



Análisis del mejor método para completar datos de precipitación en el sector alto de las cuencas de Camaná y Chili, Arequipa, Perú

Recibido: 14.07.18
Aprobado: 29.10.18

Robert Ramos Alonzo
Universidad Nacional Mayor de San Marcos
alonzoroberth@gmail.com

Miguel Alva Huayaney
Universidad Nacional Mayor de San Marcos
malvah@unmsm.edu.pe

RESUMEN

En todo el Perú se carece de estaciones meteorológicas suficientes que pudieran proporcionar datos climáticos y ante esta necesidad, en la presente investigación se analiza la aplicación de diferentes métodos con el objetivo de completar datos faltantes en cuatro estaciones pluviométricas en las zonas ubicadas entre 4000 a 4500 m.s.n.m. de la región Arequipa. Estos métodos son; regresión lineal, razones de distancia, promedios vecinales, correlación con estaciones vecinas, regresión múltiple y se ha añadido la completación por el método del Vector Regional. Los métodos se analizan a través del Coeficiente de Determinación (R^2), Error Estándar de Estimación (EEE) y el test de concordancia de Bland y Altman, logrando establecer el mejor estimador en esta región.

PALABRAS CLAVE: datos faltantes; completación de datos; coeficiente de determinación.

Analysis of the best method for completion of rainfall missing data in high areas of Camaná and Chili basins, Arequipa, Peru

ABSTRACT

In all of Peru, there are not enough weather stations that could provide climate data and in view of this need, in the present investigation the application of different methods is analyzed in order to complete missing data in four rainfall stations in the areas located between 4000 to 4500 msnm of the Arequipa region. These methods are; linear regression, distance ratios, neighborhood averages, correlation with neighboring stations, multiple regression, and completion has been added by the regional vector method. The methods are analyzed through the Determination Coefficient (R^2), Standard Estimation Error (EEE), Bland and Altman concordance test, establishing the best estimator in this region.

KEYWORDS: missing data; completion of data; coefficient of determination.

1. Introducción

El clima es un elemento esencial en el desarrollo de la vida, así por ejemplo presenta una influencia directa en los recursos hídricos, en la biodiversidad, las actividades humanas entre otros; por lo que conocer los parámetros meteorológicos como la precipitación y temperatura son importantes. Por otro lado, diversas publicaciones (CELPAL y el BID, 2014) indican que el planeta estaría sufriendo un cambio climático y que estarían afectando sobre todo a países como el nuestro. A pesar de la necesidad de contar con información actualizada no se cuenta con estaciones meteorológicas suficientes que pudieran monitorear las características climáticas, por lo que en esta investigación se presentan diversas alternativas para generar datos nuevos ante la falta de información en determinados períodos de tiempo.

En la actualidad existen diferentes formas de generar información de precipitación, que van desde datos satelitales como el TRMM (Tropical Rainfall Measuring Mission), hasta métodos estadísticos, sin embargo, aun con estas avanzadas tecnologías, los datos generados deben ser corroborados y validados con datos reales, estos últimos muchas veces son escasos por lo que resultan una limitante cuando se pretende realizar estudios climáticos y de recursos hídricos. En este contexto, esta investigación compara siete métodos para completar valores de precipitación mensual.

Antes de completar los datos, se ha realizado un análisis de consistencia, este análisis, es un proceso que implica la identificación o detección de datos anómalos, los cuales se ha realizado a través del método gráfico y el análisis de cuartiles, luego se ha realizado el análisis de doble masa y finalmente se ha corregido algunos datos dudosos. Así los métodos como razones de distancia, promedios vecinales y razones promedio se desarrollan a través de fórmulas básicas, mientras que los métodos por correlación con estaciones vecinas, regresión lineal simple y regresión múltiple se aplican bajo ciertas consideraciones estadísticas como el grado de correlación, en tanto que el Vector Regional además de esta consideración se aplica utilizando el programa Hydraccess. Para verificar el mejor método se ha considerado el Coeficiente de Determinación (R^2), Error Estándar de Estimación (EEE), test de concordancia de Bland y Altman.

2. Área de estudio

La investigación fue realizada en las zonas altas de las cuencas de Camana y Chili, emplazadas en la región Arequipa. Geográficamente, la región evaluada está comprendido entre los paralelos de $15^{\circ}17'48''$ a $16^{\circ}12'55''$ LS y los meridianos $70^{\circ}53'10''$ a $71^{\circ}24'32''$ LW. Esta región se ubica a una altitud que varía entre los 4000 a 4500 m.s.n.m., por su ubicación presenta temperaturas medias entre 4 a 6°C y la precipitación anual está próximo a 400 mm (Zúñiga, 2010).

Políticamente se ubica en los distritos de Callali, San Antonio de Chuca y San Juan de Tarucani, en las provincias de Caylloma y Arequipa, región Arequipa. La Ubicación de presenta en la Figura 1.

3. Objetivo

El objetivo de la presente investigación ha sido determinar el mejor método para la completar datos faltantes en las zonas altas de las cuencas Camaná y Chili en las provincias de Caylloma y Arequipa.

Objetivos específicos

- Determinar en qué meses del año la completación de datos es más certero.
- Recomendar que método sería el más apropiado para completar datos en esta región.

