

## MODELIZACIÓN DE LA RADIACIÓN ULTRAVIOLETA TIPO B TOTAL EN LA CIUDAD DE SAN LORENZO - PARAGUAY

### MODELING OF THE TOTAL UV-B RADIATION IN THE CITY OF SAN LORENZO - PARAGUAY

CESAR DANIEL AMARILLA<sup>1</sup>

<sup>1</sup>Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Asunción, Campus Universitario, San Lorenzo, Paraguay.  
E-mail: ceamarilla@gmail.com

**Resumen:** El desarrollo de las actividades como el comercio y la construcción en las zonas urbanas, la ganadería y agricultura en las zonas rurales son reflejos de la importancia de la energía solar en las comunidades. Por esta razón, el estudio del comportamiento de la intensidad de la radiación solar es de gran interés a nivel mundial. En esta investigación se analizan diversos modelos estadísticos obtenidos mediante la aplicación de una técnica estadística de gran importancia, el análisis de regresión lineal, sobre series de datos de dos variables radiométricas (las radiaciones global y ultravioleta tipo B total) registrados en el Laboratorio de Investigaciones de la Atmósfera y Problemas Ambientales (LIAPA) de la Facultad de Ciencias Exactas y Naturales (FACEN). El estudio de las variables radiométricas se ha hecho de forma horaria. El análisis horario permite conocer las horas del día en los cuales la intensidad de los rayos solares es mayor y en los que se deben tomar las debidas precauciones. Por ello, se construyen modelos estadísticos que relacionan al índice de transmisividad de la radiación UV-B total con el índice de claridad y la masa óptica relativa mediante la aplicación de técnicas del análisis de regresión lineal. Con todo lo expuesto, es posible informar a la población sobre las lesiones que pueden ocasionar las altas dosis de radiación solar de forma sencilla y la manera en que se pueden evitar estas lesiones.

**Palabras clave:** *análisis de regresión, variables radiométricas, índice de claridad, masa óptica relativa, índice de transmisividad de la radiación UV-B total.*

**Abstract:** The development of activities such as trade and construction in urban areas, livestock and agriculture in rural areas are reflections of the importance of solar energy in communities. For this reason, the study of the behavior of the intensity of solar radiation is of great interest globally. In this research, various statistical models obtained by applying a statistical technique of great importance are discussed, the linear regression analysis on data from two sets of radiometric variables (global radiation and ultraviolet B type total) recorded in the Research Laboratory Atmospheric and Environmental Problems (LIAPA), Faculty of Natural Sciences (FACEN). The study of radiometric variables was done hourly basis. The time analysis allows knowing the hour in which the intensity of sunlight is greater and which should take precautions. Therefore, statistical models relating the transmissivity ratio of total radiation UV-B index and optical clarity relative mass by applying linear regression analyzes were constructed. With the above, it was possible to inform people about the injuries that can cause high doses of solar radiation in a simple way and how they can avoid these injuries.

**Key words:** *regression analysis, radiometric variables, clearness index, relative optical mass, transmissivity index radiation UV-B total.*

## INTRODUCCIÓN

La radiación que llega a la superficie terrestre procedente de sol modula el clima terrestre y es un factor natural de gran importancia en la evolución de la vida en la Tierra, debido a su influencia significativa en el medio ambiente (Vallejo Delgado, 2003). Esta radiación electromagnética abarca un amplio rango de longitudes de onda y se divide en varios

intervalos espectrales en los que tiene similares características: rayos  $\gamma$ , rayos X, radiación ultravioleta (UV), radiación visible y radiación infrarroja (Antón Martínez, 2007).

La región UV se divide en las franjas ultravioleta tipo C (UV-C), tipo B (UV-B) y tipo A (UV-A). Esta radiación, que representa aproximadamente un 8,3% del espectro solar extraterrestre y un 5% de la radiación que llega a la superficie de la Tierra (Varo

Martínez, 2006), es un factor natural de gran interés para la comunidad mundial debido a que es una de las más energéticas de las capas bajas de la atmósfera, razón por la cual influye enormemente sobre ella y los seres vivos (Mateos Villán, 2008). Esta radiación abarca las longitudes de onda más cortas del espectro solar, y es responsable de diferentes procesos biológicos en los seres vivos, y fisico-químicos en la atmósfera (Román Díez, 2010).

La vida en la Tierra evoluciona dependiendo, entre otros factores, de la intensidad de radiación solar incidente sobre la superficie, esencialmente de la cantidad de radiación ultravioleta incidente sobre los organismos vivos. La radiación ultravioleta es biológicamente muy activa y posee varios efectos beneficiosos, pero resulta muy perjudicial para plantas, animales y seres humanos si se exceden ciertos límites de seguridad (Varo Martínez, 2006).

Una exposición a altas cantidades de irradiancia ultravioleta, por parte de los seres humanos, puede derivar en envejecimiento prematuro de la piel, aparición de arrugas y aumento de la probabilidad de padecer enfermedades relacionadas con la depresión del sistema inmunológico, cataratas oculares y cáncer de piel (Varo Martínez, 2006).

La radiación solar es un factor natural altamente variable en el tiempo y en el espacio. Su variabilidad horaria, diurna, mensual y anual está gobernada por parámetros astronómicos, geográficos y por las condiciones atmosféricas. Las actividades humanas afectan a la atmósfera, como por ejemplo la contaminación del aire y la destrucción de la capa de ozono, afectando también indirectamente a estas radiaciones al llegar a la superficie terrestre (Vallejo Delgado, 2003).

En el Paraguay se registran altas temperaturas durante gran parte del año y la mayor proporción de la población está expuesta a rayos solares intensos.

En cuanto a los antecedentes en el estudio de la radiación solar, a nivel país se ha revisado el de Díaz Escobar (2012). Esta investigación se llevó a cabo con el propósito de estimar el comportamiento de los valores máximos diarios de la irradiancia eritémica a través de modelos de series temporales que fueron construidos a partir de datos históricos

registrados por el Laboratorio de Investigaciones de la Atmósfera y Problemas Ambientales de la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de Asunción en la ciudad de San Lorenzo-Paraguay, además de estimar pronósticos del índice UV de acuerdo a las predicciones de la irradiancia eritémica.

A nivel de otros países se han revisado varios trabajos. Como los realizados por Varo Martínez (2006), quien con el propósito de dimensionar la influencia de la masa de aire, radiación global y columna total de ozono sobre la radiación UV-B eritémica construyó modelos estadísticos aplicando técnicas de regresión, estudiando además el comportamiento de la columna total de ozono y la intensidad de la radiación solar bajo cielo despejado, Antón Martínez (2007), que empleó técnicas del análisis de regresión simple y múltiple para la construcción de modelos estadísticos que relacionen a la radiación UV con la altura solar, nubosidad y el ozono, y Mateos Villán (2008), que con el objetivo de caracterizar y estimar la radiación solar ultravioleta horizontal (UVT), empleó modelos empíricos y estudió las principales componentes de la atmósfera que interaccionan con ella.

## MATERIALES Y MÉTODOS

### Descripción de la zona de estudio

La estación meteorológica automática, cuya instalación se encuentra ubicada en el Laboratorio de Investigaciones de la Atmósfera y Problemas Ambientales (LIAPA) de la FACEN de la UNA en el Campus Universitario de la ciudad de San Lorenzo - Paraguay, está localizada a 25,34° de latitud sur, 57,52° de longitud oeste y altitud de 123 m.snm. El LIAPA cuenta actualmente con un área aproximada de 80 m<sup>2</sup>. Las mediciones realizadas por los sensores se almacenan en un Microprocesador y Sistema de Almacenamiento de datos.

### Recolección de datos

Para medir la irradiancia solar global en una superficie horizontal, el LIAPA utiliza un piranómetro CMP 11 de marca Kipp & Zonnen. Este piranómetro utiliza la tecnología de detección de temperatura

y es adecuado para redes meteorológicas y el tiempo de respuesta reducido de 1,66 segundos (63%) cumpliendo con los requisitos para aplicaciones de energía solar. Mientras que para medir la irradiancia solar UV-B total, el LIAPA utiliza un piranómetro UVB-1. Este sensor es utilizado para medir la efectividad biológica de la radiación UV-B solar.

