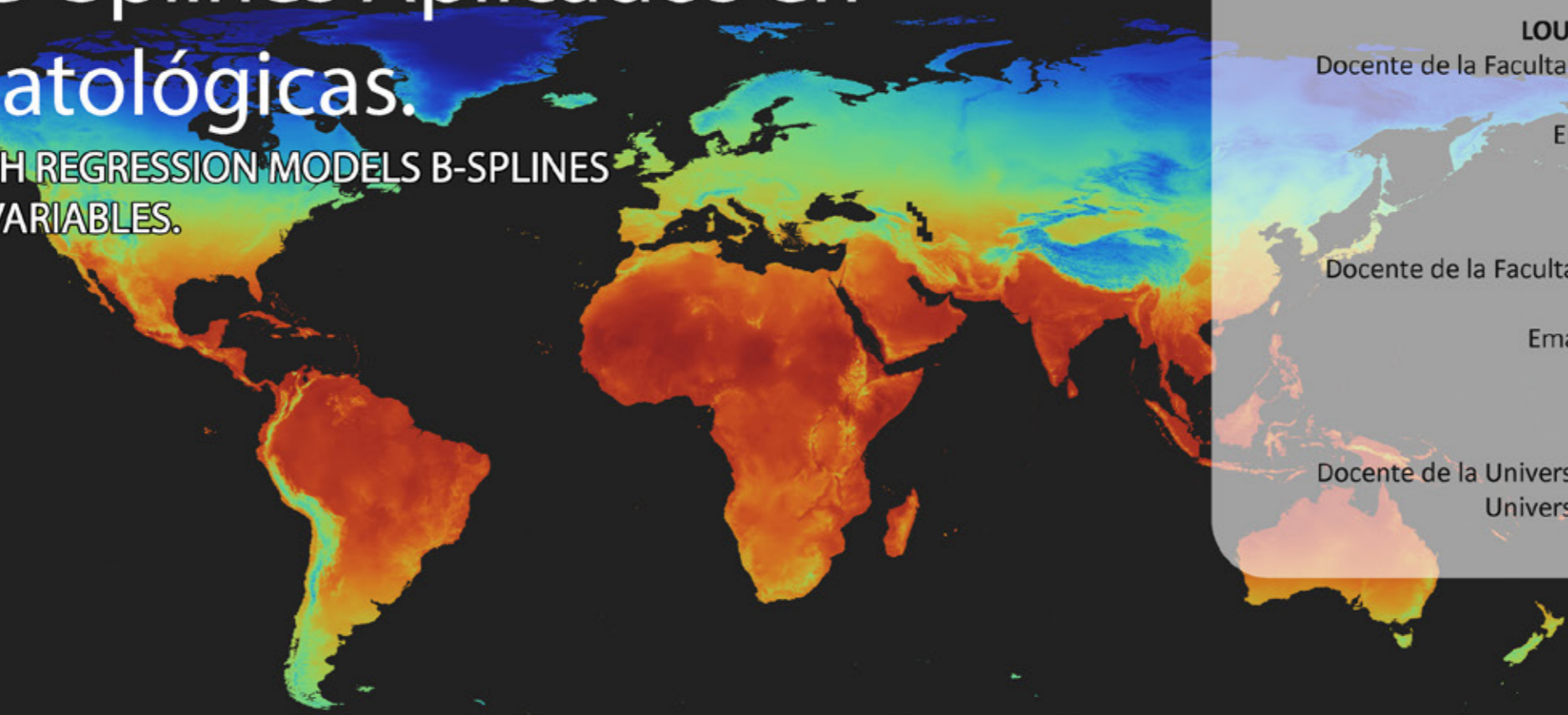


# MÉTODO DE AJUSTE CON MODELOS de Regresión B-Splines Aplicados en Variables Climatológicas.

METHOD OF ADJUSTMENT WITH REGRESSION MODELS B-SPLINES APPLIED IN CLIMATOLOGICAL VARIABLES.



**NANCY CHARIGUAMÁN**

Docente de la Escuela Superior Politécnica de Chimborazo  
Ingeniera en Estadística Informática  
Email: nchariguamán@esPOCH.edu.ec

**LOURDES DEL CARMEN - ZÚÑIGA**

Docente de la Facultad de Informática y electrónica  
ESPOCH  
Email: lzuñiga@esPOCH.edu.ec.

