

SUAVIZADO DE CURVAS MEDIANTE B-SPLINE PARA EL ANÁLISIS FUNCIONAL DE LA RADIACIÓN SOLAR GLOBAL

B-Spline Curve Smoothing In The Analysis Of Global Solar Radiation

¹Amalia Escudero*, ¹Silvia Haro

¹Escuela Superior Politécnica de Chimborazo, Escuelas de Ciencias Químicas y Física Matemática, Riobamba – Ecuador

*aescudero@epoch.edu.ec

R esumen

En este trabajo se analizaron las variaciones temporales de la radiación solar global en una localidad ecuatorial Alto Andina como elementos de un espacio de Hilbert. La base de datos corresponde a las medias registradas cada 10 minutos, formando así una matriz de orden 334 x 144 con 334 datos funcionales. Se empleó la regresión mínimo-cuadrática con base B-spline y secciones de polinomios de grado tres, mediante lo cual se identificó un óptimo suavizado con nueve funciones básicas definidas por siete nodos igualmente espaciados, para los 11 meses de las ocho estaciones meteorológicas: Alao, Cumandá, EsPOCH, Multitud, Quimiag, San Juan, Tunshi y Urbina. Se realizó la detección de funciones atípicas y bandas de confianza con 1000 remuestras Bootstrap, posteriormente se obtuvieron las curvas medias suavizadas. El estudio se realizó con las librerías fda y fda.usc del *software* libre R. Las gráficas de las curvas funcionales facilitan la comprensión del comportamiento de la radiación solar global a través del tiempo.

Palabras claves: análisis funcional, bandas de confianza, B-spline cúbico.

A bstract

In this work analyzed the temporal variations of the global solar radiation in an equatorial Alto Andina locality as elements of a Hilbert space. The database corresponds to the means recorded every 10 minutes, thus forming a matrix of order 334 x 144 with 334 functional data. Minimum-quadratic regression with B-spline base and sections of three-degree polynomials were used, by which an optimal smoothing was identified with 9 basic functions defined by 7 equally spaced nodes for the 11 months of the 8 meteorological stations: Alao, Cumandá, EsPOCH, Multitud, Quimiag, San Juan, Tunshi and Urbina. The detection of atypical functions and confidence bands was carried out with 1000 Bootstrap samples, then the smoothed curves of the means were obtained. The study was carried out with the fda and fda.usc libraries of the free software R. The graphs of the functional curves facilitate the understanding of the behavior of global solar radiation over time

Keywords: functional analysis, confidence bands, B-spline cubic.

Fecha de recepción: 09-06 -2017

Fecha de aceptación: 10-12-2017

INTRODUCCIÓN

La radiación solar incidente es la principal fuente de energía renovable en la tierra, da inicio a la cadena energética global, y es influyente en el cambio climático, por lo que es un parámetro pri-

mordial para la construcción de modelos y simulaciones en diversos campos de las ciencias, ingenierías, dimensionamiento de sistemas energéticos (1) estimaciones meteorológicas, efectos globales como la evapotranspiración, entre otros. El objetivo de esta investigación es suavizar las curvas diarias de radiación solar global (RSG) de las ocho estaciones meteorológicas (Alao, Cu-

mandá, Epoch, Multitud, Quimiag, San Juan, Tunshi y Urbina) con el fin de obtener la base funcional de los datos y determinar su comportamiento.

El método más usado en investigaciones tanto teóricas como prácticas en el tratamiento de grandes masas de datos es el método funcional (2). Siendo la radiación solar una variable que se ajusta a las condiciones de los espacios de Hilbert fácilmente tratables mediante el método funcional, por ello se dice que, un conjunto de datos funcionales de RSG $\{X_1, X_2, \dots, X_n\}$ es la observación de n variables funcionales X_1, X_2, \dots, X_n idénticamente distribuidas.

Un B-spline es una función polinomial a trozos de grado $k-1$, definida por una secuencia de n puntos de control llamados “nodos” con los cuales los valores de la función son controlados y en donde k es el orden de la más alta derivada. A la secuencia de valores de los n nodos se les llama vector nodal; tales vectores determinan las transiciones entre los polinomios a trozos. La utilización de este tipo de *splines* se ha utilizado en diferentes contextos, como son recuperación de funciones de energía para estructuras moleculares, evitar colisiones entre robots y en una extensa variedad de aplicaciones. Los B-splines cúbicos son los utilizados, pues tienen derivadas en los puntos finales y además por la necesidad de contar con la segunda derivada definida en estos.

