular son el criterio de Akaike (AIC_C) dado por

$$AIC_C = 2k \left(1 + \frac{k-1}{n-k-1} \right) - 2 \ln(L),$$

en donde n es el tamaño de muestra, k es el número de parámetros que tiene el modelo de densidad de probabilidad y L es la función de verosimilitud dada en la sección 2.4.2 y la $RECM$, dada en la expresión 3.4. Ambos estimados a partir de la componente aleatoria de las series temporales de radiación solar para los cinco modelos propuestos.

Tabla 4.1: Criterio de información de Akaike y raíz del error cuadrado medio estimados para los datos de radiación solar

Modelo	Datos Colomos		Datos IAM	
	AIC_C	$RECM$	AIC_C	$RECM$
<i>Beta generalizado</i>	2687025.49	0.007492749788821	2655112.66	0.003826120713519
<i>Exponencial</i>	2700486.35	0.000000000005597	2658977.79	0.00000000024735
<i>Gamma</i>	2677856.06	0.00000000000608	2644892.26	0.00000000005476
<i>Lognormal</i>	2719489.40	0.00000000000060	2669380.88	0.00000000000059
<i>Weibull</i>	2673318.36	0.00000000000193	2639191.65	0.00000000000675

En este caso, en la tabla 4.1 se puede observar que bajo el criterio de la $RECM$, los datos dan evidencia para suponer que el modelo lognormal es el mejor para los datos en las estaciones disponibles (valores sobresaltados); en tanto que, el modelo beta generalizado de 1er tipo es el que muestra el peor desempeño de los modelos implementados. No obstante, se debe señalar que el modelo lognormal representa el peor escenario en términos del AIC_C , ya que es el que pierde más información relativa a los datos entre los cinco modelos revisados en este trabajo.

En la misma tabla 4.1 se puede ver que con base en el AIC_C , el modelo Weibull es

el que pierde menos información relativa en contraste con los cuatro modelos de prueba (valores resaltados) y sobresale además que de las estimaciones de *RECM* el modelo Weibull es la segunda mejor alternativa. Este resultado deja una pauta importante para decidir que el modelo Weibull es el modelo más apropiado para los datos disponibles en la ZMG.

Buscando reforzar las observaciones de la tabla 4.1, en las figuras 4.9 y 4.10 se pueden observar las gráficas del histograma de la magnitud impredecible en los datos de radiación solar registrados en las estaciones del observatorio Colomos, a manera de ejemplo, incorporando los ajustes de los modelos propuestos. En donde sobresale el modelo lognormal por su proximidad al histograma de los datos, pero el modelo Weibull es una buena alternativa en contraste con los modelos bajo estudio.

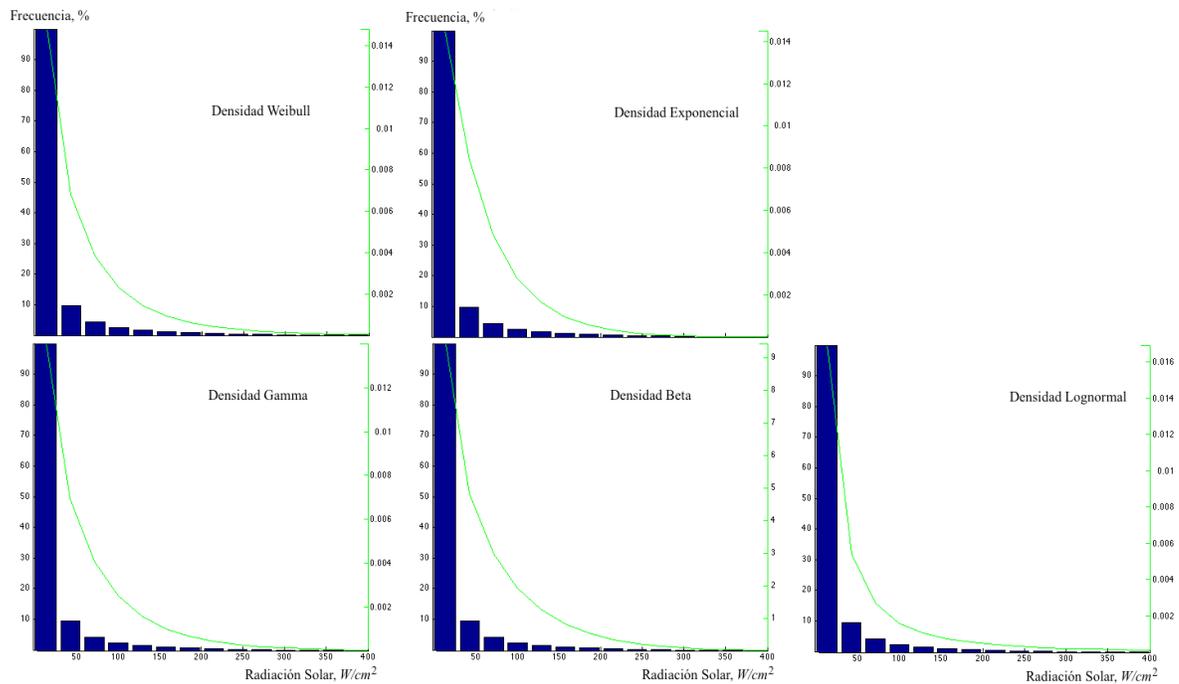


Figura 4.9: Modelos propuestos e histograma de las variaciones aleatorias registradas en el observatorio Colomos.

Sobresale además en los histogramas de las figuras 4.9 y 4.10 que no hay evidencia de un comportamiento binomial en las magnitudes aleatorias de las series de radiación solar observadas en la zona de estudio como se pudo observar en el histograma de los datos sin suprimir los efectos periódicos y de tendencia en la serie observada (ver

figura 4.6); de modo que, se puede argumentar que el comportamiento tipo bimodal observado en los datos de estudio por otros autores se puede atribuir a la influencia de los efectos periódicos y de tendencia temporal en las series de datos.

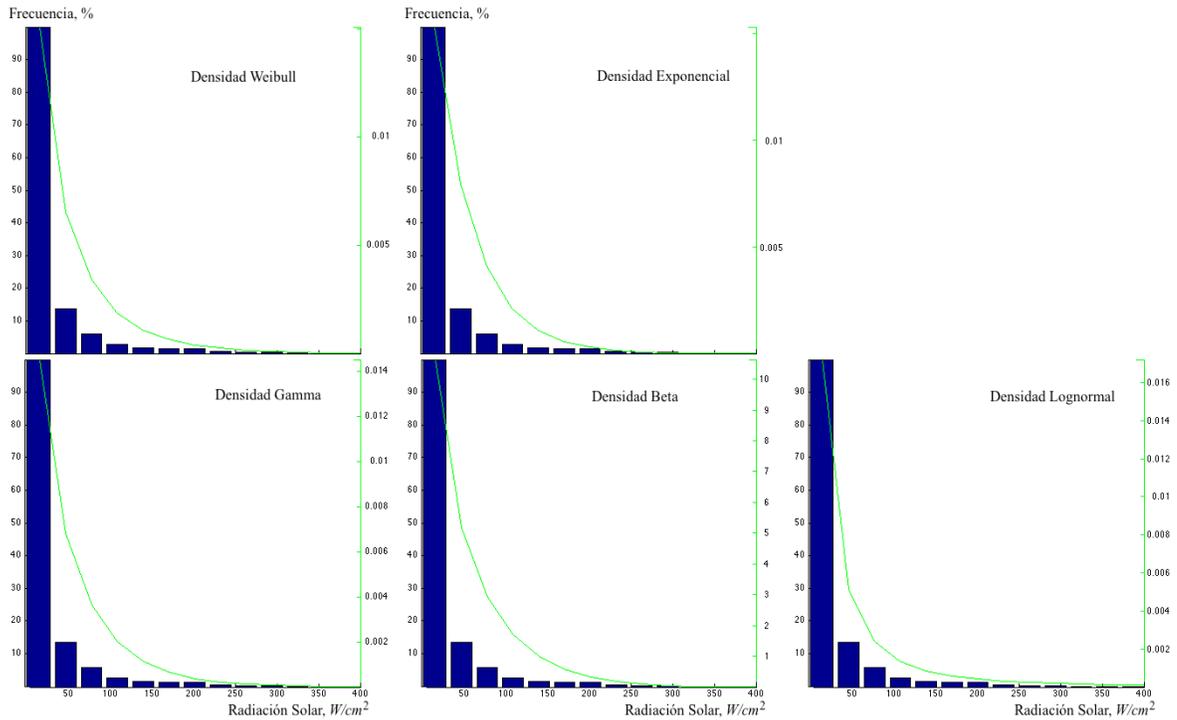


Figura 4.10: Modelos propuestos e histograma de las variaciones aleatorias registradas en el IAM.

Cabe señalar que los ejes verticales en color verde en las gráficas de ambas figuras corresponden a las magnitudes de las curvas estimadas con los datos para las densidades de probabilidad bajo estudio.

Capítulo 5

Conclusiones

En este trabajo se presentó una metodología para analizar las variaciones impredecibles en series temporales, sin la interacción de componentes típicas en series observadas en la naturaleza, incorporando parámetros que modelan efectos periódicos, el modelo de tendencia temporal y de asociación entre lecturas consecutivas; identificando claramente cada una de estas componentes y proponiendo los modelos paramétricos individuales para analizar cada componente.

Se llevó a cabo un estudio de simulación generando series temporales que reproducían el comportamiento de estas componentes típicas en la naturaleza y se identificó el efecto que produce cada una de estas en las magnitudes impredecibles de la serie, así como el impacto que se genera en las mismas componentes por su interacción conjunta y se aplicó la metodología propuesta en dos series temporales de registros de radiación solar medidos en dos estaciones automáticas instaladas en dos puntos de la Zona Metropolitana de Guadalajara.

De los resultados del estudio de simulación se encontró que:

- La componente aleatoria de una serie temporal se deforma por completo al combinarse con efectos periódicos o con efectos de tendencia temporal, perdiendo su configuración al grado de que el usuario no tiene elementos para suponer que los datos provienen de la población original.

Más aún, se observaron varios casos en donde los datos seguían un comportamiento bimodal, trimodal y con colas pesadas, cuando el modelo original de la

componente aleatoria no tenía esa forma.

- En todas las simulaciones ejecutadas, se observó que la componente aleatoria no afecta a las frecuencias significativas de la componente periódica, para cinco modelos de probabilidad (Beta Generalizado, exponencial, Weibull, gamma y Gaussiano) y en esa misma prueba, se observó que se reducen las amplitudes de las frecuencias significativas de la serie.
- Además, se observó que las frecuencias significativas de las componentes periódicas permanecían invariantes ante los efectos de tendencia, en tanto que los coeficientes del factor de tendencia se alteraban por influencia del efecto periódico.
- De las corridas de simulación, se encontró que la componente aleatoria afecta fuertemente al término α del factor de tendencia $\alpha + \beta t$, influyendo en menor grado en el parámetro β .