**MARIO PAGUAY - CUVI**

Docente de la Facultad de Informática y electrónica  
ESPOCH  
Email: mpaguay@esPOCH.edu.ec.

**ANTONIO MENESES**

Docente de la Universidad Nacional de Chimborazo  
Universidad Nacional de Chimborazo

## RESUMEN

La presente investigación desarrolla una nueva alternativa de ajuste de modelos de regresión flexibles mediante B-splines de variables climatológicas: temperatura, humedad, radiación y velocidad del viento en los meses de enero y junio que representan la época lluviosa y seca del año, respectivamente en la ciudad de Riobamba - Ecuador. Estos modelos presentan excelentes propiedades de ajuste, hecho que es verificado al observar que las longitudes de los intervalos de confianza puntuales bootstrap son más pequeñas en los modelos B-splines que en los polinómicos comúnmente utilizados en este ámbito. Además esta nueva alternativa es más precisa de acuerdo al menor error cuadrático medio (ECM) obtenido de ajustes de los dos modelos antes mencionados. Por otra parte uno de los objetivos de esta investigación es analizar las variaciones de las variables climatológicas de forma minuciosa para fines benéficos en el cuidado del medio ambiente, la flexibilidad en el ajuste de estos modelos propuestos persiguen lo antes mencionado. El tratamiento de datos y resultados se realiza con el software estadístico R.

**PALABRAS CLAVE:** variables climatológicas, B-splines, polinómicos.

## ABSTRACT

The present research develops a new alternative of adjustment of flexible regression models using B-splines of climatological variables: temperature, humidity, radiation and wind speed in the months of January and June that represent the rainy and dry season of the year, respectively. In the city of Riobamba - Ecuador. These models have excellent fit properties, which are verified by observing that the lengths of the bootstrap point confidence intervals are smaller in the B-splines than in the polynomials commonly used in this field. In addition, this new alternative is more precise according to the smaller mean square error obtained from adjustments of the two models mentioned above. On the other hand one of the objectives of this research is to analyze the variations of the climatological variables in a meticulous way for charitable purposes in the care of the environment, the flexibility in the adjustment of these proposed models pursue the aforementioned. The treatment of data and results is performed with the statistical software R.

**KEYWORDS:** Climatological variables, B-splines, polynomials.

Fecha de Ingreso: Julio 2017  
Fecha de Publicación: Agosto 2017

## INTRODUCCIÓN

Los modelos de regresión B-Splines son de tipo semiparamétrico y presentan un excelente ajuste en variables climatológicas, como se muestra en el desarrollo de este trabajo que está dividido en las siguientes partes:

Materiales y métodos, la base de datos es obtenida de la estación meteorológica que se encuentra en la Facultad de Ingeniería de la Universidad Nacional de Chimborazo.

Modelos de regresión B-splines, se presenta el modelo de regresión mediante una función en forma de una combinación lineal de polinomios cúbicos de una base B-splines más una variable aleatoria de media 0, este modelo es de tipo semi paramétrico [3, 4].

Regresión polinómica, este modelo es de tipo paramétrico es decir, su forma funcional es un polinomio que queda determinado al estimar sus parámetros mediante el método de mínimos cuadrados [5].

Intervalos de confianza puntuales bootstrap con modelos de regresión B-splines y polinómico, se desarrolla un algoritmo para calcular estos intervalos utilizando herramientas del software estadístico R [1, 2, 6].

Resultados y discusión, donde presentan gráficas de ajustes con intervalos de confianza puntuales bootstrap al 95% de confianza de los modelos B-splines y polinómicos en los meses de enero y junio, además tablas de resultados de errores cuadráticos medios de los valores de las variables climatológicas con los ajustes de los dos modelos antes mencionados, utilizando librerías del software estadístico R [6]. Por último se tiene las conclusiones de la investigación.

## MATERIALES Y MÉTODOS.

La base de datos fue obtenida de la estación meteorológica que se encuentra en la Facultad de Ingeniería de la Universidad Nacional de Chimborazo, la misma que contiene observaciones correspondientes al año 2015 de las variables climatológicas: temperatura (°C), humedad (%), radiación (Watt/m<sup>2</sup>) y velocidad del viento (m/s). A partir de estos datos, se calculan los valores en promedio en las horas de cada día y en cada mes de las variables climatológicas.

