

Estimación de datos faltantes de precipitación por el método de regresión lineal: Caso de estudio Cuenca Guadalupe, Baja California, México

Estimation rainfall missing values by linear regression method: Case study Guadalupe Basin, Baja California, Mexico

Claudia S. Herrera-Oliva*✉, José R. Campos-Gaytán*,
Fátima M. Carrillo-González**

Herrera-Oliva, C. S., Campos-Gaytán, J. R., & Carrillo-González, F. M. (2017). Estimación de datos faltantes de precipitación por el método de regresión lineal: Caso de estudio Cuenca Guadalupe, Baja California, México. *Investigación y Ciencia de la Universidad Autónoma de Aguascalientes*, 25(71), 34-44.

RESUMEN

Para completar los datos faltantes en los registros de la precipitación pluvial anual reportados por 13 estaciones climatológicas distribuidas en el área de la Cuenca Guadalupe se realizó un análisis de regresión lineal entre estaciones cercanas. Para determinar la utilidad de la inferencia estadística, se calculó el coeficiente de correlación lineal (r), en todos los casos se obtuvo un alto valor que en promedio fue $\bar{r} = 0.89$, también se calculó la eficiencia estadística (E), la cual en todos los casos analizados sugiere la viabilidad de la inferencia estadística. Como resultado principal de este análisis se presenta una base de datos de precipitación pluvial completa para el periodo 1948-2012.

Palabras clave: Cuenca Guadalupe; precipitación pluvial; completación de datos; análisis de regresión y correlación lineal; inferencia estadística; eficiencia estadística.

Keywords: Guadalupe Basin; rainfall; completion values; regression analysis and linear correlation; statistical inference; efficiency statistics.

Recibido: 12 de enero de 2017, aceptado: 19 de junio de 2017

* Facultad de Ingeniería, Arquitectura y Diseño, Universidad Autónoma de Baja California. Av. Álvaro Obregón y Julián Carrillo s/n, Colonia Nueva, C. P. 21100, Mexicali, Baja California, México. Correo electrónico: cherrera@uabc.edu.mx; rcampos@uabc.edu.mx

** Departamento de Ciencias Exactas, Universidad de Guadalajara Centro Universitario de la Costa. Av. Universidad 203, Delegación Ixtapa, C. P. 48280, Puerto Vallarta, Jalisco, México. Correo electrónico: fmacielux@gmail.mx

✉ Autor para correspondencia

ABSTRACT

To complete the missing values in the records of the annual rainfall reported by 13 climatological stations distributed in the area of the Guadalupe Basin, an analysis of linear regression between nearby stations was done. In order to determine the convenience of statistical inference, the coefficient of linear correlation (r) was calculated, in all cases a high value was obtained, which on average resulted a $\bar{r} = 0.89$. In addition, the efficiency statistics (E) was calculated, which in all the analyzed cases, suggests the feasibility of statistical inference. As the main result of this analysis a complete precipitation database for the period 1948-2012 is presented.

INTRODUCCIÓN

La precipitación pluvial se considera como la variable principal en los estudios hidrogeológicos, ya que es la fuente fundamental para el cálculo de balances hídricos y la generación de alertas tempranas por riesgo de sequía en la región. En varias investigaciones el punto de partida es la estimación de la lluvia con adecuada resolución espacial y temporal (Luna Romero, & Lavado Casimiro, 2015; Schuurmans, & Bierkens, 2007; Tapiador, Kidd, Levizzani, & Marzano, 2003; Tapiador et al., 2012).

El estado de Baja California, México, se caracteriza por poseer escasos recursos hidráulicos superficiales y una baja precipitación pluvial, sólo en una pequeña porción de su territorio se presentan

lluvias en condiciones normales que varían de 200 a 300 mm al año, mientras en el resto las precipitaciones disminuyen significativamente a 50 mm. Aunado a lo anterior, en las últimas décadas se han incrementado las actividades agrícolas, urbanas e industriales en la región, lo que implica una creciente demanda de agua, por lo que el desarrollo económico de la zona depende de su disponibilidad subterránea (Campos-Gaytán, Kretzschmar, & Herrera-Oliva, 2014).

La Cuenca Guadalupe se ubica al noroeste del estado de Baja California, en ella se localizan dos zonas agrícolas de gran importancia para la economía local. El Valle de Guadalupe se considera como la región productora de vinos más importante en México (Plata Caudillo, 2010) y el Valle de Ojos Negros de productos de agricultura de riego como alfalfa, cebollín, cebolla y sandía, entre otros (Pineda Villa, 2000). Ambos dependen completamente del agua subterránea. Sin embargo, la naturaleza esporádica de las precipitaciones y de la consecuente escasa recarga del acuífero, aunadas a su extracción para la agricultura y abastecimiento de los asentamientos urbanos en los propios valles, así como el suministro para la vecina ciudad de Ensenada, han convergido para crear una crisis en el abasto (Campos-Gaytán, 2008). Dada la complejidad de este problema, es necesario analizar la precipitación pluvial en el área de la cuenca para contribuir al conocimiento de su distribución y disponibilidad, con ello se podrán tomar las medidas necesarias para su mejor aprovechamiento.

Para realizar el análisis de la precipitación se requiere contar con una base de datos continuos, homogéneos y que abarquen el máximo intervalo temporal posible, en este trabajo el periodo de análisis fue de 1948-2012. Lamentablemente la base de datos con que se cuenta presenta importantes huecos de información debidos a la ausencia de lectura, falla del instrumento de registro, error de transcripción, etcétera, lo que limita su análisis y constituye una fuente de error, dado que las conclusiones de cualquier estudio que se realice sobre esos datos serán erróneas.

Existen métodos para efectuar el completado de valores faltantes en una serie pluviométrica. En la guía de prácticas climatológicas de la Organización Meteorológica Mundial, en inglés llamada *World Meteorological Organization* (WMO, 1983) se proponen métodos estadísticos para el relleno de valores faltantes, como regresión lineal, de la razón

y la razón normal, propuestos por Paulhus y Kohler (1952). Alfaro y Pacheco (2000) presentaron un estudio en el que aplican los de regresión, de la razón, de la razón ajustada y de la razón-normal a datos anuales de precipitación y concluyeron que el mejor es el de la regresión múltiple y que las diferencias máximas más altas se dieron con el de la razón.

Young (1992) presenta una modificación al de la razón normal, donde la misma, ponderada, es cambiada por la correlación entre las estaciones, pero Yozgatligil, Aslan, Lyigun y Batmaz (2013) reportaron que el promedio aritmético simple y la correlación entre las estaciones resultan idénticos en algunos casos, si las correlaciones de estaciones de referencia son casi las mismas. McCuen (1998) recomienda el promedio aritmético simple, cuando el valor anual en cada uno de los datos a promediar difiere por lo menos 10%, pero Yozgatligil et al. (2013) presenta estimaciones fiables si la variable no tiene variabilidad espacial y si las estaciones de referencia están altamente correlacionadas.

