

Modelo de correlación múltiple para la determinación de variables estocásticas en función de coordenadas UTM

Campos Cedeño Antonio Fermín¹

¹Universidad Técnica de Manabí, Ecuador, Av. Urbina y Che Guevara, Portoviejo.
acampos@utm.edu.ec

Abstract— to determine the average annual rainfall at a specific site in a basin, it is necessary to count with rainfall information in the study area and topographic information. Counting with base information, is comes to the filling of them data missing with the help of any of them methods statistical existing for the effect. With the full series are determined values average annual rainfall for the rainfall stations studied. On the other hand, with the geographic location of the stations and maps of the areas of study Thiessen polygons are built. Then, with Thiessen polygons, study basin area and values of average annual rainfall determines the rainfall for the area of study. Based on this procedure, may be elaborated isohyet or curves of equal rainfall, that are an input that can help also to estimate of precipitation in a specific geographical site. Here is a methodology that allows creating with the help of geographic information systems (GIS), a mathematical model to estimate very accurately the values of rainfall based only on the geographical coordinates. To achieve this objective, the basins of the Hydrographic Demarcation of Manabí (Ecuador) have been chosen to develop the indicated mathematical model, which can be applied to other basins in the world.

Index Terms— multiple linear regression, mathematical model, geographical information systems, Hydrology, rainfall

I. INTRODUCCIÓN

Por un modelo matemático se refiere a una descripción, en lenguaje matemático, de un objeto que existe en un universo no matemático y que emplea algún tipo de simbolismo matemático para expresar relaciones, proposiciones sustantivas de hechos, variables, parámetros, entidades y relaciones entre variables.

Los modelos matemáticos se llaman "fenomenológicos", cuando tratan de eventos estocásticos de tipo hidrometeorológico y variables similares que se miden regularmente en estaciones y sitios especiales diseñados para este efecto.

En la investigación "*Fundamentación hidrológica para el control y optimización del uso de los recursos hídricos de la Demarcación Hidrográfica de Manabí*" fueron determinadas las precipitaciones medias anuales sobre la base de registros de las precipitaciones mensuales para el período 1963 – 2013, las cuales sirvieron de insumo para la elaboración del mapa de isoyetas [1, 2]. Dicho mapa sirve para poder de manera gráfica estimar las precipitaciones medias anuales en cualquier sitio de la Demarcación de Manabí, lo cual tiene un carácter subjetivo

con la probabilidad de tener errores en las estimaciones. Por este motivo, se ha elaborado un modelo matemático para la determinación de las mencionadas precipitaciones de una manera objetiva.

II. MATERIALES Y MÉTODOS

Se hace un análisis de modelación espacial con 3 variables: 1) *Precipitación*, variable dependiente z ; 2) *Longitud* en el sistema UTM WGS84, variable independiente x ; y, 3) *Latitud* en el sistema UTM WGS84, variable independiente y .

Las variables x y y se toman directamente de las cartas topográficas; y la variable z , se la determina con la ayuda de los elementos de sistemas de información geográfica (SIG).

Se toma como *base matemática de modelación* a la "correlación lineal múltiple" [3], con lo cual se tiene una ecuación con 3 variables y 3 coeficientes:

$$z = a + b_1x + b_2y \quad (1)$$

Donde, para este caso, a , b_1 , b_2 – son coeficientes a encontrar; z – precipitaciones medias anuales P , mm; x – longitud Este, m; y – latitud Norte, m.

Para obtener los coeficientes a , b_1 y b_2 de la ecuación de regresión lineal múltiple (1), mediante el método de los mínimos cuadrados, se precisa resolver las siguientes 3 ecuaciones:

$$\sum z = na + b_1 \sum x + b_2 \sum y \quad (2)$$

$$\sum xz = a \sum x + b_1 \sum x^2 + b_2 \sum xy \quad (3)$$

$$\sum yz = a \sum y + b_1 \sum xy + b_2 \sum y^2 \quad (4)$$

El sistema de ecuaciones (2), (3), (4) puede ser resuelto mediante la utilización de software especializado como son: AD+, SPSS, R, Minitab, Mathematica, Excel, entre otros.