## MODELOS DE REGRESIÓN B-SPLINES

Estos modelos de regresión permiten establecer la relación entre la co variable x de diseño fijo y la respuesta Y de acuerdo a la siguiente relación matemática:

$$Y = f(x) + \varepsilon$$

donde  $\varepsilon$  es la variable aleatoria de media 0 y  $f(x)$  es una relación que tiene la siguiente expresión:

$$f(x) = a_1 B_1(x) + \dots + a_m B_m(x)$$

de esta relación,

- $m$  indica el número de funciones de la base B-splines
- $a_1, \dots, a_m$  son parámetros desconocidos que se estiman a partir de los datos
- $B_1, \dots, B_m$  son funciones conocidas de la base B-splines

Esto significa que la regresión por B-splines, reduce un problema de regresión no paramétrica a un problema paramétrico, por lo tanto estos modelos de regresión se consideran semiparamétricos [3,4].

## REGRESIÓN POLINÓMICA

Estos modelos son una generalización de los modelos de regresión lineal, y describen el comportamiento en promedio de la variable respuesta Y, condicionada por los valores de la variable independiente x de diseño fijo, cuya representación funcional polinómica es [5]:

$$Y = f(x) + \varepsilon$$

$$Y = a_0 + a_1 x + \dots + a_p x^p + \varepsilon$$

donde  $\varepsilon$  es la variable aleatoria de media 0 y  $a_0, \dots, a_p$  son parámetros que se estiman con el método de mínimos cuadrados. Es decir, de la muestra observada:

$$(x_1, Y_1), \dots, (x_n, Y_n)$$

las estimaciones  $\hat{a}_0, \dots, \hat{a}_p$  se obtienen minimizando la siguiente suma de residuos al cuadrado:

$$(Y_1 - \hat{Y}_1)^2 + \dots + (Y_n - \hat{Y}_n)^2$$

siendo:

$$\hat{Y}_i = \hat{a}_0 + \hat{a}_1 x_i + \dots + \hat{a}_p x_i^p$$

Para la regresión por B-splines y Polinómica, Y es la variable respuesta que en este caso representa a la variable climatológica y x es la variable de diseño fijo que representa la hora-día.

**INTERVALOS DE CONFIANZA PUNTUALES BOOTSTRAP CON MODELOS DE REGRESIÓN B-SPLINES Y POLINÓMICO**

Los intervalos de confianza bootstrap en cada ordenada estimada  $\hat{Y}_i$  mediante el modelo de regresión polinómico o B-splines, son calculados con el siguiente algoritmo [1,2]:

1. Calcular la estimación de la varianza residual  $\hat{\sigma}^2$  de acuerdo a:

$$\varepsilon_i = Y_i - \hat{Y}_i \text{ con } i = 1, 2, \dots, n$$

donde  $(Y_i) = f(x_i)$  está dada por la expresión polinómica o por B-splines,

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2}{n - 1}$$

siendo  $\bar{\varepsilon}$  la media de los  $\varepsilon_i$ ,  $i = 1, 2, \dots, n$ .

2. Generar un número B de muestras bootstrap que imitan la muestra original de la siguiente manera:

donde  $Y_i^* = f(x_i) + \varepsilon_i^*$  para

$$i = 1, 2, \dots, n, \text{ con } \varepsilon_i^* \sim N(0, \hat{\sigma}^2)$$

-Para cada valor de  $x_i$ ,  $i = 1, 2, \dots, n$  el intervalo de confianza bootstrap es calculado con los siguientes pasos:

-Las B estimaciones bootstrap se ordenan de forma creciente en cada  $x_i$ ,

$$\hat{Y}_i^{*(1)} \leq \dots \leq \hat{Y}_i^{*(B)}$$

-Los valores formados por los cuantiles  $\alpha/2$  y  $1-\alpha/2$  de  $Y_i^{*(b)}$ ,  $b=1, \dots, B$  en cada  $x_i$  son los límites inferior y superior del intervalo de confianza con nivel de significancia  $\alpha$ .