De modo que, con base en los resultados de simulación, el algoritmo propuesto para analizar la componente aleatoria consta de los siguientes pasos:

1. Bajo el supuesto de una componente periódica, primero se debe investigar si existe evidencia de frecuencias significativas ocultas en la serie.
2. Si se tienen efectos periódicos, una vez que se estimaron las frecuencias significativas, un segundo paso es calcular con el método de mínimos cuadrados las amplitudes correspondientes a cada una de las frecuencias estimadas.
3. En un tercer paso, se debe verificar si existe evidencia de tendencia temporal en la serie mediante el método de mínimos cuadrados.
4. Si los parámetros del modelo de tendencia son significativos, se debe retirar el efecto de tendencia y así, el paso final corresponde al análisis del comportamiento de la componente aleatoria.

Por otro lado, de los resultados obtenidos en dos series temporales de radiación solar registradas en dos estaciones automáticas instaladas en dos puntos de la ZMG se observa lo siguiente:

- No hay evidencia de un comportamiento bimodal en los datos registrados en la zona de estudio como lo han reportado en otros trabajos. Más aún, hay evidencia para afirmar que el comportamiento bimodal se puede atribuir a la influencia de los efectos periódicos y de tendencia temporal de las series.
- Se probaron cinco modelos de densidad de probabilidad sugeridos en la literatura, en específico, los modelos *beta* (beta generalizado), el *exponencial*, el modelo *gamma*, el *Weibull* y el *lognormal*. De ellos, el modelo que presenta el mejor desempeño en términos del criterio de Akaike, es la función de densidad de probabilidad de Weibull, en tanto que bajo la RECM, el mejor es el modelo lognormal. No obstante, dado que el modelo lognormal pierde más información en términos del AIC_C , se concluye que la densidad Weibull es la mejor alternativa para la componente aleatoria de las series de radiación solar observadas en la ZMG.
- Se encontró además que la Metrópoli presenta una tendencia negativa, es decir, se tiene evidencia de “obscurecimiento” en el interior de la ciudad, posiblemente atribuible a los niveles de contaminación atmosférica, mientras que en la zona boscosa (observatorio Colomos) se tiene evidencia de “brillantéz” o tendencia creciente en la incidencia de radiación solar.

Finalmente, se logra cumplir con los objetivo general y particulares planteados en este trabajo encontrando un algoritmo confiable que permite analizar la componente aleatoria de series temporales observadas en la naturaleza, dando apertura a implementar esta metodología en otras series de datos, tales como velocidad del viento, concentración de contaminantes en el aire.

Apéndice A

Programas utilizados

Los comandos y rutinas que se enlistan a continuación se ejecutaron en Mat-Lab 7.6, R2008a en una computadora con sistema operativo Linux, con procesador Intel Core 2 Duo a 2.4GHz y 2GB SDRAM.

En términos generales, se supone que los datos disponibles se almacenan en la variable y y el número de lectura en la variable x .

Tendencia

Buscando siempre establecer el caso más simple de tendencia, se pueden estimar los parámetros del modelo de tendencia determinística 2.3 mediante los comandos

```
l=ones(size(x));  
x=[1,x];  
a=inv(x'*x)*(x'*y);  
trend=x*a;  
x=x(:,2);  
y=y-x*a(2); % Elimina tendencia
```

Buscando periodicidades ocultas

Para obtener evidencia de periodicidades ocultas, se utiliza el modelo 2.6 buscando las $n/2$ (para una muestra de tamaño n) frecuencias posibles, mediante las siguientes líneas

```
N=floor(max(t)/2);
fr=linspace(1/N,.5,10000);
n=2*N;
c=y'*cos(2*pi*t*fr)/N;
s=y'*sin(2 *pi*t*fr)/N;
rp=n*(c.^2+s.^2);
li=prctile(rp,95);
[r,s]=max(rp);
rps=rp;
for i=1:3
    rps=smooth(rps);
end
```

en donde la variable li almacena la longitud a partir de la cual la frecuencia f_i se considera significativa a un nivel $\alpha = 0.95$.

Cabe mencionar que por el número de lecturas $n = 733249$ (tomadas cada 10 minutos) se tiene que, aproximadamente:

- 72 (= 6*12) corresponde a medio día,
- 144 (= 6*24) corresponde a un día,
- 2160 (= 6*24*15) corresponde a 15 días,
- 4320 (= 6*24*30) corresponde a 30 días,
- 26280 (= 6*12*365) corresponde a medio año y
- 52560 (= 6*24*365) corresponde a un año.

Ajustando los modelos de probabilidad propuestos

Una ventaja de utilizar software matemático, es que cuenta con algoritmos probados y verificados. Entre éstas, matlab cuenta con funciones que proporcionan la estimación de parámetros por máxima verosimilitud para los modelos analizados en este trabajo, de modo que; si se utilizaran lenguajes de programación, tales como *C*, *C++*, *C#*, Fortran 77, 90 o 95, se tendrían que programar las rutinas que generan números pseudoaleatorios de los modelos propuestos, así como los métodos de estimación paramétrica de máxima verosimilitud, de momentos, soluciones matriciales de ecuaciones lineales y no lineales, inversa de una matriz, entre otras operaciones necesarias en este trabajo. De esta forma, las estimaciones para cada uno de los modelos, así como el calculo del AIC_C y de la $RECM$ se obtienen mediante:

- Para el modelo Weibull se tiene


```
pdf=wblfit(yruido);
pdf=wblpdf(nn, pdf(1), pdf(2));
k=2; pdf=wblfit(yruido);
AIC=2*wbllike(pdf, yruido)+2*k*nr/(nr-k-1);
yest=wblinv(wblcdf(yruido, pdf(1), pdf(2)), pdf(1), pdf(2));
RECM=sqrt(mean((yruido-yest).^2));
```
- Para el modelo exponencial se tiene


```
pdf=expfit(yruido);
pdf=exppdf(nn, pdf);
k=1; pdf=expfit(yruido);
AIC=2*explike(pdf, yruido)+2*k*nr/(nr-k-1);
yest=expinv(expcdf(yruido, pdf), pdf);
RECM=sqrt(mean((yruido-yest).^2));
```

- Para el modelo gamma se tiene


```

pdf = gamfit(yruido);
pdf = gampdf(nn,pdf(1),pdf(2));
k = 2;pdf = gamfit(yruido);
AIC = 2*gamlike(pdf,yruido)+2*k*nr/(nr-k-1);
yest = gaminv(gamcdf(yruido,pdf(1),pdf(2)),pdf(1),pdf(2));
RECM = sqrt(mean((yruido-yest).^2));

```
- Para el modelo beta se tiene


```

ybeta = yruido/max(yruido);
pdf = betafit(ybeta);
pdf = betapdf(nn/max(yruido),pdf(1),pdf(2));
k = 2;pdf = betafit(ybeta);
AIC = -2*betalike(pdf,ybeta)+2*k*nr/(nr-k-1);
yest = betainv(betacdf(ybeta,pdf(1),pdf(2)),pdf(1),pdf(2))*max(yruido);
RECM = sqrt(mean((yruido-yest).^2));

```
- Para el modelo lognormal se tiene


```

pdf = lognfit(yruido);
pdf = lognpdf(nn,pdf(1),pdf(2));
k = 2;pdf = lognfit(yruido);
AIC = 2*lognlike(pdf,yruido)+2*k*nr/(nr-k-1);
yest = logninv(logncdf(yruido,pdf(1),pdf(2)),pdf(1),pdf(2));
RECM = sqrt(mean((yruido-yest).^2));

```

Índice de figuras

2.1. Una representación gráfica de dos rutas muestrales de un proceso estocástico.	8
2.2. Representación gráfica de un esquema que muestra el error que se comete al suponer incorrectamente un modelo.	17
2.3. Representaciones gráficas para $\beta\left(\frac{1}{3}, \frac{1}{3}\right)$, $\beta\left(\frac{1}{3}, 3\right)$, $\beta(3, 5)$, $\beta\left(3, \frac{1}{3}\right)$ y $\beta(3, 3)$. 19	
2.4. Representaciones gráficas de la funciones de densidad y de distribución, familia exponencial, para $\mu = 1/2$, $\mu = 1$ y $\mu = 2$	22
2.5. Representaciones gráficas de la funciones de densidad y de distribución de la familia Weibull, para $\gamma = 2$ y $\lambda = 1/2, 1, 2$	23
2.6. Representaciones gráficas de la funciones de densidad y de distribución de la familia gamma, para $\lambda = 2$ y $k = 1/2, 1, 2$	26
3.1. Histograma de frecuencias de una serie simulada con $\mu = 0.2$	36
3.2. Representación gráfica de la componente aleatoria de una serie temporal simulada, variable aleatoria exponencial con $\mu = 0.2$; serie completa (a), haciendo un acercamiento para $1000 \leq t \leq 2000$ (b) y para $1000 \leq t \leq 1500$ (c).	36
3.3. Representación gráfica de la componente periódica de la serie simulada para frecuencias diurna (c), mensual (b) y anual (a), todas con amplitud 20.	37

3.4. Representación gráfica de la componente periódica de la serie temporal simulada, incorporando las frecuencias diurna, mensual y semianual; serie completa (a), haciendo un acercamiento para $1000 \leq t \leq 2000$ (b) y para $1000 \leq t \leq 1500$ (c).	38
3.5. Representación gráfica de la serie temporal simulada $X_t = P_t + E_t$ para la serie completa (a) y haciendo un acercamiento para $1000 \leq t \leq 1500$ (b).	38
3.6. Representación gráfica de los periodogramas de la serie temporal simulada: (a) únicamente la componente periódica P_t y (b) la interacción de las componentes aleatoria y periódica $X_t = P_t + E_t$	39
3.7. Representación gráfica de los histogramas de la serie temporal simulada: (a) únicamente la componente aleatoria E_t y (b) la interacción de las componentes aleatoria y periódica $X_t = P_t + E_t$	40
3.8. Representación gráfica de la serie temporal simulada, (a) componente de tendencia temporal T_t , (b) componente aleatoria E_t y la interacción de las componentes aleatoria y temporal (c).	44
3.9. Histogramas de frecuencia de una variable aleatoria exponencial (a) y la interacción de las componentes aleatoria y de tendencia temporal (b).	45
3.10. Representación gráfica de la serie temporal simulada, (a) componente de tendencia temporal T_t , (b) componente periódica P_t y la interacción de las componentes periódica y temporal $P_t + T_t$ (c).	46
3.11. Representación gráfica de los periodogramas de la serie temporal simulada: (a) únicamente la componente periódica P_t y la interacción de las componentes de tendencia temporal y la periódica (b).	47
4.1. Localización de la Zona Metropolitana de Guadalajara (obtenida a partir de imágenes google).	50
4.2. Condiciones geográficas de la Zona Metropolitana de Guadalajara (Elaborado con imágenes el software libre <i>Google Earth Pro</i> Ver. 7.3.1.4507).	51
4.3. Estación meteorológica automática Davis Vantage Pro2.	54

4.4. Estaciones meteorológicas Davis georreferenciadas en la Zona Metropolitana de Guadalajara (Elaborado con imágenes el software libre <i>Google Earth Pro</i> Ver. 7.3.1.4507).	55
4.5. Series temporales de datos registrados en dos estaciones meteorológicas instaladas al interior de la ZMG: Observatorio Colomos y en el IAM.	56
4.6. Histogramas de frecuencias observadas para las dos series temporales de radiación solar recibida en dos estaciones meteorológicas instaladas en la Zona Metropolitana de Guadalajara.	57
4.7. Periodogramas estimados con los datos de radiación solar observados en el Observatorio Colomos y en el IAM.	59
4.8. Periodogramas estimados con los datos de radiación solar observados en el Observatorio Colomos y en el IAM.	60
4.9. Modelos propuestos e histograma de las variaciones aleatorias registradas en el observatorio Colomos.	62
4.10. Modelos propuestos e histograma de las variaciones aleatorias registradas en el IAM.	63