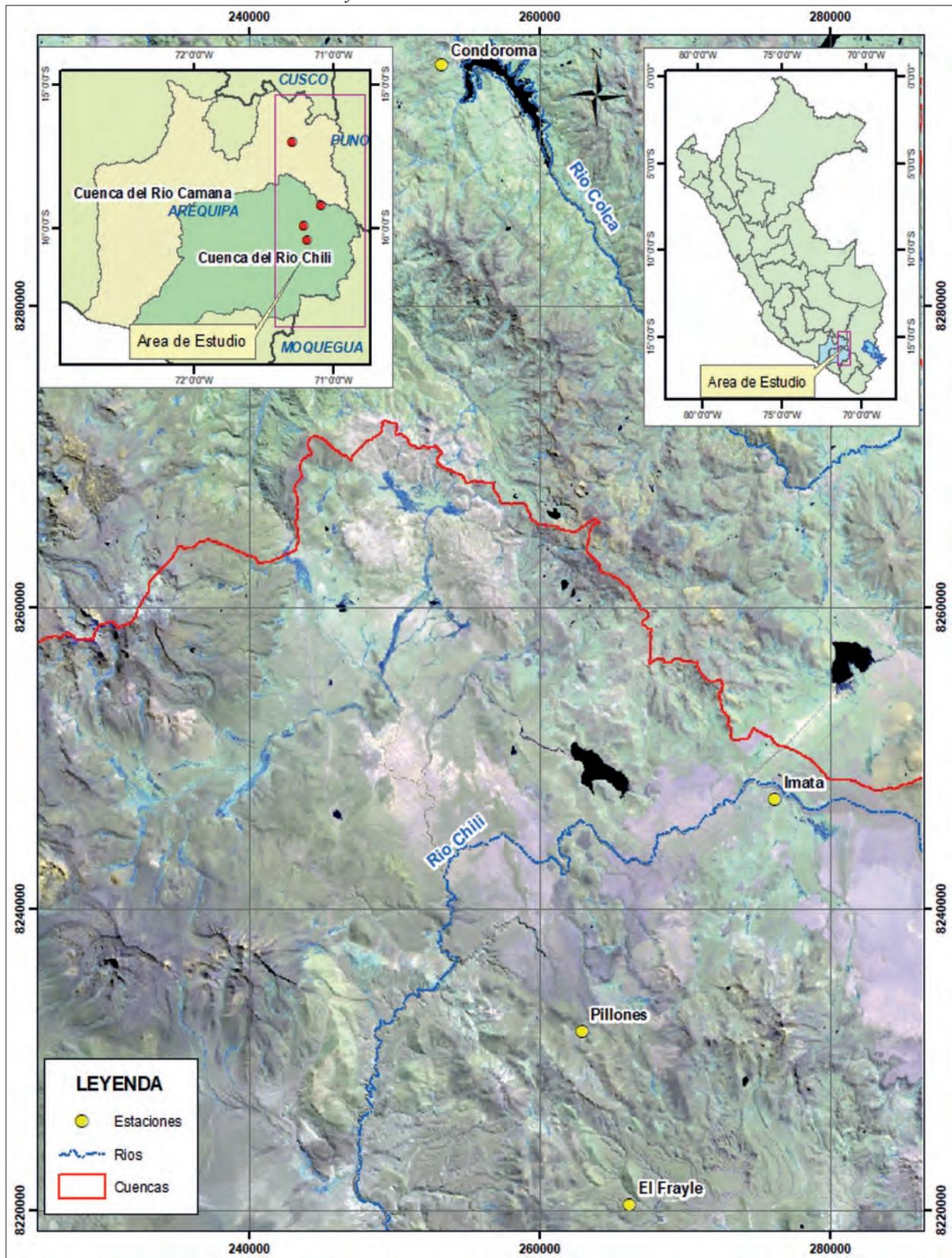
4. Metodología

Previo a la completación de datos faltantes se realizó el análisis de consistencia de datos mediante el método de doble masa y el método del Vector Regional en donde se evidencia que los valores guardan consistencia entre ellos. Una vez realizado este análisis, se procedió a completar los datos a través de los métodos antes indicados.

Los datos de precipitación provienen de cuatro estaciones meteorológicas, y son a paso mensual (Tabla 1), estos corresponden al período 1984-1997 ya que las cuatro estaciones consideradas presentan datos homogéneos en estos años; en total se analizaron 214 datos por estación. Eligiéndose al azar el



FIGURA 1. Ubicación de las cuencas Camaná y Chili



Fuente: Elaboración propia

TABLA 1. *Ubicación política y geográfica de las estaciones pluviométricas utilizadas.*

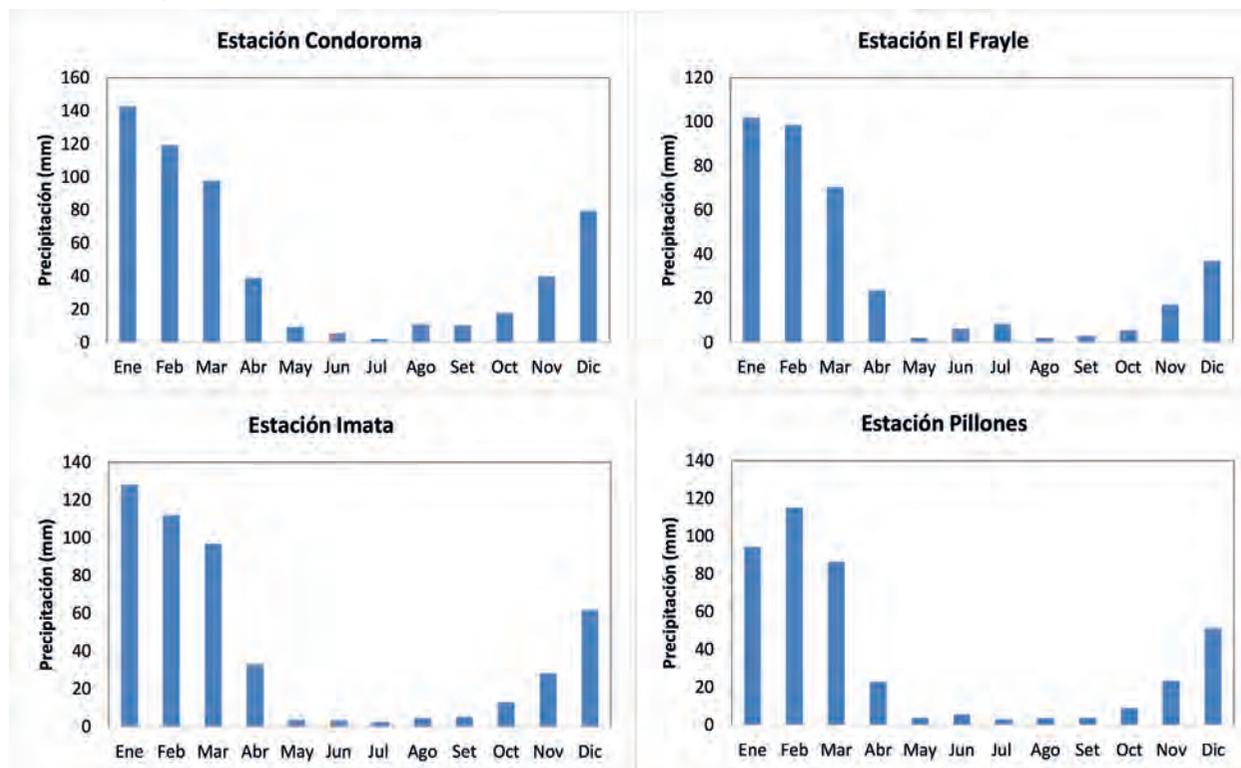
Estaciones	Departamento	Provincia	Distrito	Latitud (S)	Longitud (W)	Altitud (msnm)
Condoroma	Arequipa	Caylloma	Callalli	15°24'00"	71°18'00"	4160
El Frayle	Arequipa	Arequipa	Tarucan	16°05'04"	71°11'13"	4060
Imata	Arequipa	Caylloma	Chuca	15°50'33"	71°05'26"	4520
Pillones	Arequipa	Caylloma	Chuca	15°58'52"	71°13'00"	4450