Aparicio (2011) y Campos Aranda (1998) indican que este método puede emplearse cuando se basa en registros simultáneos de tres estaciones que se encuentren lo más cerca posible a la estación en estudio. Por su simplicidad, porque no requiere de software específico para la estimación, por la bondad y claridad de sus resultados, el método de regresión lineal es uno de los más utilizados para estimar los valores de datos faltantes de precipitación y temperatura (Eischeid, Pasteris, Diaz, Plantico, & Lott, 2000; Hubbard, 2001; Wade, 1987).

DeGaetano, Eggleston y Knapp (1995), Eischeid, Bruse, Karl y Díaz (1995), Kashani y Dinpas-hoh (2012), Kemp, Burnell, Everson y Thomson (1983), Presti, Barca y Passarella (2010), Xia, Fabian, Stohl y Winterhalter (1999), You, Hubbard y Goddard (2008), así como Young (1992) manifiestan que el método de regresión simple es superior entre los tradicionales para las variables temperatura mínima, máxima y precipitación en diferentes condiciones climáticas. Por esta razón en este trabajo se utilizó el de regresión lineal, con el propósito de completar la base de datos de la precipitación pluvial registrada en la Cuenca Guadalupe, y con ello contribuir en la realización de futuras investigaciones en diversas áreas, como son la agronomía, hidrología y climatología de la zona de estudio. Para ello se emplearon los registros de la precipitación anual, reportados por

13 estaciones climatológicas que cubren el área de la cuenca. Para determinar lo conveniente de la inferencia estadística, se calcularon el coeficiente de correlación lineal y la eficiencia estadística.

MATERIALES Y MÉTODOS

Zona de estudio

La Cuenca Guadalupe se localiza en el noroeste del estado de Baja California, aproximadamente a 37 km de la ciudad de Ensenada, entre los paralelos 31° 51' y 32° 15' de latitud norte, y los meridianos 115° 52' y 116° 51' de longitud oeste; colinda al norte con la subcuenca Las Palmas, al sur con las subcuencas Ensenada y Maneadero, al este con la subcuenca Laguna Salada, al oeste con el Océano Pacífico (Beltrán-Gómez, 2001) y cuenta con una superficie total hasta su desembocadura en el mar de aproximadamente 2,400 km² (Hernández-Rosas, & Mejía-Vázquez, 2003), como se puede ver en la figura 1.

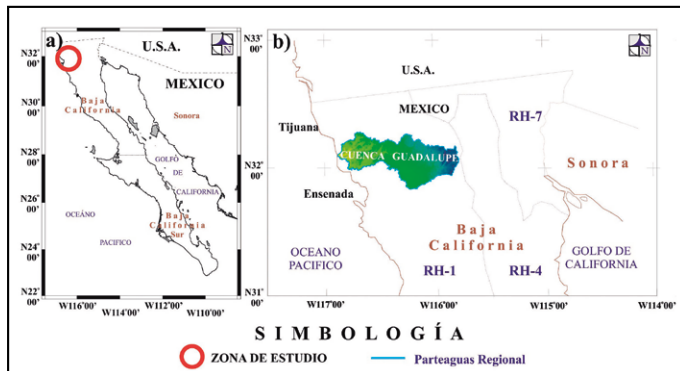


Figura 1. Localización de la zona de estudio: (a) En escala estatal. El círculo marca la zona de estudio. (b) División hidrológica de la parte norte del estado de Baja California y ubicación de la Cuenca Guadalupe. Elaboración propia.

Los escurrimientos superficiales de la Cuenca Guadalupe tienen su origen en la Sierra Juárez, pasan a través de los valles de Ojos Negros y Guadalupe y terminan en el Océano Pacífico a la altura del poblado La Misión (figura 2).

Climatología y base de datos

De acuerdo con la clasificación climática de Köppen, modificada por García (1981), el clima en la zona de estudio es semiárido, subtipo moderado, con precipitaciones principalmente en invierno (diciembre a marzo), comúnmente conocido como tipo mediterráneo, que se establece en gran

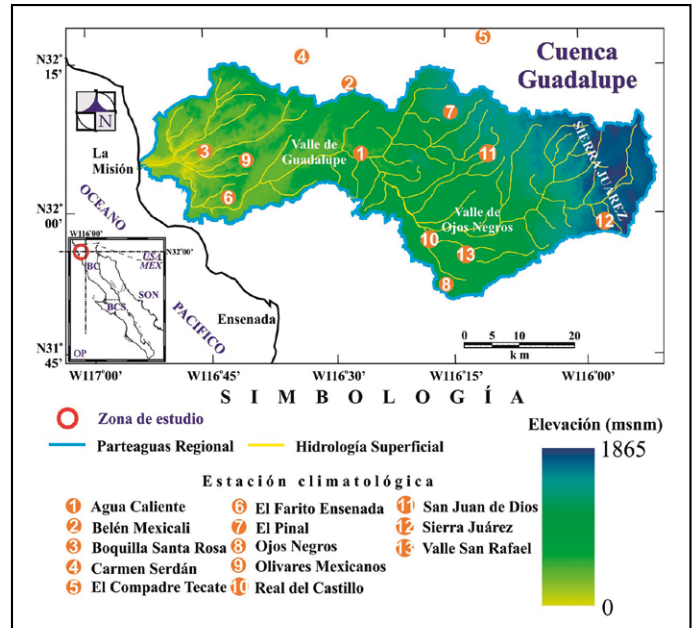


Figura 2. Cuenca Guadalupe: Imagen de la cuenca hidrográfica, la hidrología superficial y las estaciones climatológicas. Elaboración propia.

parte de la costa oeste y norte de Baja California y en la costa del sur de California. La información utilizada en este trabajo es parte del registro de datos monitoreados por las estaciones: 1. Agua Caliente, 2. Belén Mexicali, 3. Boquilla Santa Rosa, 4. Carmen Serdán, 5. El Compadre Tecate, 6. El Farito Ensenada, 7. El Pinal, 8. Ojos Negros, 9. Olivares Mexicanos, 10. Real del Castillo, 11. San Juan de Dios Norte, 12. Sierra Juárez, y 13. Valle de San Rafael, del Extractor Rápido de Información Climatológica (ERIC III), el cual proporciona información contenida en el banco de datos histórico nacional del Servicio Meteorológico Nacional (SMN) de la Comisión Nacional del Agua (CONAGUA), que se enlistan en la tabla 1.

