Entre Ciencia e Ingeniería, ISSN 1909-8367
Año 7. No. 14 - Segundo Semestre de 2013, página 32 - 37

Aplicación de la suavización Spline en la modelación de la temperatura promedio mensual del Valle del Cauca usando ponderación por diagramas de Voronoi[1]

Applying the smoothing Spline when modeling the average monthly temperature of Valle del Cauca using weighted Voronoi diagrams

A. J. Florez, J. E. Delgado, M. Mera

Recibido Febrero 19 de 2013 – Aceptado Noviembre 15 de 2013

Resumen **- Entender el comportamiento de algunos fenómenos climatológicos, en especial la temperatura, es de gran importancia para diferentes actividades humanas. Por esta razón el objetivo de este documento es la modelación de la temperatura mensual del Valle de Cauca en el periodo 1971-2002 por medio de suavización Spline ponderada. Para la modelación se tuvieron en cuenta dos estratos (Valle y Montaña), en términos de temperatura, puesto que la región del Valle del Cauca está ubicada en diferentes pisos térmicos que afectan su comportamiento. Además se hizo uso de diagramas de Voronoi para determinar el área de influencia de cada estación meteorológica que se encuentra en el departamento y así asignarle su ponderación para la modelación.**

Palabras clave **- regresión no paramétrica, suavización Spline, diagramas de Voronoi, temperatura.**

Abstract **- Understanding some weather phenomena behavior, especially temperature, is very important to many human activities. For this reason the aim of this work is to model monthly temperature in Valle del Cauca between 1971 and 2002 using Weighted Smoothing Splines. Two stratums (valley and mountain), in terms of temperature were considered during the process, since Valle del Cauca is a region located in different thermal floors, that affect its behavior. A voronoi diagram was used to determine the area of influence of each weather station located in the department and assign its modeling weight.**

Key Words **- non parametric regression, Spline smoothing, Voronoi diagrams, temperature.**

I. Introducción

La modelación de fenómenos climatológicos es de gran interés para muchas actividades humanas como la industria agropecuaria o la piscicultura, puesto que para desarrollarlas de forma óptima, es de vital importancia conocer y entender el comportamiento de dichos fenómenos. Es así como el conocimiento previo sobre el comportamiento de diferentes factores climáticos permite a los agricultores programar las temporadas de siembra y cosecha.

Particularmente la temperatura ambiental es un fenómeno complejo de analizar, por su constante cambio debido a múltiples factores, como pueden ser fenómenos naturales como el del niño o la niña, el calentamiento global, la

[1] Producto derivado del proyecto de Investigación "Estadística de Variables Climáticas en el Suroccidente Colombiano", apoyado por la Universidad del Valle a través del grupo de investigación INFERIR.

Alvaro J. Florez es profesor Auxiliar, José E. Delgado y Mauricio Mera son estadísticos, pertenecientes al grupo de investigación INFERIR de la Universidad del Valle (correos e.: alvaro.florez@correounivalle.edu.co, jose.efrain@hotmail.com, mao_m22@hotmail.com).

contaminación ambiental, entre otros. Esta dificultad en el análisis del comportamiento temporal de la temperatura genera una incertidumbre en la planificación y análisis del contexto climatológico, donde este factor toma parte fundamental, lo cual se puede evidenciar claramente en la Segunda Comunicación Nacional ante la Convención Marco de las Naciones Unidas Sobre Cambio Climático [1].

Colombia cuenta con cuatro pisos térmicos distribuidos en cálido, templado, frío y páramo, según su altura en metros sobre el nivel del mar (msnm). Esta distribución establece una especie de estratos entre pisos térmicos, donde el comportamiento de la temperatura difiere y debe ser analizado cuidadosamente, por ejemplo en la región montañosa, a mayor altura la variación de la temperatura es en promedio de 5.53ºC cada kilómetro [2]

En el departamento del Valle del Cauca las instituciones públicas encargadas de monitorear fenómenos climatológicos y medioambientales son el IDEAM (Instituto de Hidrología, Meteorología y Estudios Ambientales) y la CVC (Corporación Autónoma Regional del Valle del Cauca). Estas entidades suministran información de las variables climatológicas, entre ellas la temperatura, a través de estaciones meteorológicas ubicadas a lo largo y ancho del territorio vallecaucano.