El coeficiente de correlación múltiple R – es un parámetro adimensional, cuyo valor absoluto puede encontrarse entre 0 y 1. Mientras más se acerque a la unidad, mayor es el grado de asociatividad entre las variables; y, mientras se aproxima a cero,

la relación de lineal tiende a desaparecer. Dicho coeficiente se lo calcula con la fórmula [3, 5]:

$$R = \sqrt{\frac{r_{zx}^2 + r_{zy}^2 - 2r_{zx}r_{zy}r_{xy}}{1 - r_{xy}^2}} \quad (5)$$

Donde r_{zx} – coeficiente de correlación para las series z y x , r_{zy} – coeficiente de correlación para las series z y y , r_{xy} – coeficiente de correlación para las series x y y .

Otro parámetro estadístico que mide la correlación de datos es el **error estándar**, Mientras menor sea su valor, existe mejor ajuste de datos. Se lo estima con la fórmula:

$$s = \sqrt{\frac{\sum(Z - \check{Z})^2}{n - m - 1}} \quad (6)$$

Donde s - error estándar, Z – datos observados, \check{Z} – datos calculados con la fórmula obtenida, n - número de datos, m – número de variables independientes.

Para resolver el sistema de ecuaciones (2), (3) y (4), se debe calcular los siguientes determinantes:

$$D = \begin{vmatrix} n & \sum x & \sum y \\ \sum x & \sum x^2 & \sum xy \\ \sum y & \sum xy & \sum y^2 \end{vmatrix} \quad (7)$$

$$D_a = \begin{vmatrix} \sum z & \sum x & \sum y \\ \sum zx & \sum x^2 & \sum xy \\ \sum zy & \sum xy & \sum y^2 \end{vmatrix} \quad (8)$$

$$D_{b1} = \begin{vmatrix} n & \sum z & \sum y \\ \sum x & \sum zx & \sum xy \\ \sum y & \sum zy & \sum y^2 \end{vmatrix} \quad (9)$$

$$D_{b2} = \begin{vmatrix} n & \sum x & \sum z \\ \sum x & \sum x^2 & \sum zx \\ \sum y & \sum xy & \sum zy \end{vmatrix} \quad (10)$$

Finalmente, los coeficientes se calculan con las siguientes fórmulas:

$$a = \frac{D_a}{D}, b_1 = \frac{D_{b1}}{D}, b_2 = \frac{D_{b2}}{D} \quad (11)$$

Sistemas de Información Geográfica (SIG). La mayor parte de los elementos que existen en la naturaleza pueden ser

representados por figuras geométricas, tales como puntos, líneas y polígonos, es decir, **vectores**, o en forma de mallas con información (RASTERS). El vector y el raster (Fig. 1) son formas para representar al espacio, mismas que ayudan a interpretar a los elementos investigados en dependencia de su naturaleza.

Los principales elementos de los cuales están formados los SIG son de vectores y de rasters, mismos que se elaboran teniendo en cuenta ciertos principios y leyes de la geoinformática y de los métodos que se emplean en esta ciencia [6].