Las técnicas de medición que utiliza el instrumento emplean una combinación de filtros de vidrios de colores para bloquear toda la luz visible del sol y fósforo fluorescente sensible que convierte la luz UV-B a luz visible (Verde). La luz verde resultante es medida con un foto detector de estado sólido. Es un instrumento rústico, diseñando para asegurar la estabilidad de su operación durante largo tiempo, sin necesidad de mantenimiento en las estaciones de campo. Este sensor está estabilizado a una temperatura de 40°C.

### Modelos estadísticos y análisis de regresión

Los modelos estadísticos consisten en expresiones matemáticas que contienen variables aleatorias, fijas o matemáticas y parámetros poblacionales. Son modelos probabilísticos. Es decir, mezclan modelos matemáticos en los cuales puede haber equivocación pero no variabilidad (modelos deterministas) y componentes aleatorios.

El análisis de regresión es una técnica estadística para investigar y modelar la relación funcional entre dos o más variables (Montgomery et al., 2002).

Los modelos de regresión son funciones de la forma  $Y = f(X_1, X_2, \dots, X_p, \epsilon)$ . Estas formas funcionales incorporan componentes que se describen a continuación:

- $Y$ : variable aleatoria, denominada **variable de respuesta**
- $X_1, X_2, \dots, X_p$ : variables no aleatorias, denominadas **variables predictoras o independientes**, consideradas como fijas dentro del modelo y cuyos valores son fijados por el experimentador.
- $\beta_p$  para todo  $i=0, 1, \dots, p$ : conectores, denominados **parámetros poblacionales**, que determinan los aportes de las variables

predictoras o transformaciones de ellas en la variable de respuesta.

- $\epsilon$ : término de perturbación, denominado **error aleatorio**, que es un error que obedece a factores de caracteres aleatorios. Este error también puede obedecer variaciones de caracteres sistemáticos a causa de errores en el modelo.

La estimación de un modelo de regresión pueden ser realizada por varios métodos. Los métodos más utilizados, según Canavos (1998) son:

- **Mínimos Cuadrados**: Este método realiza las estimaciones de los parámetros intervinientes en el modelo seleccionado mediante la minimización de la suma de los cuadrados de las diferencias entre los valores observados y estimados, mediante el uso del modelo selecto, de la variable de respuesta. La diferencia del valor observado y del estimado de la variable respuesta se expresa como:

$$e_i = y_i - \hat{y}_i, \text{ para } i=1, 2, \dots, n \quad (1)$$

y recibe el nombre de **residual**, que constituye una estimación de  $\epsilon$ .

- **Máxima verosimilitud**: Este método se basa en el hecho de que los factores de perturbación  $\epsilon_i$  tienen distribuciones normales con medias cero y varianzas constante  $\sigma^2$ . Como la variable aleatoria  $Y$  depende de la distribución del factor  $\epsilon$ , entonces  $Y$  también sigue una distribución normal. Los estimadores de máxima verosimilitud se obtienen mediante la maximización de una cierta función denominada de verosimilitud.
- Los modelos obtenidos por análisis de regresión se diagnostican mediante el **análisis de los residuales**. Este análisis consiste en un procedimiento que muestra el cumplimiento de ciertos supuestos básicos que se establecen al plantear el modelo de regresión. Estos supuestos, según Hair et al. (1999) son:
- **Normalidad de los términos de error**: Establece que la distribución de probabilidad de

los errores aleatorios es normal con media cero y varianza  $\sigma^2$ .

- Homogeneidad de varianzas de los términos de error: Establece que los errores aleatorios tienen varianzas constantes. Es decir,

$$\text{Var}(\varepsilon_i) = \sigma^2, \text{ para } i=1,2,\dots,n \quad (2)$$

- Independencia de los términos de error: Establece que los errores no están correlacionados. Es decir, cada variable predictor es independiente.
- Linealidad del fenómeno: La linealidad en regresión representa el grado de cambio en la variable dependiente asociado con la variable independiente.

### Modelos de regresión simple y múltiple

Según Canavos (1998), los modelos de regresión simple proporcionan métodos a través de los cuales pueden establecerse asociaciones entre las dos variables de interés en las cuales la relación usual no es casual. Estos modelos tienen dos significados: uno que surge de la distribución conjunta de las dos variables aleatorias; el otro es empírico y nace de la necesidad de ajustar alguna función al conjunto de datos.

Cuando los modelos de regresión simple no nos proporcionan buenas estimaciones surge la necesidad de incorporar más variables en el estudio de regresión que nos permitan mejorar las estimaciones y de esta forma disminuir el error de estimación. El análisis de regresión múltiple constituye un procedimiento estadístico que nos permite relacionar en un estudio, más de dos variables. Son modelos empíricos más precisos que los de regresión simple, debido a que involucra más variables en el estudio realizado, pero son de mayor complejidad.

### Modelos polinómicos, potencial y exponencial de regresión

Según Montgomery et al. (2002), un modelo de regresión polinómico de k-ésimo orden en una variable se define como:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_k X^k + \varepsilon \quad (3)$$

donde  $\varepsilon$  es el componente aleatorio con media cero y varianza  $\sigma^2$ .

La ecuación (3) corresponde a un modelo de regresión lineal, ya que es lineal en los parámetros y en la variable de respuesta. Además, considerando a cada potencia de  $X$  como una variable predictora, puede ser tratado como un modelo de regresión lineal múltiple.

Los modelos potencial y exponencial de regresión simple se definen respectivamente como:

$$Y = \beta_0 X_1^{\beta_1} \varepsilon \quad (4)$$

$$Y = \beta_0 \exp(\beta_1 X) \varepsilon \quad (5)$$

En estos casos  $\varepsilon$  es una componente aleatoria multiplicativa y las ecuaciones (4) y (5) presentan modelos de regresión simple no lineales, ya que no son lineales en los parámetros, pero son linealizables bajo transformaciones logarítmicas.

Al modelo potencial de regresión se le denomina también doblemente logarítmico, ya que tanto la variable de respuesta como la predictora se transforman bajo logaritmos. Las linealizaciones de los modelos potencial y exponencial están dadas por las expresiones:

$$\ln(Y) = \ln(\beta_0) + \beta_1 \ln(X) + \ln(\varepsilon) \quad (6)$$

$$\ln(Y) = \ln(\beta_0) + \beta_1 X + \ln(\varepsilon) \quad (7)$$

donde  $\ln(\varepsilon)$  es una componente aleatoria con media cero y varianza  $\sigma^2$ .

Un modelo potencial de regresión también puede ser múltiple. Por ejemplo, la siguiente ecuación corresponde a un modelo potencial de regresión múltiple que relaciona a una variable respuesta con dos variables predictoras.

$$Y = \beta_0 X_1^{\beta_1} X_2^{\beta_2} \varepsilon \quad (8)$$

y cuya linealización se expresa como:

$$\ln(Y) = \ln(\beta_0) + \beta_1 \ln(X_1) + \beta_2 \ln(X_2) + \ln(\varepsilon) \quad (9)$$

donde  $\ln(\varepsilon)$  es una componente aleatoria con media cero y varianza  $\sigma^2$ .

Cuando se trabaja con un modelo potencial o exponencial de regresión, se está estimando la media geométrica y no la media aritmética.

### Estimaciones de los parámetros de un modelo de regresión lineal por Mínimos Cuadrados

Dado un modelo de regresión lineal muestral de la forma:

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i, \quad \forall i=1, 2, \dots, n \quad (10)$$

el método de mínimos cuadrados realiza las estimaciones mediante la minimización de la suma de los cuadrados de los errores. Así, minimiza respecto a los parámetros  $\beta_0, \beta_1, \dots, \beta_p$  la función de mínimos cuadrados que se expresa como:

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})]^2, \quad \text{para } i=1, 2, \dots, n \quad (11)$$

### Estimación de la varianza del error

En un modelo de regresión con  $p$  variables predictoras se estiman  $p+1$  parámetros a través de la minimización de la suma de los cuadrados de los errores. Por lo tanto, al estimar a los parámetros  $p+1$  quedan  $n-(p+1)$  características libres, *grados de libertad*, y se tiene que la estimación insesgada de  $\sigma^2$ , a través de la suma de los cuadrados de los residuales y los grados de libertad se expresa como:

$$\widehat{\sigma}^2 = \text{CME} = \frac{1}{n-p-1} \sum_{i=1}^n e_i^2 \quad (12)$$

La expresión anterior recibe el nombre de *Cuadrado Medio del Error* (CME) y su raíz cuadrada, *error estándar de estimación* (RCME).