Este trabajo, adscrito al proyecto de modelación de variables climáticas del grupo de investigación INFERIR de la Universidad del Valle, tiene como objeto analizar la evolución, en el tiempo, de la temperatura promedio mensual en el Valle del Cauca durante el periodo 1971-2002, aplicando metodología estadística no paramétrica a través de la técnica de suavización Spline ponderada. Para la ponderación se utilizaron diagramas de Voronoi, esto con el fin de determinar el sector de influencia de cada una de las estaciones meteorológicas utilizadas para la medición de la temperatura, asignándole a cada estación un peso correspondiente a su importancia relativa en términos del área total considerada. Se espera que de esta forma el modelo o mejor la curva característica estimada capte con mayor precisión la variabilidad natural de las mediciones.

II. Modelación de la temperatura

En la mayoría de los casos en que se emplean metodologías estadísticas para la modelación de la temperatura, se aplican técnicas de series temporales, uno de ellos es el modelo autorregresivo integrado de media móvil (ARIMA), o modelos de regresión lineales [1].

En pocos casos se han utilizado técnicas más avanzadas, por ejemplo [3] hace uso de modelos mixtos. El inconveniente de estas técnicas es que su validez depende del cumplimiento de numerosos supuestos que ante fenómenos de esta naturaleza es difícil que se cumplan, además de ser muy susceptibles a la presencia de múltiples valores atípicos.

Teniendo en cuenta estos inconvenientes, los métodos de modelación no paramétricos parecen ser apropiados para cumplir este propósito en términos de temperatura, dado que estas técnicas son consideradas como robustas frente a los inconvenientes anteriormente mencionados y tienen gran capacidad de adaptación a la forma natural de variabilidad de los datos.

III. Diagramas de Voronoi

Según [4], los diagramas de Voronoi son una construcción geométrica que permite realizar una partición del plano euclídeo en polígonos convexos. Este es un método de interpolación basado en la distancia euclidiana, donde los polígonos se crean al unir los puntos entre sí, trazando las mediatrices de los segmentos de unión. Las intersecciones de estas mediatrices determinan una serie de polígonos en un espacio bidimensional alrededor de un conjunto de puntos de control, de manera que el perímetro de los polígonos generados sea equidistante a los puntos vecinos designando su área de influencia.

A. Definición

Sea $Q=\{q_1,\ldots,q_n\}$ un conjunto de n puntos distintos en el plano, con coordenadas cartesianas $(x_{11}, x_{12}),\ldots, (x_{n1}, x_{n2})$ y $2 \leq n \leq \infty$. Se puede definir el diagrama de Voronoi $V(Q)$ como una subdivisión del plano en n regiones, cada una correspondiente a un punto de Q, donde se cumple que un punto cualquiera q pertenece a la región correspondiente al punto q_i perteneciente a Q, sí y solamente sí $dist(q,q_i) < dist(q,q_j)$ con $i \neq j$ para cada punto q_j perteneciente a Q, la línea que separa la región $V(Q)$ es la mediatriz entre los respectivos puntos generadores de ambas regiones de Voronoi. Se denota la distancia euclidiana entre dos puntos q_j y q_i en el plano como:

$$dist(q_i, q_j) = \sqrt{(x_{j1} - x_{i1})^2 + (x_{j2} - x_{i2})^2}$$

IV. Modelo de regresión no paramétrico

Como lo expresa [5], el objetivo del análisis de regresión, ya sea paramétrico o no paramétrico, es estimar y probar las características de la función de regresión . Esta función describe la relación entre la variable explicativa habitualmente conocida como X y la variable de respuesta generalmente denotada por Y. Si se tienen n observaciones , la curva de regresión es comúnmente modelada como:

$$Y_i = f(X_i) + \varepsilon_i, \quad i = 1,2,\ldots,n \qquad (1)$$

Donde ε es una variable aleatoria que denota la variación de Y alrededor de la función $f(X)$ que depende de los puntos, $x_1\ldots, x_n$, que representa la media de la curva de regresión . $E(Y|X = x)$ Además se debe asumir que $E(\varepsilon_i)$ y $Var(\varepsilon_i) = \sigma^2 < \infty$.

El procedimiento para estimar la función de regresión

f (X) del modelo (1) en el marco de la regresión no paramétrica se llama *suavización*.

Para el uso de estas técnicas, a diferencia de los métodos de regresión paramétrica que poseen varios supuestos en el modelo, solamente se debe asumir que sea suave, lo que nos podría decir que para el ajuste de la curva en punto determinado de x, se espera que las observaciones asociadas a los cercanos a , posean información de en el punto de interés de [6]. Teniendo en cuenta lo anterior, los métodos de suavización consisten en promedio ponderado de dependiendo de la distancia de , donde los suavizadores más comunes son los estimadores lineales que tienen la forma:

$$\hat{f}(x_i) = n^{-1} \sum_{i=1}^{n} K(x, x_i; \lambda) y_i$$

Donde $K(\mathrm{x}, \mathrm{x_i}; \lambda)$ es una colección de pesos que dependen de la técnica de suavización, la distancia entre los puntos y el punto de estimación y de un llamado *parámetro de suavización*, o *ancho de banda*, encargado de determinar el grado de suavización a los datos. Algunos de estos estimadores se basan en un ajuste local de los datos, como es la suavización por Kernel o la regresión polinómica local. Otros ajustan un modelo que incluye una parte paramétrica y otra parte no paramétrica sujeta a una penalidad por aspereza, como es el caso de los Spline.

V. Suavización Spline

En la suavización Spline para la estimación de *f,* donde se asume q$f \in w_2^m[0,1]$ $\mathrm{w_2^m}$ es un espacio de Sobolev), en el modelo (1) se realiza mediante la minimización de la siguiente expresión:

$$s_\lambda(f) = \sum_{i=1}^{n} \left(y_i - f(x_i)\right)^2 + \lambda \int \left(f''(x)\right)^2 dx \quad (2)$$

Donde, el primer término (suma de cuadrados de los residuales) representa una medida de bondad de ajuste del modelo y el segundo término es una medida de la variabilidad local de la curva (penalización por aspereza) [6]. El objetivo de la suavización Spline es producir un buen ajuste con baja variabilidad [7].

El parámetro de suavización λ en (Ec. 2), cumple la función de equilibrar la relación entre la bondad de ajuste del modelo y la suavidad del mismo [6]. Cuando λ es grande se tendrá una estimación muy suave. Mientras que cuando es pequeño la estimación hará más énfasis en la bondad de ajuste, produciendo estimaciones que interpolan los datos (caso λ=0). Como menciona [5] es posible probar que la estimación Spline (*f*) es un promedio ponderado de las respuestas $\mathrm{y_1}$, de la forma:

$$\hat{f}(x) = n^{-1} \sum_{i=1}^{n} W_{\lambda i}(x) y_i$$

A. Suavización Spline Ponderada

Esta metodología consiste en ponderar la suma de cuadrados de los residuales $\sum_{i=1}^{n}\left(y_i - f(x_i)\right)^2$ en (2) con unos pesos $w_i > 0 \;\; i = 1, \ldots, n$,. Los pesos tienen como objetivo brindar una medida de importancia relativa a cada observación para la obtención del modelo estimado. De esta manera la suavización Spline ponderada consiste en minimizar:

$$s_\lambda(f) = \sum_{i=1}^{n} w_i \left(y_i - f(x_i)\right)^2 + \lambda \int \left(f''(x)\right)^2 dx$$