Se realizó la detección de funciones atípicas y bandas de confianza con 1000 remuestras Bootstrap, las cuales son un conjunto de curvas que siguen la misma distribución que los datos originales cuya distancia funcional $d(T, T^{*j})$ es inferior al nivel de significancia α , posteriormente se obtuvieron las curvas suavizadas de las medias mensuales y por cada estación, presentado un comportamiento acompañado con mayor RSG aproximadamente de 11:00 a 13:00 en promedio, siendo además en estas horas la banda de confianza más amplia. Las funciones identificadas como atípicas se caracterizan por un comportamiento de subidas y bajas bruscas de intensidad de RSG. El estudio se realizó con las librerías *fda* y *fda.usc* del *software* libre R.

METODOLOGIA

El análisis descriptivo de datos funcionales es aquella parte de la estadística que trabaja con muestras de funciones aleatorias. El espacio más utilizado en este tipo de datos es el espacio de Hilbert $L^2(T)$ (cuadrado integrable) en el intervalo real T que está definido por:

$$L^2(T) = \left\{ f: T \rightarrow \mathbb{R}: \int_T f^2(t)dt < \infty \right\}$$

con el producto escalar dado por

$$(f, g) = \int_T f(t)g(t)dt, \quad \forall f, g \in L^2(T)$$

Para muestras de datos funcionales, se han introducido las medidas de tendencia central, dispersión y asociación. Una variable aleatoria X se dice que es una variable funcional si toma valores en un espacio funcional $L^2(T)$ cuyos valores dependen de una magnitud continua como el tiempo. Antes de realizar el análisis descriptivo de los datos se debe realizar el suavizado de las curvas muestrales $x_1(t), x_2(t), \dots, x_n(t)$, mediante el método que mejor se ajuste a las características de los datos, estos pueden ser mediante base B-splines, Fourier, Wavelets, potencias y exponenciales, polinomiales, funciones constantes, entre otras. En este estudio se usó regresión mínimo-cuadrática con base B-splines, debido a que las funciones de RSG diarias no son periódicas, a más de que puede igualar e incluso superar la eficiencia computacional de la base Fourier. Los B-splines son trozos de polinomios de grado p conectados entre sí (3).

En general, un B-spline de grado p tiene las siguientes características:

- Consiste en $p + 1$ trozos de polinomios de orden p que se unen en p nodos internos.
- Las derivadas hasta el orden $p-1$ son continuas en los puntos de unión.
- El B-spline es positivo en el dominio expandido por $p + 2$ nodos y 0 en el resto.
- Excepto en los extremos, se solapa con $2p$ trozos de polinomios de sus vecinos.
- Para cada valor de x , $p + 1$ B-splines son no nulos.
- Para construir un B-spline de grado p son necesarios $p + 2$ nodos.

Los B-splines no “padecen” de los efectos frontera comunes en otros métodos de suavizado, como en algunos suavizados tipo núcleo, en los que al extender la curva ajustada fuera del dominio de los datos ésta tiende hacia cero.

Si se denota por $t_0 < t_1 < \dots < t_m$ una partición de nodos del intervalo de observación T. Extendiendo la partición como $t_{-3} < t_{-2} < t_{-1} < t_0 < \dots < t_m < t_{m+1} < t_{m+2} < t_{m+3}$, la base B-splines de orden $p+1$ de grado p se define de la siguiente forma:

$$B_{j,1}(t) = \begin{cases} 1 & t_{j-2} \leq t < t_{j-1} \\ 0 & \text{en otro caso} \end{cases}$$

con $j=-1,0,1,\dots,m+4$

$$B_{j,p}(t) = \frac{t - t_{j-2}}{t_{j+p-2} - t_{j-2}} B_{j,p}(t) + \frac{t_{j+p-1} - t}{t_{j+p-1} - t_{j-1}} B_{j+1,p}(t)$$

$$p=1,2,\dots; \quad j=-1,0,\dots,m-p+4.$$

Cuando $p = 4$ estas funciones base son B-splines cúbicos, utilizados para la aproximación de curvas muestrales regulares. A partir de ahora, se omite el subíndice correspondiente al orden, de manera que los B-splines cúbicos se denotarán como $B_{j,4}(t) = B_j(t)$, con $j = -1,0,\dots,m+1$.

La base de los B-splines de grado p genera el espacio de los splines del mismo grado definidos como curvas formadas por polinomios a trozos de grado p con continuidad en las derivadas hasta el orden $p-1$ (3).

