

Ajustes de distribuciones probabilísticas para la variable temperatura media multianual para el departamento de Boyacá (Colombia)

Adjust of the probabilistic distribution for the multi-year average temperature variable for the department of Boyacá (Colombia)

Dayam Soret Calderón Rivera¹, Claudia Fernanda Navarrete López² y José Luis Díaz Arévalo³

Resumen

En este trabajo se presenta un estudio acerca de la selección de la mejor distribución probabilística para la variable media multianual de la temperatura en el departamento de Boyacá (Colombia), como base para futuras estimaciones y proyecciones de la variable en condiciones de incertidumbre. Se seleccionaron las distribuciones Normal, Gamma, Weibull y LogNormal para ajustar los datos; y para encontrar cual distribución ajusta mejor los datos se utilizaron los criterios de información basados en la máxima verosimilitud de Akaike (Akaike Information Criterion) y Bayesiano (Bayesian Information Criterion). Se muestran los resultados tanto en forma tabular como gráfica, así como un plano de las funciones de distribución probabilísticas más representativas en el área de estudio. Como resultado se obtiene que en general la distribución que mejor se ajusta es la Weibull.

Palabras clave: Distribuciones de probabilidad; Temperatura; Criterios de información; Boyacá.

Abstract

This paper presents a study on the selection of the best probabilistic distribution for multi-year average temperature variable in the department of Boyacá (Colombia) as the basis for future estimates and projections of the variable under uncertainty arises. Normal, Gamma, Weibull and lognormal distributions were selected to fit the data. To find which best fits the data criteria information based in maximum likelihood, Akaike (Akaike Information Criterion) and Bayesian (Bayesian Information Criterion) were used. The results are

-
- 1 Estudiante Facultad de Ingeniería Ambiental. Universidad Santo Tomás (Colombia). Carrera 9 51-11, Bogotá (colombia). E-mail: dayamcalderon@usantotomas.edu.co
 - 2 Magister en Agronomía. Universidad Santo Tomás (Colombia). Carrera 9 51-11, Bogotá (colombia). E-mail: claudianavarrete@usantotomas.edu.co
 - 3 Doctor Ingeniería Hidráulica y Medio Ambiente. Universidad Santo Tomás (Colombia). Carrera 9 51-11, Bogotá (colombia). E-mail: jluisdiaz@usantotomas.edu.co

shown in both tabular and graphical form, as well as a map of the probabilistic distribution functions most representative in the study area. As a result we obtain that the Weibull distribution is the best fits in general.

Key words: Probabilistic Distributions; Temperature; Information criterion; Boyacá.

1. Introducción

El objeto de éste trabajo es determinar la distribución probabilística de mejor ajuste para los datos medios multianuales de la variable climatológica temperatura, en el departamento de Boyacá (Colombia). Estos datos fueron tomados del catálogo de estaciones climatológicas proporcionado por el IDEAM. Se probaron las distribuciones Normal, Gamma, Weibull y LogNormal para ajustar los datos (Navarrete, 2008; Minka 2002; Aristizabal, 2012; Martínez, 2011); y para encontrar cual distribución ajusta mejor los datos se utilizaron los criterios de información de Akaike (AIC) (Akaike, 1974) y Bayesiano (BIC) (Akaike, 1978; Schwarz 1978). Un aspecto importante de los criterios de información es que estos están asociados con el método de máxima verosimilitud, el cual, en modelos causales, está basado en el supuesto de que las variables observadas siguen una distribución normal multivariante (Caballero, 2011). La idea clave del AIC es la de penalizar un exceso de parámetros ajustados, mientras que en el BIC se penaliza el número de parámetros teniendo en cuenta que el AIC podría no ser asintóticamente justificable (Caballero, 2011; Montesinos, 2011; Ramírez, 2000).

Se encontró escasa información aplicada a series de datos hidrológicos (Ayodele, 2015; Chaouche, *et.al.*, 2010; Donat, 2012; Räsänen, 2009; Hingray, 2007), entre los cuales a nivel local se destaca el trabajo presentado por Dorado *et. al.* en el 2006 donde analizaron la variabilidad espacio-temporal de la precipitación mensual y anual en el área de influencia aguas abajo del embalse de regulación y abastecimiento SARA-BRUT, ubicado en el departamento del Valle del Cauca, Colombia. Para el ajuste de las series históricas se usaron modelos de distribución de probabilidad teórica, como Gumbel, Normal, LogNormal y Wakeby, y modelos de distribución empírica, como Weibull y Landwehr. Se utilizó la prueba de Error Estándar de Ajuste (EEA) para seleccionar el modelo teórico de mejor ajuste. El modelo Wakeby presentó el mejor ajuste en el 89% del total de estaciones analizadas, el cual corresponde a un modelo probabilístico adecuado para representar los caudales de crecidas máximas diarias o instantáneas en una región. No obstante para nuestro estudio, la distribución Weibull se presenta como la

más confiable debido a su versatilidad (Behar y Grimas, 2004), (Moreno *et al.*, 2011). En su trabajo de maestría (Navarrete, 2008) encuentra que para los criterios de AIC y BIC la distribución Weibull es la que mejor describe los datos.

2. Metodología

Este trabajo fue realizado sobre la totalidad de la extensión del departamento de Boyacá a partir de la información proporcionada por el Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia (IDEAM), el cual cuenta con 35 estaciones meteorológicas en la región, las cuales se presentan en la tabla 1 junto con la extensión de la serie analizada. De estas estaciones se obtuvo el promedio multianual de la variable temperatura, y se realizaron los histogramas de frecuencia para cada una de las estaciones con la finalidad de determinar gráficamente las funciones a ajustar. No se consideró necesario realizar completado de datos ya que la variable temperatura media presenta poca varianza (0.52 en promedio) y el porcentaje de datos faltantes global es bajo (3.60%).