### Análisis de varianza y el coeficiente de determinación

El análisis de la varianza consiste en un procedimiento estadístico que divide la variabilidad total de la variable dependiente mediante la utilización de un modelo de regresión estimado, de la siguiente manera:

$$SCT = SCR + SCE \quad (13)$$

El primer miembro de la ecuación final obtenida mediante el análisis de varianza se denomina Suma de Cuadrados Total (*SCT*) y es el que mide la variabilidad total de las observaciones. Mientras

que la primera y segunda parte del segundo miembro, se denominan Suma de Cuadrados Residual (*SCE*) y Suma de Cuadrados de la Regresión (*SCR*) respectivamente. La suma cuadrado de regresión mide la parte de la variabilidad total explicada por la ecuación de regresión y la suma cuadrado de error la variabilidad no explicada por la ecuación de regresión.

Al cociente entre la suma cuadrado de regresión y la suma cuadrado total se la denomina *coeficiente de determinación* ( $R^2$ ) y es el que mide la proporción de variabilidad de la variable de respuesta explicada por los regresores considerados en el modelo de regresión. El coeficiente de determinación se expresa como:

$$R^2 = \frac{SCR}{SCT} \quad (14)$$

### La estadística de Durbin-Watson

La estadística que estudia la correlación de los errores en el tiempo originados por un modelo de regresión fue desarrollada por James Durbin y Geoffrey Watson entre 1950 y 1951. Esta estadística supone que los errores están autocorrelacionados mediante un modelo autorregresivo de primer orden. Esto es,

$$\varepsilon_t = \rho \varepsilon_{t-1} + \eta_t, \quad \text{para } t \geq 2 \quad (15)$$

donde  $\eta_t$ , denominada ruido blanco, es libre de autocorrelación y  $\rho$  es la constante que mide la correlación entre los errores y se estima mediante:

$$\hat{\rho} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2} \quad (16)$$

donde  $e_{t-1}$  y  $e_t$  son los residuos en los instantes  $t-1$  y  $t$ . La autocorrelación puede ser negativa, cero o positiva y se mide por medio del estadístico  $d$  que está dada por:

$$d \approx 2 - 2\hat{\rho} \quad (17)$$

Cuando se detecta autocorrelación, Canavos (1998) muestra la utilización de un proceso iterativo que se debe a Cochrane y Orcutt (1949), y que plantea la transformación:

$$y_t - \rho y_{t-1} = \beta_0 (1 - \rho) + \beta_1 (x_{t1} - \rho x_{(t-1)1}) + \dots + \beta_p (x_{tp} - \rho x_{(t-1)p}) \quad (18)$$

Entonces, el modelo de regresión lineal transformado libre de autocorrelación y los ajustes de los parámetros originales se expresan como:

$$y_t^* = \beta_0^* + \beta_1^* x_{t1}^* + \dots + \beta_p^* x_{tp}^* + \eta_t, \quad \text{para } t=2,3,\dots, n \quad (19)$$

$$\beta_0 = \frac{\beta_0^*}{1 - \rho} \text{ y } \text{Var}(\beta_0) = \frac{\text{Var}(\beta_0^*)}{(1 - \rho)^2} \quad (20)$$

$$\beta_i = \beta_i^* \text{ y } \text{Var}(\beta_i) = \text{Var}(\beta_i^*), \quad \text{para } i=1,2,3,\dots, p \quad (21)$$

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

Este criterio es una medida de la calidad relativa de un modelo para un conjunto dado de datos. Proporciona un medio para la selección del modelo con mejor ajuste a los datos y se basa en la entropía de la información, que ofrece una estimación relativa de la pérdida de información cuando se utiliza un modelo determinado para representar el proceso que realmente genera un conjunto de datos.

El modelo preferido es el que tiene el valor mínimo de AIC. El AIC utilizado para analizar la comparación de las bondades de ajustes de los modelos de regresión se define como:

$$AIC_p = n \log \left( \frac{SCE}{n} \right) + 2k \quad (22)$$

donde SCE es la suma cuadrado de error, k es el número de parámetros estimados.

**Validación cruzada en k iteraciones**

La validación cruzada permite estudiar la estabilidad dentro de una determinada muestra considerada para un análisis estadístico. Se utiliza preferentemente en situaciones en donde el objetivo principal es la predicción y se quiere estudiar la precisión de un modelo, construido a partir de una muestra, que se utilizará para la predicción.

Este método consiste en la división de los datos

de muestra en dos conjuntos complementarios, una denomina muestra de entrenamiento y la otra de prueba. La muestra de entrenamiento es utilizado para estimar los parámetros correspondientes de un determinado modelo y la de prueba para la validación del mismo.

**Medias de las diferencias relativas y absolutas relativas**

Las medias de las diferencias relativas (MBE) y absolutas relativas (MABE) son indicadores de la bondad de ajuste de los modelos estimados y del error cometido al realizar predicciones de la variable de respuesta a través de un modelo de ajuste respectivamente. Estos indicadores están dados en la siguiente expresión:

$$MBE = \frac{1}{n} \sum_{i=1}^n \frac{(y_i - \hat{y}_i)}{y_i} \text{ y } MABE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad (23)$$

donde  $y_i$  es un valor de la variable de respuesta obtenida a través de mediciones y  $\hat{y}_i$  es el correspondiente valor estimado a través de un modelo de ajuste.

**Variables involucradas en los modelos de regresión**

Las variables consideradas en la formulación de los modelos de regresión son construidas a partir de las variables radiométricas, radiación global y UV-B total, y de parámetros astronómicos.

- **Índice de claridad o transmisividad solar atmosférica**

Este índice refleja la influencia de los factores atenuantes, en especial el de las nubes y aerosoles, sobre la radiación global antes de incidir sobre la superficie terrestre. Se define la transmisividad solar horaria en la siguiente ecuación:

$$k_t^h = \frac{G_{med}^h}{E_{ext}^h} \quad (24)$$

$G_{med}^h$  es la irradiancia horaria de la radiación global registrada en el LIAPA y  $E_{ext}^h$  es la irradiancia

horaria de la radiación solar extraterrestre y viene dada por la expresión:

$$E_{ext}^h = I_{sc} E_0 \cos(SZA_M) \quad (25)$$

$I_{sc}$  es la constante solar extraterrestre y toma un valor de  $1366,1 \text{ Wm}^{-2}$ , dando una energía por unidad de área de  $4917960 \text{ Jm}^{-2}$  por cada hora según el espectro de Gueymard (2003),  $E_0$  constituye el factor de corrección de la excentricidad de la órbita terrestre y cuya expresión en serie de Fourier, siguiendo a Varo Martínez (2006), es:

$$E_0 = 1,00011 + 0,034221 \cos(\Gamma) + 0,00128 \text{sen}(\Gamma) + 0,000719 \cos(2\Gamma) + 0,000077 \text{sen}(2\Gamma) \quad (26)$$

con  $\Gamma$  medido en radianes, denominado *ángulo diario*, y que está relacionado con el día juliano  $n$  (sucesivamente se cuentan los días hasta 365 o 366) de la siguiente manera:

$$\Gamma = \frac{2\pi(n-1)}{361} \quad (27)$$

Finalmente,  $SZA_M$  es el ángulo cenital, en el punto medio de un intervalo de tiempo de una hora, medido en radianes. Este parámetro astronómico, siguiendo a Antón Martínez (2007), se puede calcular mediante la expresión:

$$\cos(SZA_M) = \text{sen}(\varphi)\text{sen}(\delta) + \cos(\varphi)\cos(\delta)\cos(\omega_M) \quad (28)$$

En la ecuación anterior,  $\varphi$ ,  $\delta$  y  $\omega_M$  constituyen la latitud del lugar, declinación de la Tierra y el ángulo horario en el punto medio de un intervalo de una hora. La declinación se puede estimar a través de la serie de Fourier dada por:

$$\delta = 0,00692 - 0,39991 \cos(\Gamma) + 0,07026 \text{sen}(\Gamma) - 0,00676 \cos(2\Gamma) + 0,00091 \text{sen}(2\Gamma) - 0,0027 \cos(3\Gamma) + 0,00148 \text{sen}(3\Gamma) \quad (29)$$