Para hallar la representación básica funcional, se tomaron los datos de RSG diaria, de 06:00 a 18:00 de cada estación durante los meses estudiados, formando una matriz de dimensión (12,334), se realizó un suavizado de las 334 curvas diarias de RSG mediante aproximación mínimo cuadrática con bases de B-splines cúbicos ($p = 4$), mediante simulaciones se logró un mejor ajuste con nueve funciones básicas definidas por siete nodos igualmente espaciados en el período [1,12] y

finalmente se graficaron las representaciones básicas con B-splines (*splines* de regresión) de todos los datos.

La identificación de funciones atípicas y bandas de confianza se realizó con el remuestreo de Bootstrap que permite resolver problemas relacionados con la prueba de significación estadística. Consiste en la creación un gran número de muestras con reposición de los datos observados, con el fin de aproximar la distribución de muestreo de un estadístico y de sus propiedades (4). Pueden cubrirse diferentes objetivos de análisis de datos como: desarrollo de contraste de hipótesis, construir bandas de confianza, detectar datos atípicos, entre otros (5). En este trabajo se usó para la detección de datos atípicos y para estimar las bandas de confianza para las funciones medias de RSG.

Un conjunto de confianza para el funcional $X(t)$ con nivel α , se define como el conjunto de curvas c las cuales tienen la misma distribución que $X(t)$, tal que

$$CS(x) = \{c: d(x,c) < D_\alpha\} \quad (14)$$

donde d es una distancia funcional y D_α es tal que $P(CS(x)) = \alpha$. Esto es, dada una muestra $X_1(t), X_2(t), \dots, X_n(t)$ se extraen muestras Bootstrap denotadas por $X_1^*(t), X_2^*(t), \dots, X_n^*(t)$. Entonces un conjunto del $\alpha\%$ de confianza basado en el estadístico $T(X_1(t), X_2(t), \dots, X_n(t))$ es definido calculando el valor de D_α tal que el $\alpha\%$ replicas $T(X_1(t), X_2(t), \dots, X_n(t))$ se localicen dentro de una distancia menor que D_α de su promedio, es decir D_α es el radio del conjunto de confianza Bootstrap centrado en la media (medida de profundidad) de T.

Para medir la distancia D_α , se usó la métrica de L^2 la cual para una curva $x(t)$ está dada por:

$$\|X(t)\|_2 = \left(\int X^2(t) dt \right)^{1/2}$$

(6, 7).

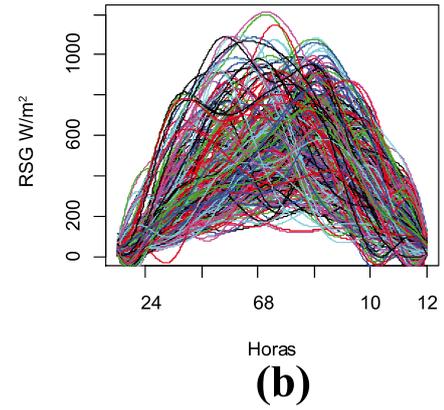
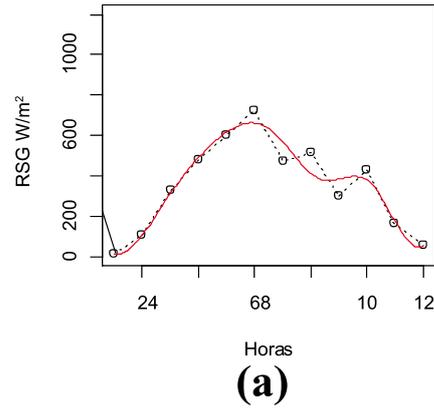
La media funcional se halló mediante:

$$\bar{x}(t) = \frac{1}{n} \sum_{i=1}^n x_i(t)$$

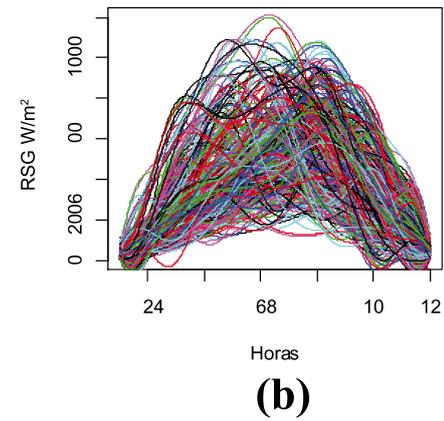
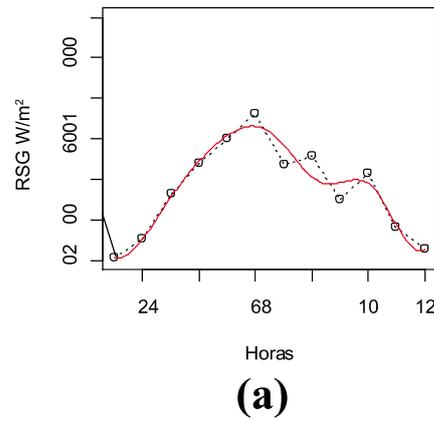
RESULTADOS Y DISCUSIÓN

Con los B-splines se obtuvieron las curvas suavizadas de los 344 días del año de cada una de las estaciones meteorológicas, mostrando así curvas suaves que facilitan la identificación de patrones en su comportamiento (fig. 1 y 2).