Con el propósito de identificar la existencia de valores atípicos en las series pluviométricas, se utilizó el diagrama de caja y bigotes para cada una de las series de datos reportados por las 13 estaciones que aparecen en la tabla 1. Del análisis de los diagramas se observó que sólo las estaciones Belén Mexicali, Boquilla Santa Rosa y El Pinal presentaron escasos valores atípicos, aproximadamente 2.0. Dado que el método de regresión lineal requiere series largas para el análisis, se decidió utilizar la serie completa.

Método de correlación lineal

El método de correlación lineal es uno de los más

Tabla 1
Coordenadas de las estaciones climatológicas

No	ESTACIÓN	Coordenadas UTM		
		Norte x	Oeste y	Altitud m.s.n.m.
1	Agua Caliente	551894.33	3551652.39	410
2	Belén Mexicali	548781.00	3551637.00	555
3	Boquilla Santa Rosa	523596.68	3547889.51	40
4	Carmen Serdán	539288.10	3566338.72	455
5	El Compadre Tecate	570190.00	3578090.00	1162
6	El Farito Ensenada	531459.51	3538599.80	250
7	El Pinal	567875.43	3560614.76	1320
8	Ojos Negros	569336.37	3525927.96	720
9	Olivares Mexicanos	531436.63	3546026.32	351
10	Real del Castillo	566155.33	3535107.45	745
11	San Juan de Dios	574728.00	3553956.00	1280
12	Sierra Juárez	599334.00	3541320.00	1545
13	Valle San Rafael	572513.40	3531492.44	721

Nota: Elaboración propia.

utilizados, se recomienda para la estimación de datos mensuales y anuales de la estación en estudio, y los de una pluviométrica cercana, que cuente con una estadística consistente y observada. Para ello se requiere establecer una regresión y correlación lineal entre una estación patrón y la que tenga carencia de datos, mediante una ecuación lineal. Son herramientas estadísticas comúnmente empleadas en hidrología para estimar datos faltantes y ampliar el registro de cierta estación climatológica, con base en la información disponible en las más cercanas. Cuanto más amplio sea el registro o serie de valores observados en otra cercana, mayores serán las estimaciones e inferencias estadísticas basadas en tales datos.

Campos Aranda (1998) menciona que el uso de las herramientas mencionadas es particularmente recomendable para incrementar los registros anuales, ya sea de lluvias, temperaturas o escurrimientos. Sin embargo, debe mencionarse que una de sus principales limitaciones es que tanto la recta de regresión como el coeficiente de correlación no son robustos, en el sentido de que resultan muy afectados por medidas particulares que se alejan mucho de la tendencia general.

El método empleado en el presente trabajo establece que dados dos registros (x_i, y_i) de manera

que x contenga más valores que y , si se desea estimar los valores faltantes de y a partir de los valores de x . Suponiendo que se tienen k parejas de valores (x_i, y_i) y además $(n - k)$ valores de y_i , entonces primero se calcula el coeficiente de correlación (r) de las k parejas definida por la ecuación (1). El coeficiente de correlación está definido en el intervalo $-1 \leq r \leq 1$, que denota una correlación total, cuando adopta el valor de cero se considera que la correlación es nula.

$$r = \frac{S_{xy}}{\sqrt{S_x^2 S_y^2}} \quad (1)$$

donde:

$$S_{xy} = \frac{1}{N} \sum x_i y_i - (\bar{x})(\bar{y}) \quad (2)$$

$$S_x^2 = \frac{\sum x_i^2}{N} - (\bar{x})^2 \quad (3)$$

$$S_y^2 = \frac{\sum y_i^2}{N} - (\bar{y})^2 \quad (4)$$

$$\bar{x} = \frac{\sum x_i}{N} \quad (5)$$

$$\bar{y} = \frac{\sum y_i}{N} \quad (6)$$

Si se toma en cuenta que el valor del coeficiente de correlación (r) se calcula con base en una muestra relativamente pequeña, se considera que es únicamente una estimación del valor poblacional (ρ) y, por tanto, es necesario investigar si existe la posibilidad de que $\rho = 0$. Para ello se emplea el estadístico Z dado por la siguiente expresión:

$$Z = \frac{\sqrt{n-3}}{2} \ln \left(\frac{1+r}{1-r} \right) \quad (7)$$

y se compara con el valor de Z_c que tiene distribución normal para cierto nivel de confianza, comúnmente 95%, para el cual $Z_c = 1.645$. Si $Z \geq Z_c$, no hay posibilidad de que (ρ) sea igual a cero y, por tanto, r es significativo a 5%, como lo establece Campos Aranda (1998).

Si el coeficiente de correlación de las k parejas resultó mayor o igual a 0.8, como en Pizarro, González, Witterssheim, Saavedra y Soto (1993), se considera aceptable inferir los $(n - k)$ valores faltantes de y , con una relación lineal definida por la ecuación (8), en la cual los parámetros m y b se evalúan con las k parejas de valores comunes x y y , según las ecuaciones (9) y (10).

$$y = b + mx \quad (8)$$

donde:

$$m = \frac{S_{xy}}{S_x^2} \quad (9)$$

$$b = \bar{y} - m\bar{x} \quad (10)$$

Aun cuando la correlación entre los registros x y y sea alta, es recomendable investigar hasta qué punto es provechoso utilizar el registro más amplio para incrementar el otro, dado que puede suceder que los valores inferidos discrepen mucho de la realidad. La eficiencia estadística (E) ayuda a determinar si mejora el valor medio del registro y , y con ello si conviene hacer la inferencia. Si el valor de (E) resulta mayor que uno, no mejorará en nada el registro de las y_i , en cambio, si (E) es menor que la unidad, convendrá hacer la inferencia de los valores faltantes a partir de las x_i . La eficiencia estadística se calcula con la siguiente expresión contenida en Campos Aranda (1998):

$$E = 1 - r^2 \left(\frac{n-m}{n} \right) + \frac{n-m}{n(m-3)} (1 - r^2) \quad (11)$$

donde:

m = número total de datos de y ; n = número total de datos de x ; r = coeficiente de correlación.

De acuerdo con Cárdenas Montoya (1994), los métodos de regresión no consideran las características físicas y geomorfológicas de las cuencas hidrográficas y sus implicaciones sobre los resultados obtenidos, solo hacen consideraciones estadísticas. No obstante, Allen, Pereira, Raes y Smith (2006) recomiendan que para la aplicación de este método se seleccione una serie de datos con un comportamiento similar, esto es, dentro de la misma área de influencia topoclimática, a la serie que tiene los datos faltantes.