En ángulo horario se estima mediante la expresión

$$\omega_M = \frac{\pi}{12} t - \pi \quad (30)$$

Con  $t$  denominado *tiempo solar verdadero* y cuya expresión está dada por:

$$t = t_{civil} + \frac{L_{ref} - L_{local}}{15} + \frac{E_t}{60} \quad (31)$$

$t_{civil}$ ,  $L_{ref}$  y  $L_{local}$  constituyen la hora local, la longitud del mediano de referencia y la longitud del mediano del lugar respectivamente.  $E_t$  es denominada la *ecuación del tiempo* y es estimada por serie de Fourier mediante la expresión:

$$E_t = 229,18 [0,000075 + 0,001868 \cos(\Gamma) - 0,032077 \text{sen}(\Gamma) - 0,014615 \cos(2\Gamma) - 0,040849 \text{sen}(2\Gamma)] \quad (32)$$

#### • Índice de transmisividad de la UV-B total

Este índice mide la modulación que sufre la radiación UV-B total en su paso por la atmósfera. Se define la transmisividad ultravioleta tipo B total en el intervalo de tiempo de una hora de la siguiente manera:

$$k_{UVB}^h = \frac{E_{UVB_{med}}^h}{E_{UVB_{ext}}^h} \quad (33)$$

donde  $E_{UVB_{med}}^h$  es la irradiancia total en el intervalo de una hora de la radiación UV-B total registrada en la estación LIAPA y  $E_{UVB_{ext}}^h$  constituye la irradiancia total en el intervalo de una hora de la UV-B total extraterrestre, y que puede ser calculada considerando, la parte de la constante solar que corresponda a la banda de la UV-B total y la ecuación:

$$E_{UVB_{ext}}^h = I_{sc_{UVB}} E_0 \cos(SZA_M) \quad (34)$$

donde  $I_{sc_{UVB}}$  es la constante solar correspondiente a la irradiancia de la UV-B total y que toma un valor de  $21,30 \text{ Wm}^{-2}$  aproximadamente, dando una energía por unidad de área de  $76680 \text{ Jm}^{-2}$  por cada hora según el espectro de Gueymard (2003).

#### • Altura solar y masa óptica relativa

La altura solar es un factor de gran influencia sobre la radiación solar a través del tiempo. Esto se debe a que la cantidad de absorbentes y dispersores de la irradiancia ultravioleta tipo B total presentes

en la atmósfera es diferente para cada altura solar (Antón Martínez, 2007). Las revisiones bibliográficas de las investigaciones hechas sobre radiación solar muestran que el indicador de la influencia de la altura solar sobre la radiación UV-B total es la **masa óptica relativa** simbolizada por  $m_r$  y que se determina mediante la expresión:

$$m_r = \frac{1}{\cos(SZA)} \quad (35)$$

### Software informático

Para la construcción de los modelos estadísticos y el análisis de regresión, este trabajo ha requerido la utilización de los paquetes estadísticos R y SPSS.

## RESULTADOS Y DISCUSIÓN

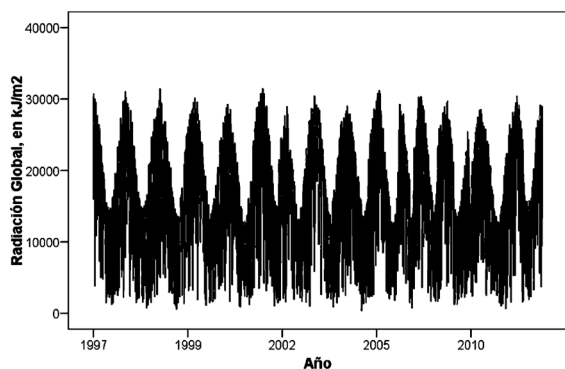
### Caracterización de las evoluciones temporales de las variables radiométricas

Para cada variable radiométrica, el análisis exploratorio de los datos da cuenta de los resultados:

- **Evolución temporal de la radiación global**

La variación temporal de la radiación global a lo largo del periodo 1997 al 2012 presenta un comportamiento sinusoidal con picos máximos y mínimos en los meses de verano e invierno respectivamente. Este comportamiento se debe a la variación de la altura solar a lo largo del año.

Los valores horarios de la irradiancia global que muestran intensidades elevadas se registraron entre



**Figura 1.** Evolución diaria de la radiación global, en  $kJm^{-2}$ , en el periodo 1997 al 2012.

9:00 y 15:00 horas. Los valores máximos observados fueron 3980 y 3989  $kJm^{-2}$  y se registraron entre las 11:00 y 13:00 horas el 27 de enero de 1997 y 14 de diciembre de 2012 respectivamente. El menor valor horario fue registrado el 14 de noviembre de 2001 entre las 7:00 y 8:00 horas, y fue de 5  $kJm^{-2}$ .

Se constatan dispersiones horarias importantes en los distintos intervalos horarios, debido a que tanto las desviaciones estándar y los rangos intercuartílicos son un tanto altas con relación a las medias y medianas. El coeficiente de variación en todos los casos es mayor al 40%.

En cuanto al análisis de los valores máximos y mínimos horarios, se constatan que las diferencias entre los valores mínimos y los percentiles 5%; y las diferencias entre los valores máximos y los percentiles 95% son muy elevadas. Entonces, los valores mínimos y máximos no son representativos dentro de las distribuciones horarias de la radiación global, pudiendo representar valores atípicos en la distribución horaria general de datos de esta variable.

- **Evolución temporal de la radiación UV-B total**

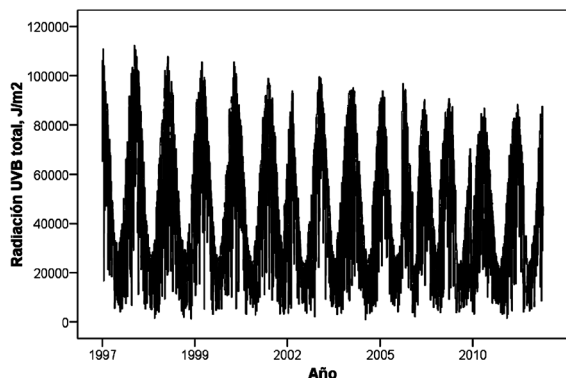
Al igual en el caso de la radiación global, la radiación UV-B total presenta un comportamiento del tipo sinusoidal con picos máximos y mínimos en los meses de verano e invierno respectivamente.

Las intensidades horarias más altas de irradiancia UV-B total horaria se registraron entre las 9:00 y 15:00 horas. Los valores máximos encontrados fueron de 18193 y 18026  $Jm^{-2}$ ; que se registraron los días 11 y 13 de enero de 1997 en los intervalos horarios de 12:00 a 13:00 horas y de 11:00 a 12:00 horas, respectivamente. El menor valor horario fue de 11  $Jm^{-2}$ , registrado entre las 17:00 y 18:00 horas del día 21 de diciembre de 2010.

Se observaron grandes dispersiones en los distintos intervalos horarios, debido a que tanto las desviaciones estándares y los rangos intercuartílicos son altos con relación a las medias y medianas. El coeficiente de variación en todos los casos es mayor al 40%, llegando a casi 70% en algunos casos.

En el análisis de los valores máximos y mínimos





**Figura 2.** Evolución diaria de la radiación UV-B total, en  $Jm^{-2}$ , en el periodo 1997 al 2012.

horarios, se constatan que las diferencias, entre los valores mínimos y los percentiles 5% por un lado y las diferencias entre los valores máximos y los percentiles 95% por otro, son muy elevadas. Entonces, los mínimos y máximos no son representativos dentro de las distribuciones horarias de la radiación UV-B total, pudiendo representar valores atípicos en la distribución general de datos de esta variable.

### Aplicación del análisis de Regresión

#### Modelos de regresión lineal simple

Para poder determinar la influencia de la altura solar sobre la UV-B total en un instante dado, el cielo debe estar despejado. Una de las maneras para determinar si el cielo está despejado, es a través del índice de claridad. Según Román Díez (2010), la condición del cielo según el índice de claridad horario está dada en la Tabla 1.