**ERROR CUADRÁTICO MEDIO (ECM)**

La medida de precisión entre los valores observados de la variable climatológica y los ajustes de los modelos se define mediante el error cuadrático medio [5],

$$ECM = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}$$

donde n es el tamaño de la muestra observada (n=24),  $Y_i$  valores observados e  $\hat{Y}_i$  valores estimados por los modelos B-splines o polinómico.

**RESULTADOS Y DISCUSIÓN**

Ajustes de las variables climatológicas con modelos de regresión B-splines y polinómicos en la época lluviosa (mes de enero).

Modelos de regresión en R,

B-splines: `gam(Y ~s( x))`

polinómico de grado 7: `lm(Y ~poly( x,7))`

**Variable climatológica temperatura**

•x: 24 horas del día.

•Y: Temperatura promedio en cada hora de cada día del mes de enero.

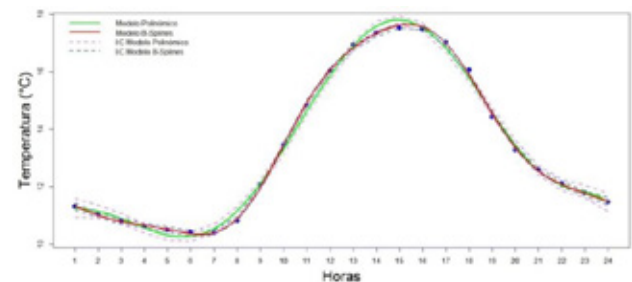


Figura 1: Ajuste de los modelos de regresión polinómico y B-splines de la variable climatológica temperatura versus horas (mes de enero) e intervalos de confianza puntuales bootstrap al 95% de confianza.

Modelo de ajuste	ECM
B-splines	0.0049
Polinómico de grado 7	0.0262

Tabla 1: Precisión entre los valores observados de la temperatura (mes de enero) con los ajustes de los modelos de regresión B-splines y Polinómico de grado 7, mediante el error cuadrático medio

**Variable climatológica radiación**

Modelos de regresión B-splines y polinómico de grado 7:

- x: 24 horas del día.
- Y: Radiación promedio en cada hora de cada día del mes de enero.

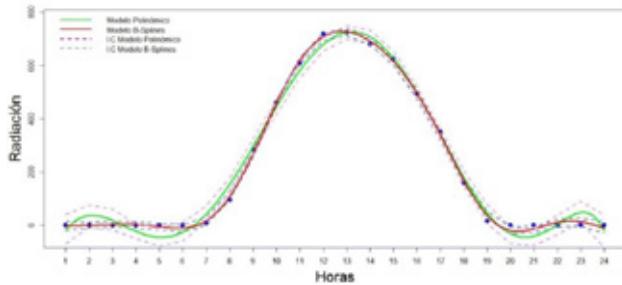


Figura 2: Ajuste de los modelos de regresión polinómico y B-splines de la variable climatológica radiación versus horas (mes de enero) e intervalos de confianza puntuales bootstrap al 95% de confianza.

Modelo de ajuste	ECM
B-splines	121.7585
Polinómico de grado 7	858.6313

Tabla 2: Precisión entre los valores observados de la radiación (mes de enero) con los ajustes de los modelos de regresión B-splines y Polinómico de grado 7, mediante el error cuadrático medio

**Variable climatológica humedad**

Modelos de regresión B-splines y polinómico de grado 7:

- x: 24 horas del día.
- Y: Humedad promedio en cada hora de cada día del mes de enero.

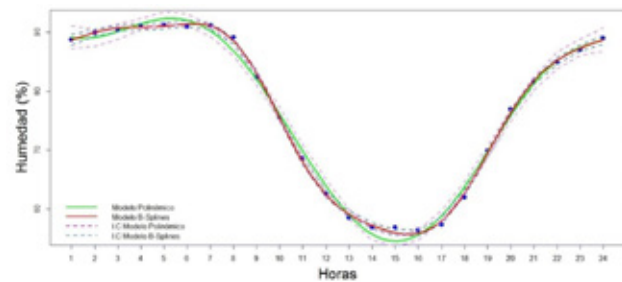


Figura 3: Ajuste de los modelos de regresión polinómico y B-splines de la variable climatológica humedad versus horas (mes de enero) e intervalos de confianza puntuales bootstrap al 95% de confianza.