Fuente: SENAMHI

13% de los datos, en donde se supuso como datos no existentes (Tabla 2).

Los datos analizados y que fueron retirados corresponden a los mismos años. El procedimiento de cálculo de completación de datos faltantes se realizó para los meses comprendidos entre octubre hasta mayo, meses caracterizados por presentar considera-

bles valores de precipitación, mientras que el resto de los meses no presentan precipitación alguna. Así mismo los datos que luego fueron retirados se consideraron reales puesto que sus valores fueron sometidos a un análisis de consistencia, razón por lo cual se consideran como el valor real, por lo que fue el referente para verificar los valores obtenidos.

FIGURA 2. *Precipitación mensual en las estaciones consideradas*

Fuente: Elaboración propia

TABLA 2. Valores pluviométricos que fueron retirados

Año	Mes	Condorama	El Frayle	Imata	Pillones
1984	Mar	200.8	87.3	135.5	173.7
	Dic	105.7	37.9	69.9	79.4
1985	May	27.4	3	14.5	3.5
	Nov	56	14	80.4	16.1
1986	Abr	89.5	12.4	46.6	26.6
	May	4	0.6	0.9	6
1987	Feb	45	13.1	65.7	24.9
	Oct	14.4	7.5	15.1	16.6
	Dic	20.5	0.2	13.1	22.8
1988	Ene	220.7	122.6	200.7	210.2
	Feb	95.7	27.8	29	40.4
1989	Mar	123	72.8	109.2	79
	Nov	13.2	10.2	15.8	72.6
1990	Abr	35	8.9	23.6	6.2
	Dic	116.2	45.6	96.6	51.8
1991	Feb	148.3	65.3	74.9	52.2
	Feb	75.3	9.9	31.4	8.2
1992	Oct	23.8	2.6	7.6	9
	Dic	87.5	43.2	105	163.9
1993	Mar	198.1	54.6	136.4	126.6
	Nov	60.4	4.5	42.1	10.5
1994	May	16	0	3.2	0
	Dic	116.6	37	79	42.4
1995	Abr	48.3	5.7	14.8	9.1
	Nov	20.1	12.7	41.5	9.1
1996	Ene	70	37.4	137.6	57.6
	Nov	33.3	8.6	19.2	26.2
1997	Ene	120	69	117.1	85.5
	Mar	79.4	51.1	52.5	75.2

Fuente: Elaboración propia

5. Métodos para completar datos faltantes

5.1. Por regresión lineal

Es un método que presenta gran aceptación, y para su aplicación es necesario que los datos analizados presenten un buen grado de correlación. Se puede aplicar tanto a datos diarios, mensuales y anuales.

Para aplicar este método, se estableció una regresión lineal simple, entre los valores de precipitación de la estación que carecía con parte de su información y de otras que si contaban con una data completa. Para un mejor análisis las estaciones fueron organizadas desde la 1º, 2º y 3º según su cercanía (Tabla 3).

$$y = a + b * x \quad (1)$$

TABLA 3. Esquema para la aplicación del método de regresión lineal

Estación con datos faltantes (Y)	Estaciones utilizadas (X), para la estimación de precipitación (Y)		
	1° estación mas cercana	2° estación mas cercana	3° estación mas cercana
Condorama	Imata	Pillones	El Frayle
El Frayle	Pillones	Imata	Condorama
Imata	Pillones	El Frayle	Condorama
Pillones	El Frayle	Imata	Condorama

Fuente: Elaboración propia

5.2. Por razones de distancias

Este método tiene su aplicación en zonas de relieve homogéneo en donde dominen las planicies. Y para obtener mejores resultados las estaciones elegidas tienen que estar distribuidas en forma lineal (Pizarro, 1993), y se denota mediante la siguiente ecuación.

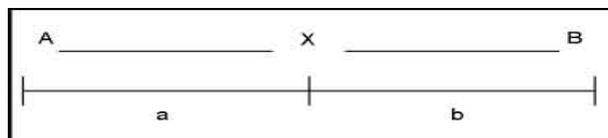


Figura 2. Ejemplo aplicado para completar datos, por razones de distancia. (Fuente: Pizarro et al., 1993).

$$PX = PA + a * [(PB - PA)/(a + b)]x \quad (2)$$

Donde:

PX, PA, PB, representan la precipitación para las estaciones X, A y B, respectivamente y a y b representan las distancias.

Este método puede ser aplicado cuando se desea completar datos de una estación, y esta debe estar entre dos estaciones que presenten información de un período considerable. Por lo que solo dos estaciones podían completar sus datos a partir de sus vecinos más próximos (Tabla 4). Para establecer distancias entre las estaciones, estas fueron ubicadas sobre una imagen de satélite a una escala apropiada, esta escala ha permitido visualizar con claridad la ubicación de las estaciones.

TABLA 4. Estaciones consideradas para aplicar el método por razones de distancias.

Estación sin datos	Estaciones con datos completos	
	Estación A	Estación B
Imata	Pillones (20 km)	Condorama (54 km)
Pillones	El Frayle (12 km)	Imata (20 km)

Fuente: Elaboración propia

5.3. Por promedios vecinales

Al igual que el método anterior, para aplicar este método, las estaciones deben estar ubicadas en un mismo tipo de relieve con un dominio de planicies o llanuras. Para poder completar los datos de una estación que carece de la misma, este debe estar en una zona central respecto a las otras que cuentan con datos, su cálculo es un promedio de las estaciones con datos completos.