Según la Tabla 1, para proponer y estimar las funciones que relacionen a la UV-B total con la altura solar se deben utilizar datos para los cuales

**Tabla 1.** Condición del cielo según el índice de claridad horario  $k_t^h$ .

Criterio	Condición de cielo
$0 < k_t^h \leq 0,34$	Hora nublada
$0,34 < k_t^h \leq 0,75$	Hora parcialmente nublada
$k_t^h \geq 0,75$	Hora de cielo claro

$$k_t^h \geq 0,75 .$$

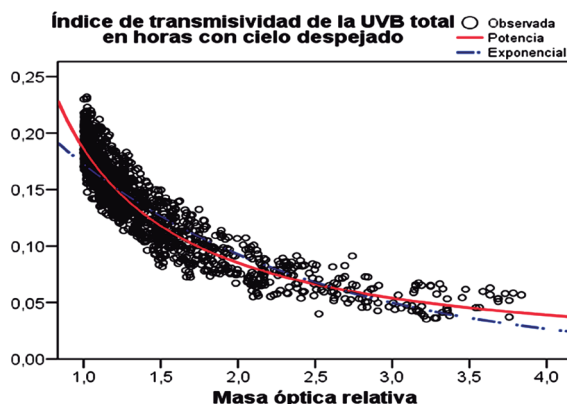
Para establecer la influencia de la altura solar sobre la variación de la intensidad de la radiación UV-B total en horas con cielo despejado en este trabajo se proponen modelos de los tipos potencial y exponencial, ambos linealizables mediante la transformación logarítmica, siguiendo las recomendaciones de Murillo et al. (2003) y Antón Martínez (2007) para el primer caso y la de Dubrovsky et al. (2000) y Antón Martínez (2007) para el segundo.

$$k_{UVB}^h = \beta_0 m_r^{\beta_1} \varepsilon_h \tag{36}$$

$$k_{UVB}^h = \beta_0 \exp(\beta_1 m_r) \varepsilon_h \tag{37}$$

En la Figura 3 se puede observar la dispersión de los datos referentes al índice de transmisividad de la UV-B total en función de la masa óptica relativa. Las curvas azul y roja representan a los modelos potencial y exponencial estimados por mínimos cuadrados ordinarios.

En la Tabla 2 se puede observar que los coefi-



**Figura 3.**  $k_{UVB}^h$  versus  $m_r$ , en horas con cielos despejados, en el periodo 1997 al 2011.

cientes de determinación para los modelos potencial y exponencial son superiores a los 80%. Además, el menor valor en el error estándar de estimación lo presenta el modelo potencial. Entonces, el modelo que muestra el mejor ajuste es el potencial.

Por su parte, para establecer la influencia horaria de la nubosidad, columna total de ozono y la carga

**Tabla 2.** Estadísticos  $R^2$  y  $RCME$  para los modelos potencial y exponencial que relacionan a las variables  $k_{UVB}^h$  y  $m_r$  en horas con cielo despejado.

Modelos	$n$	$R^2$	$RCME$
Potencial	1162	0,880	0,103
Exponencial	1162	0,835	0,120

de aerosoles en conjunto sobre la radiación UV-B total, el ángulo cenital debe ser prácticamente constante. Entonces, para proponer las funciones que relacionen a los índices de transmisividades  $k_{UVB}^h$  y  $k_t^h$  se seleccionan datos horarios en intervalos cenitales de longitud 5°. Así Antón Martínez (2007) y Mateos Villán (2008) entre otros proponen modelos polinómicos de hasta tercer grado y potencial linealizable mediante la transformación logarítmica. Además de los modelos polinómicos y potencial, en esta investigación, se propone también un modelo exponencial linealizable mediante la transformación logarítmica.

$$k_{UVB}^h = \beta_0 + \beta_1 k_t^h + \beta_2 (k_t^h)^2 + \beta_3 (k_t^h)^3 + \varepsilon_h \quad (38)$$

$$k_{UVB}^h = \beta_0 (k_t^h)^{\beta_1} \varepsilon_h \quad (39)$$

$$k_{UVB}^h = \beta_0 \exp(\beta_1 k_t^h) \varepsilon_h \quad (40)$$

La Tablas 3 muestra los valores estimados de los estadísticos  $R^2$ ,  $AIC$  y  $RCME$ , en intervalos cenitales de longitud 5°, obtenidos para los modelos polinómicos, potencial y exponencial horarios. Se observa que, en los distintos intervalos cenitales considerados, el mayor valor del  $R^2$  y el menor valor en el  $AIC$ , en casi todos los casos, se presenta para el polinomio de grado tres. Además, el mayor valor para  $R^2$  superior al 90% en casi todos los casos, se presenta en el modelo potencial y el menor valor en el estadístico  $RCME$ , inferior a 0,02, se obtiene para el polinomio de grado tres en todos los casos analizados. No obstante, el modelo potencial también muestra un valor bajo, inferior a 0,3, del estadístico  $RCME$ . Entonces, los modelos polinómico de grado tres y potencial muestran buenos ajustes.

Como el modelo polinómico cúbico tiene cuatro

parámetros a estimar y el potencial solamente dos, por el principio de parsimonia, en este trabajo se selecciona al modelo potencial como el adecuado para estudiar la influencia de la nubosidad y aerosoles en conjunto sobre la radiación UV-B total.

**Modelo de regresión múltiple**

El análisis de regresión lineal simple realizado en la sección anterior, reveló que el índice de transmisividad horario de la UV-B total se relacionaba de mejor manera con el índice de claridad y la masa óptica relativa a través de modelos potenciales.

Antón Martínez (2007) y Mateos Villán (2008), en estudios realizados en las ciudades españolas de Badajoz y Valladolid propusieron con muy buenos resultados modelos potenciales múltiples para explicar la variabilidad del índice de transmisividad horaria de la UV-B total por medio del índice de claridad y la masa óptica relativa. Entonces, según el análisis de regresión simple realizada en la sección anterior y la bibliografía consultada, se propone el siguiente modelo de regresión múltiple:

$$k_{UVB}^h = \beta_0 m_r^{\beta_1} (k_t^h)^{\beta_2} \varepsilon_h \quad (41)$$

cuya linealización mediante la transformación logarítmica se expresa como:

$$\ln(k_{UVB}^h) = \ln(\beta_0) + \beta_1 \ln(m_r) + \beta_2 \ln(k_t^h) + \ln(\varepsilon_h) \quad (42)$$

La ecuación (32) modela la media geométrica de la variable  $k_{UVB}^h$ . En esta ecuación la variable  $\ln(\varepsilon_h)$  tiene distribución normal con media cero y varianza uno. El constaste de hipótesis global que se plantea sobre los parámetros de conexión, es:

$$H_0 : \beta_1 = \beta_2 = 0$$

$$H_1 : \beta_1 \neq 0 \text{ o } \beta_2 \neq 0$$

El resumen del análisis de varianza realizado en base a la muestra tomada se presenta en la Tabla 4. Se observa que el valor obtenido para el p-valor es de 2,00E-16, que es mucho menor que 0,05. Por lo tanto, la regresión múltiple es estadísticamente significativa. Entonces, podemos decir que alguna

**Tabla 3.** Estadísticos  $R^2$ ,  $AIC$  y  $RCME$ , en intervalos cenitales de longitud  $5^\circ$ , para los modelos polinómicos, potencial y exponencial que relacionan a los índices de transmisividades horarios.

