

Evaluación de métodos hidrológicos para la completación de datos faltantes de precipitación en estaciones de la cuenca Jetepeque, Perú

Eduardo Luna Romero^{1*}; Waldo Lavado Casimiro²

Universidad Nacional Agraria La Molina, Perú¹; Servicio Nacional de Meteorología (SENAMHI), Perú²
info@agrosur.ec

Resumen. El presente trabajo evalúa la aplicabilidad de distintos métodos de completación de datos faltantes de la precipitación, en seis estaciones pluviométricas durante el periodo de 1992-2007 en la cuenca del Río Jequetepeque, Perú. Primero, se realizó el Análisis Exploratorio de Datos (AED) para determinar la existencia de valores atípicos de la información obtenida de la base datos del ANA (Autoridad nacional del Agua). Estos métodos corresponden a regresión lineal, razones de distancias, promedios vecinales, razones promedio y correlaciones con estaciones vecinas, propuestos por UNESCO-ROSTLAC, en 1982, en el marco del balance hídrico para América del Sur. Además se agregó el método de completación por regresión múltiple y el método del vector regional. Los métodos se analizan a través del coeficiente de determinación (R^2), error estándar de estimación (EEE), test de concordancia de Bland y Alamant (ACBA) y análisis de varianza (ANOVA), con los que se determina que método presenta mejor ajuste para la cuenca del río Jequetepeque. Según el ACBA, el mejor método de completación de datos para las estaciones de San Juan, Asunción y Contumaza es el vector regional, en las estaciones Levis y Magdalena el mejor método es el de regresión lineal, la regresión múltiple se presenta como mejor resultado en la estación Llapa, en los valores obtenidos del R^2 y EEE el mejor método es el vector regional en todas las estaciones.

Palabras claves: Cuenca Rio Jequetepeque, precipitación, completación de datos, Vector regional

ABSTRACT

This study evaluates the applicability of different methods of completion of missing data of rainfall in six meteorological stations for the period of 1992-2007 in Jequetepeque River Basin, Peru. First, the exploratory data analysis (AED) was performed to determine the existence of outliers of the information obtained from the database of the ANA (National Water Authority). These methods correspond to linear regression distances reasons, neighborhood average, average rates and correlations with neighboring stations, proposed by UNESCO-ROSTLAC, in 1982, under the water balance for South America. The method further completion by multiple regression and the method of regional vector was added. The methods are analyzed using the coefficient of determination (R^2), standard error of estimate (EEE), test matching Bland and Alamant (ACBA) and analysis of variance (ANOVA), with which it is determined that method has better fit Jequetepeque for the river basin. According to ACBA, the best method of completion of data for the stations of San Juan, Asunción and Contumaza is the regional vector, in Levis and Magdalena stations the best method is linear regression, multiple regression is

presented as best result the Llapa station in the values obtained from R^2 and EEE best regional vector method is in all seasons.

Keywords: Jequetepeque River Basin, rainfall, completion data, regional vector

INTRODUCCIÓN

La precipitación se define como la variable clave en los sistemas hidrológicos. En los diferentes estudios de investigación, el punto de partida es la estimación de la precipitación con adecuada resolución espacial y temporal (Tapiador et al., 2003; Tapiador et al., 2012; Schuurmans and Bierkens, 2007). La información de la precipitación exacta es de suma importancia para numerosas aplicaciones incluyendo la agronomía, hidrología, meteorología y climatología (Kidd and Levizanni, 2009; Hou et al., 2013, Scheel et al., 2010), entre ellas el poder pronosticar las máximas avenidas (Collier, 2007).