Se aplicaron las pruebas de Mann-Kendall (Mann, 1945; Kendall, 1975) y Rachas con la finalidad de validar supuestos de homogeneidad y aleatoriedad de las estaciones muestreadas. Los resultados obtenidos mostraron que la mitad de las estaciones presentan un comportamiento homogéneo y aleatorio. Como el objetivo del trabajo es presentar las funciones de distribución de mejor ajuste a las series de datos, no se realizaron transformaciones a las mismas.

Por medio de los software R⁴ y ParameterSolver⁵ se procesaron los datos y se estimaron los parámetros, con el fin de hallar las distribuciones de probabilidad y las funciones de máxima verosimilitud, las cuales fueron utilizadas para calcular los criterios de información Akaike (AIC) y Bayesiano (BIC); con base en estos se escogió la distribución de mejor ajuste. El AIC fue propuesto por Akaike, H. (1974) como un estimador insesgado asintótico de la información de Kullback-Leibler esperada entre un modelo candidato ajustado y el verdadero modelo. El BIC fue derivado por Akaike (1978) y Schwarz (1978) como una aproximación a

Tabla 1. Estaciones climatológicas

Nombre	Código	Años	Nombre	Código	Años
PTO BOYACA	23115010	39	TUNEL EL	35095030	42
OTANCHE	23125080	39	UPTC	24035130	45
Buena Vista	23125100	20	SURBATA BONZA	24035120	60
APTO FURATENA	23125140	23	TUNGUAVITA	24035170	46
INST AGR SANTA SOFÍA	24015090	28	CHITA	24035250	42
VILLA CARMEN	24015220	46	SATIVANORTE	24035320	39
VILLA DE LEIVA	24015300	34	NUEVO COLON	35075010	48
SIERRA LA	24025030	47	INST AGR MACANAL	35075040	32
CUSAGUI	24035010	28	CAMPOHERMOSO	35085050	27
BETEITIVA	24035010	23	CARDON	35235010	39
COPA LA	24035040	22	BELENCITO	24035150	47
GUICAN	24035070	23	COCUY	24035240	40
APTO A LLERAS C	24035340	32	ANDALUCIA	24035350	18
CHISCAS	24035310	40	VILLA LUISA	35075030	29
RONDON	35085020	42	VIVERO	35085040	30
SAN RAFAEL	24035180	34	CORINTO	35195050	29
SUTATENZA	35075020	44	TUNEBIA	37035010	40
BOAVITA	24035330	36			

Fuente: Catálogo red estaciones hidroclimáticas - IDEAM.

una transformación de la probabilidad posterior de un modelo candidato.

Las distribuciones de probabilidad y los criterios se describen a continuación:

Distribuciones de probabilidad

Distribución normal: esta distribución de probabilidad continua es la más importante y la más utilizada, también llamada curva de Gauss. Esta dada por dos parámetros de la función $f(x, \mu, \sigma)$, donde μ representa la desviación estándar y μ la media de la muestra

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1)$$

Distribución Gamma: es un caso especial de la distribución Pearson tipo III donde sus parámetros deben ser diferentes de 0. Esta distribución está dada por: α (α) la función gamma de α , β como el parámetro de escala y β es el parámetro de forma.

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{(\alpha-1)} e^{-\frac{x}{\beta}} \quad (2)$$

donde $\alpha = \frac{1}{(\frac{\sigma}{\mu})^2}$ y $\beta = \frac{\mu}{\sigma^2}$

Distribución Weibull: esta distribución es muy versátil y es utilizada para modelar amplia gama de aplicaciones, describiéndose según los parámetros de escala y localización α , forma m y β como origen de la distribución.

$$f(x) = \frac{m}{\beta} \left(\frac{x-\alpha}{\beta}\right)^{m-1} e^{-\left(\frac{x-\alpha}{\beta}\right)^m} \quad (3)$$

donde $\alpha = \frac{1}{n} \sum_{i=1}^n (x_i^\beta)^{1/\beta}$ y $\beta = \frac{\mu}{\sigma}$

Distribución Log normal: al igual que la distribución normal está dada por los dos parámetros donde σ representa la desviación estándar y μ la media de la muestra

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\left(\frac{\ln(x)-\mu}{\sigma}\right)^2} \quad (4)$$

Máxima verosimilitud

La función de máxima de verosimilitud está dada por la máxima probabilidad de que ocurra o se dé una determinada muestra si es cierta la estimación

que hemos efectuado, o el estimador que hemos planteado. La densidad conjunta para esas observaciones es denotada como $L = f(x_i, p)$, y la máxima verosimilitud puede ser obtenida tomando la

derivada de L en relación con la probabilidad p igualando la función a 0, para lo cual es mejor tomar logaritmos de la función L y derivarlo; n es el número de observaciones.