VI. Selección del parámetro de suavización

Como se mencionó anteriormente la importancia de calcular el valor adecuado del parámetro de suavización radica principalmente en que si λ es muy pequeño la estimación de f(*x*) estaría dada por la suma de cuadrados de los residuales generando mayor variabilidad, mientras que si λ es muy grande f(*x*) tenderá a interpolar los datos aumentando el sesgo [8].

Para la selección de f(*x*) se han propuesto varios métodos entre los que se destacan la Validación Cruzada y la Validación Cruzada Generalizada.

A. Validación Cruzada (CV):

La idea de este método es en encontrar un que minimice el error cuadrático medio de la función estimada (*f*(*x*) - *f*(*x*)). Este método se basa en la predicción de cada respuesta en el punto $\mathrm{x_i}$ por medio del ajuste de la curva con las observaciones restantes $(x_j, y_j, i \neq j)$. La función de validación cruzada está definida como:

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \hat{f}_i(x_i)\right)^2 \qquad (3)$$

Donde $\hat{f}_\mathrm{i}(x_\mathrm{i})$ mes la estimación por medio de Spline, omitiendo la observación $(x_\mathrm{i}, y_\mathrm{i})$, en el punto x_i. Por lo tanto el método consiste en encontrar el valor λ de que minimice (3). La Validación Cruzada Generalizada (GCV) es una modificación del CV para simplificar los cálculos de (3) [9]

VII. Estimación de la varianza

Como lo menciona [8], la estimación de la varianza de forma análoga con un modelo lineal, basándose en la suma de cuadrados de los residuales, no es válida debido a que las estimaciones hechas por técnicas no paramétricas son sesgadas, lo que tendrá como efecto la sobreestimación de la varianza. Por esta razón, en el contexto de regresión no paramétrica se han propuesto varios estimadores de donde se destacan los estimadores basados en diferencias, que tienen la particularidad de no depender de ningún parámetro de suavización y tampoco asume ninguna distribución de los errores. Entre estos estimadores se destacan los propuestos por [10], [11] y [12].

[11] proponen una interpolación lineal para el cálculo

de unos seudo residuales. Estos se obtienen tomando una tripleta consecutiva de puntos de diseño , uniendo las dos observaciones de los límites y , por medio de una línea recta, para luego calcular la diferencia entre esta línea recta y la observación del medio (*xi, yi*), de la siguiente manera:

$$\varepsilon_i = \frac{x_{i+1} - x_i}{x_{i+1} - x_{i-1}} y_{i-1} + \frac{x_i - x_{i-1}}{x_{i+1} - x_{i-1}} y_{i+1} - y_i$$
$$= a_i y_{i-1} + b_i y_{i+1} - y_i$$

Este estimador queda definido como:

$$\sigma^2_{GSJ} = \frac{1}{n-2} \sum_{i=3}^{n} c_i^2 \, \varepsilon_i^2 \qquad (4)$$

[13] y [14] en estudios de simulación donde se compararon varios estimadores de varianza bajo diferentes escenarios propuestos muestran que el estimador de [11] (También denominado estimador basado en diferencias ordinarias para el caso de un diseño equidistante) presentaba mejor comportamiento, en la mayoría de los escenarios planteados, que los estimadores propuestos por [12].