Índice de tablas

3.1. Errores observados al comparar las frecuencias simuladas f_j y las frecuencias estimadas $\hat{f}_{j_{X_t}}$, en la interacción de componentes periódicas con magnitudes aleatorias.	42
3.2. Errores observados al comparar las amplitudes simuladas a_j y las estimadas $a_{j_{X_t}}$, en la interacción de componentes periódicas con magnitudes aleatorias.	42
3.3. Errores observados al comparar las densidades empíricas de probabilidad para la magnitud E_t y la interacción de componentes periódicas con magnitudes aleatorias X_t	43
4.1. Criterio de información de Akaike y raíz del error cuadrado medio estimados para los datos de radiación solar	61

Bibliografía

- [1] SENER (2013) *Prospectiva de energías renovables 2013-2027*, Secretaría de Energía, México, Recuperado el 20 de mayo 2014, [http://www.sener.gob.mx/res/PE_y_DT/pub/2013/Prospectiva_Energias_Renovables_\\$2013-2027\\$.pdf](http://www.sener.gob.mx/res/PE_y_DT/pub/2013/Prospectiva_Energias_Renovables_$2013-2027$.pdf)
- [2] REN21 (2010) *Renewables 2010: Global Status Report*, REN21, Secretariat c/o UNEP, Paris, Recuperado en enero 2014, http://www.ren21.net/Portals/0/documents/activities/gsr/REN21_GSR_2010_full_revised%20Sept2010.pdf
- [3] REN21 (2014) *Renewables 2014: Global Status Report*, REN21, Secretariat c/o UNEP, Paris, ISBN9783981593426, Recuperado en julio 2014, http://www.ren21.net/Portals/0/documents/Resources/GSR/2014/GSR2014_full_report_low_res.pdf
- [4] C.A. Gueymard, S.M. Wilcox (2011) *Assessment of spatial and temporal variability in the US solar resource from radiometric measurements and predictions from models using ground-based or satellite data*, Sol. Energy 85(5), pp. 1068–1084.
- [5] F. Vignola, C. Grover, N. Lemon, A. McMahan (2012) *Building a bankable solar radiation dataset*, Sol. Energy 86(8), pp. 2218–2229.
- [6] T. Soubdhan, R. Emilion, R. Calif (2008) *Classification of daily solar radiation distributions using a mixture of Dirichlet distributions*. Article soumis à Solar Energy. Recuperado en mayo 15 de 2014. <https://hal.archives-ouvertes.fr/hal-00289567/document>.

- [7] McKnight, T.L.; Hess, D. (2000). *Climate Zones and Types: The Köppen System*, Physical Geography: A Landscape Appreciation. Upper Saddle River, NJ. Prentice Hall. pp. 205–211.
- [8] Assunção, H. F.; Escobedo, J. F.; Oliveira, A. P. (2003) *Modelling frequency distributions of 5 minute-averaged solar radiation indexes using Beta probability functions*, Theor. Appl. Climatol. 75, pp. 213–224.
- [9] M. Jurado, J.M. Caridad, V. Ruiz (1995) *Statistical distribution of the clearness index with radiation data integrated over five minute intervals*, Solar Energy, 55, pp. 469–473.
- [10] Chang, T. P. (2010) *Investigation on Frequency Distribution of Global Radiation Using Different Probability Density Functions*, International Journal of Applied Science and Engineering, 8, 2, pp. 99-107.
- [11] Hayashi, Y. (1981) *Space-Time Spectral Analysis and its Applications to Atmospheric Waves*, Journal of the Meteorological Society of Japan, 60, 1, pp. 156-171.
- [12] Zhao, J.; Qu, W. (1995) *Study of the dynamical mechanism of seasonal variation of Earth's rotational velocity*, J. of Geophysical Research, 100, B7, pp. 12719-12730.
- [13] Wilks, D.S. (2002). *Realizations of daily weather in forecast seasonal climate*, Journal of Hydrometeorology, 3, pp. 80-92.
- [14] Sánchez R.; Calderón H.; Fregoso J.; Brito E.; Rujano M.; Sánchez S. (2010) *Análisis y estimación de datos perdidos en investigación aplicada*, Editorial Universitaria, Universidad de Guadalajara.
- [15] Bendat, J. S.; Piersol, A. G. (2010) *Random Data: Analysis and Measurement Procedures, 4th Edition*, John Wiley & Sons, Inc., New York.
- [16] Bhattacharya, R. N.; Gupta, V. K.; Waymire, E. (1983) *The Hurst effect under trends*, J. Appl. Prob. 20, pp. 649-662.
- [17] Sánchez Gómez, R. (2001) *Análisis de tendencia en excedencias sobre un umbral alto, con aplicación en ozono urbano*, Disertación doctoral, Colegio de Postgraduados.

- [18] Martínez, J.; Fernández, A. (2004) *Cambio Climático, una visión desde México*, Instituto Nacional de Ecología, Secretaría del Medio Ambiente y recursos Naturales, México.
- [19] Mudelsee, M. (2010) *Climate time series analysis: classical statistics and bootstrap methods*, Springer, Dordrecht, <http://www.manfredmudelsee.com/book/>.
- [20] Gluhovsky, A. (2011) *Statistical inference from atmospheric time series: detecting trends and coherent structures*, *Nonlin. Processes Geophys.*, 18, pp. 537–544.
- [21] Wilks, D.S.; Wilby, R.L. (1999) *The weather generation game: a review of stochastic weather models*, *Progress in Physical Geography*, 23, pp. 329-357.
- [22] Karaki, H.; Chedid, R.B.; Ramadan, R. (1999) *Probabilistic Performance Assessment of Autonomous Solar-Wind Energy Conversion Systems*, *IEEE Transactions on Energy Conversion*, 14, 3, pp. 766-772.
- [23] Villanueva, D.; Feijóo, A.; Pazos, J. L. (2013) *Multivariate Weibull Distribution for Wind Speed and Wind Power Behavior Assessment*, *Resources*, 2, pp. 370-384.
- [24] VanMarke, E. (1988) *Random Fields, Analysis and Synthesis*. The MIT Press, Cambridge, Massachusetts, London, England.
- [25] Smith, R. L. (1999) *Time Series*, Department of Statistics, University of North Carolina, Chapel Hill, NC 27599-3260, Version 1.0.
- [26] Bloomfiel, P. (2000) *Fourier Analysis of Time Series*, John Wiley & Sons, Inc., New York.
- [27] Chatfield, C. (2003) *The Analysis of Time Series*. Chapman Hall/CRC, sixth edition.
- [28] Brockwell, P.J.; Davis, R. A. (2002) *Introduction to time series and forecasting 2nd ed.*, Springer-Verlag New York, Inc.
- [29] Stuart A. & Ord, K. (1994) *Kendall's advanced theory of statistics, sixth edition, Vol1, Distribution Theory*, Arnold Ed., London, England.

- [30] Rohatgi, V.K. (2005) *An Introduction To Probability And Statistical*, 2Ed., John Wiley & Sons Inc.
- [31] Casella, G. & Berger, R.L. (2002) *Statistical Inference*, 2Ed., Duxbury Advanced Series, Thomson Learning.
- [32] Papoulis, A. (1991) *Probability, Random Variables, and Stochastics Processes*, McGraw-Hill, USA.
- [33] Bailey, N.T.J. (1964) *The elements of Stochastics Processes with applications to the natural sciences*, John Wiley & Sons, Inc.
- [34] Kannan, N.T.J. (1979) *An Introduction to Stochastics Processes*, Elsevier North Holland Series.
- [35] Tijms, H.C. (2003) *A first Course in Stochastics Models*, John Wiley.
- [36] Lawless, J. F. (2002) *Statistical Models and Methods for Lifetime Data*, Editorial Wiley-VCH, USA.
- [37] Harvey, H. H. (1979) *The Acid Deposition Problem and Emerging Research Needs in the Toxicology of Fishes*, Proc. 5th Annual Aquatic Toxicity Workshop, Hamilton, Ontario, 7-9, 1978, Fish. Afar. Ser. Tech. Rep. 862, 1, pp. 15-128.
- [38] Driscoll, C.T.; Lawrence, G.B.; Bulger, A.J.; Butler, T.J.; Cronan, C.S.; Eagar, C.; Lambert, K.F.; Likens, G.E.; Stoddard, J.L.; Weathers, K.C. (2001) *Acid Rain Revisited: advances in scientific understanding since the passage of the 1970 and 1990, Clean Air Act Amendments*. Hubbard Brook Research Foundation. Science LinksTM Publication. Vol. 1, no. 1.
- [39] García, M. E.; *et al.* (2006) *Influencia de los contaminantes SO₂ y NO₂ en la formación de lluvia ácida en la Zona Metropolitana de Guadalajara, Jalisco.*, e-Gnosis, Universidad de Guadalajara, ISSN 1665 – 5745, pp. 1-16.
- [40] Singh, A; Agrawal, M. (2008) *Acid rain and its ecological consequences*, Journal of Environmental Biology, 29, 1, pp. 15-24, Recuperado el 20 de mayo de 2014, http://jeb.co.in/journal_issues/200801_jan08/paper_02.pdf.

- [41] National Research Council. (1991) *Rethinking the Ozone Problem in Urban Regional Air Pollution*, National Academy Press, Washington, D.C.
- [42] Lippmann M. (1992) *Environmental Toxicants. Human Exposures; Their Health Effects*, Van Nostrand Reinhold, NY.
- [43] Gobierno del Estado de Jalisco, Secretaría del Medio Ambiente, Recursos Naturales y Pesca, Secretaría de Salud (1997) *Programa para mejoramiento la calidad del aire en la Zona Metropolitana de Guadalajara 1997-2001*, Gobierno del Estado de Jalisco.
- [44] IPCC (2007) *Cambio climático 2007: Informe de síntesis*, Contribución de los Grupos de trabajo I, II y III al Cuarto Informe de evaluación del Grupo Intergubernamental de Expertos sobre el Cambio Climático [Equipo de redacción principal: Pachauri, R.K. y Reisinger, A. (directores de la publicación)]. IPCC, Ginebra, Suiza.
- [45] NRC (2010). *Advancing the Science of Climate Change*. National Research Council. The National Academies Press, Washington, DC, USA.
- [46] Farmer, G.T. (2015) *Modern Climate Change Science, Chapter 2 "Status of Climate Change Research"*, SpringerBriefs in Environmental Science, Springer International Publishing, ISBN 9783319092218.
- [47] Office of the Historian (2013), *Oil Embargo, 1973–1974*, Bureau of Public Affairs, United States Department of State. Recuperado el 20 de mayo de 2014, <http://history.state.gov/milestones/1969-1976/oil-embargo>.
- [48] Alatorre, C. (2009) *Energías Renovables para el Desarrollo Sustentable en México*, Secretaría de Energía, Deutsche Gesellschaft für Technische Zusammenarbeit, MBZ.
- [49] Suehrcke, H; McCormick, P.G. (1989) *Solar radiation utilizability*, Solar Energy, 43, pp. 339-345.
- [50] Gansler, R.A.; Klein, S.A.; Beckman, W.A. (1995) *Investigation of minute solar radiation data*, Solar Energy, 55, pp. 21-27.