RESULTADOS

Como se mencionó anteriormente, con objeto de completar los datos faltantes en los registros de precipitación pluvial de 13 estaciones climatológicas distribuidas en el área de la Cuenca Guadalupe, en el presente trabajo se realizó un análisis de regresión lineal entre estaciones cercanas y para determinar la conveniencia de la inferencia estadística, se calcularon el coeficiente de correlación lineal y la eficiencia estadística, tales resultados se presentan a continuación.

En la figura 3 se muestran los diagramas de dispersión de los datos de precipitación reportados por pares de estaciones cercanas y sus correspondientes rectas de regresión. En todos los casos el ajuste lineal posee valores de $m > 0$, lo que sugiere que existe una fuerte correlación positiva entre las variables, a pesar de que las muestras analizadas son pequeñas. En la tabla 2 se presentan los resultados de los parámetros estadísticos que sugieren que es recomendable hacer la inferencia estadística de los datos faltantes mediante el análisis de regresión lineal.

Una vez realizado el ajuste de regresión lineal, es importante disponer de una medida que pondere la bondad del ajuste y permita decidir si el ajuste lineal es suficiente o se deben buscar modelos alternativos. Como medida de bondad del ajuste se utiliza el coeficiente de determinación r^2 , que indica la proporción del ajuste que se ha conseguido con

Tabla 2
Parámetros estadísticos

x_i	y_i	r	r^2	k	$\geq Zc$	E
Boquilla Santa Rosa	Agua Caliente	0.90	0.81	23	6.6	0.0999
Agua Caliente	Boquilla Santa Rosa	0.91	0.83	23	6.6	0.0297
El Pinal	Belén Mexicali	0.82	0.67	24	5.3	0.1083
Agua Caliente	Carmen Serdán	0.87	0.76	13	4.2	0.1376
Boquilla Santa Rosa	El Farito Ensenada	0.95	0.90	7	3.7	0.0810
Boquilla Santa Rosa	El Compadre Tecate	0.80	0.64	15	3.8	0.1839
Ojos Negros	El Pinal	0.82	0.67	23	5.1	0.1351
Boquilla Santa Rosa	Olivares Mexicanos	0.90	0.81	33	8.0	0.0219
Agua Caliente	Ojos Negros	1.00	1	33	-----	0.0000
Ojos Negros	Real del Castillo	0.95	0.90	11	5.2	0.0869
Ojos Negros	San Juan de Dios	0.92	0.85	25	7.4	0.0691
San Juan de Dios	Sierra Juárez	0.92	0.85	21	6.7	0.0816
Ojos Negros	Valle San Rafael	0.80	0.64	19	4.39	0.1988

Nota: Estos datos son el resultado del análisis de correlación lineal, eficiencia estadística entre estaciones cercanas. Elaboración propia.

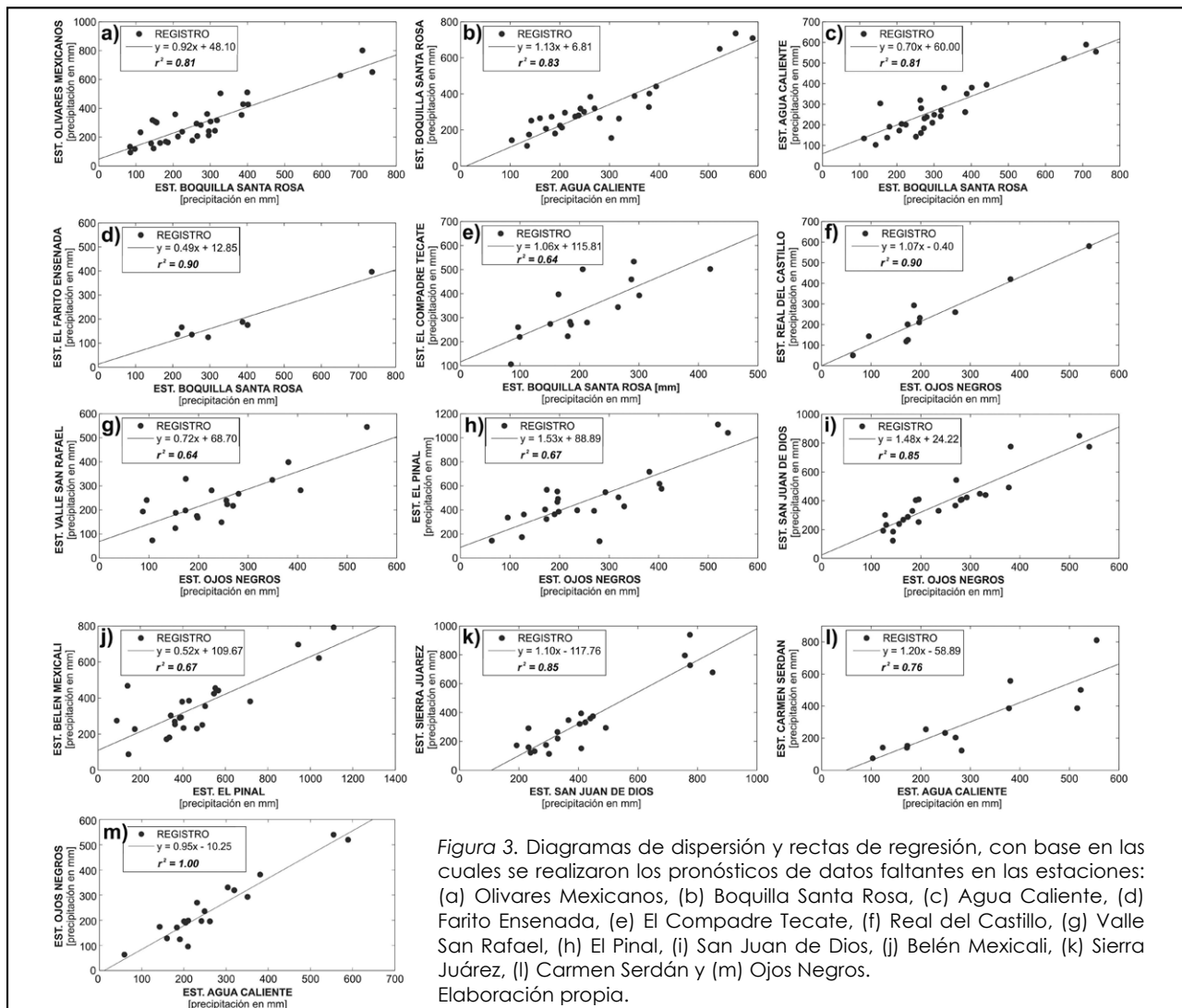


Figura 3. Diagramas de dispersión y rectas de regresión, con base en las cuales se realizaron los pronósticos de datos faltantes en las estaciones: (a) Olivares Mexicanos, (b) Boquilla Santa Rosa, (c) Agua Caliente, (d) Farito Ensenada, (e) El Compadre Tecate, (f) Real del Castillo, (g) Valle San Rafael, (h) El Pinal, (i) San Juan de Dios, (j) Belén Mexicali, (k) Sierra Juárez, (l) Carmen Serdán y (m) Ojos Negros. Elaboración propia.

el modelo lineal. Cuanto más se acerque r^2 a 1, mejor es la estimación.