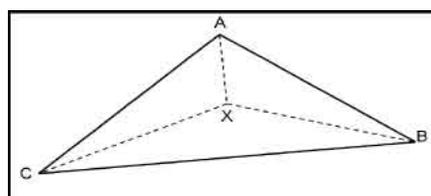


Figura 3. Esquema de la distribución de estaciones para la completación de datos por promedios vecinales (Pizarro, 1993)

Donde; X, A, B y C representan la disposición espacial de cuatro estaciones pluviométricas y donde la estación X es la que presenta carencia de información. Dada esta condición para el cálculo de los datos faltantes, solo fue posible aplicarlo a uno de las cuatro estaciones (Tabla 5).

TABLA 5. Esquema para la aplicación del método por promedios vecinales

Estación sin datos (PX)	Estaciones con datos completos		
	Estación A	Estación B	Estación C
Pillones	El Frayle	Imata	Condoroma

Fuente: Elaboración propia

$$PX = \sum_{i=1}^n \frac{Pi}{n} \quad (3)$$

Donde; Pi = Precipitación de la estación *i* en el período de estudio; n = Número total de estaciones.

5.4. Por razones promedio

A continuación, se presenta el detalle.

TABLA 6. Relación de estaciones para completar datos por el método por razones promedio

Estación sin datos (PX)	Estaciones con datos completos		
	Estación A	Estación B	Estación N
Condoroma	Imata	Pillones	El Frayle
El Frayle	Pillones	Imata	Condoroma
Imata	Pillones	El Frayle	Condoroma
Pillones	El Frayle	Imata	Condoroma

Fuente: Elaboración propia

5.5. Por correlación con estaciones vecinas

En este método se utilizan las precipitaciones estimadas a partir de correlaciones entre la estación con carencia de información y cada una de las estaciones vecinas a los que se asocian los coeficientes de correlación respectivos (Unesco-Rostlac, 1982), su expresión es según la ecuación (5).

$$PX = \frac{PXA * rXA + PXB * rXB + \dots + PXN * rXN}{rXA + rXB + \dots + rXN} \quad (5)$$

$$PX = Px * \left[\frac{PA}{Pa} + \frac{PB}{Pb} + \frac{PN}{Pn} \right] x \quad (4)$$

Donde; Px, Pa, Pb y Pn = promedio de las precipitaciones registradas en cada estación, mientras que PX, PA, PB y PN representan las precipitaciones mensuales en las tres estaciones en el período completo incluido la estación PX (Pizarro, 1993). Los resultados que se obtienen son producto de la relación entre los valores de precipitación por cada año y el promedio del período analizado (Tabla 6).

Donde; PXi = valor estimado de precipitación en X, a partir de las regresiones con cada una de las estaciones, rXi = coeficiente de correlación entre los registros de la estación X, y cada una de las estaciones consideradas, A, B y N son las estaciones analizadas. Para la aplicación de este método primero se estimó el coeficiente de correlación que es la relación que existe entre dos variables (entre los datos de la estación PX y cada una de las otras estaciones), en la Tabla 7 se presenta el cálculo de los valores entre cada una de las estaciones.

TABLA 7. Relación de estaciones para aplicar el método por correlación con estaciones vecinas

Estación sin datos (PX)	Estaciones con datos completos		
	Estación A	Estación B	Estación N
Condoroma	Imata	Pillones	El Frayle
El Frayle	Pillones	Imata	Condoroma
Imata	Pillones	El Frayle	Condoroma
Pillones	El Frayle	Imata	Condoroma

Fuente: Elaboración propia

5.6. Por regresiones múltiples

Este método, tiene el mismo principio del método de regresión simple. Para su estimación se considera la siguiente expresión matemática (6).

$$Y = a + bX1 + cX2 + dX3 + nX4 \quad (6)$$

Donde; Y = valor de precipitación estimada para la estación con carencia de información; X1 = valor de precipitación con información completa. *a, b, c, n* = constantes de regresión. Este método tiene la capacidad de analizar los datos de varias estaciones a la vez y someterlos a un análisis de consistencia y por tanto permite verificar su grado de correlación entre ellas. El detalle de la conformación de estaciones se presenta en la Tabla 8.

TABLA 8. Estaciones consideradas para la aplicación del método por regresiones múltiples

Estación sin datos (PX)	Estaciones con datos completos		
	Estación A	Estación B	Estación N
Condorama	Imata	Pillones	El Frayle
El Frayle	Pillones	Imata	Condorama
Imata	Pillones	El Frayle	Condorama
Pillones	El Frayle	Imata	Condorama

Fuente: Elaboración propia

5.7. Por el método del vector regional (MVR)

El método del vector regional consiste en lo siguiente: en lugar de comparar dos por dos estaciones por correlación o doble masa como en otros casos, aquí se elabora una estación ficticia que sea una «especie de promedio» de todas las estaciones de la zona con la cual se compara cada una de las estaciones. El MVR emplea para el cálculo de esta estación «Vector» el concepto de precipitación media extendida al período de trabajo. Se emplea el método de mínimos cuadrados para encontrar los índices pluviométricos regionales anuales «Zi» y la precipitación media extendida «Pj» el que se muestra en la ecuación matemática 7 (Espinoza, 2005).

$$\sum_{i=1}^n \sum_{j=1}^m \left(\frac{P_{ij}}{P_i} - Z_i \right) \quad (7)$$

Donde *i* es el índice de año, *j* el índice de estación, *N* el número de años y *M* el número de estaciones «Pij» es la precipitación anual en la estación «j» el año «i», «Pj» es la precipitación media extendida al período de *N* años y finalmente «Zi» es el índice pluviométrico regional del año «i» (Hiez, 1977, y Brunet Moret, 1979).

Para el cálculo de los valores se utilizó el paquete computacional HYDRACCESS utilizando las cuatro estaciones que se muestran en la tabla 9.