Los pluviómetros son la fuente principal de datos de precipitación, los datos pluviométricos tienen la ventaja de representar la medición directa de la lluvia (Scofield and Kuligowski, 2003). El Perú está dividido de norte a sur por la cordillera de los Andes en tres vertientes, una dreña hacia el océano Pacífico (VP), otra hacia el océano Atlántico (VA) a través del río Amazonas y hacia el sur al lago Titicaca (VT) (Lavado et al., 2012). El Servicio Nacional de Meteorología e Hidrología (SENAMHI, <http://www.senamhi.gob.pe>) es el encargado de monitorear y registrar los datos de la red hidrometeorológicos. De acuerdo al ministerio de agricultura (<http://www.minag.gob.pe>) la agricultura a lo largo de la VP se ha desarrollado gracias a la irrigación que utiliza el 80% de la disponibilidad de agua fresca, y, precisamente es en ésta vertiente es donde se concentra la mayor cantidad de estaciones meteorológicas. La VP está conformada por 53 cuencas hidrográficas, se caracterizan por ser cuencas pequeñas y sus aguas fluyen de este a oeste, desde los Andes hacia el océano Pacífico (Lavado et al., 2012).

Cuando en los registros pluviométricos existe una carencia de datos, existen diversos métodos para la completación de datos que parten desde la metodología descrita por la UNESCO-ROSTLAC (1982) en su guía metodológica para la elaboración del balance hídrico de América del Sur (Pizarro et al., 2009), hasta el uso de programas estadísticos y

metodología más compleja (Ramesh, 2014). La metodología establecida por la UNESCO-ROSTLAC (1982) es más factible de utilizar, se basan en fórmulas matemáticas simples, en donde se establecen relaciones entre estaciones patrones o cercanas (con datos completos) y la estación con carencia de información pluviométrica.

En el territorio peruano al igual que en el resto del planeta, la red pluviométrica se distribuye de manera desigual. En la actualidad, la precipitación puede ser estimada a través de la información registrada por satélites, como es el caso del TRMM (Tropical Rainfall Measuring Mission) y sus algoritmos (por ejemplo: 3B42V7, 3B42RT), pero cuya información debe ser validada con la información in situ. La calidad de los resultados de estos estudios es altamente dependiente de la calidad de la información disponible por los pluviómetros (Scheel et al., 2010).

En el contexto de una metodología tradicional y no muy compleja, la presente investigación compara siete métodos de completación de datos para la estimación de precipitación puntual; los cinco primeros métodos bajo los lineamientos y metodología establecida por UNESCO-ROSTLAC (1982): Regresión Lineal (RL), por Razón de Distancias (RD), Promedios vecinales (PV), Razones promedio (RP) y Correlación con estaciones vecinas (CEV); el sexto método es el de Regresión Múltiple (RM) (Pizarro et al, 2009); y el último método utilizado es el Vector Regional (VR) (Hiez, 1977 y Brunet Moret, 1979) en la cuenca del río Jequetepeque. El objetivo principal del presente trabajo fue evaluar la calidad de la predicción de dichos métodos, para su posterior recomendación.

MATERIALES Y MÉTODOS

Área de estudio

La cuenca del río Jequetepeque está ubicada en la costa norte del Perú, entre las coordenadas $6,81^{\circ}$ - $7,44^{\circ}$ S y $78,35^{\circ}$ - $79,59^{\circ}$ O. Tiene una superficie de $3\,956\text{ km}^2$. El río Jequetepeque recorre de este a oeste, desde la cordillera de los Andes hacia el océano Pacífico. Con del modelo digital de elevación (DEM) proporcionado por la Administración Nacional de Aeronáutica y del Espacio (NASA) a través de la Misión de Radar Topográfico del Transbordador Espacial

(SRTM) se puede observar su accidentada topografía que va desde el nivel del mar hasta su punto más alto de los 4189 msnm (Figura 1).