SZA	n	Comparaciones entre los modelos			Comparaciones entre modelos		
		Polinómicos	$R^2$	AIC	Modelos	$R^2$	RCME
15-20	1579	Lineal	0,885	-5633,620	Cúbico	0,886	0,016
		Cuadrático	0,886	-5640,974	Potencial	0,958	0,128
		Cúbico	0,886	-5639,040	Exponencial	0,787	0,288
20-25	2382	Lineal	0,895	-8597,153	Cúbico	0,896	0,016
		Cuadrático	0,896	-8604,018	Potencial	0,966	0,126
		Cúbico	0,896	-8603,081	Exponencial	0,791	0,313
25-30	1892	Lineal	0,853	-6587,395	Cúbico	0,854	0,018
		Cuadrático	0,853	-6586,655	Potencial	0,944	0,162
		Cúbico	0,854	-6591,443	Exponencial	0,778	0,322
30-35	3087	Lineal	0,855	-11060,167	Cúbico	0,834	0,016
		Cuadrático	0,858	-11080,965	Potencial	0,948	0,155
		Cúbico	0,834	-11084,119	Exponencial	0,745	0,343
35-40	3108	Lineal	0,855	-11220,964	Cúbico	0,858	0,015
		Cuadrático	0,855	-11218,911	Potencial	0,942	0,162
		Cúbico	0,858	-11250,667	Exponencial	0,780	0,317
40-45	3484	Lineal	0,832	-12657,539	Cúbico	0,840	0,015
		Cuadrático	0,832	-12657,615	Potencial	0,928	0,178
		Cúbico	0,840	-12730,653	Exponencial	0,762	0,325
45-50	5366	Lineal	0,849	-24227,734	Cúbico	0,856	0,012
		Cuadrático	0,852	-20467,690	Potencial	0,942	0,159
		Cúbico	0,856	-20542,175	Exponencial	0,749	0,329
50-55	4259	Lineal	0,846	-16451,857	Cúbico	0,856	0,011
		Cuadrático	0,846	-16449,872	Potencial	0,935	0,172
		Cúbico	0,856	-16557,281	Exponencial	0,767	0,326
55-60	3782	Lineal	0,792	-14790,224	Cúbico	0,819	0,010
		Cuadrático	0,792	-14790,681	Potencial	0,905	0,198
		Cúbico	0,819	-15015,929	Exponencial	0,727	0,335
60-65	3808	Lineal	0,795	-15371,694	Cúbico	0,795	0,010
		Cuadrático	0,795	-15369,859	Potencial	0,912	0,198
		Cúbico	0,816	-15547,619	Exponencial	0,719	0,355
70-75	3730	Lineal	0,738	-16035,590	Cúbico	0,764	0,007
		Cuadrático	0,739	-16045,065	Potencial	0,851	0,257
		Cúbico	0,764	-16202,080	Exponencial	0,662	0,486

de las variables predictoras colabora en la explicación del comportamiento del índice de transmisividad horario de la UV-B total. Los contrastes de hipótesis individuales que se plantean sobre  $\beta_1$  y  $\beta_2$  son las siguientes:

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0 \text{ para } i = 1, 2$$

La Tabla 5 presenta el resumen del análisis realizado en base a la muestra tomada para los contrastes de hipótesis anteriores. Se puede observar que los valores del p-valor son iguales a  $2,00E-16$ , que es muy inferior al 0,05. Por lo tanto, se puede decir que

**Tabla 4.** Análisis de varianza para el modelo de regresión lineal que relaciona al índice de transmisividad horario de la UV-B total con el índice de claridad y la masa óptica relativa.

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrados medios	F-cal	p-valor
Regresión	27603,172	2	13801,586	408724,123	2,00E-16
Residual	1425,629	42219	0,034		
Total	29028,802	42221			

cada variable predictora colabora de manera significativa en presencia de la otra sobre la explicación del comportamiento del índice de transmisividad de la UV-B total.

El valor obtenido del estadístico de Durbin-Watson es  $d = 0,601$  que indica la presencia de autocorrelación positiva. Para dar solución a la autocorrelación positiva detectada, se realizan las estimaciones de los parámetros  $\ln(\beta_0)$ ,  $\beta_1$  y  $\beta_2$  mediante las relaciones:

$$\ln(\beta_0) = \frac{\ln(\beta_0^*)}{1-\rho}, \beta_1 = \beta_1^* \text{ y } \beta_2 = \beta_2^* \quad (43)$$

$\ln(\beta_0^*)$ ,  $\beta_1^*$  y  $\beta_2^*$  son parámetros correspondientes al modelo de regresión lineal múltiple transformado:

$$\ln(k_{UVB}^h)^* = \ln(\beta_0^*) + \beta_1^* \ln(m_r)^* + \beta_2^* \ln(k_t^h)^* + \ln(\varepsilon_h)^* \quad (44)$$

**Tabla 5.** Contrastes de hipótesis realizado sobre los parámetros  $\beta_1$  y  $\beta_2$  del modelo de regresión lineal múltiple.

Coefficientes	Estimación	Error estándar	t-cal	p-valor
$\beta_1$	0,8569	0,0012	712,395	2,00E-16
$\beta_2$	-1,0806	0,0026	-418,734	2,00E-16

donde  $\ln(\varepsilon_h)^*$  es un ruido blanco y  $\ln(k_{UVB}^h)^*$ ,  $\ln(m_r)^*$  y  $\ln(k_t^h)^*$  están dadas por:

$$\ln(k_{UVB}^h)^* = \ln(k_{UVB}^h)_t - \rho \ln(k_{UVB}^h)_{t-1} \quad (45)$$

$$\ln(k_t^h)^* = \ln(k_t^h) - \rho \ln(k_{t-1}^h) \quad (46)$$

$\ln(m_r)^* = \ln(m_r)_t - \rho \ln(m_r)_{t-1}$  (47) y  $\rho$  el coeficiente de correlación entre los errores.

La Tabla 6 presenta el resumen de las estimaciones por mínimos cuadrados ordinarios de los parámetros intervinientes en el modelo lineal múltiple presentado en la ecuación (34). El nuevo valor del estadístico  $d$  de Durbin-Watson indica ausencia de

**Tabla 6.** Estimaciones de los parámetros  $\ln(k_{UVB}^h)^*$ ,  $\beta_1^*$  y  $\beta_2^*$ , además de los cálculos de los estadísticos  $R^2$ ,  $RCME$  y  $d$  para el modelo lineal múltiple transformado.

Coefficientes	Estimación	Error estándar	$R^2$	$CCME$	$d$
$\ln(\beta_0^*)$	-0,4552	0,0008	0,937	0,1308	2,241
$\beta_1^*$	0,856	0,0014			
$\beta_2^*$	-1,0339	0,0027			

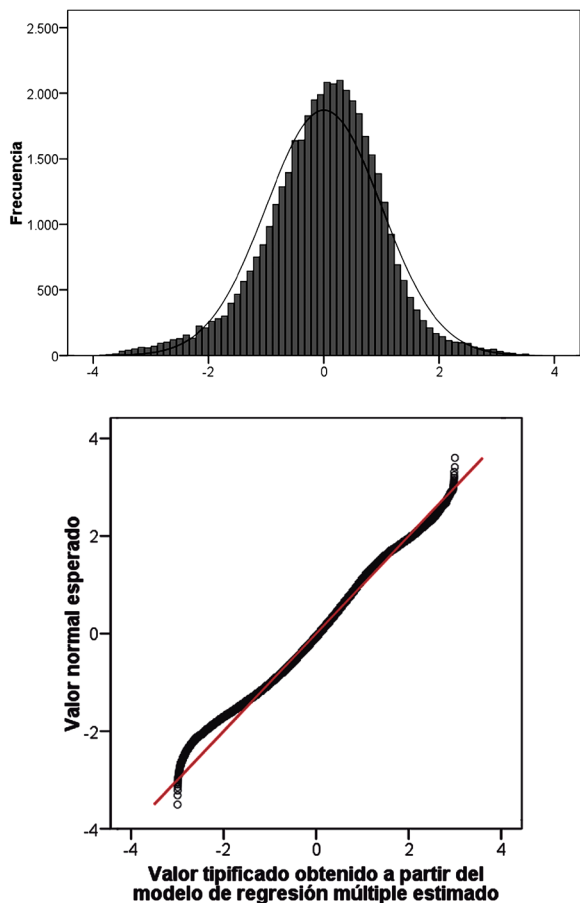


Figura 4. Histograma de los residuos estandarizados y gráfico Q-Q plot normal de regresión, resultantes en la estimación del modelo de regresión lineal múltiple.

autocorrelación. Las nuevas estimaciones de los parámetros  $\ln(k_{UVB}^h)$ ,  $\beta_1^*$  y  $\beta_2^*$  con sus intervalos de confianza a un 95% de fiabilidad, se presentan en la Tabla 7.