TABLA 9. Estaciones consideradas para la aplicación del método del Vector Regional

Estación sin datos (PX)	Estaciones con datos completos		
	Estación A	Estación B	Estación N
Condorama	Imata	Pillones	El Frayle
El Frayle	Pillones	Imata	Condorama
Imata	Pillones	El Frayle	Condorama
Pillones	El Frayle	Imata	Condorama

Fuente: Elaboración propia

6. Análisis comparativo entre el dato real y los obtenidos por los diferentes métodos

Luego de la aplicación de los diferentes métodos, se obtuvieron los resultados para cada caso, estos fueron comparados con los datos reales que previamente fueron retirados. Los resultados son diversos puesto que varios de los métodos son más exigentes para su aplicación, y si bien otros también permiten su aplicación sin embargo al no pasar por un filtro, sus valores se alejan en mayor o menor medida respecto al dato real. Es preciso recalcar que para aplicar estos métodos los datos de cada estación deben corresponder a un mismo período de análisis.

En esta primera visión de los resultados se evidencia que el mejor estimador es el método del Vector Regional, seguido del método de regresión múltiple y finalmente el método que presenta un estimador más bajo es el de Regresión Simple.

Si bien en una primera revisión se evidencia a los mejores métodos para completar datos en esta

región del país, sin embargo, es necesario aplicar algunos métodos estadísticos para poder establecer que método es el que mejor estima los datos reales, y estos son: el Coeficiente de Determinación (R^2), Error Estándar de Estimación (EEE); que permite calcular la disparidad promedio entre los valores reales de precipitación y los estimados, en donde los valores cercanos a cero indican la buena descripción de la precipitación (Caro, 2001). Test de Concordancia de Bland y Altman (ACBA); se basa en las diferencias promedio existentes entre el valor real y el estimado, y la comparación de las gráficas residuales (Bland y Altman, 1999; Dunn y Roberts, 1999). La ecuación de test de concordancia es la siguiente:

$$LC = dp \pm 1.96 * SD \quad (8)$$

Donde LC es el límite de concordancia; dp son las diferencias promedio y SD es la desviación estándar.

TABLA 10. Datos reales y datos calculados

Año	Dato Real		Calculo por diversos metodos					
	Pillones	RL	RD	PV	RP	EV	RM	VR
1984	136.6	74.9	118.3	117.8	152.2	122.7	134.2	127.4
1985	24.6	45.5	28.1	39.4	30.8	37.0	9.1	36.8
1986	111.6	92.1	115.2	100.3	145.1	104.7	121.6	125.4
1987	107.7	77.4	110.0	106.9	148.9	111.0	98.5	118.1
1988	210.2	97.3	151.9	155.3	209.4	158.0	175.8	193.3
1989	84.4	85.5	109.5	115.6	124.7	115.2	95.7	102.3
1990	91.2	66.7	81.8	88.6	118.2	89.0	81.4	103.1
1991	53.1	71.9	88.6	91.2	90.5	92.3	61.2	66.1
1992	20.8	59.7	47.5	65.2	48.2	60.1	32.4	17.6
1993	143.5	99.4	147.9	148.5	195.4	150.9	169.7	164.9
1994	178.1	123.1	185.0	188.1	201.2	188.8	216.2	185.1
1995	69.3	57.8	51.5	94.8	78.9	83.6	62.0	81.2
1996	57.6	53.4	75.0	74.2	62.1	79.0	78.2	65.9
1997	85.5	69.7	87.0	107.5	103.6	104.1	88.7	95.4

Fuente: Elaboración propia

RL: Regresión Lineal, RD: Razones de Distancia, PV: Promedios Vecinales, RP: Razones Promedio, EV: Estaciones Vecinas, RM: Regresión Múltiple, VR: Vector Regional.

7. Resultados

En las tablas desde el 11 hasta la 18 se muestran los valores obtenidos de acuerdo a cada método estadístico analizado como son el coeficiente de determinación (R^2) y el error estándar de estimación (EEE).

Los resultados siguen un patrón común, aunque presentan ligeras diferencias, por lo que estos se analizan según cada estación y cada mes.

TABLA 11. *Test estadístico R^2 y EEE – Enero*

Método	Pillones			El Frayle			Imata			Condoroma		
	R2	EEE	LC									
Regresión Lineal	0.65	38.5	123.7	0.62	23.7	63.2	0.65	32.2	94.8	0.39	39.3	126.5
Razones de Distancia	0.76	33.4	68.6	-	-	-	0.71	29.6	55.9	-	-	-
Promedios Vecinales	0.75	33.8	70.2	-	-	-	-	-	-	-	-	-
Razones Promedio	0.68	36.6	85.0	0.60	24.4	48.9	0.67	31.5	62.2	0.49	35.9	82.9
Estaciones Vecinas	0.77	32.8	63.5	0.60	24.4	43.1	0.71	29.3	53.0	0.48	36.2	87.7
Regresión Múltiple	0.79	30.2	50.2	0.64	20.5	33.6	0.71	29.2	55.4	0.53	34.6	65.7
Vector Regional	0.78	31.5	54.2	0.63	21.6	38.2	0.88	14.7	27.4	0.52	36.7	72.3

Fuente: Elaboración propia

TABLA 12. *Test estadístico R^2 y EEE – Febrero*

Método	Pillones			El Frayle			Imata			Condoroma		
	R2	EE	LC									
Regresión Lineal	0.58	40.09	74.36	0.74	26.22	51.06	0.61	43.08	66.73	0.67	37.19	87.64
Razones de Distancia	0.64	38.41	69.45	-	-	-	0.72	36.39	78.44	-	-	-
Promedios Vecinales	0.71	34.41	58.09	-	-	-	-	-	-	-	-	-
Razones Promedio	0.72	33.71	52.76	0.88	18.04	24.35	0.80	30.55	50.75	0.72	34.46	71.32
Estaciones Vecinas	0.72	33.89	59.61	0.79	23.69	20.13	0.81	29.82	52.09	0.73	33.62	67.09
Regresión Múltiple	0.73	33.25	63.10	0.80	23.30	44.22	0.84	27.32	54.84	0.75	32.22	61.63
Vector Regional	0.75	32.30	57.54	0.88	16.84	25.92	0.90	22.10	39.23	0.83	28.17	49.47