En la tabla 3 se presentan los datos de la precipitación pluvial anual observada en la Cuenca Guadalupe durante el periodo 1948-2012, monitoreada por las 13 estaciones ya mencionadas; asimismo, se muestran las estimaciones realizadas. Como puede observarse, esta información presenta

importantes huecos de información, en algunas de las estaciones el periodo de registro está por debajo de la mitad del periodo de años analizado en este trabajo, tal es el caso de las estaciones Carmen Serdán, El Compadre Tecate, El Farito Ensenada y Real del Castillo, que tan solo reportan 18, 18, 9 y 11 años, respectivamente, de los 64 años que comprende todo el periodo analizado.

Tabla 3.

Datos completos de la precipitación anual y estimaciones, en la cuenca Guadalupe, en el periodo 1948-2012

	1	2	3	4	5	6	7	8	9	10	11	12	13
AÑO	AC	BM	BSR	CS	ECT	EFE	EP	ON	OM	RC	SJ	SJD	VSR
1948	118.3	197.9	83.3	83.1	203.3	52.8	144.0	36.0	83.3	88.7	101.2	77.1	94.6
1949	188.9	237.3	184.2	167.8	282.2	102.3	463.0	244.5	184.2	196.7	300.4	383.6	244.7
1950	129.7	204.3	99.5	96.7	219.4	60.8	231.9	93.5	99.5	106.1	58.5	161.7	136.0
1951	261.1	277.5	287.3	254.5	458.8	152.8	615.8	344.4	287.3	307.0	460.5	530.5	316.7
1952	353.9	329.1	419.8	365.8	502.4	217.7	645.5	363.8	419.8	448.8	491.6	559.0	330.6
1953	118.2	197.9	83.2	83.0	203.2	52.8	231.2	93.0	83.2	88.6	57.7	160.9	135.7
1954	190.1	237.9	185.9	169.3	270.5	103.1	500.3	268.9	163.0	174.0	339.5	419.5	262.3
1955	165.6	224.3	150.9	139.9	273.4	85.9	283.0	126.9	309.1	330.3	112.0	210.8	160.1
1956	119.4	198.5	84.8	84.4	105.8	53.6	190.6	66.5	94.0	100.2	15.2	122.0	116.6
1957	264.0	279.1	291.4	257.9	533.0	154.8	517.7	280.3	360.8	385.7	324.2	405.5	270.5
1958	203.8	245.5	205.4	185.7	501.0	112.6	504.3	271.5	358.2	382.9	474.8	543.6	264.2
1959	127.8	203.2	96.8	94.4	260.0	59.4	309.2	144.0	118.0	125.9	84.4	185.5	172.4
1960	175.4	229.7	164.8	151.6	396.7	92.8	340.7	164.6	159.4	170.2	174.5	268.1	187.2
1961	118.7	198.2	83.9	83.6	203.9	53.1	327.9	156.2	133.5	142.4	421.8	239.0	181.2
1962	161.2	221.8	144.5	134.5	268.2	82.8	367.3	182.0	318.0	339.9	16.3	38.9	199.7
1963	138.5	209.2	112.1	107.3	233.8	66.9	308.6	143.6	233.5	249.4	391.0	123.0	172.1
1964	158.5	220.3	140.7	131.3	264.1	80.9	287.8	130.0	156.5	167.1	159.0	232.2	162.3
1965	339.5	321.1	399.3	348.5	538.3	207.7	666.5	377.5	510.5	545.8	293.6	491.9	340.5
1966	267.1	280.8	295.8	261.6	428.5	156.9	368.9	183.0	241.4	257.9	265.1	329.2	200.5
1967	278.6	287.2	312.3	275.5	446.0	165.0	521.1	282.5	244.2	260.9	394.2	409.0	272.1
1968	163.1	222.9	147.3	136.9	271.1	84.2	287.8	130.0	121.8	129.9	290.6	232.0	162.3
1969	261.7	455.1	384.0	255.2	522.0	200.2	552.6	195.5	354.2	378.6	151.0	409.5	209.5
1970	238.0	303.1	280.6	226.7	412.4	149.5	342.9	215.0	239.3	230.5	174.5	291.0	68.7
1971	160.0	269.4	264.8	133.1	343.6	141.8	361.6	128.0	207.5	221.6	113.6	301.0	160.9
1972	190.5	227.6	180.3	169.7	222.2	100.3	173.5	124.0	169.0	180.4	172.1	192.1	158.0
1973	249.0	379.7	300.5	239.9	392.3	159.2	396.6	235.7	308.5	329.7	219.1	330.0	238.4
1974	204.3	253.5	212.8	186.3	279.3	137.1	362.6	189.8	204.5	218.4	321.0	404.0	205.4
1975	200.4	230.6	224.3	181.6	352.8	166.2	466.3	195.5	237.5	253.7	131.5	252.5	209.5
1976	350.9	424.5	388.0	362.2	526.3	188.3	546.5	292.6	428.5	458.1	331.6	423.0	279.4
1977	319.4	354.6	263.0	324.4	393.8	140.9	505.0	319.0	294.0	314.2	374.0	448.5	298.4
1978	589.2	791.6	709.0	648.2	866.5	359.4	1110.3	519.7	801.1	856.8	677.9	851.0	442.9
1979	304.0	385.0	155.5	305.9	279.8	88.2	428.8	330.5	301.3	322.0	358.9	440.1	306.7
1980	522.7	696.8	649.8	218.4	803.8	330.4	943.2	186.5	626.8	292.5	795.3	757.8	203.0

Continuación de la tabla 3.