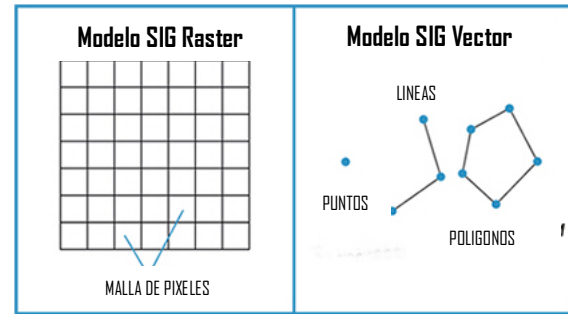


Fig. 1 — Raster y vector

Con los sistemas de información geográfica se encuentran estrechamente enlazadas las tecnologías de geoinformación, a las cuales se las puede definir como *el conjunto de programas y de medios tecnológicos para la obtención de nuevos tipos de información sobre el mundo circundante*. Las mencionadas tecnologías están desarrolladas para incrementar la efectividad de los procesos de dirección, conservación y presentación de la información, constituyéndose en un soporte fundamental en la toma de decisiones [4]. De esta forma, dentro las aplicaciones de código abierto, se tiene como principales **QGIS** y **GvSIG**, mismos que se pueden bajar gratuitamente de los sitios <http://www.qgis.org/en/site/> y <http://www.gvsig.com/en>, respectivamente. Por otro lado, en el mercado existe la aplicación comercial **ArcGis**, misma que tiene gran difusión a nivel mundial.

Información base. Para el desarrollo del modelo matemático; en calidad de información base se tiene: 1) Cartas digitales de la Demarcación Hidrográfica de Manabí con su correspondiente subdivisión en microcuencas según la metodología Pfafstetter, 2) Precipitaciones medias multianuales de 34 estaciones pluviométricas de la provincia de Manabí.

En el Fig. 2 se presentan la estaciones pluviométricas con sus códigos, precipitaciones medias anuales en milímetros, isoyetas y la codificación Pfafstetter de las microcuencas de la Demarcación de Manabí; y, en la Tabla 1 los principales datos de su ubicación geográfica.

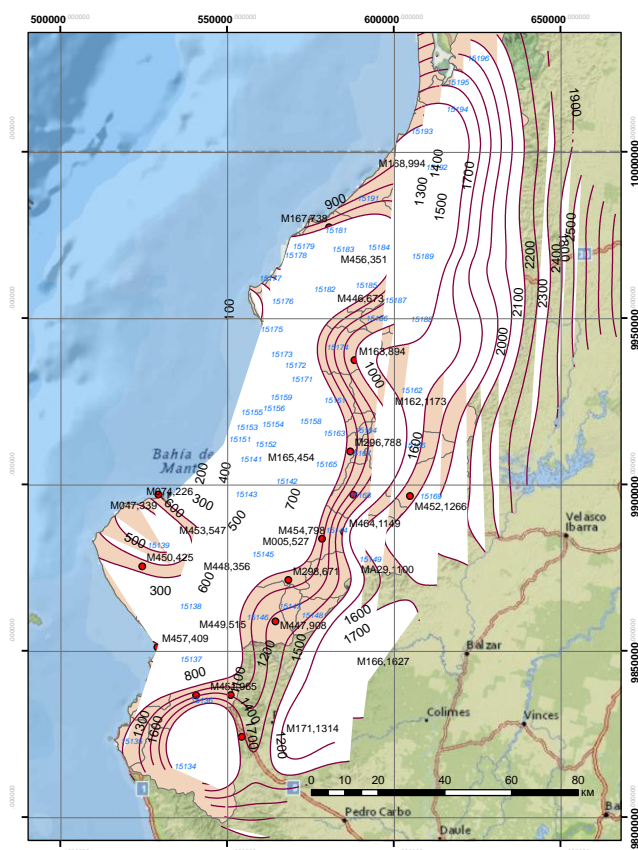


Fig. 2 — Precipitaciones medias anuales de la Demarcación Hidrográfica de Manabí

TABLE I
RAINFALL STATIONS, GEOGRAPHIC LOCATION, ANNUAL AVERAGE RAINFALL

Nº	Código estación	X-UTM	Y-UTM	Precipitación (mm)
1	M005	559523	9884982	528.0
2	M006	671167	9878373	2156.7
3	M026	684860	9947353	2768.8
4	M047	529608	9896745	397.2
5	M074	535232	9894995	270.2
6	M160	671939	9968948	2650.3
7	M162	599186	9922067	1233.4
8	M163	588400	9937145	1190.7
9	M165	561350	9905400	454.1
10	M166	587791	9845734	1657.2
11	M167	580800	9977125	778.8
12	M168	605098	9993552	1036.9
13	M169	540911	9836412	990.6
14	M171	566617	9823940	1308.6
15	M296	587159	9909725	847.6
16	M297	579744	9926307	705.8
17	M298	568607	9871041	859.8
18	M446	593441	9959038	767.5