- [51] Tovar, J.; Olmo, F.J.; Batlles, F.J.; Alados-Arboledas, L. (2001) *Dependence of one-minute global irradiance probability density distributions on hourly irradiation*, Elsevier – Energy, 26, pp. 659-668.
- [52] Skartveit, A.; Olseth, J.A.; *The probability density and autocorrelation of short-term global and beam irradiance*, Solar Energy 1992, 49, pp. 477–487.
- [53] Zaharim, A.; Mahir, R.A.; Pei, G.T.; Kamaruzzaman, S. (2009) *Time Series Analysis of Solar Radiation Data in the Tropics*, European Journal of Scientific Research, Vol.25 No.4, pp.672-678.
- [54] Cervantes, S. (2015) *Proyectan tres parques de energía eólica en BC*, Recuperado en junio de 2014, <http://eleconomista.com.mx/estados/2015/01/25/proyectan-tres-parques-energia-eolica-bc>.
- [55] Elhadidy, M.A.; Shaahid, S.M.; (2000) *Parametric study of hybrid (wind+solar+diesel) power generating systems*, Renew Energy, Vol. 21, No 2, pp. 129–139.
- [56] Tina, G.; Gagliano, S. (2008) *Probability Analysis of Weather Data for Energy Assessment of Hybrid Solar/Wind Power System*, 4th IASME/WSEAS International Conference on ENERGY, ENVIRONMENT, ECOSYSTEMS and SUSTAINABLE DEVELOPMENT (EEESD'08), Algarve, Portugal.
- [57] Kronratjev, K.Y. (1969) *Radiation in the Atmosphere*, Academic Press, NY.
- [58] Notton, G.; Cristofari, C.; Poggi, P.; Muselli, M. (2002) *Calculation of solar irradiance profiles from hourly data to simulate energy systems behaviour*. *Renewable Energy*, 27, pp. 123–142.
- [59] Suehrcke, H.; McCormick, P.G.; (1988) *The frequency distribution of instantaneous insolation values*, Solar Energy, 40, pp. 413–422.
- [60] Tovar, J.; Olmo, F.J.; Alados-Arboledas, L. (1998) *One-minute global irradiance probability density distributions conditioned to the optical air mass*, Solar Energy, 62, pp. 387–393.
- [61] Woyte A.; Belmans R.; Nijs J. (2007) *Fluctuations in instantaneous clearness index: Analysis and statistics*, Solar Energy, 81, 2, pp. 195-206.

- [62] Woyte, A.; Thong, V.V.; Belmans, R.; Nijs, J.C. (2006) *Voltage fluctuations on distribution level introduced by photovoltaic systems*, *IEEE Transactions on Energy Conversion*, 21, 1, Pages 202-209.
- [63] Gates D.M. (1962) *Energy Exchange in the Atmosphere*, Harper, Row Monographs, NY.
- [64] Seinfeld J.H.; Pandis S.N. (1998) *Atmospheric Chemistry; Physics: Air Pollution to Climate Change*, John Wiley & Sons, NY.
- [65] Bourne N.E. (1984) *Atmospheric dispersion, in Handbook of Air Pollution Technology*, John Wiley & Sons.
- [66] Turner, D. B. (1970) *Workbook of Atmospheric Dispersion Estimates*, EPA Publications AP-26, Washington, D.C.
- [67] Godish Thad (2004) *Air Quality 4th Edition*, Lewis Publishers.
- [68] Jauregui E. and Godinez L.Cruz F. (1992), *Aspects of the island in Guadalajara, México*, *Atmospheric Environment*, 26B(3), pp. 391–396”.
- [69] Tereshchenko I.E. y Sánchez-Gómez R. (1996), *Acerca de la estructura vertical de la atmósfera sobre la Zona Metropolitana de Guadalajara*, GEOS, Unión Geofísica Mexicana, boletín informativo, época II, volumen 16, 4, pp. 176.
- [70] Tereshchenko I.E. y Sánchez-Gómez R. (1997), *Resultados de los estudios de las inversiones térmicas sobre la ciudad de Guadalajara*, GEOS, Unión Geofísica Mexicana, boletín informativo, época II, volumen 17, 4, pp. 195.
- [71] R.T. Pinker, B. Zhang, E.G. Dutton (2005) *Do satellites detect trends in surface solar radiation?*, *Science* 308, pp. 850-854.
- [72] J. Ye, F. Li, G. Sun, A. Guo (2009) *Solar dimming and its impact on estimating solar radiation from diurnal temperature range in China, 1961-2007*, *Theor. Appl. Climatology*, Springer-Verlag, 101, pp. 137-142.
- [73] R. Sánchez, S. Sánchez, R.J. Sánchez, J. Alcalá, G.R. Aguirre (2015) *Análisis espacial estimando regiones de riesgo por contaminantes que exceden con más*

frecuencia la norma en Guadalajara, México, Revista Iberoamericana de Ciencias,
Vol. 2, No. 2, pp. 65-85.

Artículo Publicado

Investigando variaciones aleatorias de radiación solar en Guadalajara, México

Rubén Sánchez Gómez¹, Gilberto Rubén Aguirre Salas², Silvia Sánchez Díaz³ y Jaime Alcalá Gutiérrez⁴
Departamento de Física¹, Instituto de Astronomía y Meteorología^{2,4}, Departamento de Ciencias Naturales y Exactas³
Universidad de Guadalajara
Guadalajara, Jalisco; México
rubensg@cencar.udg.mx, raguirre@astro.iam.udg.mx, silviasandi@profesores.valles.udg.mx

Abstract— This paper aims to describe the stochastic variation of solar radiation time series observed in the metropolitan zone of Guadalajara. We use an 11-year (2003-2014) database of measures every ten minutes on two automatic stations, the first one installed in an urban area, within the city and the second one in forest area. We review separately the seasonal variation, time-dependent effect on consecutive observations, and temporal variations of short and long term. In long-term variations we found solar radiation dimming effect in the urban area and in short-term variation we found that the Weibull distribution model shows good fit to the observed data in the study area.

Keyword— *Time series analysis, seasonal variation, long-term trend, short-term variations, probability model.*

Resumen— Este trabajo tiene como objetivo describir la variación estocástica de series temporales de radiación solar observada en la zona metropolitana de Guadalajara. Se utiliza una base de datos de once años de lecturas cada diez minutos registradas en dos estaciones automáticas, la primera en un área urbana, al interior de la ciudad y la segunda en una área forestal. Se revisaron por separado efectos periódicos, dependencia temporal en lecturas consecutivas, y variaciones a corto y largo plazo. En las variaciones a largo plazo se encontró evidencia de oscurecimiento en el área urbana, mientras que en las de corto plazo, se encontró que el modelo de distribución Weibull muestra un buen ajuste a los datos observados en la zona de estudio.

Palabras claves— *Análisis de series de tiempo, variación estacional, tendencia a largo plazo, variaciones a corto plazo, modelo de probabilidad.*

I. INTRODUCCIÓN

La radiación solar es una de las fuentes energéticas de origen renovable más importante, que además de ser una de las mejores alternativas para reducir el impacto ambiental ante el uso indiscriminado de recursos convencionales (SENER, 2013; REN21, 2014), se ha convertido en una fuente atractiva y rentable con el desarrollo de colectores fotovoltaicos cada vez más potentes. No obstante, cualquier inversión en aplicaciones de energía alternativa requiere de una estimación confiable del recurso solar esperado en el punto de interés (Gueymard y Wilcox, 2011; Vignola, Grover, Lemon y McMahan, 2012), ya que la radiación solar está sujeta a variaciones impredecibles pues una parte importante puede ser absorbida o dispersada por moléculas del aire como aerosoles o nubes, mientras que otro porcentaje se recibe directamente en los colectores fotovoltaicos. Por tanto, conocer y comprender la variabilidad de la radiación solar a corto plazo es de vital importancia para contar con los elementos que permitan estimar el potencial de cualquier sistema que convierta a la radiación solar en energía.

Para entender la importancia de estos cambios impredecibles, se debe señalar que la eficiencia de los dispositivos fotovoltaicos no se puede estimar utilizando lecturas horarias. Trabajos como el de Soubdhan, Emilion y Calif (2008) muestran que las discrepancias en intervalos de corto plazo son decisivas en las diferencias observadas en estudios de simulación. Más aún, estas variaciones pueden alcanzar hasta $700\text{W}/\text{m}^2$ en intervalos de tiempo de algunos segundos a minutos, por las condiciones adversas que interfieren el flujo de radiación solar, tales como nubosidad, tamaño de la nube, velocidad de traslado o número de nubes, entre otros casos, y estas fluctuaciones repercuten negativamente en el

desempeño de los colectores que convierten la radiación solar en electricidad. Más aún, para climas tropicales se incrementa la complejidad de estos sistemas que interactúan en la atmósfera, generando escenarios impredecibles en las variaciones de radiación solar a corto plazo, por cambios bruscos en las condiciones meteorológicas locales.

Entender esta componente impredecible de la radiación solar ha sido de interés en diversos trabajos, destacando el uso de familias de probabilidad o mezclas de distribuciones con soporte no negativo, implementado directamente a los datos o a bien usando alguna transformación. Por ejemplo, Assunção, Escobedo y Oliveira (2003) modifican los datos de radiación solar en un índice de difusión y le ajustan una distribución Beta a registros observados en el área de Botucatu, Brasil; presentan además un análisis detallado del índice para diferentes rangos en $[0,1]$; y en otro caso, Soubdhan, Emilion y Calif (2008) combinan distribuciones Dirichlet investigando el comportamiento diario de un índice de claridad. No obstante, en ambos trabajos se argumenta un desempeño bimodal en la distribución de las observaciones de radiación, señalando que tal comportamiento se puede atribuir a lecturas por bloques de registros con cielo despejado contra escenarios con atmósfera difusa.

Por otro lado, Jurado, Caridad y Ruiz (1995) investigan los registros de radiación solar y detectan de forma semejante un comportamiento bimodal en la distribución de los datos para intervalos menores a 60 minutos y por su parte, Chang (2010) analiza de forma similar lecturas observadas en seis estaciones meteorológicas instaladas en Taiwan, con el fin de encontrar un modelo de probabilidad que tenga el mejor ajuste a los datos. Hace un análisis comparación del desempeño de las distribuciones Weibull, logística, normal y lognormal, sin detectar un comportamiento binomial en la distribución de los datos observados. Esta diferencia en el desempeño de los datos, entre binomial y no binomial, deja duda respecto a la metodología utilizada, dando apertura a buscar una herramienta diferente que permita investigar las variaciones impredecibles de radiación solar con técnicas de aproximación que permitan el análisis de las características típicas de cualquier magnitud observada en la naturaleza.

Bajo este contexto y desde un punto de vista práctico, es natural pensar que para cualquier conjunto de datos observados en la naturaleza, si el grado de incertidumbre es muy pequeño, los modelos numéricos determinísticos son suficientes para analizar su desempeño, si el grado de desorden es suficientemente grande, lo usual es implementar modelos probabilísticos clásicos, no obstante, si los datos muestran evidencia de patrones complejos como variaciones a corto plazo, interdependencia en lecturas consecutivas, efectos estacionales o periódicos y variaciones a largo plazo, las técnicas determinísticas convencionales son ineficientes y la estadística convencional es insuficiente. Es entonces cuando el análisis de series temporales es el campo de acción.