Normal

$$\ln(x)L(x, \mu, \sigma^2) = -n \ln(x) (\sigma\sqrt{2\pi}) \frac{\sum_{i=1}^n (x_i - \mu)}{2\sigma^2} \quad (5)$$

Gamma

$$\ln(x)L(x, \alpha, \beta) = n\alpha \ln(\beta) - n \ln[\Gamma(\alpha)] + (\alpha - 1) \sum_{i=1}^n (\ln(x_i)) - \beta \sum_{i=1}^n (x_i) \quad (6)$$

Weibull

$$\ln(x)L(x, \alpha, \beta) = n[\ln(\alpha) - \alpha \ln(\beta)] + (\alpha - 1) \sum_{i=1}^n (\ln(x_i)) - \frac{1}{\beta^\alpha} \sum_{i=1}^n (x_i^\alpha) \quad (7)$$

Log-normal

$$\ln(x)L(x, \mu, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \sum_{i=1}^n (\ln(x_i)) - \frac{\sum_{i=1}^n (\ln(x_i)^2)}{2\sigma^2} + \frac{\sum_{i=1}^n (\mu \ln(x_i))}{\sigma^2} - \frac{n\mu^2}{2\sigma^2} \quad (8)$$

Calculo AIC y BIC

El criterio de información Akaike propuesto en 1974 y utilizado como un estimador insesgado, precisa que el modelo de probabilidad de menor valor AIC es el que se selecciona como el mejor al que se le ajustan los datos. La función está dada por la maximización del logaritmo de la máxima verosimilitud denotado como $(\ln L)$ y K es el número de parámetros de la función de probabilidad (parámetros en el modelo):

$$AIC = 2K - 2\ln(L)$$

La estructura del AIC está compuesta por la maximización del logaritmo de verosimilitud, es decir, como componente de la falta de ajuste del modelo y K como el número de parámetros estimados dentro del modelo como componente de penalidad. La penalidad es una medida de la complejidad o compensación por el sesgo debido a la falta de ajuste cuando los estimadores de máxima verosimilitud son empleados (Bozdogan, 1987 en Acuña *et. al.*, 2012).

BIC sirve para la selección del modelo entre un conjunto finito de modelos, está estrechamente relacionado con el criterio AIC y se basa en parte en la función de probabilidad. Para mejorar la inconsistencia del cri-

terio AIC, Akaike 1978 y Schwarz 1978 presentaron un criterio de selección de modelos desde la perspectiva bayesiana. Schwarz estableció que la solución de bayes consiste en seleccionar el modelo con una alta probabilidad a posteriori. El criterio de información bayesiana (BIC) se define como:

$$BIC = 2\ln(L) + K\ln(n)$$

Mapa de resultados

Una vez obtenidos los criterios AIC y BIC para cada una de las distribuciones por estación se procedió a realizar el mapa de área de influencia como se muestra en la figura 2.

3. Resultados

Después de obtener las distribuciones de probabilidad para cada una de las estaciones y de la aplicación de los criterios AIC y BIC se observó que la distribución dominante en el departamento de Boyacá es la Weibull ajustándose a 21 estaciones de las 35 muestreadas (60%), en segundo lugar se encontraron las distribuciones Normal y Gamma con 6 estaciones cada una (17%) y por último la distribución Lognormal con 2 estaciones (6%) (tabla 2).

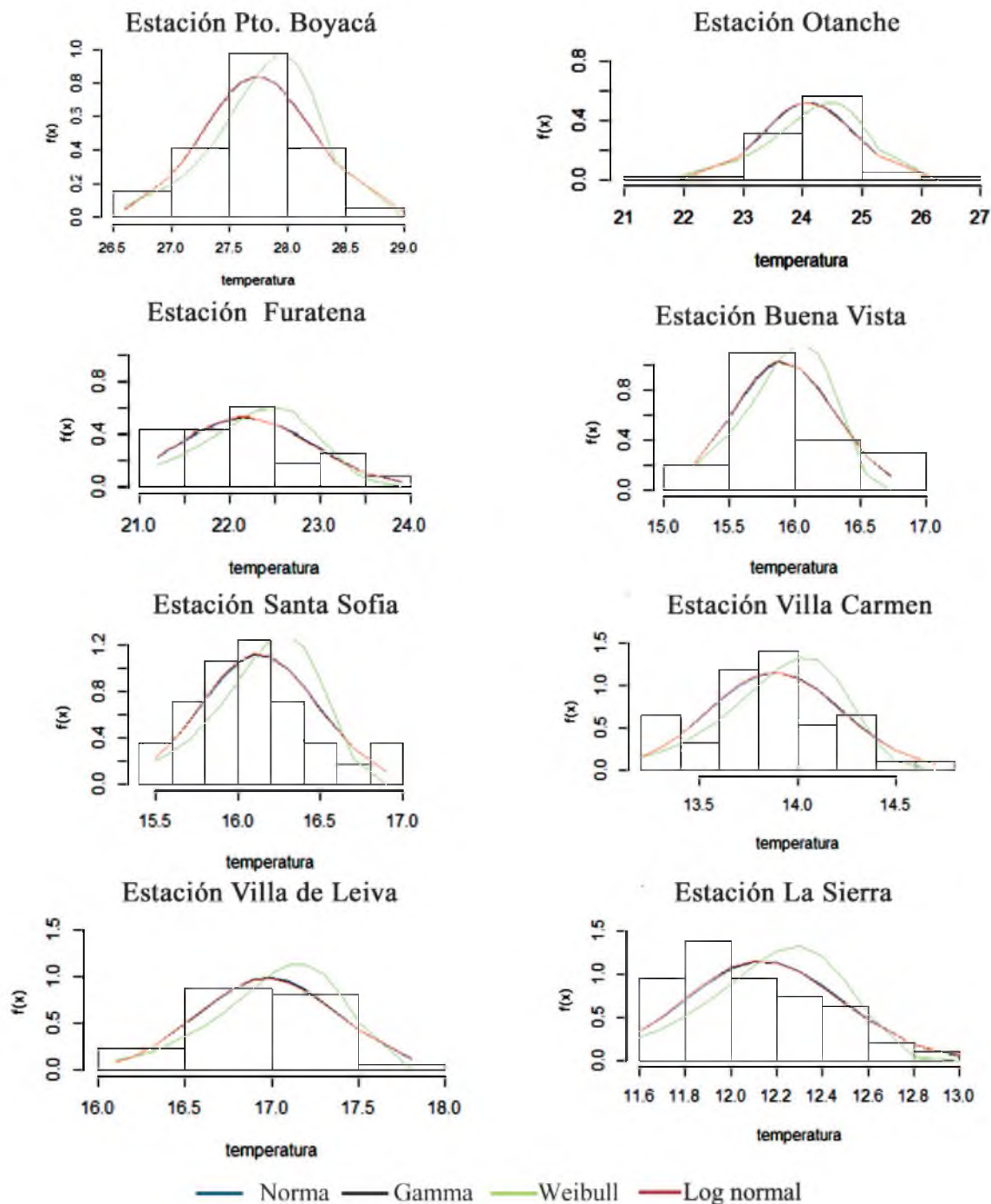
Tabla 2. Criterios de ajuste de distribución