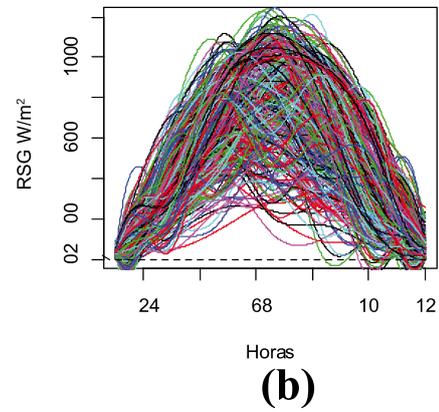
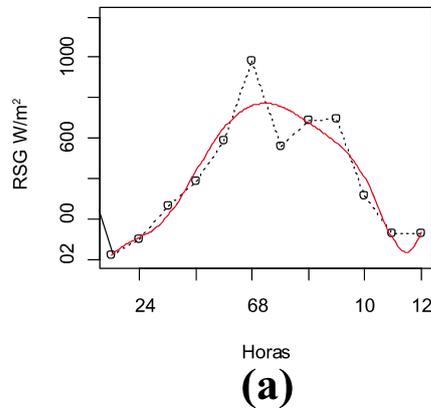
Alao



Cumandá



Espoch



Multitud

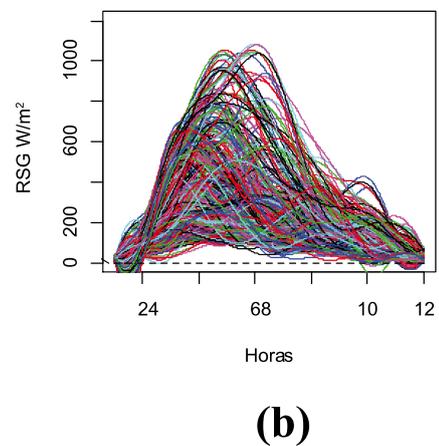
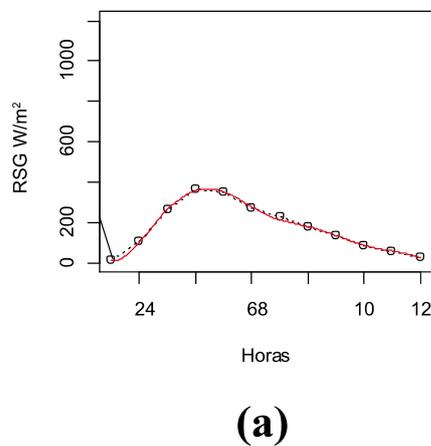


Fig. 1. Representación básica con B-spline cúbico: (a) suavizado para un día (curva roja) y (b) suavizado para los 334 días (curvas de colores) de las estaciones Alao, Cumandá, Espoch y Multitud