VIII. Bandas de variabilidad

Para la construcción de intervalos de confianza de la función de regresión $f(x)$s necesario que la estimación de la función $\hat{f}(x)$ sea normalmente distribuida y tenga una estimación de la varianza de $\hat{f}(x)$. Basándose en el teorema central del límite, la condición de normalidad puede cumplirse [8] y la estimación de la varianza de $\hat{f}(x)$se realiza de la siguiente forma:

$$\text{Var}\left(\hat{f}(x)\right) = \left(\sum_{i=1}^{n} W_{\lambda i}^2(x)\right) \sigma^2$$

El problema al realizar una estimación por intervalos de confianza está en que $\hat{f}(x)$ es sesgada, por lo cual no es posible calcular una cantidad pivotal para construir el intervalo [8].

Una alternativa que permite observar el grado de variabilidad de la estimación no paramétrica, sin tener que calcular el sesgo de estimación, es por medio de bandas de confianza, las cuales se construyen de la siguiente forma:

$$\hat{f}(x) \pm 2Var\left(\hat{f}(x)\right) \qquad (5)$$

La diferencia con los intervalos de confianza radica en que las bandas indican intervalos de confianza puntuales para $\text{E}\left(\hat{f}(x)\right)$ en lugar de $f(x)$, por cual se debe tener precaución en su interpretación [8].

IX. Resultados

Dado que las estaciones meteorológicas se encuentran en diferentes pisos térmicos, desde 920*msnm* hasta 1950*msnm*, del departamento del Valle de Cauca, fue necesario, para la modelación, dividir las observaciones en dos estratos, definidos como Valle y Montaña. Los estratos están conformados por las mediciones hechas en las estaciones que se encuentran a menos o en 1100*msnm* y el estrato montaña por las estaciones que se encuentran a más de 1100*msnm*. En el estrato Valle se encuentran 15 estaciones entre 920*msnm* y 1100*msnm*, mientras que en estrato Montaña hay 13 estaciones entre 1233*msnm* y 1950*msnm*.

En la Fig. 1 se relaciona la distribución de las observaciones de temperatura para cada estación meteorológica, se observa que la mayor dispersión en el estrato Valle se presenta en la estación Zaragoza ubicada a 925*msnm* (etiqueta v925), mientras que en el estrato Montaña la estación Monteloro (etiqueta v1861) es la que presenta mayor variabilidad.

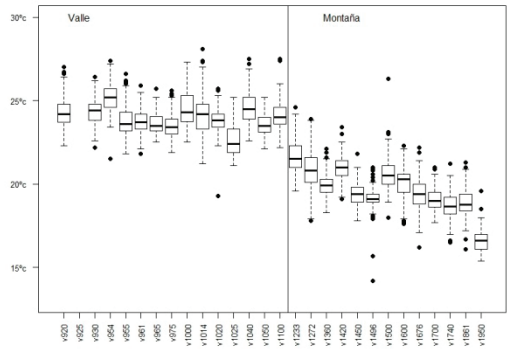

Fig. 1 Diagrama de cajas y alambres de las temperaturas mensuales para cada estación según su altitud.

En [3] todas las observaciones de la estación Zaragoza y las pertenecientes a la estación Monteloro, durante el periodo Enero 1971-Noviembre 1999, fueron descartadas del estudio, esto tras mostrar con métodos estadísticos, que el comportamiento de las temperaturas en estas estaciones durante esos periodos es errático al compararlo con el comportamiento de las estaciones más próximas en el mismo periodo, basándose en el supuesto que es poco probable que observaciones que estén cercanas en el espacio presenten comportamientos distintos.

La Fig. (2) muestra la temperatura promedio mensual de las estaciones por cada estrato en el Valle del Cauca, donde se puede apreciar que las estaciones del estrato que lleva el mismo nombre tienen promedios de temperatura mayores que los presentados en la Montaña, no se observa ninguna tendencia creciente o decreciente durante el periodo estudiado, aunque si un comportamiento parecido en los estratos, puesto que cuando la temperatura en el Valle aumenta también lo hace en la Montaña, aunque no en la misma magnitud.