De esta forma, el propósito de este trabajo es aplicar las herramientas estadísticas del análisis de series de tiempo para caracterizar los registros de radiación solar observados en la zona de estudio, analizando por separado las características típicas de cualquier serie meteorológica, como lo son: efectos estacionales, variaciones a largo plazo, dependencia temporal y variaciones a corto plazo.

A. Características de la zona de estudio

La Zona Metropolitana de Guadalajara (ZMG) está integrada por los municipios interiores de Guadalajara, Tlaquepaque, Tonalá, Zapopan, Tlajomulco de Zúñiga y El Salto, adicionando a Ixtlahuacán de los Membrillos y Juanacatlán como municipios exteriores (Delimitaciones de las zonas metropolitanas de México, 2005), concentra a más del 50% de la población del estado y ha experimentado un crecimiento acelerado en las últimas décadas.

Geográficamente, se sitúa en la cuenca del Valle del Río Grande de Santiago, en los Valles de Atemajac y la Planicie de Tonalá, entre las zonas montañosas de la Sierra Madre Occidental y el Cinturón Volcánico Transmexicano. Se localiza en la región centro – occidente del país, en el centro del estado de Jalisco, a una latitud y longitud promedio de $20^{\circ} 39' 54''N$ y $103^{\circ} 18' 42''E$, respectivamente,

y una altitud promedio de 1,540 metros sobre el nivel medio del mar, lo que la expone a la influencia de sistemas anticiclónicos generados tanto en el Golfo de México como en el Océano Pacífico (figura 1).



Fig. 1. Localización de la Zona Metropolitana de Guadalajara.

Está dentro del llamado cinturón tropical, por lo que la mayor parte del año se encuentra afectada por la afluencia de aire marítimo (Programa para mejoramiento la calidad del aire en la ZMG, 1997-2001) y recibe una abundante radiación solar debido a su latitud $20^{\circ}N$, ubicándose en el llamado “cinturón solar” de energía y sometida a una radiación solar que va de 5.6 a $6.1KWh/m^2$ (IBP, 2015). De esta forma, el clima en la ZMG se puede clasificar como semicálido (climas de transición entre cálidos y templados), subhúmedo, con temperatura anual promedio por encima de los $18^{\circ}C$ y con periodos marcados de lluvias y pocas lluvias (Davydova et al., 1999). Entonces, la ZMG representa un punto de interés para investigar variaciones a corto plazo en los registros de radiación solar, esperando encontrar perturbaciones atribuibles a condiciones típicas de patrones complejos de variación.

II. MATERIALES Y MÉTODOS

Al interior de la ZMG se cuenta con dos estaciones meteorológicas automáticas tipo Davis, modelo Vantage Pro2: la primer estación esta instalada en el interior del Observatorio “Colomos”, ubicado al interior del Parque Colomos, en una zona boscosa, semiurbanizada, a una longitud de $-103^{\circ} 23' 33''N$, latitud $20^{\circ} 42' 24''E$ y a una altitud de $1559m$ sobre el nivel medio del mar. Está a cargo de la Comisión Nacional del Agua (CNA), que provee información al Servicio Meteorológico Nacional (SMN). Por otro lado, la segunda estación se encuentra instalada en un ambiente urbanizado, al interior del Instituto de Astronomía y Meteorología (IAM) de la Universidad de Guadalajara, a una longitud de $-103^{\circ} 23' 03''N$, latitud $20^{\circ} 40' 29''E$ y a una altitud de $1592m$ sobre el nivel medio del mar (figura 2).

Ambas estaciones generan registros cada diez minutos de temperatura (promedio, mínima y máxima), humedad relativa, punto de rocío, precipitación, dirección – velocidad del viento y radiación, entre otras. Utilizan un radio intercepción de espectro amplio de frecuencia y son capaces de recibir hasta $300m$ de distancia, los sensores de temperatura y radiación solar se ubican dentro de un protector solar, por lo que se tiene fiabilidad y precisión en sus lecturas. Se alimenta de energía eléctrica, pero

cuentan con un sistema de alternativo de tres baterías que funcionan como reserva durante cortos breves de energía; no obstante, ambas estaciones tienen problemas de lecturas perdidas desde su instalación.

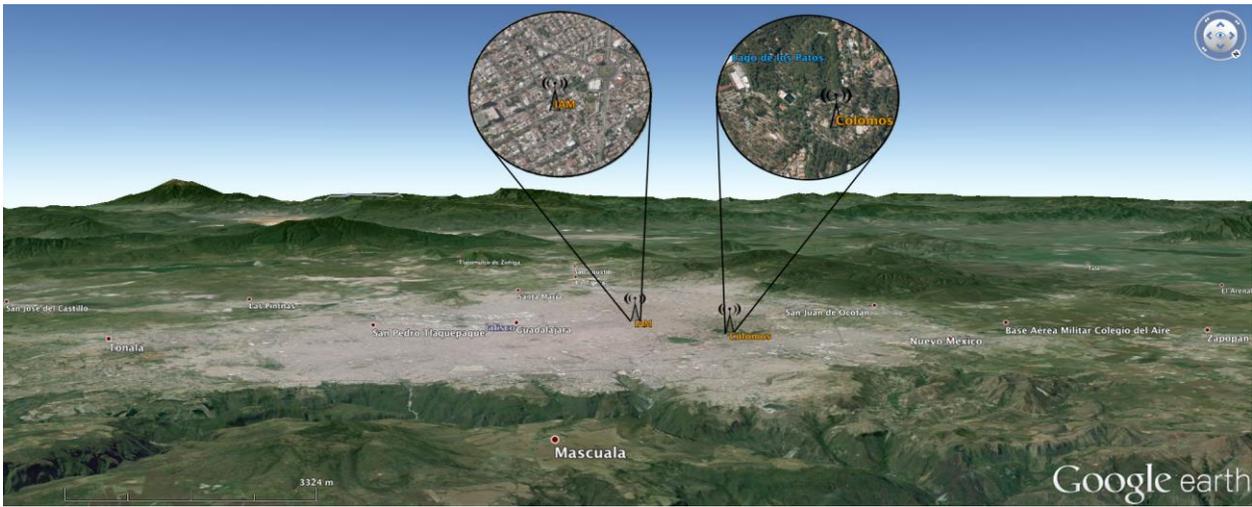


Fig. 2. Estaciones meteorológicas en la Zona Metropolitana de Guadalajara: Observatorio Colomos y el IAM.

Las series de datos disponibles van desde el 28 de febrero del 2000 para la estación del observatorio Colomos, mientras que los registros en la estación del IAM inicia el 23 de junio del 2003, ambas series muestran periodos largos de datos perdidos (figura 3).

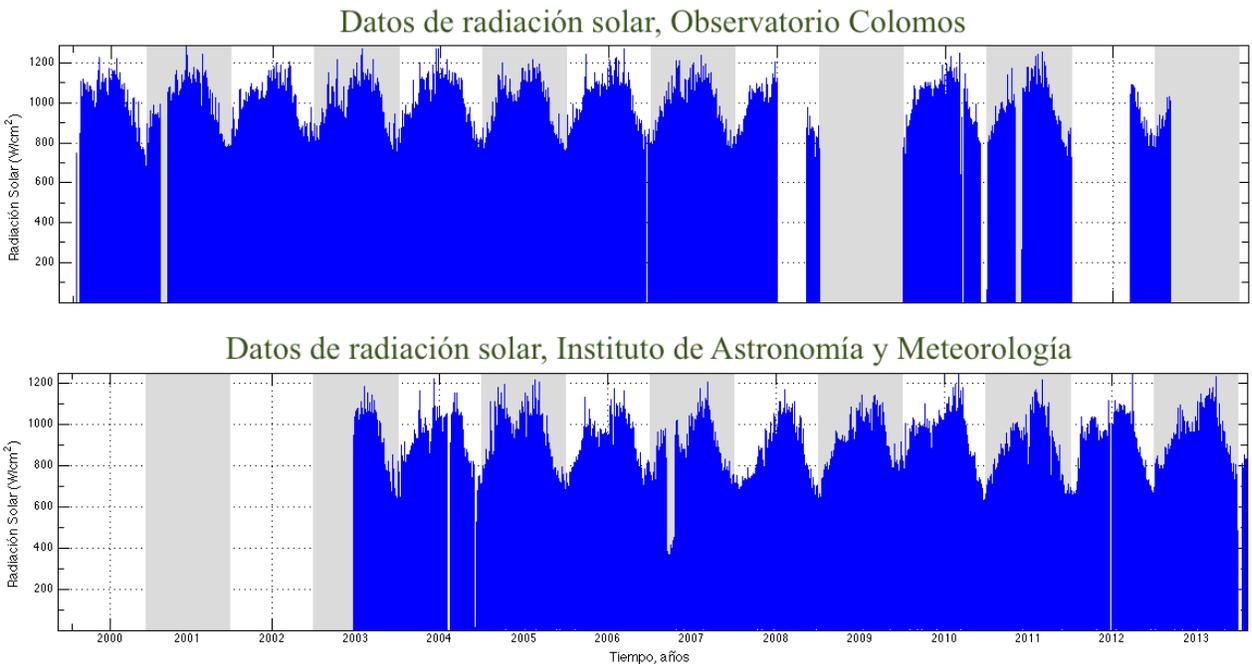


Fig. 3. Datos de Radiación Solar.

De este modo, en este trabajo se muestran los resultados obtenidos con las bases de datos de ambas estaciones, desde sus fechas respectivas de inicio hasta el siete de marzo de 2013, fecha en que no se

cuenta con datos de la estación del observatorio Colomos, teniendo una muestra de once años de información en promedio.

A. Análisis de series de tiempo

La metodología utilizada se fundamenta en el término de *series de tiempo*, que básicamente se refiere a una rama de la estadística en donde se estudian observaciones recolectadas de forma secuencial en puntos temporales igualmente espaciados, pero pueden ser lecturas en periodos de tiempo diferentes (Smith, 1999; Brockwell y Davis, 2002), de modo que se analizan las lecturas $\{x_1, x_2, \dots, x_n\}$, observadas en puntos discretos de tiempo $\{t_1, t_2, \dots, t_n\}$ con $t_1 \leq t_2 \leq \dots \leq t_n$.

Desde un punto de vista más generalizado, cualquier serie de tiempo es una *realización* de un *proceso estocástico*, definido por una magnitud $X(t,w)$ que depende de una variable determinística t (no muestra cambios impredecibles) y una variable aleatoria w impredecible. Por tanto, para un punto arbitrario en el tiempo, el valor x_t observado en el momento t es una magnitud aleatoria que sigue una función de probabilidad $p_t(x_t)$, con media $\mu(t)$ y autocovarianza $\gamma(h)$ dados por:

$$\begin{aligned} \mu(t) &= E[X_t] & (1) \\ \gamma(h) &= \text{Cov}\{(X_t - m_t)(X_{t+h} - m_{t+h})\}. \end{aligned}$$

en donde h representa un incremento de tiempo ($h \geq 0$) y tanto la función de probabilidad $p_t(x_t)$ como la media y la autocovarianza pueden ser diferentes para cada punto de tiempo t . Además, el caso particular en que $h = 0$ (no hay incremento en el tiempo) corresponde a la varianza de la variable aleatoria w observada en un tiempo t específico y se escribe como $\gamma_0 = \gamma(0)$.