19	M447	564710	9858637	1024.3
20	M448	541813	9872580	378.4
21	M449	545983	9860943	530.1
22	M450	524785	9875161	443.1
23	M451	551325	9836471	994.6
24	M452	605084	9896272	1472.6
25	M453	534613	9883481	609.9
26	M454	578716	9883443	891.3
27	M455	540758	9847496	459.0
28	M456	582963	9969540	480.3
29	M457	529325	9850844	416.0
30	M458	554628	9823913	1131.9
31	M459	545421	9825480	1671.5
32	M462	588084	9896706	1058.6
33	M464	585361	9885407	1234.5
34	MA29	589006	9876563	1287.8

Procesamiento de datos. Se precisa la obtención de una ecuación para cada microcuenca con el siguiente procedimiento de cálculo: 1) Con los datos de las precipitaciones medias anuales y la ubicación geográfica de las estaciones pluviométricas elaborar un mapa *raster de precipitación* para las Demarcación Hidrográfica de Manabí, 2) En el raster creado ubicar suficientes puntos arbitrarios de tal forma que se cubran las áreas de cada una de las 56 microcuencas, 3) Mediante las técnicas de interpolación de los SIG, para todos los puntos trazados, estimar los correspondientes valores de precipitación, 4) Para cada microcuenca exportar a formato Excel los datos tanto de precipitación, como de coordenadas UTM, de los puntos que se encuentran en su interior, 5) Mediante la correlación lineal múltiple obtener los coeficientes de la ecuación (1) para cada microcuenca.

III. RESULTADOS Y DISCUSIÓN

En la **Fig. 3** se presenta el raster de las precipitaciones medias anuales generado por el SIG, y en la **Tabla 2**, los valores de los coeficientes de la ecuación (1) para las 56 microcuencas para el cálculo de las precipitaciones medias anuales en cualquier lugar geográfico de la Demarcación Hidrográfica de Manabí.

Ejemplo de aplicación. Estimar la precipitación media anual para la microcuenca 15174 correspondiente a las coordenadas UTM 588567 E, 9938603 N.

De la **Tabla 2** tomamos los coeficientes a , b_1 y b_2 para la microcuenca 15174. La ecuación para la estimación de la precipitación sería:

$$z = 38688.983 + 0.0341316x - 0.0057961y \quad (12)$$

Reemplazando los valores de las coordenadas dadas se obtiene que la precipitación es igual a 1173 mm, lo cual se puede corroborar gráficamente con las isoyetas (**Fig. 2**).