Nombre de la estación	AIC				BIC			
	Normal	Gamma	Weibull	Log normal	Normal	Gamma	Weibull	Log normal
PTO BOYACA	234,7532	237,8511	205,2972	239,3974	231,4261	234,524	201,9701	236,0703
OTANCHE	340,4676	351,6669	277,1768	356,6061	337,1405	348,339	273,8497	353,2790
BUENA VISTA	58,12972	58,25276	59,98834	58,94445	56,13826	56,26130	57,99688	56,95299
APTO FURATENA	63,93424	63,03625	77,63782	63,01264	61,66325	60,76526	75,36683	60,74165
INST SANTA SOFIA	78,27692	78,43149	86,12450	79,21007	75,61252	75,76708	83,46009	76,54567
VILLA CARMEN	163,8532	165,5435	162,7736	166,4026	167,5105	169,2008	166,4309	170,0599
VILLA DE LEIVA	161,6651	164,0032	149,5635	163,8859	158,6124	160,9505	146,5107	160,8332
SIERRA LA	98,33291	97,61236	120,10444	98,06344	94,63261	93,91206	116,40415	94,36314
CUSAGUI	125,6314	127,9963	114,8585	127,8248	122,9670	125,3319	112,1941	125,1604
BETEITIVA	92,21152	93,08549	188,54432	92,99109	89,94054	90,81450	186,27333	90,72010
COPA LA	67,33759	66,47118	77,71301	66,28266	65,15551	64,28910	75,53093	64,10058
GUICAN	42,03726	41,40550	54,73516	41,52515	39,76627	39,13451	52,46418	39,25417
APTO A LLERAS C	193,5991	198,8904	168,3923	203,7743	190,6677	195,9589	165,4609	200,8428
CHISCAS	272,5039	280,2488	231,4761	278,6346	269,1262	276,8711	228,0984	275,2568
RONDON	125,7769	125,6248	141,0327	126,3589	122,3015	122,1495	137,5574	122,8835
SAN RAFAEL	137,8840	139,1904	138,7613	139,5146	134,8313	136,1377	135,7086	136,4618
SUTATENZA	194,1656	195,8829	195,0512	198,1723	190,5973	192,3146	191,4828	194,6039
TUNEL EL	132,3041	132,9302	149,5230	135,3521	128,8287	129,4549	146,0476	131,8768
UPTC	198,2894	201,6581	181,5968	203,8836	194,6761	198,0448	177,9835	200,2703
SURBATA BONZA	212,0912	211,0693	235,2744	213,5415	207,9025	206,8807	231,0857	209,3528
TUNGUAVITA	213,9447	216,3391	191,1763	221,1755	210,2875	212,6818	187,5190	217,5183
CHITA	249,9439	257,7754	159,1011	264,3868	246,4686	254,3001	155,6258	260,9114
SATIVANORTE	173,5771	177,2878	139,4639	177,5658	170,2499	173,9607	136,1368	174,2386
NUEVO COLON	181,1275	182,5239	152,1226	185,2750	184,8699	186,2663	155,8650	189,0174
INST AGR MACANAL	184,3095	187,8479	105,5642	188,5308	181,3780	184,9164	102,6327	185,5994
CAMPOHERMOSO	30,31309	29,96005	41,77312	30,03026	27,72141	27,36837	39,18145	27,43858
CARDON	151,3815	155,5726	108,7090	161,1811	148,0544	152,2455	105,3819	157,8540
BELENCITO	146,7130	144,3126	165,2663	144,8319	143,0127	140,6123	161,5660	141,1316
COCUY	216,3229	227,7881	161,7740	230,4454	212,9452	224,4103	158,3962	227,0676
ANDALUCIA	48,51900	48,53407	43,08183	48,15724	46,73825	46,75332	41,30108	46,37649
VILLA LUISA	115,41367	116,23370	98,95385	120,41404	112,67908	113,49911	96,21925	117,67945
VIVERO	111,7589	111,8932	106,7041	113,4887	108,9565	109,0908	103,9017	110,6863
CORINTO	221,7676	229,7986	183,0025	224,4988	219,0330	227,0640	180,2679	221,7643
TUNEBIA	131,1866	131,5875	109,3331	132,2281	127,8088	128,2097	105,9553	128,8504
BOAVITA	128,6716	129,3392	108,2066	130,4852	125,5045	126,1722	105,0396	127,3182