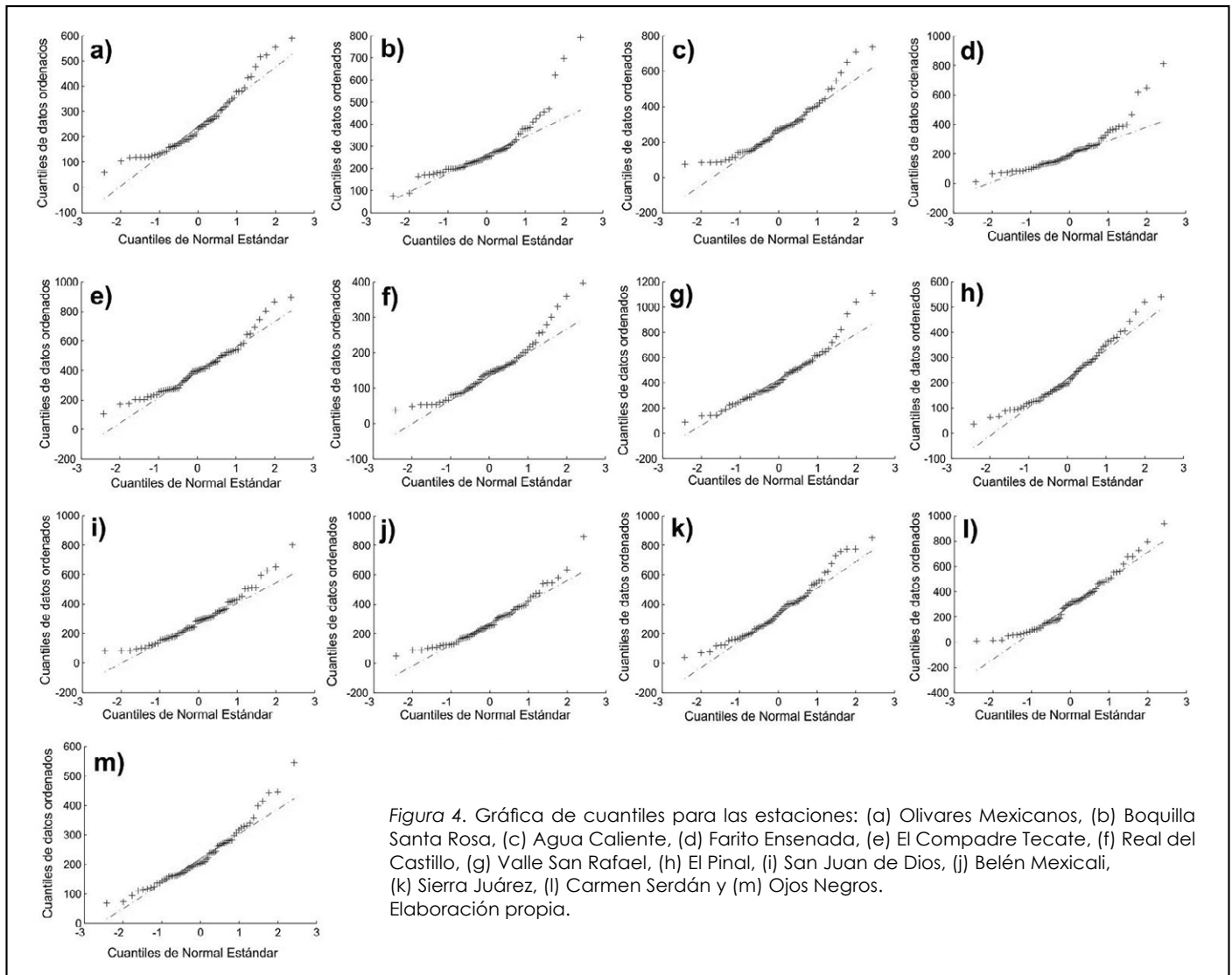
1981	231.1	293.0	274.3	112.1	405.8	146.4	391.6	269.8	283.9	259.3	346.8	366.5	216.9
1982	380.7	381.2	401.5	231.0	540.6	175.8	717.0	381.3	426.6	419.8	727.0	775.6	398.3
1983	554.8	621.7	735.8	810.8	894.9	397.0	1040.7	539.8	651.3	580.1	938.8	774.8	544.7
1984	142.5	171.6	251.5	112.1	381.6	135.6	323.3	173.7	176.3	200.3	476.0	288.0	197.6
1985	241.6	250.7	317.9	231.0	452.0	167.8	491.4	196.8	316.6	209.5	507.7	71.5	174.6
1986	210.3	289.9	295.6	254.8	428.3	124.7	385.3	198.3	212.1	231.5	385.7	315.7	167.6
1987	379.8	441.9	327.2	396.9	461.8	172.3	567.0	174.3	503.4	125.0	187.9	280.4	329.0
1988	209.7	181.3	243.7	192.8	363.4	131.4	336.0	95.5	288.1	142.5	61.7	164.6	240.5
1989	59.0	88.0	73.5	11.9	174.8	48.0	143.7	63.2	82.5	50.1	9.9	117.1	114.2
1990	183.3	233.6	213.9	161.1	330.4	116.8	403.3	170.9	180.5	117.5	182.5	275.4	191.7
1991	306.4	467.9	353.0	308.8	484.5	184.9	139.4	280.8	367.6	393.0	358.6	437.0	267.1
1992	438.1	409.1	501.7	466.8	649.3	257.8	575.7	405.9	510.2	545.5	559.1	621.0	281.4
1993	433.6	74.5	496.7	71.5	643.7	255.4	617.4	401.7	505.3	540.3	552.2	614.7	357.9
1994	249.1	274.5	288.2	232.5	412.7	153.2	88.7	226.4	305.6	326.6	271.4	357.0	281.1
1995	378.0	306.3	433.8	386.3	574.1	224.6	622.6	348.9	445.1	475.9	467.6	537.0	324.5
1996	172.9	199.6	202.1	152.5	317.3	111.0	324.5	154.0	223.1	238.3	155.4	250.6	187.6
1997	282.3	256.5	325.7	123.0	454.3	171.6	483.5	257.9	341.5	365.0	321.9	403.4	222.8
1998	515.7	377.9	589.4	386.5	746.4	300.8	822.8	479.7	594.2	635.4	677.2	729.3	414.1
1999	123.2	173.8	146.0	141.0	255.1	83.5	252.3	106.8	169.3	180.7	79.7	181.2	73.1
2000	172.0	199.1	206.7	140.1	334.1	113.3	323.2	153.2	237.6	253.8	154.0	249.4	123.7
2001	280.5	255.6	265.5	65.0	396.4	142.1	480.9	256.2	291.5	311.5	319.2	400.9	239.4
2002	103.1	163.3	143.1	74.5	170.5	37.7	223.1	87.7	179.3	191.4	49.2	153.1	193.1
2003	270.2	250.2	319.2	203.5	429.0	157.1	465.9	246.4	340.7	364.2	303.5	386.5	148.7
2004	394.4	314.8	441.5	370.5	583.0	228.3	646.5	364.4	452.9	389.5	492.6	559.9	239.7
2005	183.5	205.1	272.8	220.0	404.2	145.7	339.9	164.1	298.2	175.2	171.5	265.4	119.8
2006	137.9	181.4	174.1	144.0	299.5	97.3	273.6	120.8	202.0	128.8	102.1	201.7	110.3
2007	134.0	179.4	111.7	99.5	233.4	66.7	268.0	117.1	101.5	124.8	96.2	196.3	152.6
2008	334.6	283.7	384.8	252.5	522.9	200.6	559.5	307.6	415.1	328.8	401.5	476.4	203.8
2009	117.0	170.5	139.0	223.0	262.3	80.1	243.3	100.9	130.0	107.6	70.3	172.5	143.9
2010	476.6	357.5	545.2	617.5	692.9	279.2	765.9	442.5	412.5	473.1	617.7	674.7	446.2
2011	320.6	276.4	369.0	146.5	506.1	192.8	539.2	294.3	351.5	314.5	554.0	456.9	274.4
2012	240.8	234.9	278.8	230.1	410.6	148.6	423.2	218.5	303.0	233.4	326.0	345.4	165.0