1) Investigando Tendencia en series de tiempo

Si el valor de la media $\mu(t)$ crecer o decrece a largo plazo, se dice que la serie muestra efectos de tendencia temporal y sin pérdida de generalidad, se puede representar mediante el modelo aditivo

$$X_t = T_t + I_t$$

en donde T_t corresponde a tendencia e I_t representa la componente impredecible; Además, en el caso más simple, se puede suponer un modelo de tendencia determinística lineal, dado por

$$T_t = m + a t \quad (2)$$

de modo que el problema se reduce a resolver el sistema (2) para los valores discretizados de tiempo $\{t_1, t_2, \dots, t_n\}$ y cuya solución se puede obtiene mediante mínimos cuadrados (Rohatgi, 2005; Sánchez *et al.*, 2010). De modo que, si los datos de radiación solar observados en la zona de estudio muestran efectos de tendencia, se tendrá que el valor estimado de a con los datos deberá ser diferente de cero con cierto nivel de confianza.

2) Investigando efectos estacionales en series de tiempo

Para analizar efectos estacionales, la herramienta básica es el *análisis espectral*, *análisis armónico* o *análisis de Fourier*, de la serie de tiempo. Es un instrumento estadístico que permite descomponer la información disponible en una suma de funciones trigonométricas

$$X_t = \mu + A_i \cos 2\pi f_i t + E_i \sin 2\pi f_i t + I_t \quad (3)$$

para $i = 1, 2, \dots, n / 2$ y en donde los A_i y E_i son amplitudes, las f_i son frecuencias y μ es el valor promedio de los datos. En donde, para una serie de tamaño n , el análisis de efectos estacionales consiste

en encontrar las periodicidades “ocultas”, suponiendo frecuencias enteras que corresponden a periodos n/j con $j = 1, 2, \dots, n$. Estos armónicos respecto a la expansión de datos $f_j = j/n$ se conocen como frecuencias de Fourier (Bloomfield, 2000) de modo que, las únicas frecuencias que deben considerarse son aquellas que satisfacen $0 < f < 1/2$, ya que se repiten en ambos lados del eje de frecuencias (sección 2.5, Bloomfield, 2000).

Así, bajo este argumento el estimador más simple para densidades espectrales es el periodograma, dado por

$$I(f_j) = n R^2(f_j) \tag{4}$$

con $R^2(f_j) = (\hat{A}_j^2 + \hat{E}_j^2)$ y bajo la hipótesis nula $A_j = E_j = 0$ se tiene que $I(f_j) \sim 4\sigma^2$ se distribuye exponencial con media uno (Smith, 1999; Bloomfield, 2000; Sánchez *et al.*, 2010).

Implementando entonces esta herramienta en los datos de radiación solar, se puede esperar obtener frecuencias significativas para efectos estacionales diurnos, semianuales (efectos de temporal), anuales (efectos globales a gran escala).

3) Investigando dependencia en lecturas consecutivas

En consistencia con los modelos lineales, es común establecer alguna medida de asociación entre lecturas consecutivas mediante el coeficiente de auto-correlación γ_h para $h = 1, 2, \dots, k$ retrasos y mostrar evidencia de asociación entre la lectura x_t y los retrasos previos $x_{t-1}, x_{t-2}, \dots, x_{t-k}$, bajo el supuesto de que la lectura actual depende en cierta medida de los registros previos, de modo que un modelo que incorpora esta asociación temporal está dado por

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_k X_{t-k} + I_t \tag{5}$$

en donde se busca verificar la hipótesis $a_1 = 0, \dots, a_k = 0$ contra la alternativa $a_1 \neq 0, \dots, a_k \neq 0$, para los datos de radiación solar observados en la zona de estudio.

Cabe señalar que la ventaja de suponer modelos lineales en cada uno de los efectos típicos de series temporales (tendencia, estacionalidad, asociación con lecturas previas) permite estimar los parámetros mediante mínimos cuadrados y eliminar los efectos significativos mediante restas simples de modo que, en (5) se tendría el efecto impredecible restando los retrasos significativos, es decir

$$I_t = X_t - a_1 X_{t-1} - a_2 X_{t-2} - \dots - a_k X_{t-k}$$

y así, se puede investigar el comportamiento de la componente impredecible I_t eliminando los efectos de tendencia, estacionales y de dependencia entre lecturas consecutivas.

B. Modelos probabilísticos de soporte no negativo

Revisando la literatura, Karaki *et al.* (1999) muestra un estudio probabilístico del desempeño de sistemas de conversión de energía solar suponiendo un modelo *beta* de soporte $[0,1]$ y considera que la función de densidad Weibull representa adecuadamente los datos de velocidad del viento; Assunção Escobedo y Oliveira (2003) señalan que la densidad *beta* es un modelo adecuado para analizar la componente irregular de datos de radiación solar en Botucatu, Brasil; Chang (2010) en cambio, un poco más precavido, muestra un análisis en el que incorpora la densidad Weibull, la normal, la logística y la log-normal en su estudio, concluyendo que la densidad log-normal muestra mejor desempeño en términos de error cuadrado medio y de la prueba Kolmogorov-Smirnov; mientras que Villanueva *et al.* (2013) evalúa el desempeño de la distribución Weibull multivariada para representar de manera conjunta la dirección y la velocidad del viento. Lamentablemente, ninguno de estos trabajos dan evidencia de eliminar o incorporar efectos de tendencia o periodicidad en las series temporales que analizaron; más

aún, los histogramas que muestran en sus trabajos dan evidencia de un comportamiento bimodal en los datos o de bien, con densidades de colas muy pesadas, lo que deja evidencia del efecto de componentes cíclicas en los datos. Así, en un paso previo, en este trabajo se reducen los efectos significativos de tendencia temporal, estacionalidad y dependencia temporal, y una vez que se tenga únicamente la variación impredecible o aleatoria, se implementan cinco modelos de soporte no negativo en las bases de datos de radiación solar en la zona de estudio. En específico los modelos que se analizan son:

- Densidad *beta generalizado 1er tipo* con densidad de probabilidad (Aguirre, 2016)

$$\beta G_X(x; a, b) = x^{a-1} [c-x]^{b-1} / [c^{a+b-1} B(a, b)],$$

para $a, b > 0$, $0 < x < c$ y el término $B(a, b)$ es la *función beta* definida por $B(a, b) = \{\Gamma(a)\Gamma(b)\} / \Gamma(a+b)$, en donde $\Gamma(\bullet)$ es la *función gamma* (Lawless, 2002).

- El *exponencial*, con densidad de probabilidad

$$f_X(x; a) = \exp(-x/a) / a,$$

para $a > 0$, $x > 0$ (Johnson, Kotz y Balakrishnan, 1994; Lawless, 2002; Rohatgi, 2003; Aguirre, 2016).

- La densidad *gamma*, cuya densidad de probabilidad se escribe como

$$f_X(x; a, b) = x^{a-1} \exp(-x/b) / [\Gamma(a) b^a],$$

para $a, b > 0$, $x > 0$ (Johnson, Kotz y Balakrishnan, 1994; Lawless, 2002; Rohatgi, 2003; Aguirre, 2016).

- El *lognormal*, dado por

$$f_X(x; a, b) = [x (2\pi)^{1/2} b]^{-1} \exp(-[\log(x) - a]^2 / [2\sigma^2]),$$

para $\sigma > 0$, $x > 0$ (Johnson, Kotz y Balakrishnan, 1994; Rohatgi, 2003; Aguirre, 2016).

- El modelo *Weibull*, con densidad de probabilidad

$$f_X(x; a, b) = ab(ax)^{b-1} \exp(-[ax]^b),$$

para $a, b > 0$, $x > 0$ (Johnson, Kotz y Balakrishnan, 1994; Rohatgi, 2003; Aguirre, 2016).

1) Criterios de bondad de ajuste

Para identificar el modelo de probabilidad con mejor ajuste a un conjunto de datos, en la literatura estadística se han propuesto varias medidas, entre las que destacan el criterio de información de Akaike (AIC_C), que proporciona una estimación de la información relativa que se pierde con un modelo específico; de modo tal que, el mejor será aquel que presente el menor valor del AIC_C para los datos. Está dado por

$$AIC_C = -2\ln(L(\hat{a})) + 2Kn/(n - K - 1),$$

en donde \hat{a} corresponde al estimador paramétrico de a , n es el tamaño de la muestra, K es el número de parámetros que tiene el modelo que se analiza y L es la log-verosimilitud evaluada en los parámetros estimados.

Otro criterio de uso cotidiano es la Raíz cuadrada del Error Cuadrado Medio ($RECM$), dada por

$$RECM^2 = \frac{[x_1 - x_{1.est}]^2 + [x_2 - x_{2.est}]^2 + \dots + [x_n - x_{n.est}]^2}{n},$$

en donde x_t son los registros observados en el tiempo t , $x_{t.est}$ es el valor puntual estimado para el modelo a prueba y n es el tamaño de la muestra. En este caso, dado que se trata de distancia, el mejor modelo para los datos será aquel que tenga el valor más pequeño.

III. RESULTADOS

Analizando las series temporales de radiación solar en la ZMG, una observación general en lo que respecta al efecto de tendencia, es que los datos no dan evidencia para afirmar algún efecto significativo. Sin embargo, destaca que aplicando el modelo (2), se tiene que para la zona boscosa o semiurbana $\hat{a}_{Colomos} = 0.000010653$, es decir, en se observa un efecto de brillantés (se incrementa la cantidad de radiación recibida en superficie por unidad de tiempo), mientras que para la zona urbanizada se observa un efecto de obscurecimiento con $\hat{a}_{IAM} = -0.000014071$ (se reduce la cantidad de radiación recibida en superficie por unidad de tiempo). Este último resultado coincide con reportes de autores como Pinker, Zhang, Dutton (2005), Ye, Li, Sun y Guo (2009), quienes señalan un efecto llamado como “*obscurecimiento global*”, asociado fuertemente al cambio climático. No obstante, detectar obscurecimiento en una zona urbana y brillantés en la zona boscosa, a menos de cinco kilómetros de distancia, puede reflejar la influencia de niveles elevados en la concentración de contaminantes como el ozono o partículas menores a $10\mu m$ en la ZMG y sobre el IAM (Sánchez *et al.* 2015).

Revisando los efectos estacionales en ambas series de tiempo, se tiene evidencia de frecuencias significativas a 12 ($f_1=0.01389$) y 24 horas ($f_2=0.006948$), 15 ($f_3=0.000463$) y 30 días ($f_4=0.000231$) y de aproximadamente 6 meses ($f_5=0.000038$) con un intervalo de confianza del 95%. En éste caso, en las gráficas de la figura 4 se pueden observar dos representaciones de periodogramas estimados para las dos series disponibles, en ellas se incluyen los intervalos de confianza respectivos, destacando la evidencia de frecuencias significativas.