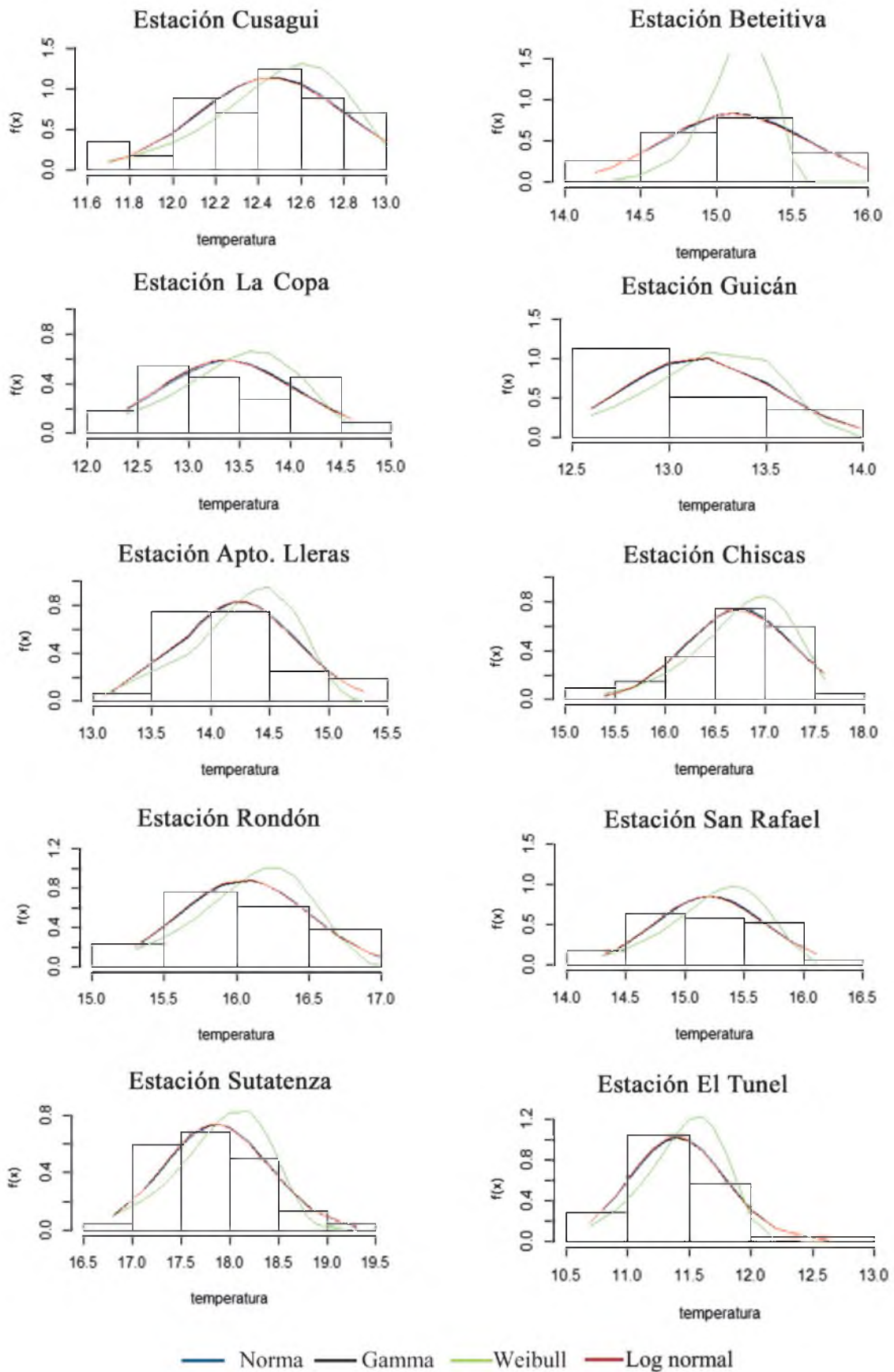
Fuente: Autores. Fecha 30 de Octubre de 2015

A continuación se presentan las figuras de las distribuciones de probabilidad para cada una de las estaciones climatológicas del estudio, en color verde se puede apreciar la distribución Weibull.