Fuente: Elaboración propia

TABLA 13. *Test estadístico R^2 y EEE – Marzo*

Método	Pillones			El Frayle			Imata			Condoroma		
	R2	EE	LC	R2	EE	LC	R2	EE	LC	R2	EE	LC
Regresión Lineal	0.50	24.432547	22.45	0.77	4.18	13.32	0.69	5.36	15.37	0.34	11.60	38.30
Razones de Distancia	0.57	24.429895	11.95	-	-	-	0.76	4.78	12.54	0.42	10.30	34.10
Promedios Vecinales	0.68	24.085188	11.80	-	-	-	-	-	-	-	-	-
Razones Promedio	0.69	24.251688	15.68	0.86	3.41	9.75	0.78	4.75	11.31	0.47	9.96	29.91
Estaciones Vecinas	0.63	16.349451	11.72	0.81	2.59	3.82	0.75	5.00	13.88	0.55	10.00	27.92
Regresión Múltiple	0.80	12.6636	8.03	0.84	2.11	3.36	0.79	4.14	8.65	0.72	6.78	15.24
Vector Regional	0.88	24.5487	7.95	0.89	3.04	3.10	0.85	4.26	6.73	0.69	9.13	24.38

Fuente: Elaboración propia

TABLA 14. *Test estadístico R^2 y EEE – Abril*

Método	Pillones			El Frayle			Imata			Condoroma		
	R2	EE	LC	R2	EE	LC	R2	EE	LC	R2	EE	LC
Regresión Lineal	0.37	21.72	61.36	0.67	9.58	9.75	0.66	15.42	28.59	0.49	14.01	46.21
Razones de Distancia	0.43	21.72	53.90	-	-	-	0.69	25.45	42.16	-	-	-
Promedios Vecinales	0.46	21.41	49.91	-	-	-	-	-	-	-	-	-
Razones Promedio	0.49	21.56	50.69	0.68	11.09	18.39	0.73	20.13	39.80	0.55	21.89	49.92
Estaciones Vecinas	0.55	14.53	27.63	0.65	12.04	18.77	0.77	16.56	35.93	0.61	21.92	39.54
Regresión Múltiple	0.58	11.26	27.31	0.77	7.67	6.51	0.83	13.28	24.48	0.76	18.61	35.80
Vector Regional	0.61	21.82	33.36	0.84	4.77	3.59	0.87	11.54	18.52	0.68	12.59	31.39

Fuente: Elaboración propia

TABLA 15. *Test estadístico R² y EEE - Mayo*

Metodo	Pillones			El Frayle			Imata			Condorama		
	R2	EE	LC	R2	EE	LC	R2	EE	LC	R2	EE	LC
Regresión Lineal	0.41	21.17	59.83	0.60	9.34	9.50	0.53	15.03	27.87	0.53	20.41	45.06
Razones de Distancia	0.44	21.17	52.56	-	-	-	0.57	24.81	41.11	-	-	-
Promedios Vecinales	0.61	20.87	48.67	-	-	-	-	-	-	-	-	-
Razones Promedio	0.48	21.02	49.43	0.67	10.81	17.93	0.56	19.62	38.81	0.59	23.03	48.67
Estaciones Vecinas	0.59	14.17	26.94	0.71	11.74	18.30	0.65	16.15	35.03	0.64	17.80	38.55
Regresión Múltiple	0.66	10.98	26.63	0.77	7.48	6.35	0.80	12.94	23.87	0.71	17.56	34.90
Vector Regional	0.73	21.28	32.52	0.82	4.65	3.50	0.79	11.25	18.06	0.67	16.28	30.61

Fuente: Elaboración propia

TABLA 16. *Test estadístico R² y EEE - Octubre*

Metodo	Pillones			El Frayle			Imata			Condorama		
	R2	EE	LC	R2	EE	LC	R2	EE	LC	R2	EE	LC
Regresión Lineal	0.46	10.16	24.95	0.68	4.65	14.80	0.68	5.96	17.08	-	-	-
Razones de Distancia	0.61	6.37	13.28	-	-	-	0.74	5.31	13.93	-	-	-
Promedios Vecinales	0.68	5.79	13.11	-	-	-	-	-	-	-	-	-
Razones Promedio	0.59	6.48	17.43	0.79	3.79	10.84	0.76	5.27	12.57	0.40	11.07	33.23
Estaciones Vecinas	0.53	6.99	13.03	0.88	2.87	4.24	0.72	5.55	15.42	0.39	11.11	31.02
Regresión Múltiple	0.79	4.63	8.92	0.92	2.35	3.73	0.81	4.60	9.61	0.72	7.54	16.93
Vector Regional	0.86	5.42	8.84	0.88	3.38	4.75	0.93	4.74	7.48	0.64	10.15	27.09

Fuente: Elaboración propia

TABLA 17. *Test estadístico R² y EEE - Noviembre*

Metodo	Pillones			El Frayle			Imata			Condorama		
	R2	EE	LC	R2	EE	LC	R2	EE	LC	R2	EE	LC
Regresión Lineal	0.30	27.15	76.71	0.66	11.98	12.19	0.66	19.27	35.73	0.51	26.17	57.77
Razones de Distancia	0.35	27.14	67.38	-	-	-	0.41	31.81	52.70	-	-	-
Promedios Vecinales	0.40	26.76	62.39	-	-	-	-	-	-	-	-	-
Razones Promedio	0.32	26.95	63.37	0.55	13.86	22.99	0.42	25.16	49.75	0.37	29.52	62.40
Estaciones Vecinas	0.55	18.17	34.53	0.47	15.05	23.46	0.61	20.70	44.91	0.62	22.83	49.42
Regresión Múltiple	0.73	14.07	34.14	0.78	9.59	8.14	0.75	16.59	30.61	0.63	22.51	44.75
Vector Regional	0.49	27.28	41.69	0.93	5.96	4.49	0.82	14.42	23.15	0.69	20.88	39.24