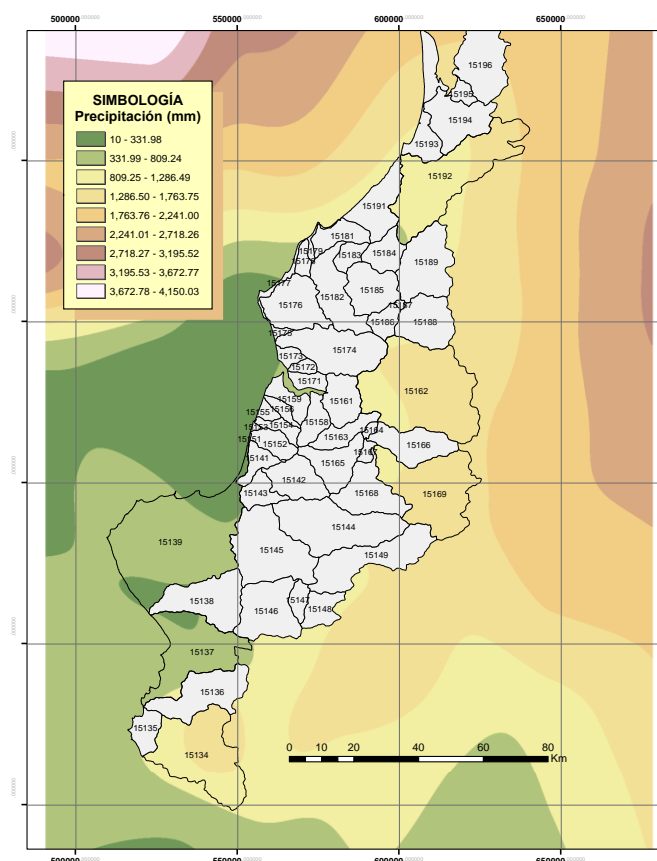


Fig. 3 — Raster de precipitaciones medias anuales de la Demarcación Hidrográfica de Manabí

TABLE II

VALORES DE LOS COEFICIENTES DE LAS ECUACIONES PARA DETERMINAR LAS PRECIPITACIONES MEDIAS ANUALES DE LA DEMARCACIÓN HIDROGRÁFICA DE MANABÍ

Nº	Código de cuenca Pfafstetter	a	b_1	b_2	R
1	15134	-256793.99	0.0203	0.0252	0.817
2	15135	34391.54	0.0198	-0.0045	0.962
3	15136	481201.28	0.0205	-0.0500	0.928
4	15137	210107.75	0.0083	-0.0217	0.935
5	15138	53644.13	0.0140	-0.0062	0.900
6	15139	-21570.60	-0.0066	0.0026	0.895
7	15141	3990.94	0.0238	-0.0017	0.996
8	15142	31060.12	0.0093	-0.0036	0.934
9	15143	30130.91	0.0188	-0.0041	0.970
10	15144	63627.57	0.0329	-0.0083	0.981
11	15145	137089.49	0.0129	-0.0145	0.985
12	15146	63761.55	0.0208	-0.0076	0.973
13	15147	146077.05	0.0043	-0.0150	0.997
14	15148	171370.71	0.0184	-0.0183	0.993
15	15149	-18891.26	0.0207	0.0008	0.976

16	15151	-29460.50	0.0194	0.0019	1.000
17	15152	49671.91	0.0129	-0.0057	0.986
18	15153	-13352.34	0.0150	0.0005	0.999
19	15154	47419.94	0.0093	-0.0053	0.996
20	15155	20142.53	0.0094	-0.0025	0.999
21	15156	26278.35	0.0083	-0.0031	1.000
22	15158	-18859.25	0.0139	0.0012	0.908
23	15159	-12273.00	0.0122	0.0006	0.993
24	15161	-186150.62	0.0355	0.0167	0.997
25	15162	4857.58	0.0131	-0.0012	0.924
26	15163	-71330.30	0.0278	0.0056	0.991
27	15164	-75154.79	0.0310	0.0058	1.000
28	15165	27703.66	0.0234	-0.0041	0.982
29	15166	28108.33	0.0223	-0.0041	0.997
30	15167	26903.27	0.0308	-0.0045	0.999
31	15168	162597.75	0.0286	-0.0180	0.986
32	15169	65321.57	0.0181	-0.0076	0.975
33	15171	-129670.51	0.0277	0.0115	0.996
34	15172	-58369.00	0.0303	0.0042	0.997
35	15173	52136.81	0.0305	-0.0069	0.998
36	15174	38688.98	0.0341	-0.0058	0.993
37	15175	103563.83	0.0326	-0.0122	0.999
38	15176	10462.80	0.0280	-0.0026	0.954
39	15177	-217037.38	0.0226	0.0205	0.987
40	15178	-372721.08	0.0005	0.0374	0.996
41	15179	-446460.64	-0.0146	0.0457	0.999
42	15181	-411614.93	-0.0134	0.0421	0.992
43	15182	82896.64	0.0286	-0.0099	0.864
44	15183	-202158.78	0.0011	0.0203	0.990
45	15184	-34165.33	0.0218	0.0022	0.972
46	15185	246825.88	0.0237	-0.0261	0.989
47	15186	265622.65	0.0150	-0.0275	0.996
48	15187	222396.95	0.0212	-0.0235	1.000
49	15188	75603.58	0.0126	-0.0082	0.954
50	15189	93988.42	0.0298	-0.0111	0.989
51	15191	-247398.93	-0.0100	0.0255	0.982
52	15192	-84988.51	0.0175	0.0075	0.978
53	15193	-194177.77	-0.0077	0.0200	0.995
54	15194	-162562.75	0.0126	0.0156	0.950
55	15195	-169326.79	0.0090	0.0165	0.988
56	15196	-153805.63	0.0111	0.0148	0.974