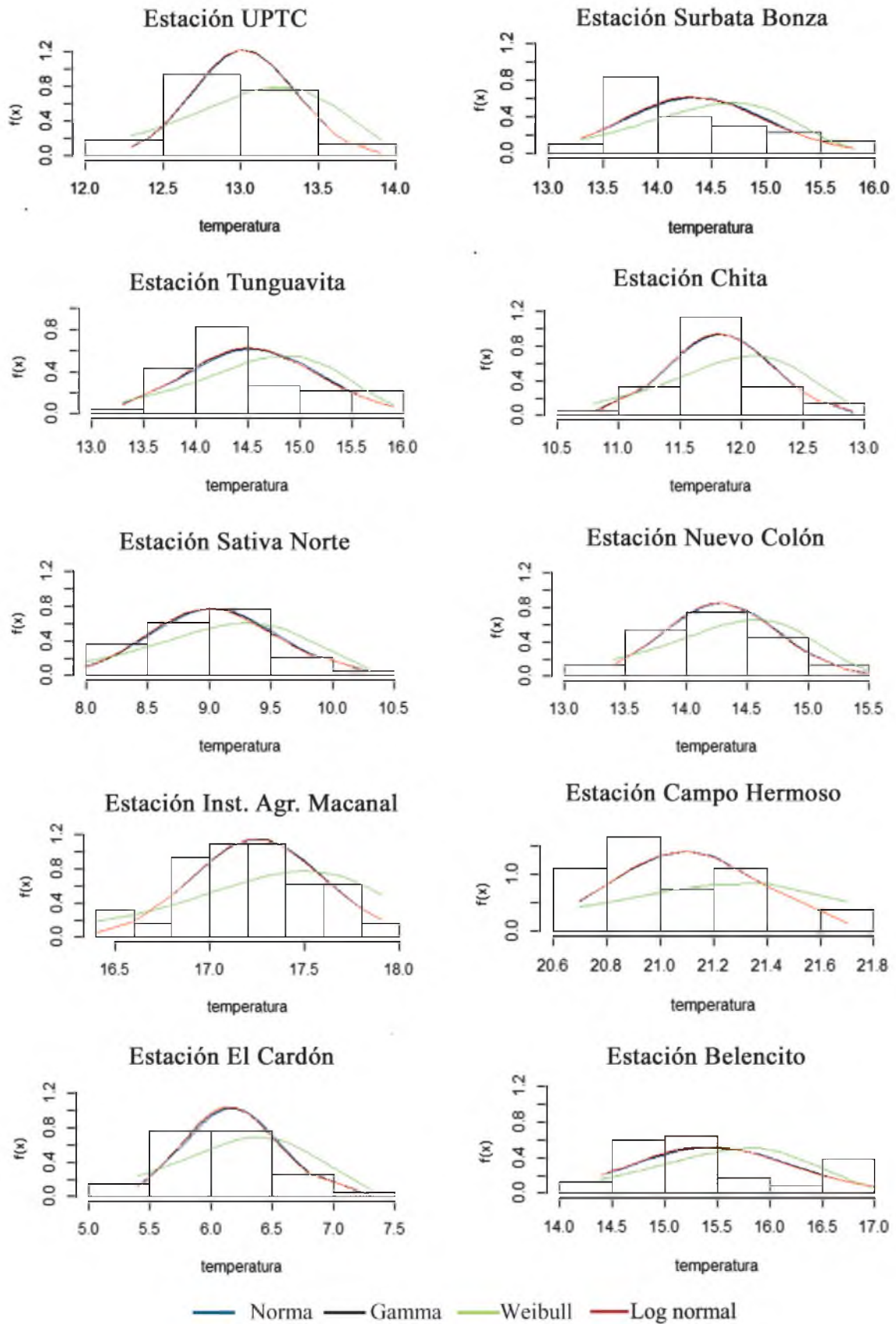
Figura 1. Distribuciones de probabilidad de la variable temperatura para cada una de la estaciones.