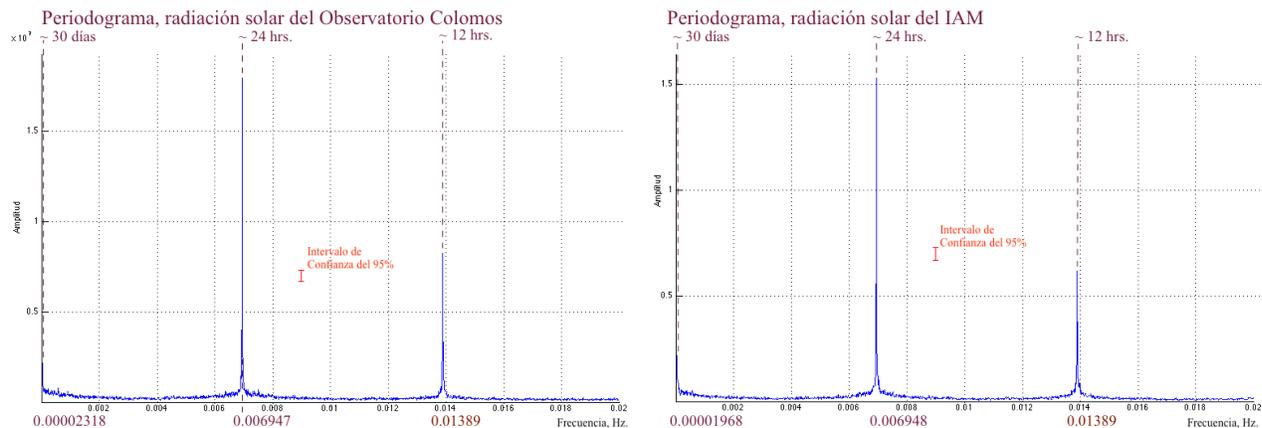


Fig. 4. Periodogramas estimados con los datos de radiación solar observados en el Observatorio Colomos y en el IAM.

Una vez identificadas los periodos significativos, el siguiente paso es aplicar un filtro para eliminar los efectos estacionales y en este caso, se pueden suprimir aquellas menores a $0.014Hz$ y con ello retirar el efecto en las ambas series de datos. Así, se rescata la variación impredecible reduciendo los efectos periódicos y de tendencia temporal.

En un paso siguiente, para verificar evidencia de efectos de asociación entre lecturas consecutivas, en la figura 5 se pueden observar los coeficientes de auto-correlación estimados para ambas series temporales de radiación solar. Destaca principalmente que para retrasos de diez minutos, la correlación estimada en ambas estaciones es menor a 0.60 y además, en observaciones horarias la auto-correlación estimada es menor a 0.45, lo que hace suponer que la asociación entre lecturas consecutivas no es un factor relevante para las magnitudes impredecibles de las series.

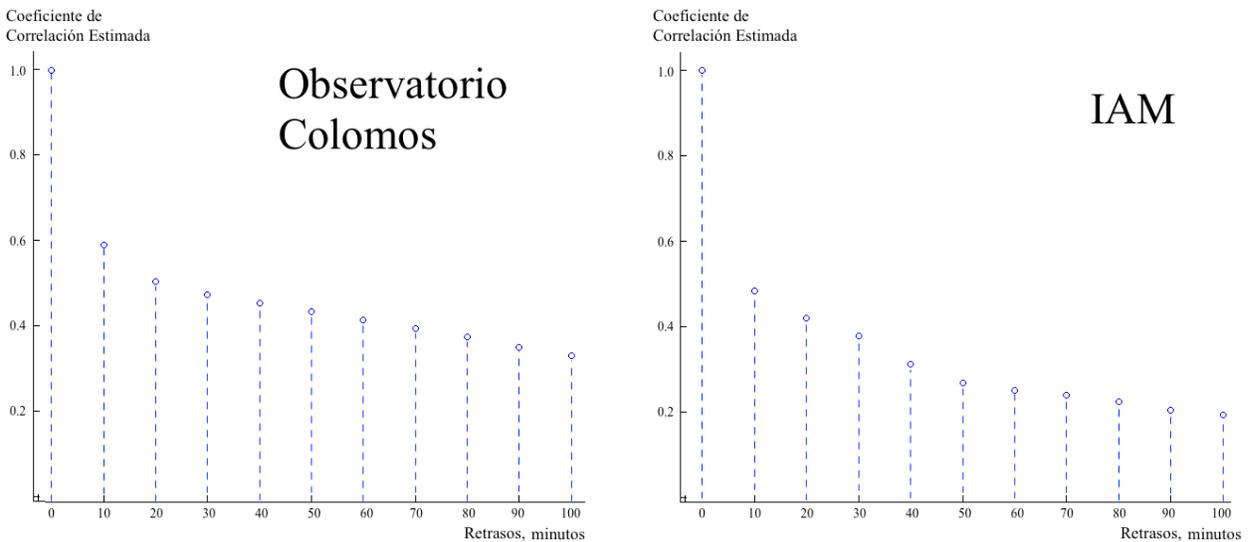


Fig. 5. Periodogramas estimados con los datos de radiación solar observados en el Observatorio Colomos y en el IAM.

Aplicando los modelos en las variaciones aleatorias de las series una vez que se suprimieron los efectos estacionarios y de tendencia temporal, en la tabla 1 se presentan los AIC_C y las $RECM$ estimados con los datos de radiación solar para los cinco modelos propuestos.

En este caso se puede observar que bajo el criterio de la $RECM$, los datos dan evidencia para creer que el modelo *lognormal* es el mejor para los datos en las estaciones disponibles (valores sobresaltados); en tanto que, el modelo *beta generalizado* de 1er tipo es el que muestra el peor desempeño de los modelos implementados. No obstante, se debe señalar además que el modelo *lognormal* representa el peor escenario en términos del AIC_C , ya que es el que pierde más información relativa a los datos entre los cinco modelos revisados en este trabajo.

Por otro lado, en la misma tabla I se puede ver que con base en el AIC_C , el modelo *Weibull* es el que pierde menos información relativa en contraste con los cuatro modelos de prueba (valores sobresaltados) y sobresale además que de las estimaciones de $RECM$ el modelo *Weibull* es la segunda mejor alternativa. Por lo que, este resultado deja una pauta importante para decidirse por el modelo más apropiado para los datos disponibles en la ZMG.

Tabla I. Criterio de información de Akaike y raíz del error cuadrado medio estimados para los datos de radiación solar.

Modelo	Criterio de información de Akaike y Raíz del Error Cuadrado Medio.			
	Datos registrados en el Observatorio Colomos		Datos registrados en el Instituto de Astronomía y Meteorología	
	AIC_C	$RECM$	AIC_C	$RECM$
Beta Generalizado de 1er tipo	2687025.4877	0.007492749788821	2655112.6546	0.003826120713519
Exponencial	2700486.3486	0.000000000005597	2658977.7893	0.00000000024735
Gamma	2677856.0570	0.00000000000608	2644892.2622	0.00000000005476
Lognormal	2719489.3963	0.00000000000060	2669380.8834	0.00000000000059
Weibull	2673318.3623	0.00000000000193	2639191.6490	0.00000000000675

Buscando reforzar las observaciones de la tabla 1, en las figuras 6 y 7 se pueden observar las gráficas del histograma de la magnitud impredecible en los datos de radiación solar registrados en las estaciones del observatorio Colomos, a manera de ejemplo, incorporando los ajustes de los modelos propuestos. En donde sobresale el modelo *lognormal* por su proximidad al histograma de los datos, pero el modelo Weibull es la segunda mejor alternativa en contraste con los modelos bajo estudio.

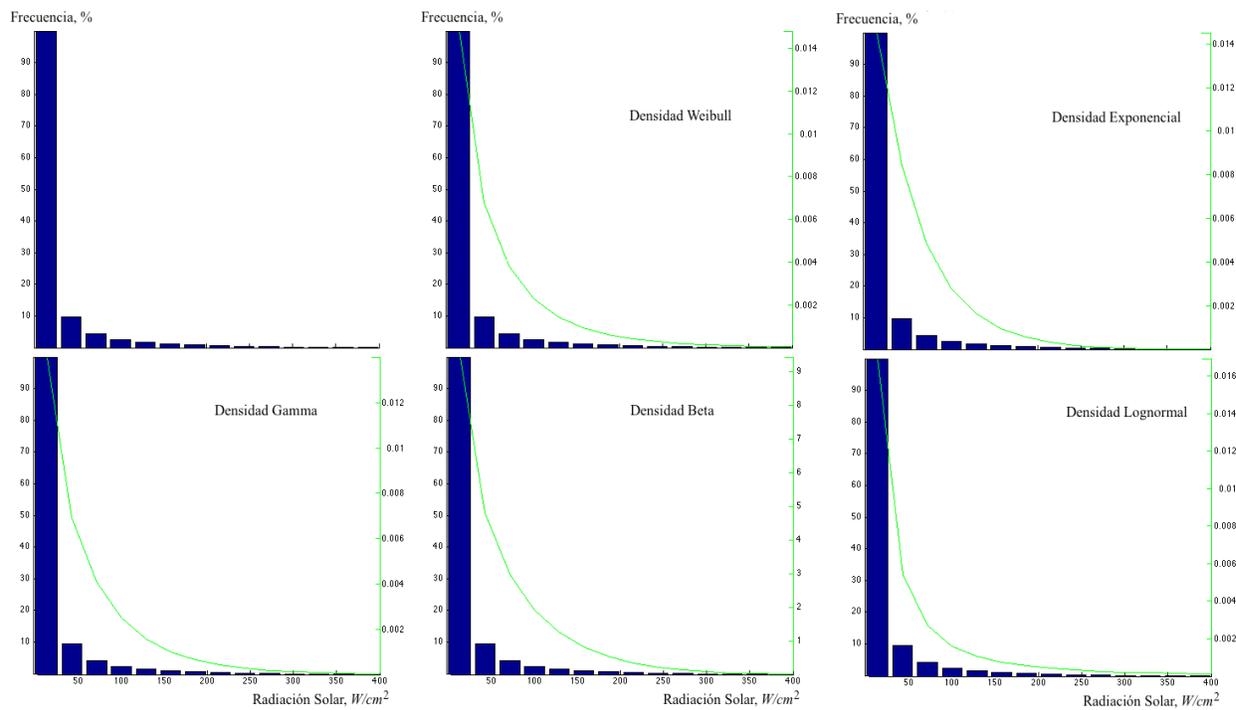


Fig. 6. Modelos propuesto e histograma de las variaciones aleatorias registradas en el observatorio Colomos.

Cabe señalar que en ambas figuras se pueden identificar ejes verticales en color verde, que corresponden a las magnitudes de las curvas estimadas con los datos para las densidades de probabilidad bajo estudio.