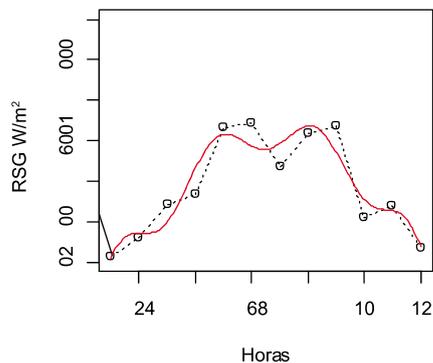
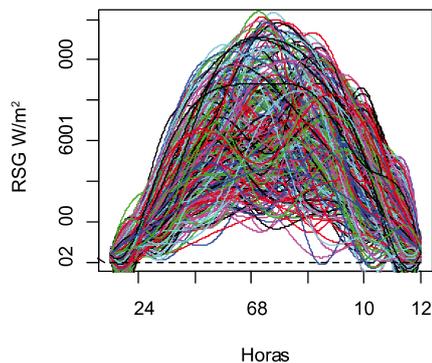
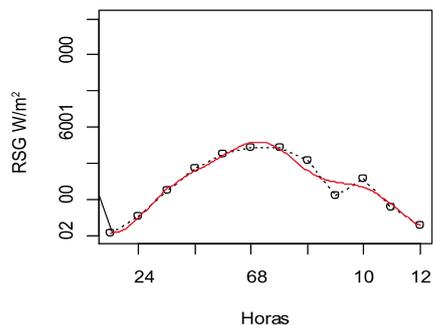
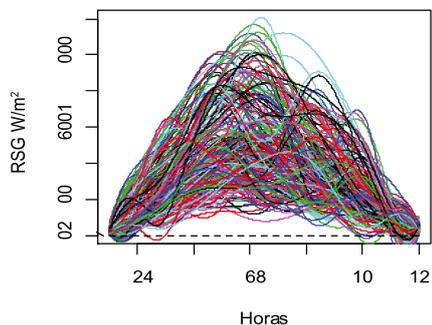
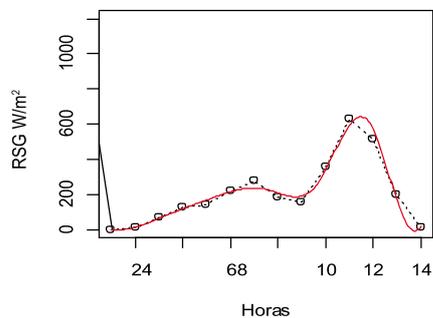
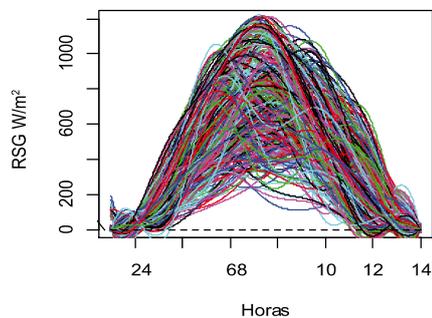
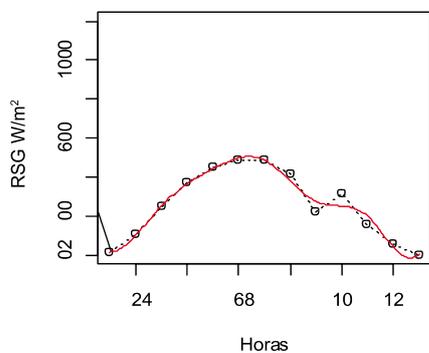
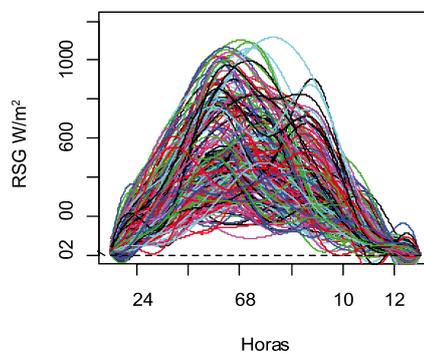
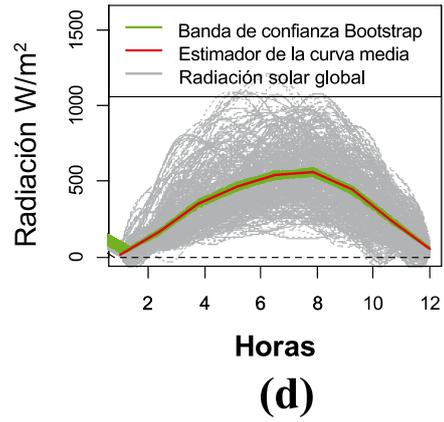
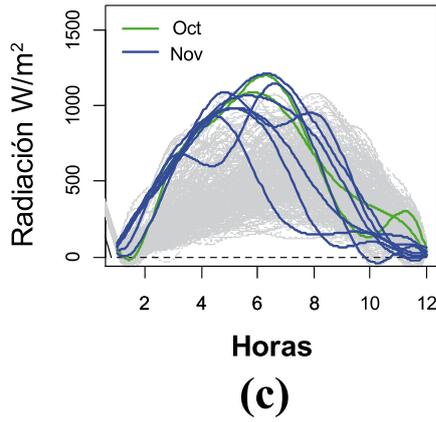
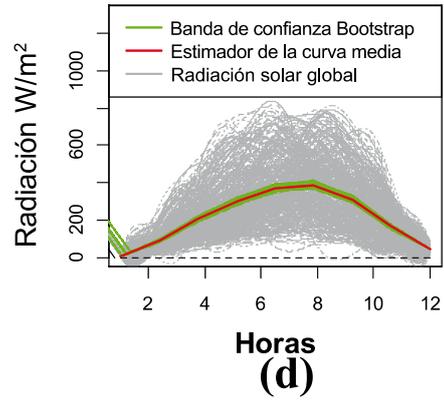
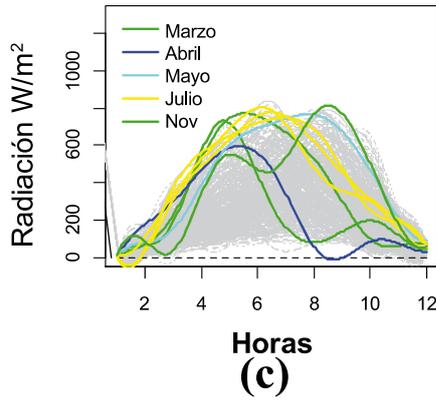
Quimiag**(a)****(b)****San Juan****(a)****(b)****Tunshi****(a)****(b)****Urbina****(a)****(b)**