Fuente: Elaboración propia

TABLA 18. *Test estadístico R² y EEE - Diciembre*

Metodo	Pillones			El Frayle			Imata			Condorama		
	R2	EE	LC	R2	EE	LC	R2	EE	LC	R2	EE	LC
Regresión Lineal	0.70	29.86	84.38	0.61	13.18	13.40	0.54	21.20	39.31	0.67	28.78	63.54
Razones de Distancia	0.65	29.86	74.12	-	-	-	0.61	34.99	57.97	-	-	-
Promedios Vecinales	0.68	29.44	68.63	-	-	-	-	-	-	-	-	-
Razones Promedio	0.78	29.64	69.70	0.65	15.25	25.29	0.68	27.68	54.73	0.69	32.48	68.64
Estaciones Vecinas	0.76	19.98	37.99	0.79	16.55	25.81	0.77	22.77	49.41	0.66	25.11	54.36
Regresión Múltiple	0.86	15.48	37.55	0.83	10.55	8.95	0.89	18.25	33.67	0.72	24.76	49.22
Vector Regional	0.80	30.00	45.86	0.89	6.56	4.94	0.84	15.87	25.47	0.76	22.96	43.17

Fuente: Elaboración propia

De acuerdo al análisis según cada mes, se evidencia un claro mejor estimador que es el método del Vector Regional, seguido del método de regresión múltiple. Sin embargo, según cada mes los coeficientes de determinación serán mucho más altos en algunos meses respecto a otros, lo mismo ocurre entre cada estación, de este modo se evidencia que los mejores ajustes presentan la estación Imata, y los ajustes más bajos presenta la estación Condorama.

TABLA 19. Valores de test de concordancia para cada metodo de completación, para la estación Pillones

Metodos	ACBA
Vector Regional	7.95
Regresión Multiple	8.03
Estaciones Vecinas	11.72
Promedios Vecinales	11.80
Razones de Distancia	11.95
Razones Promedio	15.68
Regresión Lineal	22.45

Fuente: Elaboración propia

En la Tabla 19, se muestra que en la estación Pillones, el método del Vector Regional es el que mejor estima y sus valores son muy parecidos a los datos reales, con una diferencia promedio de 7.95 mm, seguido del Metodo de Regresión Multiple con 8.03 mm y el método de Regresión Lineal es el que presenta el valor más alto con 22.45 mm.

TABLA 20. Valores de test de concordancia para cada metodo de completación, para la estación El Frayle

Metodos	ACBA
Vector Regional	3.10
Regresión Multiple	3.36
Estaciones Vecinas	3.82
Razones Promedio	9.75
Regresión Lineal	13.32

Fuente: Elaboración propia

En la Tabla 20, se muestra que en la estación El Frayle, el método del Vector Regional es el mejor estimador, con una diferencia promedio de 3.1 mm, seguido del método de Regresión Múltiple con 3.36 mm, y el método por Regresión Simple es el que muestra los valores más altos con 13.32 mm.

TABLA 21. Valores de test de concordancia para cada metodo de completación, para la estación Imata

Metodos	ACBA
Regresión Multiple	15.24
Vector Regional	24.38
Estaciones Vecinas	27.92
Razones Promedio	29.91
Razones de Distancia	34.00
Regresión Lineal	38.00

Fuente: Elaboración propia

En la Tabla 21, se muestra que en la estación Imata, el método del Vector Regional es el que más se acerca a los valores reales con una diferencia promedio de 6.73 mm, seguido del método de Regresión Múltiple con 8.65 mm, y el método por Regresión Simple es el que presenta los valores más elevados con 15.37 mm.

TABLA 22. Valores de test de concordancia para cada metodo de completación, para la estación Condoroma.

Metodos	ACBA
Regresión Multiple	15.24
Vector Regional	24.38
Estaciones Vecinas	27.92
Razones Promedio	29.91
Razones de Distancia	34.00
Regresión Lineal	38.00

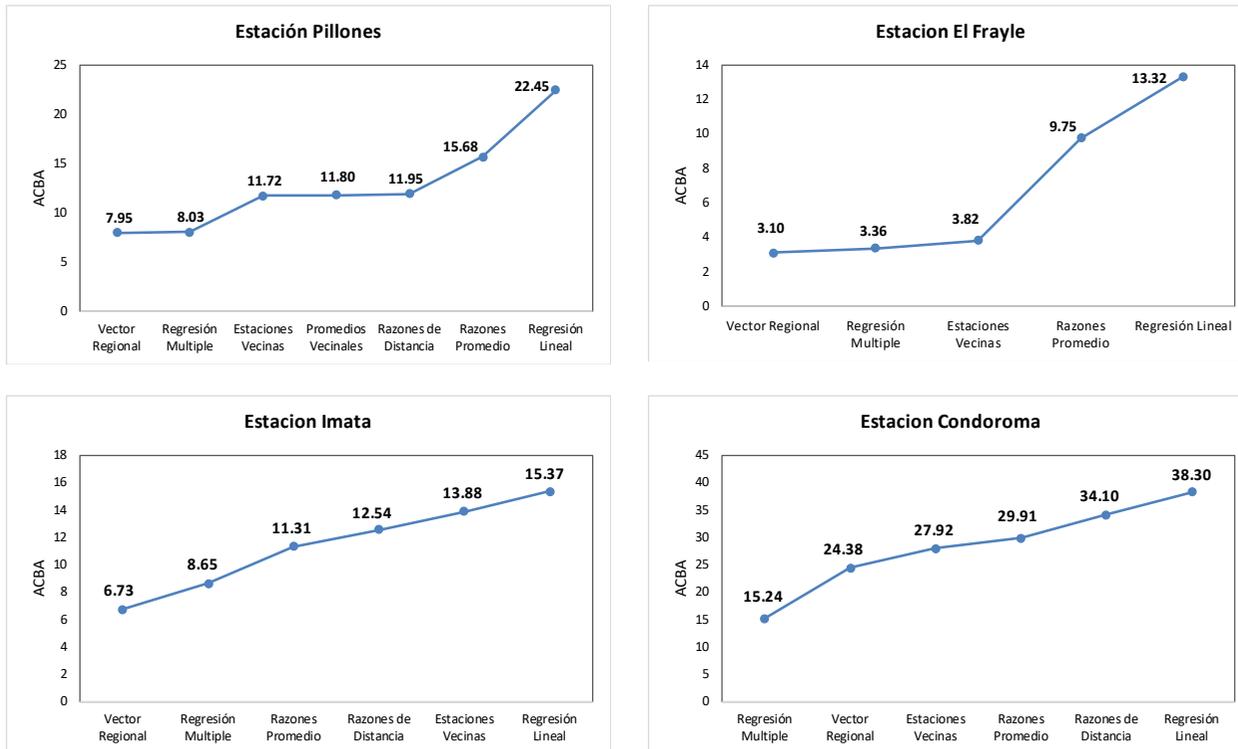
Fuente: Elaboración propia

En la Tabla 22, se muestra que en la estación Condoroma, el método de Regresión Multiple es el que muestra los mejores resultados con una diferencia promedio de 15.24 mm, seguido del método del Vector Regional con 24.38 mm.