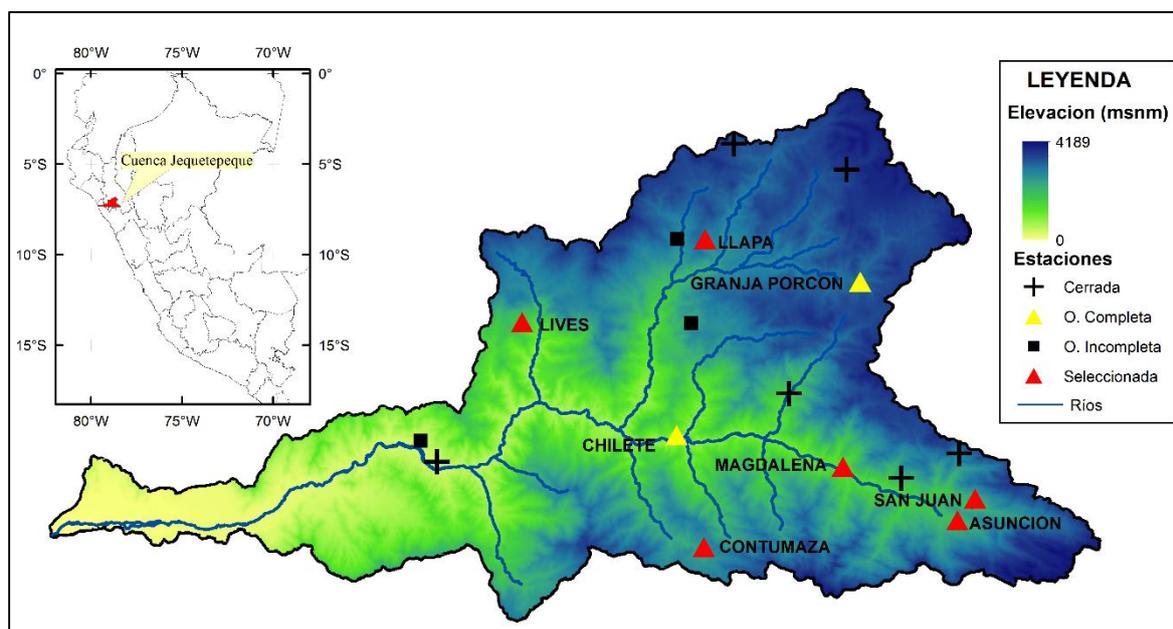


Figura 1. Mapa de ubicación de la cuenca del río Jequetepeque y ubicación de estaciones pluviométricas

La precipitación

Los datos de precipitación mensual de la cuenca del río Jequetepeque, fueron obtenidos de la página web de la Autoridad Nacional del Agua (ANA, <http://www.ana.gob.pe>), existiendo 14 estaciones pluviométricas (1989-2009), de las cuales se seleccionó 6 estaciones en estado operativo y con registro completo de 1992-2007 (Tabla 1).

Tabla 1. Estaciones pluviométricas de la Cuenca del Río Jequetepeque, serie 1992 – 2007

Nombre	Coordenadas UTM		Altitud msnm
	Este	Norte	
Contumaza	741024.07	9186971.85	2452
Lives	717223.23	9216579.05	2000
Llapa	741217.00	9227534.59	2798
Magdalena	759208.99	9197562.86	1316
San Juan	776501.91	9193345.42	2353
Asunción	774181.28	9190487.43	2229

Las precipitaciones se concentran en los meses de verano austral (Enero-Febrero-Marzo) (Figura 2), donde precipita entre un 50% y 69% del total anual. La estación de estiaje se produce en los meses de Junio-Julio-Agosto.