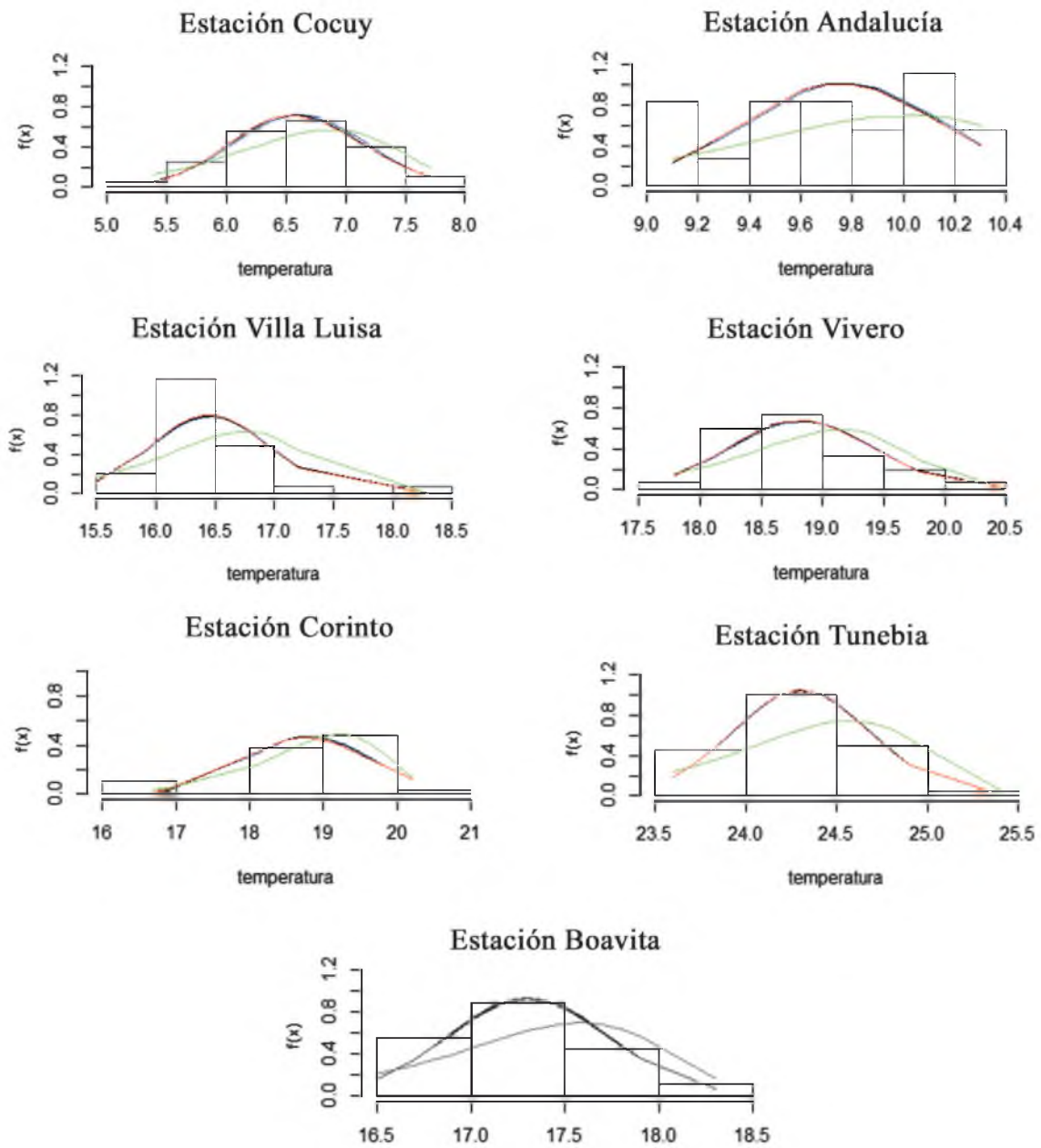
Continuación figura 1



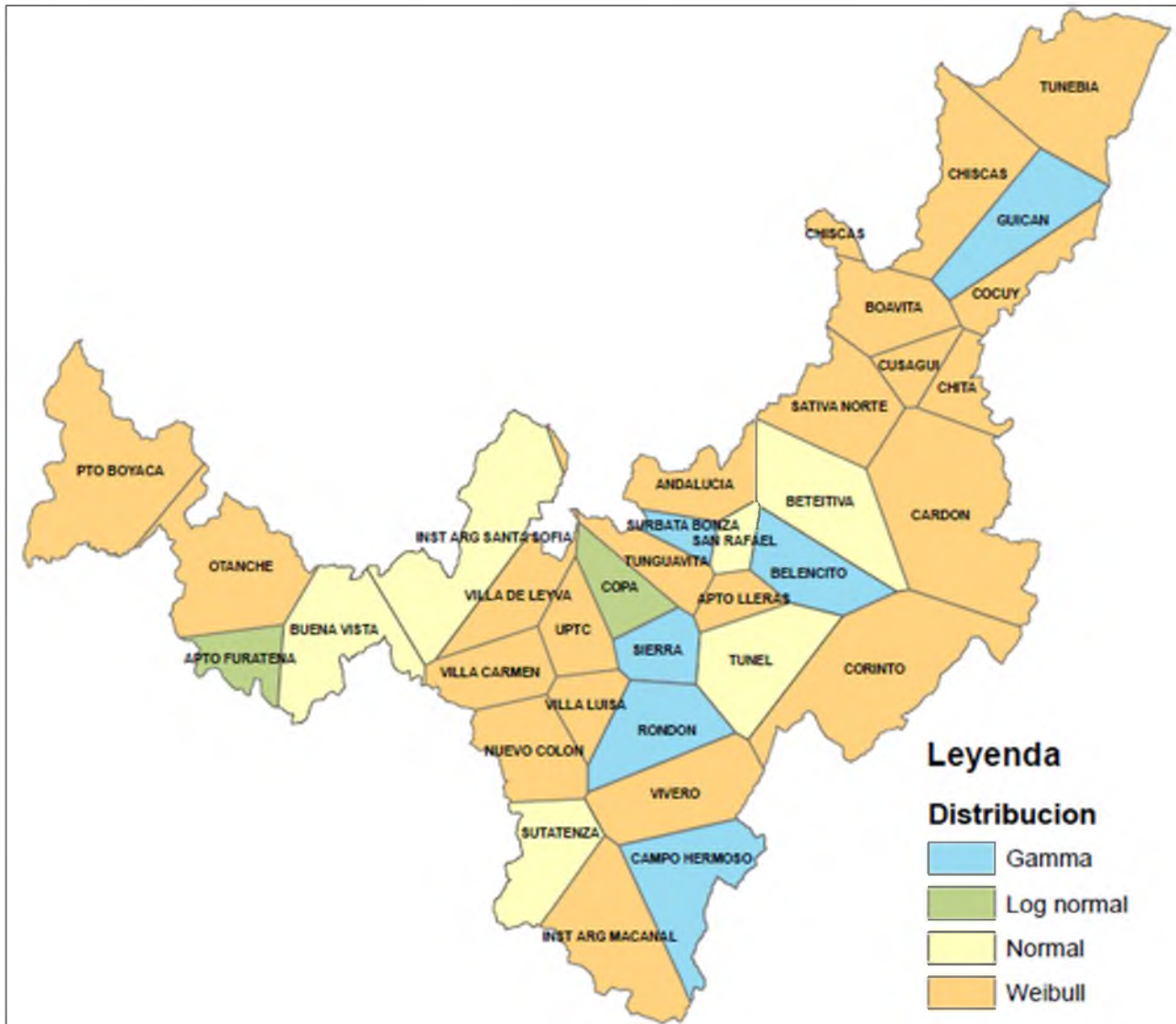
Continuación figura 1



Continuación figura 1



Fuente: Autores. Fecha 30 de noviembre de 2015.



Fuente: Autores. Fecha 30 de noviembre de 2015.

Figura 2. Área de influencia de distribuciones en el departamento de Boyacá.