Nota: Estos datos corresponden a la precipitación anual registrada en el periodo 1948-2012 por las estaciones climatológicas: Agua Caliente (AC), Belén Mexicali (BM), Boquilla Santa Rosa (BSR), Carmen Serdán (CS), El Compadre Tecate (ECT), El Farito Ensenada (EFE), El Pinal (EP), Ojos Negros (ON), Olivares Mexicanos (OM), Real del Castillo (RC), Sierra Juárez (SJ), San Juan de Dios (SJD), y Valle San Rafael (VSR). Los datos en redondas corresponden a la precipitación observada y en negritas corresponden a la estimada.
Elaboración propia.

Es importante mencionar que el número de estaciones incluidas en la estimación contribuyó significativamente a la calidad de la misma, dado que, en caso de que la estación vecina no contara con información en el periodo de interés, el número de estaciones incluidas permitió contar con otras estaciones de apoyo para completar los registros. Un ejemplo de esta situación se observa en las estimaciones obtenidas para las estaciones El

Compadre Tecate, San Juan de Dios, Sierra Juárez y Carmen Serdán.

Una de las suposiciones que se hacen al usar regresión lineal es que los valores obtenidos para la variable (y) deben estar distribuidos normalmente. En la figura 4 se muestra el Gráfico Q-Q, en él se representan los cuantiles correspondientes a la distribución Normal estándar versus los cuantiles empíricos obtenidos de los datos ordenados.



DISCUSIÓN

En todos los casos la correlación lineal (r) entre los registros x y y es alta, además en la mayoría de los casos resultó $Z \geq Z_c$, lo que sugiere que no hay posibilidad de que (ρ) sea igual a cero y, por tanto, el coeficiente de correlación r es significativo a 5% de significancia (ver tabla 2). En el caso particular en el que $r = 1$, se indetermina el valor Z , pero el valor de r por sí solo sugiere un alto nivel de confianza en la inferencia de datos faltantes. También puede apreciarse que en todos los casos el valor (E) es menor que la unidad. Por tanto, de acuerdo con los tres criterios, es conveniente hacer la inferencia de los valores faltantes a partir de las X_i .

Para comprobar la calidad de los ajustes, al verificar los supuestos de normalidad, se observó que la mayoría de los casos analizados muestran asimetría hacia la izquierda, lo que sugiere que los datos no se ajustan a una distribución Normal (ver figura 4); no obstante, el coeficiente de determinación observado promedio es $\bar{r}^2 = 0.80$, lo que ratifica la bondad del ajuste.

Por otra parte, se compararon los resultados obtenidos por el método de regresión lineal con los obtenidos por el método de la razón (no mostrados aquí), de esta comparación se observó que las máximas diferencias relativas entre los valores obtenidos por ambos métodos corresponden a las estimaciones para la estación Sierra Juárez y son del orden de 35%, mientras que las estimaciones obte-

nidas para la estación Real del Castillo resultaron idénticas. Para el resto de las estaciones se observaron diferencias relativas promedio del orden de 16%. También se hizo una comparación entre los valores reales y los estimados por ambos métodos, como medida de desempeño se calcularon las diferencias promedio (dp) existentes entre el valor real y estimado y la desviación estándar (sd) de dichas diferencias. Los resultados indican que el valor promedio de sd observado por el método de la razón es 1.5 veces mayor que el observado por el método de regresión.

CONCLUSIONES

Se estimaron los datos faltantes en los registros de la precipitación pluvial anual reportados por 13 estaciones climatológicas distribuidas en el área de la Cuenca Guadalupe en el periodo de 1948-2012 mediante un análisis de regresión lineal, entre estaciones cercanas. Las pruebas estadísticas realizadas; es decir, el coeficiente de correlación lineal promedio ($\bar{r} = 0.89$), y la eficiencia estadística promedio ($\bar{E} \sim 1$), confirman la viabilidad de la

inferencia estadística. El valor promedio observado del coeficiente de determinación $\bar{r}^2 = 0.80$ confirma la bondad de los resultados.

Del contraste entre las estimaciones observadas por el método de regresión lineal y el método de la razón se observó un caso en que las estimaciones resultaron idénticas y en que las mayores diferencias correspondían a las estimaciones para las estaciones más distantes. Para comparar el desempeño de ambos métodos se calcularon las desviaciones estándar de las diferencias promedio entre los valores reales y los valores estimados y resultó que las estimaciones obtenidas por el método de regresión lineal son 1.5 veces mejores que las del método de la razón.

Con base en estos resultados, este trabajo proporciona una base de datos de precipitación pluvial completa, homogénea y estadísticamente confiable para el periodo de análisis (1948-2012), la cual puede ser usada en futuras investigaciones, entre ellas, el análisis del comportamiento del agua superficial y subterránea en la Cuenca Guadalupe.