Modelo de ajuste	ECM
B-splines	0.2078
Polinómico de grado 7	1.1430

Tabla 3: Precisión entre los valores observados de la humedad (mes de enero) con los ajustes de los modelos de regresión B-splines y Polinómico de grado 7, mediante el error cuadrático medio

**Variable climatológica velocidad del viento**

Modelos de regresión B-splines y polinómico de grado 7:

- x: 24 horas del día.
- Y: Velocidad del viento promedio en cada hora de cada día del mes de enero.

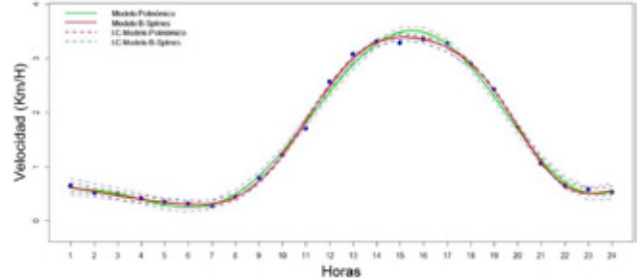


Figura 4: Ajuste de los modelos de regresión

polinómico y B-splines de la variable climatológica velocidad del viento versus horas (mes de enero) e intervalos de confianza puntuales bootstrap al 95% de confianza.

Modelo de ajuste	ECM
B-splines	0.0026
Polinómico de grado 7	0.0080

Tabla 4: Precisión entre los valores observados de la velocidad del viento (mes de enero) con los ajustes de los modelos de regresión B-splines y Polinómico de grado 7, mediante el error cuadrático medio

**Ajustes de las variables climatológicas con modelos de regresión B-splines y polinómicos en la época seca (mes de junio).**

**Variable climatológica temperatura**

Modelos de regresión B-splines y polinómico de grado 7:

- x: 24 horas del día.
- Y: Temperatura promedio en cada hora de cada día del mes de junio.

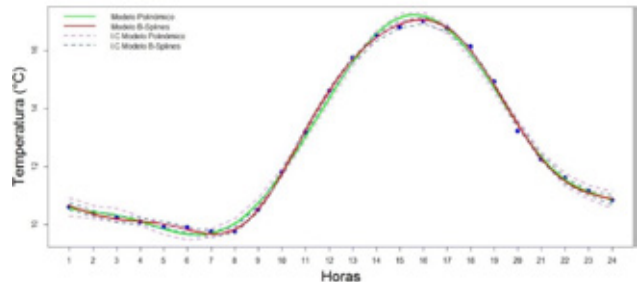


Figura 5: Ajuste de los modelos de regresión polinómico y B-splines de la variable climatológica temperatura versus horas (mes de junio) e intervalos de confianza puntuales bootstrap al 95% de confianza.

Modelo de ajuste	ECM
B-splines	0.0074
Polinómico de grado 7	0.0273

Tabla 5: Precisión entre los valores observados de la temperatura (mes de junio) con los ajustes de los modelos de regresión B-splines y Polinómico de grado 7, mediante el error cuadrático medio

**Variable climatológica radiación**

Modelos de regresión B-splines y polinómico de grado 7:

- x: 24 horas del día.
- Y: Radiación promedio en cada hora de cada día del mes de junio.

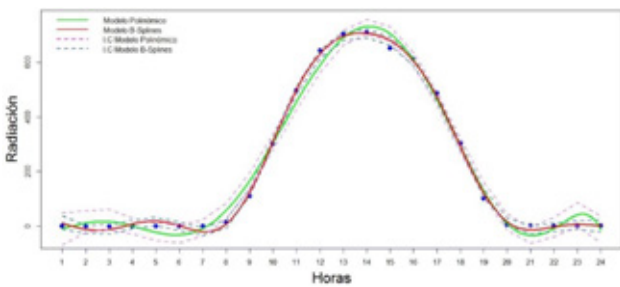


Figura 6: Ajuste de los modelos de regresión polinómico y B-splines de la variable climatológica radiación versus horas (mes de junio) e intervalos de confianza puntuales bootstrap al 95% de confianza.

Modelo de ajuste	ECM
B-splines	160.1293
Polinómico de grado 7	833.4673

Tabla 6: Precisión entre los valores observados de la radiación (mes de junio) con los ajustes de los modelos de regresión B-splines y Polinómico de grado 7, mediante el error cuadrático medio

**Variable climatológica humedad**

Modelos de regresión B-splines y polinómico de grado 7:

- x: 24 horas del día.
- Y: Humedad promedio en cada hora de cada día del mes de junio.