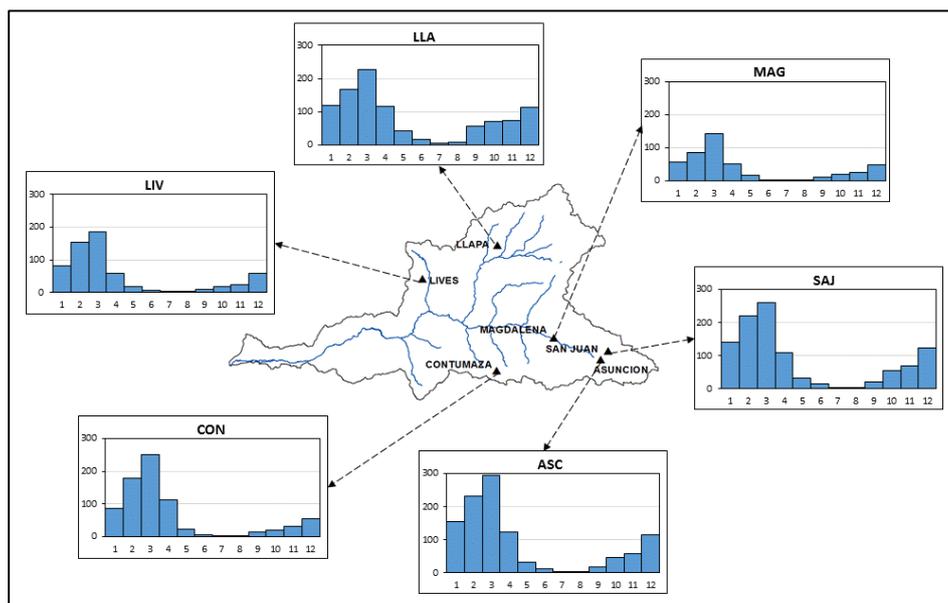


Figura 2. Precipitación media mensual para el periodo 1992-2007 (mm/mes) de las seis estaciones pluviométricas analizadas

Análisis exploratorio de datos (AED)

Antes de utilizar los diferentes métodos de completación de datos, se realizó el AED a las 8 estaciones seleccionadas en un inicio como operativas y completas, este análisis de la base de datos fue realizado con el método del vector regional (MVR) (Hiez, 1977 y Brunet Moret, 1979). La idea básica del MVR es el crear una estación ficticia (un vector) de todas las estaciones de una zona homogénea, que se compara con cada estación, en lugar de comparar estaciones pares por correlación o doble masa (Hiez, 1977; Vauchel, 2005). Con el MVR y bajo el criterio de la desviación estándar de desvíos ($DED < 0.20$) se decidió eliminar las estaciones de Granja Porcon (3000 msnm) y Chilite (850 msnm).

Descripción de los métodos de completación de información de datos faltantes

Método de regresión lineal

Este método es uno de los más utilizados; se recomienda para la estimación de datos mensuales y anuales. Se requiere establecer una regresión y correlación lineal con la estación patrón consistente, mediante una ecuación lineal de dos variables (1):

$$y = a + bx \quad (1)$$

Donde; y es la precipitación estimada, x es la precipitación de la estación patrón, a y b son las constantes de la regresión lineal.

A través del coeficiente de correlación (R) se determina la mejor estación vecina más cercana como patrón. En términos hidrológicos, se considera aceptable un valor de $-0.8 \leq R \leq 0.8$ (UNESCO-ROSTLAC, 1982; Pizarro *et al*, 1993; 2009).

Método por razones de distancias

Se utiliza para la estimación de datos menores de un año, en zonas planas; las estaciones deben tener una disposición espacial lineal como la que se muestra en la Figura 3. Donde A y B esté completa su información y la estación X está incompleta, en este caso se aplica la ecuación (2).

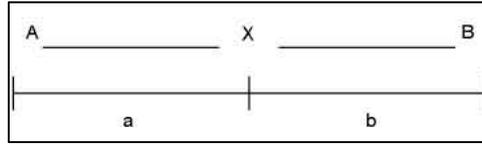


Fig. 3. Disposición espacial para la completación por razones de distancia. (Fuente: Pizarro et al, 1993).

$$PX = PA + a \times \frac{(PB-PA)}{(a+b)} \quad (2)$$

Donde; PX , PA , PB es la precipitación de las estaciones X , A , B , y, a y b es la distancia con relación a X .

Método por promedios vecinales

Se utiliza para la completación de datos menores de un año, en zonas planas no montañosas. Se debe considerar la distribución espacial en donde la estación con carencia de datos, queda ubicada al centro de tres estaciones con estadística completa (Figura 4), en este caso se aplica la ecuación (3).

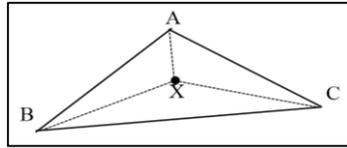


Fig. 4. Disposición espacial para la completación de datos por promedios vecinales (Pizarro et al, 1993)

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

Donde; P_i es la precipitación de las estaciones A , B , C , y n es el número de estaciones.