Luego de haber aplicado el test de concordancia de Bland y Altman (ACBA), para cada estación; este método muestra que, al comparar los valores obtenidos con los datos reales, el que muestra el mejor ajuste y por tanto el mejor estimador sería, aquellos en donde las diferencias promedio son menores. En la Figura 5 se muestran que los métodos del Vector Regional y Regresión Múltiple son los mejores estimadores.

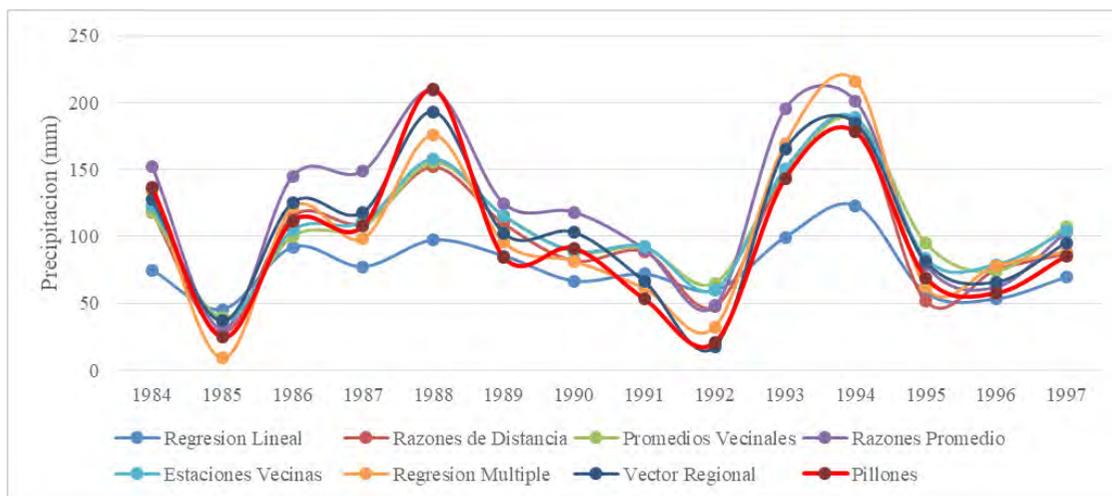


FIGURA 5. Diferencia promedio del análisis de concordancia de Bland y Altman (ACBA)



Fuente: Elaboración propia

FIGURA 6. Análisis gráfico entre el dato real y los obtenidos



Fuente: Elaboración propia

Conclusiones

Del análisis de los siete métodos, el que obtuvo mejores ajustes y por tanto el mejor estimador fue el método del Vector Regional y en segundo lugar el método de regresión lineal múltiple. Muy por el contrario, el método de regresión simple no obtuvo buenos resultados frente a los otros métodos.

De los 32 datos, se ha contabilizado que en 20 veces el método del Vector Regional es el mejor estimador, esto significa que en más del 65% de las veces el mejor estimador es el vector regional, mientras en 12 veces el método de Regresión Múltiple es el mejor estimador que representa el 35%.

Los diferentes métodos presentan buenos resultados en los meses entre octubre hasta mayo, mientras en los otros meses los resultados se presentan bastante alejados de los datos reales ya que frecuentemente en estos meses no se evidencian precipitación alguna y en años eventuales se pueden presentar algunas lluvias, estos al ser analizados generar datos poco representativos.

En la región analizada se sugiere que, para completar datos, se considere al método del Vector Regional, y en dependiendo de los meses del año el método regresión múltiple.

Bibliografía

- APARICIO, F. (2003). *Fundamentos de hidrología de superficie*. México: Editorial Limusa S.A. 303 pp.
- BLAND, J., ALTMAN, D. (1999). *Statistical methods in medical research. Measuring agreement in methods comparative studies*. Vol. 8 (2) pp. 35-160.
- BRUNET-MORET (1979). Homogénéisation des précipitations. Bureau Central Hydrologique de Orstom, Serie Hydrologie. 26(3): 147-170.
- COMISIÓN Económica para América Latina y el Caribe, Banco Interamericano de Desarrollo (2014). *La Económica del Cambio Climático en el Perú*. 17-18 p. Lima, Perú.
- ESPIÑOZA, J. (2005). *Manual escrito con motivo del curso taller. Utilización del Método del Vector Regional con Hydracces*. Senamhi-Perú. Colaboración de Philippe Vaulchel (IRD Perú) y Josyane Ronchail (UP7-Locean, Paris). 9 p.
- PIZARRO, R., GONZÁLEZ, P., WITTERSSHEIN, 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. Universidad de Talca, Talca, Chile. 135 p.
- PIZARRO R., AUSENSI P., ARAVENA D., SANGUESA C., LEON L., BALOCCHI F., (2009). Evaluación de métodos hidrológicos para la completación de datos faltantes de precipitación en estaciones de la región del Maule, Chile. *Aqua-LAC* - Vol. 1 - Nº 2. pp. 172-185. Chile.
- SMITH, D., KNIVETON, E., (1997). Statistical modeling approach to passive microwave rainfall retrieval. *Journal of Applied Meteorology*: Vol. 37, No. 2, 1997. 135-154.
- UNESCO-ROSTLAC (1982). *Guía metodológica para la elaboración del balance hídrico de América de sur*. Oficina Regional de Ciencias y Tecnología de la Unesco para América Latina y el Caribe. Montevideo. Uruguay. 129 pp. España. 745 pp.
- ZÚÑIGA S. (2010). *Climatología de la Reserva Nacional de Salinas y Aguada Blanca, Suroeste del Perú*. Ministerio del Ambiente. 264 pp.