LIST OF SYMBOLS

- a Constante de correlación múltiple
- b_1 Constante de correlación múltiple
- b_2 Constante de correlación múltiple
- D Determinante

m	Número de variables independientes
n	Número de datos analizados
r_{zx}	Coefficiente de correlación parcial
R	Coefficiente de correlación múltiple
s	Error estándar
x	Coordenada UTM – Este
y	Coordenada UTM – Norte
z	Precipitación media anual

IV. CONCLUSIONES

A. Por primera vez se han obtenido 56 ecuaciones para la estimación de las precipitaciones medias anuales en función de las coordenadas geográficas.

B. Se considera a las ecuaciones obtenidas muy confiables puesto que se tienen elevados valores de coeficientes de correlación múltiple en los límites de 0.8170 – 0.9997.

V. REFERENCIAS Y BIBLIOGRAFÍA

[1] Campos A.F. Runoff of the hydrographic demarcation of Manabi (Ecuador)/ // Proceedings of the VIII International Scientific and Practical Conference "Engineering Systems - 2015". — 2015.pp. 291-295.

[2] Campos, A. F., Sinichenko, E. K., Gritsuk, I. I. Morphometric characterization parameters of river basin of Manabi (Ecuador)/ // VESTNIK № 1 - Peoples' Friendship University of Russia. — 2015.pp. 10.

[3] Rummyantsev, V. K. Hydrological calculations in hydrotechnical constructions /. — Moscú: Peoples' Friendship University of Russia, 1992. — pp. 100.

[4] Zhurkin, I. G., Shaytura, S. V. Sistemas de Información Geográfica /. — Moscú: KUDITS-PRESS, 2009. — pp. 312.

[5] Universidad de Córdoba. Correlación múltiple y correlación canónica [Electronic Resource] / Universidad de Córdoba. — 2015. — Access mode: http://www.uco.es/zootecniaygestion/img/pictorex/06_19_26_8_correlacion_multiple.pdf.

[6] Wikimedia, F. Sistemas de Información Geográfica [Electronic Resource] / Wikimedia, F. — 2015. — Access mode: <http://dic.academic.ru/dic.nsf/ruwiki/22609>.



Campos Cedeño A.F. Doctor of Philosophy in Technical Sciences, Master of Sciences in Ingeniería, Magister en Administración Ambiental, Ingeniero Civil, Naturópata, Docente Principal de la Universidad Técnica de Manabí por 21 años. Ha impartido e imparte disciplinas relacionadas con la Mecánica de Fluidos, Hidráulica e Hidrología de Ingeniería.

Autor de varios libros y artículos científicos para revistas indexadas de la Federación Rusa y de Ecuador.