La Figura 4 presenta el histograma de los residuos estandarizados observados frente a la curva

normal y el comportamiento de dichos residuos frente a los valores de los cuantiles normales esperados. Se puede ver que los residuos se ubican en su gran mayoría sobre la curva normal y que además, se presenta una muy leve asimetría negativa. Según lo observado, podemos asumir que se verifica de manera aproximada el supuesto de normalidad.

La Figura 5 muestra las regresiones parciales entre  $\ln(k_{UVB}^h)$  con cada regresora. Se observa la

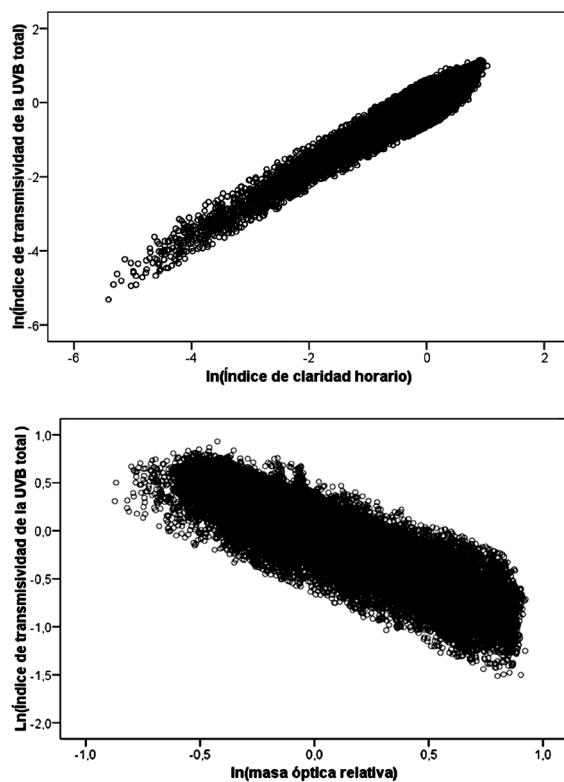


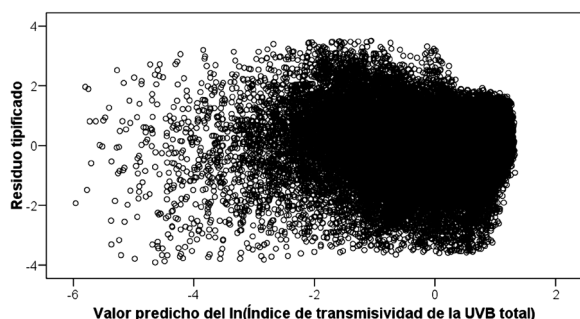
Figura 5. Regresiones parciales entre el logaritmo del índice de transmisividad horario de la UVB total con los logaritmos del índice de claridad horario y de la masa óptica relativa.

Tabla 7. Estimaciones de los parámetros  $\ln(\beta_0)$ ,  $\beta_1$  y  $\beta_2$ , utilizando las estimaciones de los parámetros del modelo lineal múltiple transformado.

Coefficientes	Estimación	Error estándar	Intervalo de confianza al 95%
$\ln(\beta_0)$	-1,4944	0,0027	(-1,4996;-1,4892)
$\beta_1$	0,856	0,0014	(0,8506;0,8614)
$\beta_2$	-1,0339	0,0027	(-1,0389;-1,0283)

correcta relación lineal entre la variable  $\ln(k_{UVB}^h)$  con cada una de las regresoras. Así, se confirma el cumplimiento del supuesto de linealidad del fenómeno.

La Figura 6 muestra la dispersión de los residuos estudentizados frente a los valores predichos. Se constata a través de esta figura, que la dispersión de los residuos estudentizados no presentan ninguna



**Figura 6.** Residuos estudentizados versus valores predichos obtenidos en a través del modelo de regresión lineal múltiple estimado.

**Tabla 8.** Tolerancia y factor de inflación de la varianza correspondientes a las variables regresoras presentadas en el modelo de regresión múltiple.

Regresora	Tolerancia	FIV
$\ln(k_t^h)$	0,967	1,034
$\ln(m_r)$	0,967	1,034

tendencia, confirmando el cumplimiento de los supuestos de homogeneidad de varianzas e independencia de los errores.

En cuanto al análisis de la multicolinealidad entre las variables regresoras, según los valores de la tolerancia y el factor de inflación de la varianza (FIV) observados en la Tabla 8, podemos decir que se detecta en un nivel muy bajo. Así, el valor  $1,034 < 5$  del FIV para ambos casos, indica una colinealidad irrelevante.

El análisis de la estabilidad dentro de la muestra considerada se realiza a través del método de la validación cruzada de k iteraciones. En la Tabla 9 se resumen las estimaciones realizadas

**Tabla 9.** Estimaciones a través de las muestras de entrenamiento para el modelo de regresión lineal múltiple, en 15 iteraciones.

k	n <sub>e</sub>	Estimaciones			R <sup>2</sup>	RCME
		$\ln(\beta_0)$	$\beta_1$	$\beta_0$		
1	38881	-1,5056	0,8531	-1,0318	0,938	0,1297
2	38803	-1,5030	0,8566	-1,0311	0,937	0,1317
3	38869	-1,4984	0,8570	-1,0323	0,938	0,1311
4	38722	-1,4890	0,8561	-1,0316	0,936	0,1322
5	38748	-1,4973	0,8555	-1,0322	0,937	0,1320
6	39703	-1,4951	0,8551	-1,0324	0,938	0,1294
7	39657	-1,4891	0,8564	-1,0329	0,938	0,1307
8	38885	-1,4945	0,8556	-1,0352	0,937	0,1313
9	38970	-1,4895	0,8576	-1,0368	0,938	0,1296
10	40392	-1,4931	0,8568	-1,0348	0,938	0,1310
11	39998	-1,4900	0,8584	-1,0347	0,937	0,1312
12	39561	-1,4918	0,8558	-1,0345	0,937	0,1309
13	40120	-1,4904	0,8553	-1,0353	0,937	0,1310
14	41048	-1,4918	0,8549	-1,0339	0,937	0,1306
15	38749	-1,4883	0,8561	-1,0341	0,938	0,1295

a través de las muestras de entrenamiento en cada iteración. Se observa en las distintas iteraciones como los valores estimados de los parámetros involucrados en el modelo de regresión múltiple no presentan grandes diferencias. Con lo cual, se puede apreciar una muy buena estabilidad dentro de la muestra considerada. La Tabla 10 presenta el resumen de las validaciones realizadas a través de las muestras de prueba. Los valores absolutos observados del promedio de las diferencias relativas  $MBE$  son inferiores al 5% en todos

los casos, indicando de este modo, ligeros sub o sobre-estimaciones en las distintas validaciones realizadas. Los valores del promedio del valor absoluto de las diferencias relativas  $MABE$  indican que se ha estimado al índice de transmisividad horaria de la UV-B total en las distintas iteraciones con errores inferiores al 7%.

A continuación, se comparan los datos del índice de transmisividad horario de la UV-B total, correspondientes al año 2012, construidos a partir de mediciones realizadas en el LIAPA,

**Tabla 10.** Validaciones a través de las muestras de prueba para el modelo de regresión lineal múltiple, en 15 iteraciones.

$k$	Año	$n_p$	$MBE(\%)$	$MABE(\%)$
1	1997	3341	4,693	6,799
2	1998	3419	4,620	6,542
3	1999	3353	2,758	6,478
4	2000	3500	1,928	5,131
5	2001	3474	1,502	5,510
6	2002	2519	0,282	5,841
7	2003	2565	-2,608	5,380
8	2004	3337	0,871	4,752
9	2005	3252	-0,500	5,009
10	2006	1830	0,208	4,121
11	2007	2224	-0,874	5,057
12	2008	2661	-1,118	4,009
13	2009	2102	-2,518	4,526
14	2010	1174	-3,926	5,137
15	2011	3473	-2,252	4,513

en el periodo 1 de enero al 31 de diciembre, y no fueron utilizadas en el proceso de estimación, con los datos obtenidos a partir de predicciones, para el mismo año, hechas es base al modelo de regresión múltiple estimado. Se cuentan con 2525 casos en total.