Se observa que el rango de la temperatura parece ser mayor en el estrato Montaña en todo el periodo de estudio. También se pueden ver puntos alejados de la nube de observaciones, sobre todo en la parte inferior donde parece formarse un tercer estrato (formado por las temperaturas observadas en la estación La Teresita, ubicada a 1950msnm)

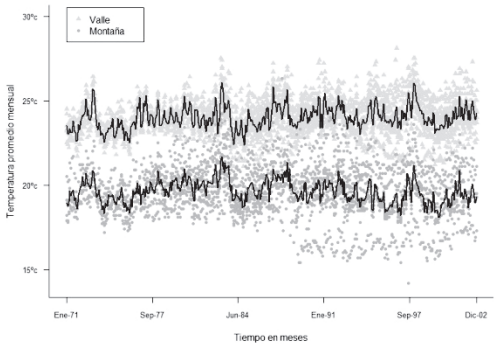

Fig. 2 Temperatura promedio mensual por cada estrato en el Valle del cauca.

A. Diagramas de Voronoi

La Fig. 3 muestra el croquis del mapa del departamento del Valle del Cauca y en su interior se pueden observar las regiones de Voronoi que se han formado para cada estación meteorológica; cada uno de estos polígonos representa la región de influencia de cada estación meteorológica según su estrato.

Para la eliminación de las regiones donde la altura no estaba dentro de las consideradas en este estudio, se crearon estaciones de monitoreo ficticias, tomando como referencia información del IDEAM y la CVC sobre los puntos más altos ubicados en el departamento y luego se creó la región de Voronoi, determinando de este modo su área de influencia para posteriormente proceder a extraerla.

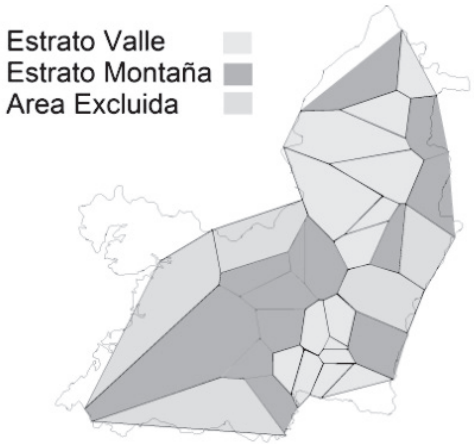

Fig. 3 Temperatura promedio mensual por cada estrato en el Valle del cauca.

Finalmente el proceso de la creación del diagrama de Voronoi permite detallar el cubrimiento que se tuvo sobre el área total considerada del Valle del Cauca y principalmente calcular el área de influencia correspondiente a cada región de Voronoi, obteniendo el área de cada polígono formado, para después ser usado como vector de pesos para los splines, logrando así mayor precisión y eficiencia en la estimación de las curvas típicas.

B. Modelación Spline

La estimación de las curvas típicas de la temperatura promedio mensual del Valle del Cauca entre los años 1971 a 2002 por medio de Spline (los anchos de banda fueron estimados por validación cruzada) se muestra en la Fig. (4), se puede observar en ambos casos, haciendo referencia a los dos estratos, que las curvas estimadas siguen de buena manera el comportamiento de los datos, teniendo en cuenta la alta variabilidad que estos presentan.

Lo anterior se observa sobre todo en el estrato Montaña, puesto que este presenta la mayor variabilidad en la temperatura entre las estaciones de muestreo. Para ambos estratos se pueden observar picos de temperatura altos y bajos, casi que similares en ambos casos, esto se debe al efecto de fenómenos como el de El Niño y La Niña que afectan de manera significativa su comportamiento.