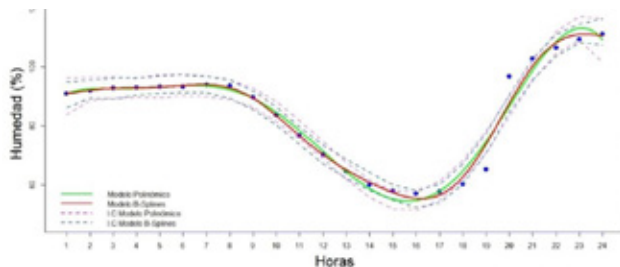


Figura 7: Ajuste de los modelos de regresión polinómico y B-splines de la variable climatológica humedad versus horas (mes de junio) e intervalos de confianza puntuales bootstrap al 95% de confianza.

Modelo de ajuste	ECM
B-splines	7.5731
Polinómico de grado 7	11.2522

Tabla 7: Precisión entre los valores observados de la humedad (mes de junio) con los ajustes de los modelos de regresión B-splines y Polinómico de grado 7, mediante el error cuadrático medio

**Variable climatológica velocidad del viento**

Modelos de regresión B-splines y polinómico de grado 7:

- x: 24 horas del día.
- Y: Velocidad del viento promedio en cada hora de cada día del mes de junio.

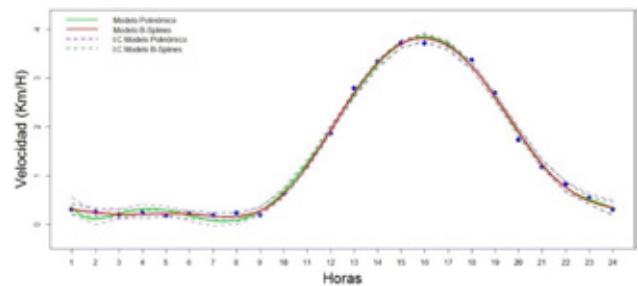


Figura 8: Ajuste de los modelos de regresión polinómico y B-splines de la variable climatológica velocidad del viento versus horas (mes de junio) e intervalos de confianza puntuales bootstrap al 95% de confianza.

Modelo de ajuste	ECM
B-splines	0.0045
Polinómico de grado 7	0.0076

Tabla 8: Precisión entre los valores observados de la velocidad del viento (mes de junio) con los ajustes de los modelos de regresión B-splines y Polinómico de grado 7, mediante el error cuadrático medio

En las figuras de la 1 a la 8 se observan que los ajustes de los modelos B-splines están más próximos a los valores de las variables climatológicas con respecto a los modelos polinómicos, esto también se puede observar en los intervalos de confianza puntuales bootstrap de los modelos B-splines que son de longitud menor a los intervalos de confianza de los modelos polinómicos.

Los errores cuadráticos medios observados en las tablas de la 1 a la 8 son menores en los modelos B-splines a los modelos polinómicos, esto prueba que los ajustes de las variables climatológicas mediante los modelos B-splines son más precisos.

## CONCLUSIONES

El método propuesto por B-Splines es una alternativa de ajuste de las variables climatológicas que presenta adecuadas propiedades que pueden ser aprovechadas para investigar de mejor manera las variaciones medio ambientales; además el modelo de regresión propuesto B-Splines es confiable científicamente y preciso en el ajuste de variables climatológicas en comparación con modelos de regresión polinómicos comúnmente utilizados.

## REFERENCIAS

1. Davison, A.C. and Hinkley, D.V. Bootstrap Methods and their Application. Cambridge University Press; 1997.
2. Efron, B. and Tibshirani, R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* 1; 1986.
3. Gu. RKPAC and its application: Fitting smoothing spline models, Proc. Statistical Computing Section, Amer. Statist. Assoc., pp. 42-51; 1998.
4. Hastie, T.J., Tibshirani, R.J. Generalized Additive Models. Chapman & Hall; 1990.
5. Johnson, R. Probabilidad y estadística para ingenieros. Vol 1. 8a ed. México: Pearson educación; 2012.
6. Rizzo, M.L. Statistical Computing with R. Chapman&Hall/CRC. 1a ed; 2008.