Fig. 2. Representación básica con B-spline cúbico: (a) suavizado para un día (curva roja) y (b) suavizado para los 334 días (curvas de colores) de las estaciones Quimiag, San Juan, Tunshi y Urbina

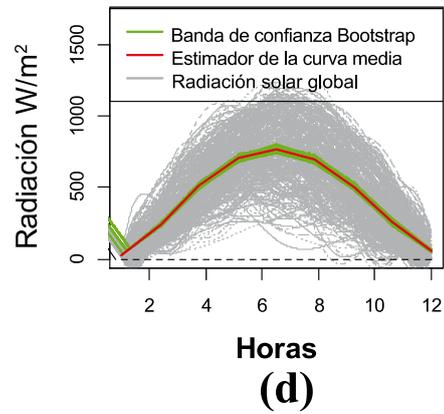
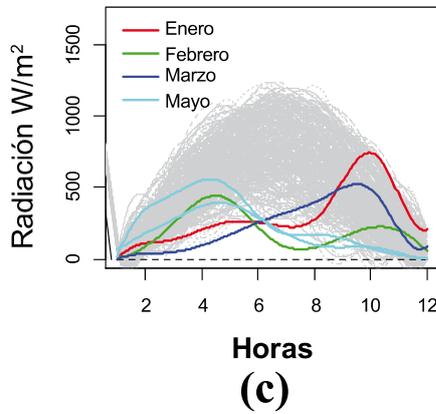
Alao



Cumandá



Espoch



Multitud

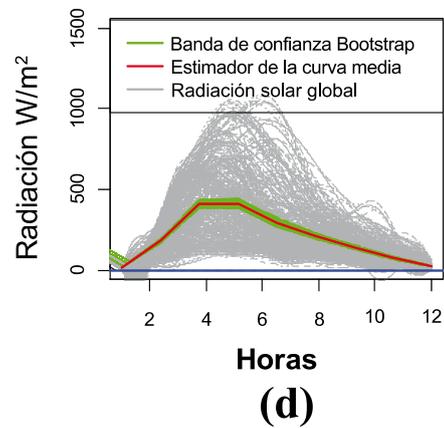
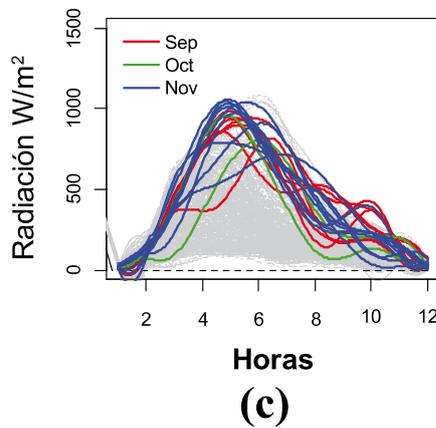


Fig. 3. Curvas atípicas (c) diferenciadas cada mes por color, y estimación de curvas medias mensuales por cada estación con su banda de confianza (d), estaciones de Alao, Cumandá, Espoch y Multitud