En la Fig. (4) también se observan las bandas de variabilidad de la modelación de la temperatura mensual para cada estrato, estas fueron construidas por medio de (5) con una estimación de varianza usando el estimador (4). Es importante destacar que las bandas de confianza atrapan en su interior un alto porcentaje de observaciones, pues esto da buena señal de que el modelo estimado es adecuado, también se puede visualizar que algunos puntos quedan muy distantes de las bandas, pero se debe tener en cuenta que el trabajo [15] sobre observaciones atípicas para este mismo conjunto de datos, demostró la presencia de observaciones de este tipo, por lo cual, dichas observaciones o algunas de ellas que se encuentran muy alejadas podrían ser observaciones atípicas y por tanto estarían lógicamente lejos de la estimación de las curvas típicas y de sus respectivas bandas de confianza.

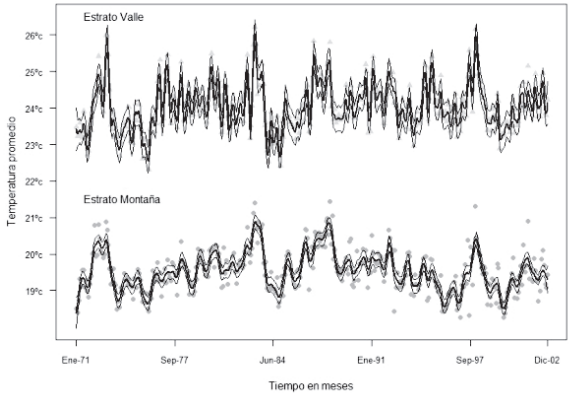

Fig. 4 Curvas típicas y bandas de confianza para la temperatura promedio mensual en cada estrato.

X. Conclusiones y sugerencias

Los diagramas de Voronoi son una alternativa muy eficiente y practica como método de interpolación, puesto que nos ayudaron a determinar el área de influencia de cada estación meteorológica y fueron fundamentales como elemento ponderador en la estimación de los Splines para disminuir el efecto del muestreo irregular.

Los resultados obtenidos invitan a pensar que la

estimación de la curva típica a través de la suavización spline ponderada por diagramas de Voronoi es una metodología útil y eficiente para modelación de la temperatura promedio mensual del Valle del Cauca para el periodo 1971-2002. Además se mostró que el comportamiento de la temperatura en el Valle del Cauca varía entre los pisos térmicos cálido y templado formando dos estratos, por lo que fue necesario evaluarlos y modelarlos por separado.

Uno de los motivos por los cuales se propuso el uso de técnicas de modelación no paramétricas fue el alto número de observaciones faltantes durante el periodo de estudio, puesto que se observó que ningún mes tiene las observaciones para todas las estaciones de muestreo, también que el estrato con más datos faltantes fue el de Montaña. El alto número de datos faltantes restringe el uso de algunos métodos paramétricos de modelación como es los modelos de series de tiempo ARIMA.

Aunque la suavización Spline proporcionó buenos resultados para la modelación de la temperatura, esta se podría mejorar con la inclusión en el modelo de posibles variables que estén altamente relacionadas con la temperatura, como puede ser el índice de oscilación sur (OIS). También se podrían observar los resultados de otras técnicas de regresión no paramétrica, como por ejemplo la regresión lineal local.

Teniendo en cuenta que el uso de los diagramas de Voronoi se debió principalmente a su simpleza, se propone el uso de métodos más avanzados para determinar el área de influencia de las estaciones meteorológicas y observar si se presentan cambios significativos en los resultados.

Referencias

[1] Segunda comunicación nacional ante la convención marco de las naciones unidas sobre cambio climático. SCN. 2010.

[2] Eslava, J. *"Climatología del Pacífico colombiano"*, Academia colombiana de ciencias geofisicas. Bogotá. Colección Eratóstenes, 1994.

[3] Andrade, M. *"Monthly Average Temperature Modelling for Valle del Cauca(Colombia)"*. Ph.D. dissertation, The University of Reading, 2009.

[4] Boots, B., Okabe, A., Sugihara, K., and Chiu, S.N. *"Spatial tessellations: concepts and applications of Voronoi diagrams"*. West Sussex, England. Wiley, 2000.