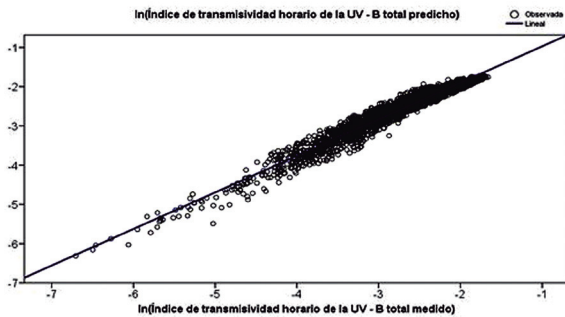
Las asociaciones de los valores medidos y predichos correspondientes al año 2012 de la variable  $\ln(k_{UVB}^h)$  se observan en el diagrama

de dispersión presentado en la Figura 7. La línea azul corresponde a la regresión lineal simple entre las magnitudes medidas y predichas, y cuya expresión está dada por:

$$\ln(k_{UVB}^{h(predicho)}) = -0,0416 + 0,9309 \ln(k_{UVB}^{h(medido)})$$

con  $R^2 = 0,96$  (38)

Dados los resultados obtenidos en el proceso



**Figura 7.** Diagrama de dispersión que refleja la asociación entre los valores medidos y predichos, correspondientes al año 2012, del logaritmo del índice de transmisividad horario de la UV-B total.

de diagnóstico y validación, el modelo de regresión lineal múltiple estimado que relaciona a los índices de transmisividades horarios y la altura solar está dado por la expresión:

$$\widehat{k_{UVB}^h} = 0,2244 m_r^{-1,0336} (k_t^h)^{0,8560}$$

con  $R^2 = 0,94$  (39)

En la ecuación (39), la variable  $\widehat{k_{UVB}^h}$  es la media geométrica estimada de la variable  $k_{UVB}^h$  a partir de las variables  $k_t^h$  y  $m_r$ . Este modelo indica que con cada incremento del 1% de la variable  $k_t^h$ , manteniéndose constante la variable  $m_r$ , se estima que la variable  $k_{UVB}^h$  se incrementa en promedio en un 0,86% y que con cada incremento del 1% de la variable  $m_r$ , manteniéndose constante la variable  $k_t^h$ , se estima que la variable  $k_{UVB}^h$  disminuye en un 1,03%.

Entonces, según la estimación del modelo presentado en la ecuación (39), el máximo valor que alcanza la media geométrica del índice de transmisividad horario de la UV-B total en FACEN es 0,2244 cuando el índice de claridad y la masa óptica relativa alcanzan el valor 1. Además, entre el índice de claridad y la masa óptica relativa explican un 94% de la variabilidad total experimentada por el índice de transmisividad horario de la UV-B total.

## CONCLUSIONES

Con el fin de determinar modelos empíricos que

ayuden a entender y estimar el comportamiento de la radiación UV-B total, se proponen modelos de regresión polinómico, potencial y exponencial. Con la ayuda de los software estadísticos R y SPSS se estiman, por el método de mínimos cuadrados ordinarios, los diversos modelos horarios propuestos. Según los datos horarios disponibles, los modelos estimados arrojaron valores altos en los coeficientes de determinación y bajos valores en los errores estándares de estimación en la mayoría de los casos. A través del Criterio de Información de Akaike, se selecciona el modelo polinómico de mejor ajuste. De entre los modelos polinómico, potencial y exponencial propuestos, se selecciona el que presenta menor valor en el error estándar de estimación. El modelo de regresión seleccionado para su posterior validación es del tipo potencial.

El resultado del proceso de validación del modelo potencial seleccionado a través del análisis de los residuales, validación cruzada y utilización de nuevas muestras, arroja valores en los coeficientes de determinación superiores al 93%. El error estándar de estimación para el modelo que relaciona a los índices de transmisividades horarios y la masa óptica relativa es de 0,1308. Por lo tanto, la dependencia potencial estudiada arroja muy buenos resultados.

En resumen, en base a la técnica estadística que se ha aplicado en la búsqueda de posibles relaciones funcionales entre las radiaciones global y UV-B total y la altura solar, se constata que es factible predecir la intensidad de la irradiancia de la UV-B total en la ciudad de San Lorenzo-Paraguay mediante el uso de modelos potenciales de regresión linealizables bajo la transformación logarítmica. Cabe destacar que antes de hacer uso de estos modelos potenciales, debe tenerse en cuenta que la interpretación de un modelo de regresión lineal, bajo transformación logarítmica en la variable respuesta, debe hacerse en términos de la media geométrica o, equivalentemente, de la mediana, en lugar de en términos de la media aritmética.

Este trabajo multidisciplinario propone un modelo estadístico validado que muestran el comportamiento de la intensidad en superficie de la radiación UV-B total en la ciudad de San Lorenzo-Paraguay



durante los distintos meses del año, además de posibilitar la predicción de esta magnitud en los meses de mayor riesgo y de esta manera adoptar una correcta prevención social pública. Es decir, la utilidad de estos modelos de predicción es la de facilitar el control del comportamiento temporal de la radiación UV-B total a partir de mediciones registradas en la estación LIAPA.

### AGRADECIMIENTOS

A mis padres y mi familia por el apoyo de siempre. En especial a mi abuela Avelina por haberme guiado desde siempre por el largo sendero de la vida y a mi hermana Máxima por darme su apoyo y confianza incondicional para el logro de mis metas y seguir mis ideales.

Al Prof. MSc. Jose Barrera Gómez por su orientación, paciencia y comprensión durante la realización de esta investigación.

Al Laboratorio de Investigaciones de la Atmósfera y Problemas Ambientales por cederme tan amablemente los datos que fueron utilizados para la construcción de la base de datos y posterior análisis descriptivo e inferencial.

### REFERENCIAS BIBLIOGRÁFICAS

- ANTÓN MARTÍNEZ, M. (2007). Modelos empíricos para la estimación de la Irradiancia Solar Ultravioleta. 212 páginas. Tesis Doctoral. Universidad de Extremadura, España.
- CANAVOS, G. C. (1988). Probabilidad y Estadística. Aplicaciones y Métodos. Primera edición en español. McGraw-Hill/Interamericana. México. 651 páginas.
- DÍAZ ESCOBAR, V. I. (2012). Estimación de la Irradiancia Eritémica basada en Modelos de Series Temporales y determinación del Índice UV en el Departamento Central del Paraguay. Tesis de Maestría. Facultad de Ciencias Exactas y Naturales - UNa.
- DUBROVSKY, M. (2000). Analysis of UV-B irradiances measured simultaneously at two stations in the Czech Republic. *Journal Geophysical Research*, Nº D4, Vol. 105. Pp. 4907 - 4913.
- HAIR, J.F., ANDERSON, R. E., TATHAM, R.L. y BLACK, W. C. (1999). Análisis Multivariante. Quinta edición. Prentice Hall Iberia. Madrid. 799 páginas.
- MATEOS VILLÁN, D. D. (2008). Medida, Análisis Climatológico y Modelización de la Radiación Solar Ultravioleta Total en Valladolid. 111 páginas. Tesis de Maestría. Universidad de Valladolid, España.
- MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. (2002). Introducción al análisis de regresión lineal. 1ra. Edición en Español. Compañía Editorial Continental. México. 588 páginas.
- MURILLO, W.; CAÑADA, J. & PEDRÓS, G. (2003). Correlation between global irradiation in Valencia and Córdoba (Spain). *Inglaterra Renewable Energy*, issue 3, Vol. 28. Pp. 409 - 418.
- ROMÁN DIEZ, R. (2010). Efectos de los Componentes Atmosféricos sobre la Radiación Eritémica y Modelo propuesto para la reconstrucción de series. 111 páginas. Tesis de Maestría. Universidad de Valladolid, España.
- SPSS Inc. Released 2002. SPSS for Windows, Version 11.0. Chicago: SPSS Inc.
- VALLEJO DELGADO, L. (2003). Índice Ultravioleta. Departamento de Física. 72 páginas. Universidad de Antofagasta, Chile.
- VARO MARTÍNEZ, M. M. (2006). Modelización de la Radiación Ultravioleta Solar. 329 páginas. Tesis Doctoral. Universidad de Córdoba, España.
- COCHRANE, D., & ORCUTT, G. H. (1949). Application of least squares regression to relationships containing auto-correlated error terms. *Journal of the American Statistical Association*, 44(245), 32-61.