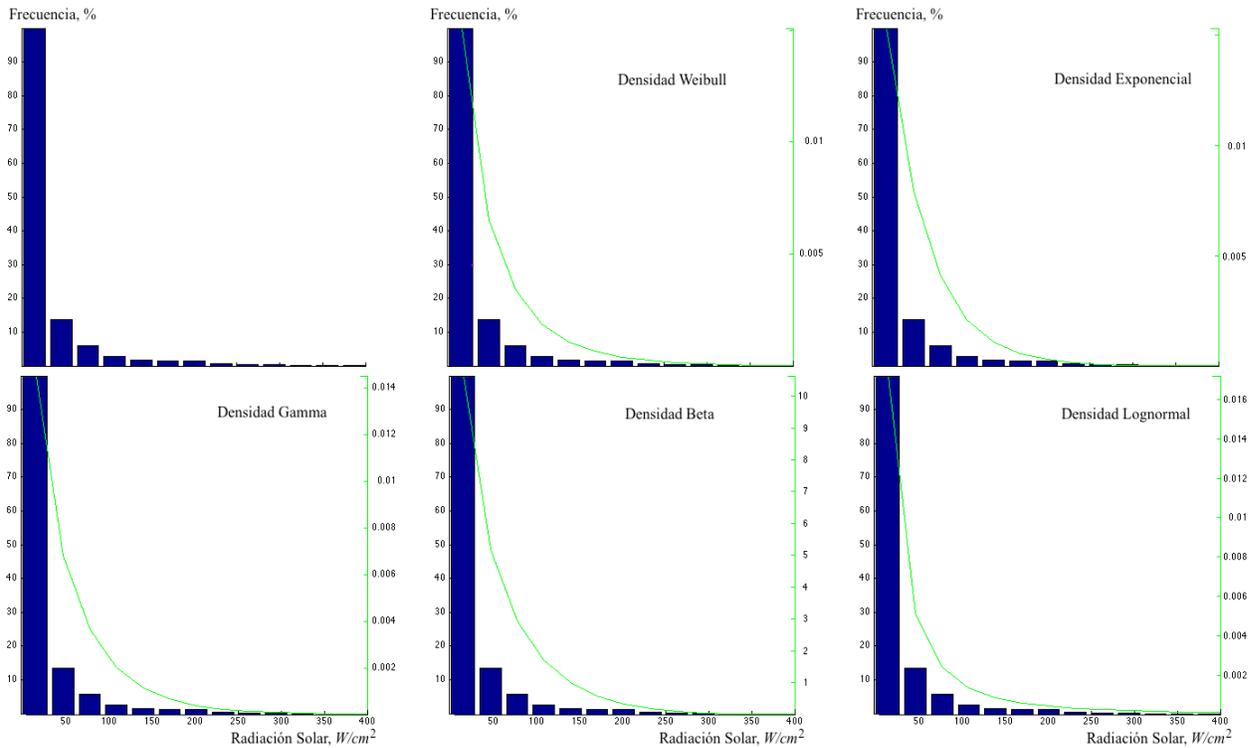


Fig. 7. Modelos propuesto e histograma de las variaciones aleatorias registradas en el IAM.

Sobresale además en los histogramas de las figuras 6 y 7 que no hay evidencia de un comportamiento binomial en las magnitudes aleatorias de las series de radiación solar observadas en la zona de estudio.

Por otro lado, a manera de ejemplo en la figura 7 se presenta el histograma de los datos sin suprimir los efectos periódicos y de tendencia en la serie observada, de modo que, en la gráfica se puede observar un comportamiento tipo bimodal en los datos, por lo que se concluye que este comportamiento se atribuye a la influencia de los efectos periódicos y de tendencia temporal en las series de datos.

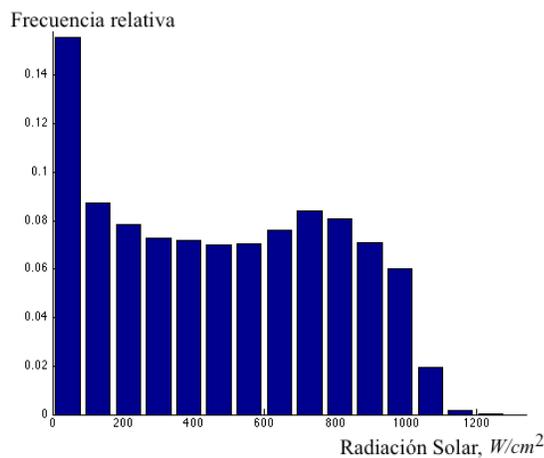


Fig. 8. Histograma de frecuencia de la serie de datos registrados en el observatorio Colomos.

IV. DISCUSIÓN DE RESULTADOS Y CONCLUSIONES

En este estudio se presenta una metodología para suprimir el efecto periódico, de tendencia temporal y de asociación entre lecturas consecutivas, con el fin de investigar de forma aislada las variaciones impredecibles o aleatorias en series temporales. Ésto permite entender el comportamiento de las series temporales de radiación solar en la zona, para disponer de los elementos de decisión necesarios para invertir en fuentes que aprovechen la energía solar.

Se probaron cinco modelos de densidad de probabilidad sugeridos en la literatura, en específico, los modelos *beta* (*beta generalizado de 1er tipo*), el *exponencial*, el modelo *gamma*, el *lognormal* y el *Weibull* y se hicieron pruebas de desempeño utilizando conjuntos de datos de radiación solar registradas en dos estaciones automáticas instaladas en la ZMG.

Los resultados dan evidencia para creer que el modelo *Weibull* es el mejor para entender el efecto impredecible en las series temporales de datos de radiación solar; aportando la menor pérdida de información en términos del AIC_C y ofreciendo la segunda mejora alternativa con base en la *RECM*.

Por otro lado, destacando de manera contundente que la variación impredecible no sigue un comportamiento bimodal en los datos registrados en la zona de estudio y este punto es relevante ya que se tienen elementos para afirmar que el comportamiento bimodal se puede atribuir a la influencia de los efectos periódicos y de tendencia temporal de las series.

Finalmente, el análisis de tendencia muestra un efecto decreciente en la intensidad de la radiación solar recibida en zona urbana, es decir, se tiene evidencia para creer que se presenta el fenómeno de “obscurecimiento” en el interior de la ciudad, posiblemente atribuible a los elevados niveles de contaminación atmosférica. En tanto que, para la zona boscosa (observatorio Colomos) se tiene evidencia de “brillantéz”, esto es, la tendencia temporal muestra un crecimiento moderado en la incidencia de radiación solar; esto coincide con los últimos reportes publicados en la literatura científica con datos de radiación solar registrados en otras regiones del mundo.

RECONOCIMIENTOS

Los autores quieren expresar su reconocimiento al apoyo brindado por el personal del Observatorio Colomos y del Instituto de Astronomía y Meteorología de la UdeG cuyas bases de datos son el sustento del presente trabajo.

REFERENCIAS

- [1] SENER (2013) “Prospectiva de energías renovables 2013-2027”, Sec. de Energía, México. Recuperado en mayo 20 de 2014. www.sener.gob.mx/res/PE_y_DT/pub/2013/Prospectiva_Energias_Renovables_2013-2027.pdf.
- [2] REN21 (2014) “Renewables 2014: Global Status Report”, Secretariat c/o UNEP, Paris. Recuperado en julio 15 de 2014, http://www.ren21.net/Portals/0/documents/Resources/GSR/2014/GSR2014_full.
- [3] C.A. Gueymard, S.M. Wilcox (2011) “Assessment of spatial and temporal variability in the US solar resource from radiometric measurements and predictions from models using ground-based or satellite data”, *Sol. Energy* 85(5), pp. 1068–1084.
- [4] F. Vignola, C. Grover, N. Lemon, A. McMahan (2012) “Building a bankable solar radiation dataset”, *Sol. Energy* 86(8), pp. 2218–2229.

- [5] T. Soubdhan, R. Emilion, R. Calif (2008) "Classification of daily solar radiation distributions using a mixture of Dirichlet distributions". Article soumis à Solar Energy. Recuperado en mayo 15 de 2014. <https://hal.archives-ouvertes.fr/hal-00289567/document>.
- [6] H.F. Assunção, J.F. Escobedo, and A.P. Oliveira (2003) "Modelling frequency distributions of 5 minute-averaged solar radiation indexes using Beta probability functions", *Theor. Appl. Climatol.*, 75, pp. 213–224.
- [7] M. Jurado, J.M. Caridad, V. Ruiz (1995) "Statistical distribution of the clearness index with radiation data integrated over five minute intervals", *Solar Energy*, 55, pp. 469–473
- [8] T.P. Chang (2010), "Investigation on Frequency Distribution of Global Radiation Using Different Probability Density Functions", *Int. J. Appl. Sci. Eng.*, Vol. 8, No. 2, pp. 99–107.
- [9] GEJ/SEMARNAP/SS (1997) Gobierno del Estado de Jalisco, Secretaría del Medio Ambiente, Recursos Naturales y Pesca, Secretaría Salud, "Programa para mejoramiento la calidad del aire en la Zona Metropolitana de Guadalajara 1997-2001".
- [10] IBP (2015) "Mexico Energy Policy, Laws and Regulations Handbook Volume 1 Strategic Information and Basic Laws", International Business Promotion, Inc., ISBN:1312958766, 9781312958760, pp. 284.
- [11] V. Davydova, Y. Skiba, S. Bulgakov y A. Martínez (1999) "Modelación Matemática de los Niveles de Contaminación en la Ciudad de Guadalajara, Jalisco, México. Parte I". *Microclima y Monitoreo de la Contaminación. Rev. Internacional de Contaminación Ambiental.* pp. 103–111.
- [12] R. L. Smith (1999) "Time Series", Department of Statistics, University of North Carolina, Chapel Hill, NC 27599-3260, Version 1.0.
- [13] P. J. Brockwell, R. A. Davis (2002) "Introduction to time series and forecasting 2nd ed.", Springer-Verlag New York, Inc.
- [14] V. K. Rohatgi (2005) "An Introduction To Probability And Statistical, 2E", John Wiley & Sons Inc.
- [15] R. Sánchez, H. Calderón, J. Fregoso, E. Brito, M. Rujano, S. Sánchez (2010) "Análisis y estimación de datos perdidos en investigación aplicada", Editorial Universitaria, U. de G.
- [16] P. Bloomfiel (2000) "Fourier Analysis of Time Series", John Wiley & Sons, Inc., New York.
- [17] H. Karaki, R.B. Chedid, R. Ramadan (1999) "Probabilistic Performance Assessment of Autonomous Solar-Wind Energy Conversion Systems", *IEEE Transactions on Energy Conversion*, 14, 3, pp. 766-772.
- [18] D. Villanueva, A. Feijóo, J.L. Pazos (2013) "Multivariate Weibull Distribution for Wind Speed and Wind Power Behavior Assessment", *Resources*, 2, pp. 370-384.
- [19] R.G. Aguirre (2016) "Evaluando el uso de energías eólica – solar en la Zona Metropolitana de Guadalajara, México", Tesis Doctoral, CUValles, UdeG.
- [20] J.F. Lawless (2002) "Statistical Models and Methods for Lifetime Data", Editorial Wiley-VCH, USA.
- [21] N.L. Johnson, S. Kotz, N. Balakrishnan (1994) "Continuous Univariate Distributions, Volume 1, 2nd Edition", John Wiley & Sons, Inc.
- [22] P. R. Cristian (1994) "The Bayesian Choice, A Decision Theoretic Motivation", Springer-Verlag.
- [23] A. O'Hagan, J. Forster (2004) "Kendall's Advanced Theory of Statistics, Vol 2B, Bayesian Inference", Arnold.
- [24] R.T. Pinker, B. Zhang, E.G. Dutton (2005) "Do satellites detect trends in surface solar radiation?", *Science* 308, pp. 850-854.
- [25] J. Ye, F. Li, G. Sun, A. Guo (2009) "Solar dimming and its impact on estimating solar radiation from diurnal temperature range in China, 1961-2007", *Theor. Appl. Climatology*, Springer-Verlag, 101, pp. 137-142.
- [26] R. Sánchez, S. Sánchez, R.J. Sánchez, J. Alcalá, G.R. Aguirre (2015) "Análisis espacial estimando regiones de riesgo por contaminantes que exceden con más frecuencia la norma en Guadalajara, México", *Revista Iberoamericana de Ciencias*, Vol. 2, No. 2, pp. 65-85