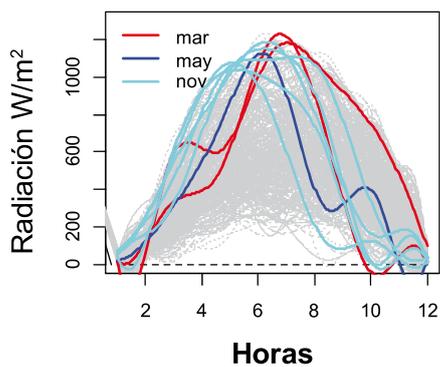
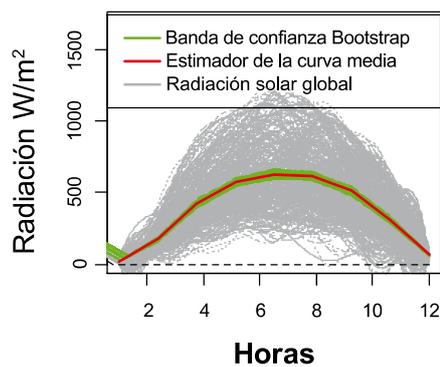
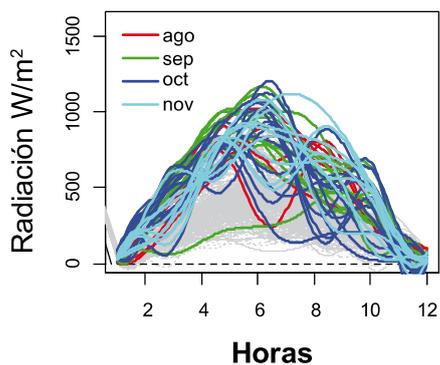
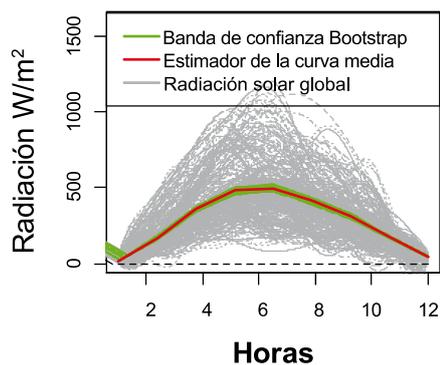
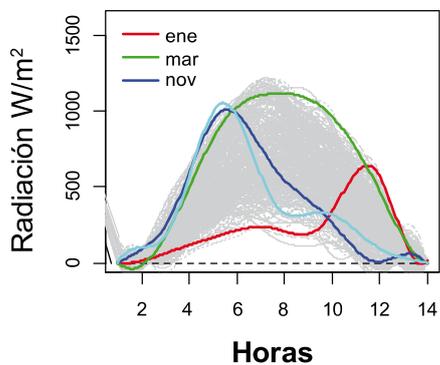
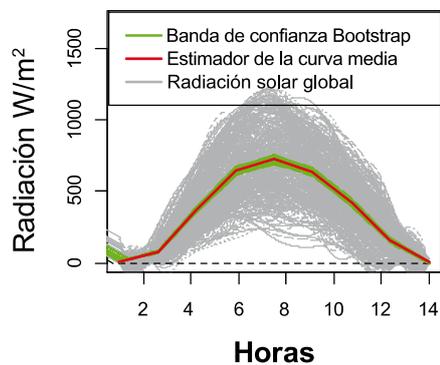
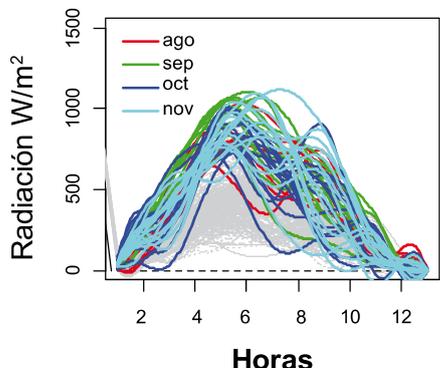
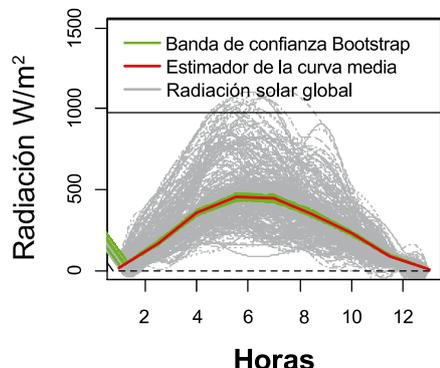
Quimiag**(c)****(d)****San Juan****(c)****(d)****Tunshi****(c)****(d)****Urbina****(c)****(d)**

Fig. 4. Curvas atípicas (c) diferenciadas cada mes por color. Y estimación de curvas medias mensuales por cada estación con su banda de confianza (d), estaciones Quimiag, San Juan, Tunshi y Urbina