REFERENCIAS

- Alfaro, R., & Pacheco, R. (2000). Aplicación de algunos métodos de relleno a series anuales de lluvia de diferentes regiones de Costa Rica. *Tópicos Meteorológicos y Oceanográficos*, 7(1), 1-20.
- Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (2006). *Evapotranspiración del cultivo-Guías para la determinación de los requerimientos de agua de los cultivos*. (Estudio FAO riego y drenaje. N° 56) (pp. 227-240). Roma, Italia: Organización de las Naciones Unidas para la Agricultura y la Alimentación.
- Aparicio, F. J. (2011). *Fundamentos de hidrología de superficie*. D. F., México: Limusa.
- Beltrán-Gómez, L. (2001). *Balance geohidrológico del acuífero BC-07 Valle de Guadalupe, Municipio de Ensenada*. B. C., México: CONAGUA, Gerencia Regional en la Península de Baja California, Subgerencia Regional Técnica.
- Campos Aranda, D. F. (1998). *Procesos del ciclo hidrológico* (3ª ed.). San Luis Potosí, México: Universidad Autónoma de San Luis Potosí, Facultad de Ingeniería.
- Campos-Gaytán, J. R. (2008). *Simulación del flujo de agua subterránea en el acuífero del Valle de Guadalupe, Baja California, México*. Tesis de Doctorado. B. C., México: Centro de Investigación Científica y de Educación Superior de Ensenada.
- _____, Kretzschmar, T., & Herrera-Oliva, C. S. (2014). Future groundwater extraction scenarios for an aquifer in a semiarid environment: case study of Guadalupe Valley Aquifer, Baja California, Northwest Mexico. *Environmental monitoring and assessment*, 186(11), 7961-7985.
- Cárdenas Montoya, J. I. (1994). Importancia de los registros hidrológicos en el diseño y proyección de estructuras hidráulicas. Los métodos de regresión (Segunda parte). *Revista de Ingeniería Civil*, 2(1), 50-65.
- DeGaetano, A. T., Eggleston, K. L., & Knapp, W. W. (1995). A method to estimate missing daily maximum and minimum temperature observations. *Journal of Applied Meteorology*, 34(2), 371-380. doi: 10.1175/1520-0450-34.2.371
- Eischeid, J. K., Bruse, B. C., Karl, T. R., & Díaz, H. F. (1995). The quality control of long-term climatological data using objective data analysis. *Journal of Applied Meteorology*, 34(12), 2787-2795.
- Eischeid, J. K., Pasteris, P. A., Diaz, H. F., Plantico, M. S., & Lott, N. J. (2000). Creating a serially complete, national daily time series of temperature and precipitation for the Western United

- States. *Journal of Applied Meteorology*, 39(9), 1580-1591. doi: 10.1175/1520-0450(2000)039<1580:CASCND>2.0.CO;2
- García, E. (1981). *Modificaciones al sistema de clasificación climática de Koppen: para adaptarlo a las condiciones de la República Mexicana*. México: Editorial Larios.
 - Hernández-Rosas, M. T., & Mejía-Vázquez, R. (2003). Relación de aguas superficiales y subterráneas del acuífero BC-07, Valle de Guadalupe, Municipio de Ensenada, Baja California. *XIII Congreso Nacional de Hidráulica*. Ensenada, B. C., México: Comisión Nacional del Agua, Gerencia de Aguas Subterráneas, Subgerencia Regional Técnica, Jefatura de Aguas Subterráneas.
 - Hubbard, K. G. (2001). Multiple station quality control procedures. *Automated weather stations for applications in agriculture and water resources management*. High Plains Regional Climate Center, AGM-3 WMO/TD No. 1074. Lincoln, US.
 - Kashani, M. H., & Dinpashoh, Y. (2012). Evaluation of efficiency of different estimation methods for missing climatological data. *Stochastic Environmental Research and Risk Assessment*, 26(1), 59-71.
 - Kemp, W. P., Burnell, D. G., Everson, D. O., & Thomson, A. J. (1983). Estimating missing daily maximum and minimum temperatures. *Journal of Climate and Applied Meteorology*, 22, 1587-1593.
 - Luna Romero, E., & Lavado Casimiro, W. (2015). Evaluación de métodos hidrológicos para la completación de datos faltantes de precipitación en estaciones de la cuenca Jetepeque, Perú. *Revista Tecnológica ESPOL-RTE*, 28(3), 42-52.
 - McCuen, R. H. (1998). *Hydrologic analysis and design*. (3ª. ed.). NJ, US: Prentice-Hall.
 - Paulhus, J. L. H., & Kohler, M. A. (1952). Interpolation of missing precipitation records. *Monthly Weather Review*, 80(8), 129-133.
 - Pineda Villa, A. (2000). *Cambios ambientales y uso sustentable del agua en el valle de Ojos Negros*. B. C., México. Tesis de Maestría. México: Universidad Autónoma de Baja California Unidad Ensenada, Facultad de Ciencias.
 - Plata Caudillo, J. A. (2010). *Aislamiento y evaluación in vitro del efecto de Trichoderma spp. nativas sobre hongos patógenos de la madera de vid aislados en la región vitivinícola de Ensenada, Baja California*. Tesis de Maestría. B. C., México: CICESE.
 - Pizarro, R., González, P., Witterssheim, M., Saavedra, J., & Soto, C. (1993). *Elementos técnicos de hidrología III. Proyecto regional mayor sobre uso y conservación de recursos hídricos en áreas rurales de América latina y el Caribe*. Talca, Chile: Universidad de Talca.
 - Presti, R. L., Barca, E., & Passarella, G. (2010). A methodology for treating missing data applied to daily rainfall data in the Candelaro River Basin (Italy). *Environmental Monitoring and Assessment*, 160(1-4), 1-22.
 - Schuurmans, J. M., & Bierkens, M. F. P. (2007). Effect of spatial distribution of daily rainfall on interior catchment response of a distributed hydrological model. *Hydrology and Earth System Sciences*, 11(2), 677-693.
 - Tapiador, F. J., Kidd, C., Levizzani, V., & Marzano, F. (2003). A neural networks-based fusion technique to estimate half-hourly rainfall estimates at 0.1° resolution from satellite passive microwave and infrared data. *Journal of Applied Meteorology*, 43(4), 576-594.
 - Tapiador, F. J., Turk, F. J., Petersen, W., Hou, A. Y., García-Ortega, E., Machado, L. A. T.,... de Castro, M. (2012). Global precipitation measurement: Methods, datasets and applications. *Atmospheric Research*, 104-105, 70-97.
 - Wade, C. G. (1987). A quality control program for surface mesometeorological data. *Journal of Atmospheric and Oceanic Technology*, 4(3), 435-453.
 - World Meteorological Organization. (1983). *Guide to climatological practices (WMO-No100)*. Ginebra, Suiza: Autor.
 - Xia, Y., Fabian, P., Stohl, A., & Winterhalter, M. (1999). Forest climatology: estimation of missing values for Bavaria, Germany. *Agricultural and Forest Meteorology*, 96, 131-144.
 - You, J., Hubbard, K. G., & Goddard, S. (2008). Comparison of methods for spatially estimating station temperatures in a quality control system. *International Journal of Climatology*, 28(6), 777-787.
 - Young, K. C. (1992). A three-way model for interpolating monthly precipitation values. *Monthly Weather Review*, 120(11), 2561-2569.
 - Yozgatligil, C., Aslan, S., Iyigun, C., & Batmaz, I. (2013). Comparison of missing value imputation methods in time series: The case of Turkish meteorological data. *Theoretical and Applied Climatology*, 112(12), 143-167.