[5] Olaya, *"J. Suavización y regresión no paramétrica: una alternativa de modelación estadística"*. XV Simposio de Estadistica. Universidad Nacional de Bogotá.

[6] Eubank, R. *"Nonparametric Regression and Spline Smoothing"*. New York. Marcel Dekker, 1999.

[7] Härdle, W. *"Smoothing Techniques With Implementation in S"*. New York. Springer-Verlag, 1991.

[8] Bowman, W. and Azzalini, A. *"Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations"*. New York. Oxford University Press, 1997.

[9] Hastie, T. and Tibshirani, R. *"Generalized Additive Models"*. Chapman & Hall, London (1990)

[10] Rice, J. *"Bandwidth Choice for Nonparametric Regression". The Annals of Statistics*, vol. 12, no. 4, pp. 1215-1230. 1984.

[11] Gasser, T., Sroka, L. and Jennen-Steinmetz, C. *"Residual Variance and Residual Pattern in Nonlinear Regression". Biometrika*, vol. 73, no. 3, pp. 625-633. 1986

[12] Hall, P., Kay, J. W. and Titterington, D. M. *"Asymptotically Optimal Difference-Based Estimation of Variance in Nonparametric Regression". Biometrika,* vol. 77, no.3, pp. 521-528. 1990

[13] Dette, H., Munk, A. and Wagner, T. *"Estimating the Variance in Nonparametric Regression-What is a Reasonable Choice?"* 4, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 60, no. 4, pp. 751-764. 1998.

[14] Florez, A. y Olaya, J. *"Estudio de simulación para comparar varios estimadores de varianza en el marco de la regresión no paramétrica"*. Comunicaciones en Estadística, vol. 7, n 1, pp 49-66. 2014

[15] Andrade, M. y Longford, N. *"Outliers in mixed models for monthly average temperatures". Austrian Journal of Statistics*, vol. 39, n. 3, pp. 203-221. 201

Alvaro José Flórez. Nació en Cali, Colombia, el 8 de octubre de 1983. Se graduó de estadístico en el año 2007 en la Universidad del Valle de Cali, Colombia. Actualmente ejerce como profesor auxiliar de la Escuela de Estadística de la Universidad del Valle, Cali, Colombia. Algunas de sus publicaciones son: "Modelación de la concentración atmosférica de CO usando regresión no paramétrica con bandas de variabilidad no homogéneas". Ingeniería y Competitividad, Volumen 16, No. 1, p. 259 - 267 (2014) y "Estudio de simulación para comparar varios estimadores de varianza en el marco de la regresión no paramétrica". Comunicaciones en Estadística, Vol. 7, No. 1, pp. 49–66 (2014).

El profesor pertenece al Grupo de Investigación en Estadística Aplicada (INFERIR) de la Universidad del Valle, donde el área actual de estudio es el análisis de datos longitudinales.

Mauricio Mera Hoyos. Nació en Cali, Colombia. Culmino sus estudios de estadística en el año 2011 en la Universidad del Valle con sede en Cali, Colombia. Ha ejercido su profesión diversas áreas entre las cuales se encuentran la Investigación de Mercados en Colombia y Centro América y el sector Energético Colombiano; actualmente se encuentra vinculado al sector financiero a través del el Banco de Occidente, Cali, Colombia.

En desarrollo de su trabajo académico hizo parte del grupo Grupo de Investigación en Estadística Aplicada (INFERIR) de la Universidad del Valle.

José Efraín Delgado. Nació en Cali, Colombia, el 15 de Marzo de 1983. Se graduó de estadístico en el año 2011 en la Universidad del Valle de Cali, Colombia. Actualmente ejerce como docente en la fundación Santa Isabel de Hungría, Cali, Colombia. Entre sus funciones se encuentran además de las labores pedagógicas, la medición y evaluación de las actividades realizadas en las diferentes gestiones de la institución, en el marco del proceso de calidad institucional.