Mediante la función `outliers.depth.trim` de R basado en medidas de profundidad y remuestreo de Bootstrap, se identificaron curvas atípicas, las cuales se muestran en cada estación diferenciadas mensualmente por color (fig. 3 y 4, literales c), las estaciones con mayor número de curvas atípicas fueron San Juan y Urbina, muchas de estas curvas coinciden con las fechas en las que se realizó mantenimiento de los sensores, motivo por el cual se las separó del análisis, dando lugar a la reducción del sesgo en los resultados. En general, la RSG diaria muestra un comportamiento acampanado, característica aproximadamente similar en todas las estaciones. La estación de Multitud es la única con muy notoria asimetría positiva. La banda de confianza de la curva media (fig. 3 y 4, literales d) muestra aumento en el ancho de banda aproximadamente de 10:00 a 15:00 que corresponden a las horas con mayor RSG diaria y alcanzaron hasta 1200 W/m^2 . Se halló funciones medias diarias hasta 900 W/m^2 en las estaciones Quimiag, Epoch y Tunshi; de 700 W/m^2 en las estaciones de Alao, Multitud, San Juan y Urbina; y de 600 W/m^2 en la estación de Cumandá, valores muy asociados a la posición geográfica en las que se encuentran ubicadas las estaciones.

CONCLUSIONES

Se realizó el estudio de la RSG diaria de las ocho estaciones meteorológicas, calculando los promedios generales

de cada una con su respectiva banda de confianza.

Las curvas B-splines suavizan los datos originales con nueve funciones básicas definidas por siete nodos igualmente espaciados en el período [1,12], mostrando así curvas sin picos que reducen la variabilidad del comportamiento de la RSG diaria, al mismo tiempo que facilitan la comprensión de dicha variable a través del tiempo.

El comportamiento diario medio, las bandas de confianza y las curvas atípicas de cada estación tienen características propias debido a la zona geográfica en las que están ubicadas, dando lugar a la presencia de microclimas en la provincia de Chimborazo, localidad ecuatorial Alto Andina.

AGRADECIMIENTO

Centro de Energías Alternativas y Ambiente (CEAA), Grupo Facultad de Ciencias.

Referencias

1. Monné C, Alonso S, Palacín F. Evaluación de una instalación de refrigeración por absorción con energía solar. Información Tecnológica [Internet]. 2011 [citado 04 de abril del 2017]; 22(3): 39-44. Disponible en: [10.4067/50718-07642011000300006](http://dx.doi.org/10.4067/50718-07642011000300006).
2. Aguilera A, Aguilera C. Introducción al análisis de datos funcionales con R. III Jornadas para la Enseñanza y Aprendizaje de la Estadística y de la Investigación Operativa [sidertación]. Cádiz: Dpto. Estadística Investigación Operativa Universidad de Granada. 2013.
3. Aguilera M. Estimación penalizada con datos funcionales. [tesis de maestría]. Granada: Universidad de Granada; 2009.
4. Ledesma, R. Introducción al Bootstrap. Desarrollo de un ejemplo acompañado de *software* de aplicación. Tutor Quan Methods Psychol, 4(2): 51-60.
5. Tarrío J. Aportaciones a la clasificación de curvas con datos funcionales y a la modelización del crecimiento de grietas por fatiga. Aplicaciones a la ciencia de los materiales. [Internet]. 2010 [citado 03 de agosto de 2016]; Disponible en : http://eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto_394.pdf.
6. Rivera, D. Cartas de Control para datos funcionales. [Internet]. 2011 [citado el 01 de enero 2016]; disponible en: http://probayestadistica.cimat.mx/sites/default/files/PDFs/TE446_RiveraGarcia.pdf.
7. Escudero A, Recalde C, Haro S, Meneses M. 2016. Spline cúbico para el tratamiento funcional de la radiación solar global. Información tecnológica. 27(2), 153-162.

Bibliografía

- Antonanzas F, Sodupe E, Sanz A, Fernandez R, Martinez F. Technical feasibility assessment of integrated solar combined cycle power plants in Ciudad Real (Spain) and Las Vegas (USA). In Proc. 16th International Congress on Project Engineering. Valencia, Spain. 2012; 18 (2).
- Bulut H, Büyükalaca O. 2007. Simple model for the generation of daily global solar-radiation data in Turkey. *Applied Energy*. 84(5), 477-491.
- Corach G, Andrucho E. Análisis Funcional. En C. G. E.. Buenos Aires, Argentina: Universidad de Buenos Aires. 1997.
- Febrero B, Oviedo M. 2012. Statistical computing in functional data analysis: The R package fda.usc. *Journal of Statistical Software*. 51(4), 28.
- Freiman R, Munís G. 2001. Trimmed means for functional data. *Sociedad de Estadística e Investigación Operativa*. 10(2), 7-13.