4. Conclusiones

- Gráficamente las distribuciones Normal, Gamma y Log normal presentan comportamientos muy similares para la variable temperatura media; la distribución Weibull difiere de estos comportamientos tanto gráficamente así como cuando se aplican los criterios AIC y BIC, tal como se puede apreciar en la figura 1.
- El ajuste según el logaritmo de la función de máxima verosimilitud muestra que distribución Weibull es la más adecuada en la mayoría de las estaciones seguida de las funciones Normal y Gamma.
- Los métodos propuestos en el presente trabajo permiten ajustar y determinar de manera sencilla distribuciones probabilísticas continuas aplicadas a la variable de temperatura media, permitiendo futuras estimaciones y proyecciones de la variable en condiciones de incertidumbre.
- El 60 % de las estaciones presentan un ajuste a la distribución Weibull, el cual puede estar basado en la versatilidad de dicha función, la cual puede adoptar diferentes formas según sus parámetros.
- Las estaciones con menor cantidad de datos tienden ajustarse a distribuciones diferentes a la Weibull, por el contrario cuando las estaciones

presentan mayor cantidad de datos estas se ajustan mejor a la función de probabilidad Weibull.

El tamaño de la muestra influye directamente en la confiabilidad de los datos en el cual se trabajo con un mínimo de 18 y un máximo de 60 años para las estaciones analizadas.

5. Referencias bibliográficas

- Acuña, C., J.A.; Dominguez, C., A.H.; Toro, O., E.M. 2012 Una comparación entre métodos estadísticos clásicos y técnicas metaheurísticas en el modelamiento estadístico. *Scientia et Technica* Año XVII, No 50, Abril de 2012. Universidad Tecnológica de Pereira. ISSN 0122-1701
- Akaike. H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), pp. 716-723.
- Akaike. H. 1978. A Bayesian analysis of the minimum AIC procedure. *The Annals of Statistics*, 30(1), pp. 9-14, 1978.
- Aristizabal, R. J. 2012. Estimating the parameters of the three-parameter lognormal distribution. Florida International University. FIU Electronic Theses and Dissertations. Paper 575.
- Ayodele, T.R.; Ogunjuyigbe, A.S.O. 2015. Prediction of monthly average global solar radiation based on statistical distribution of clearness index. *Energy* 90. pp. 1733-1742.
- Behar, G., R.; Grima, C., P. 2004 55 Respuestas a dudas típicas de estadística. Ediciones Díaz de Santos, S. A.
- Bozdogan. H. 1987. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(3), pp. 345-379, 1987.
- Caballero, D.; F.F. 2011. Selección de modelos mediante criterios de información en análisis factorial. Aspectos teóricos y computacionales. Tesis Doctoral. Universidad de Granada. Departamento de Estadística e I.O. ISBN 978-84-694-5742-9.
- Chaouche, K.; Neppel, L.; Dieulin, C.; Pujol, N.; Ladouche, B.; Martin, E.; Salas, D.; Caballero, Y. 2010. Analyses of precipitation, temperature and evapotranspiration in a French Mediterranean region in the context of climate change. *C. R. Geoscience* 342. 234-243.
- Donat, M.G.; Alexander, L.V. 2012. The shifting probability distribution of global daytime and nighttime temperatures. *GEOPHYSICAL RESEARCH LETTERS*, VOL. 39, L14707, doi:10.1029/2012GL052459, 2012.
- Dorado, D., J.; Burbano, C., J.C.; Molina, T., J.M.; Carvajal, E., Y.; Aristizabal, H.F. 2006. Ajuste de modelos probabilísticos para el estudio de la variabilidad espacio-temporal de la precipitación: caso de estudio sistema Sara-Brut. *Meteorología Colombiana*. No. 10. pp. 60-75. Marzo 2006. ISSN 0124-6984.
- Hingray, B.; Mezghani, A.; Buishand, T.A. 2007. Development of probability distributions for regional climate change for uncertain global mean warming and uncertain scaling relationship. *Hydrology & Earth Systems Sciences*, 11(3), pp. 1097-1114.
- Kendall, M.G. 1975. Rank Correlation Methods, 4th edition, Charles Griffin, London.
- Mann, H.B. 1945. Non-parametric tests against trend, *Econometrica* 13, pp.163-171.
- Martínez, F., L. Métodos de inferencia para la distribución Weibull: aplicación en fiabilidad industrial. Trabajo fin de máster. Máster en Técnicas Estadísticas. Universidad de Vigo.
- Minka, T. P. 2002. Estimating a Gamma distribution. Microsoft Research, Cambridge, UK, Tech. Rep.
- Montesinos, L., A. 2011. Estudio del AIC y BIC en la selección de modelos de vida con datos censurados. Tesina de Maestría. Centro de Investigación en Matemáticas, A.C. Guanajuato, Gto.
- Moreno, R. S., Aguilar, A. R., Hernández, E. F., & Soto, F. P. 2011. Aplicaciones de la distribución weibull en ingeniería de confiabilidad. Memoria del XXI Coloquio Mexicano de Economía Matemática y Econometría, p. 148.
- Navarrete, L., C.F. 2008. Distribuição de probabilidade e dimensionamento amostral para tamanho de partícula em gramíneas forrageiras. Teses de Mestra. Universidade de São Paulo. Escola Superior de Agricultura Luiz de Queiroz.

Ramírez, Á. S. 2000. Criterios para la selección de modelos estadísticos. Banco Central de Costa Rica. DIE-NT-07-00.

Räisänen, J. 2009. Probability distributions of monthly-to-annual mean temperature and precipitation in a changing climate (CES Climate Modelling and Scenarios Deliverable D2.4, task I). Department of

Physics, P.O. Box 48, FI-00014 University of Helsinki, Finland. AVAILABLE FROM: http://www.atm.helsinki.fi/~jaraisan/CES_D2.4/CES_D2.4_task1.html

Schwarz. G. 1978. Estimating the dimension of a model. The Annals of Statistics, 6(2), pp. 461-464.