Método por razones promedio

Es complementario al método anterior para zonas montañosas, donde las precipitaciones de A , B y C diferirán generalmente en más de un 10%, entonces es necesario aplicar la ecuación (4) (UNESCO-ROSTLAC, 1982; Pizarro et al., 1993; 2009):

$$PX = \frac{\overline{PX}}{N} \times \left[\frac{PA}{\overline{PA}} + \frac{PB}{\overline{PB}} + \dots + \frac{PN}{\overline{PN}} \right] \quad (4)$$

Donde; \overline{PA} , \overline{PB} , \overline{PN} y \overline{PX} son el promedio de las precipitaciones anuales registradas en un período común para las N estaciones y PA , PB ,... PN es la precipitación en las N estaciones durante el período que falta en X .

Método por correlación con estaciones vecinas

Se utilizan las precipitaciones estimadas a partir de correlaciones entre la estación con datos faltantes y cada una de las estaciones vecinas, a los que se asocian los coeficientes de correlación respectivos (UNESCO-ROSTLAC, 1982). El coeficiente de correlación no debe superar la barrera del valor $\pm 0,8$ (CAZALAC, 2005). Su expresión matemática definida en la ecuación (5) es la siguiente:

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

Donde; PX es el valor estimado de precipitación, PXi son los valores estimados de precipitación en X , a partir de las regresiones con cada una de las i estaciones y rXi son los coeficientes de correlación entre los registros de la estación X y cada una de las i estaciones.

Método por regresiones múltiples

Es una extensión del método de regresión simple y su forma de estimar la precipitación se basa en la expresión matemática (6):

$$Y = a + bX_1 + cX_2 + \dots + nX_1 \quad (6)$$

Donde; y es la precipitación estimada, xi es la precipitación de la estación completas, y , a , b , c y n son las constantes de la regresión.

Método del vector regional.

Es un método de cálculo orientado a tres tareas definidas: la crítica de datos, la homogenización y la extensión-completación de datos de precipitación. El vector regional elabora una estación ficticia (vector) que sea una “especie de promedio” de todas las estaciones de la zona. El vector se calcula bajo el concepto de precipitación media extendida, salvando los problemas del peso de las estaciones más lluviosas sobre las menos lluviosas. Se emplea el método de mínimos cuadrados para encontrar los índices pluviométricos regionales anuales Zi y la precipitación media extendida Pj , esto se logra al minimizar la expresión (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 del año, j es el índice de la estación, N es el número de años, M es el número de estaciones. P_{ij} es la precipitación anual en la estación j el año i , P_j es la precipitación media extendida al período de N años y finalmente Z_i es el índice pluviométrico regional del año i . Para ejecutar este método se hizo uso del paquete computacional HYDRACCES, y proporciona información diversa en hojas de Microsoft Excel masa (Vauchel, 2005).

Completación de datos

Los datos de precipitación mensual corresponden a 6 estaciones para un periodo de 15 años (1992-2007), se obtuvo un total de 180 datos mensuales por estación. Se hizo una selección aleatoria del 20% de los datos por estación, los que fueron considerados como faltantes, con el fin de generar vacíos de información, para posteriormente establecer las comparaciones entre las precipitaciones reales y las estimadas. Se debe señalar que para poder utilizar de forma comparativa los métodos, se estableció que los meses en los cuales se eliminó la información, debían ser los mismos en cada estación.

Para la aplicación del método de RL, se determinó el R con las tres estaciones más cercanas a la estación planteada como incompleta. Para el método de RD se consideraron las menores distancias con relación a la estación de datos faltantes. El método de PV considera una disposición espacial de tipo triangular, la única estación que cumplió con esta condición fue la estación Magdalena, y a su vez se le aplicó el método de RP. Las correlaciones con estaciones vecinas se realizaron en dos escenarios; el primero, con dos estaciones y, el segundo, con tres estaciones, con el fin de establecer si el número de estaciones incluidas en la estimación, tiene relación con la calidad de ésta (Pizarro et al., 1993; 2009). En el método de RM se utilizó la misma combinación del CEV y bajo el mismo escenario, lo que busca es establecer si la cantidad y calidad de las estimaciones hechas con el método de RL, se ve afectada por la cantidad de estaciones incluidas.

Análisis comparativo

Realizadas las estimaciones de los datos faltantes, se hizo una comparación entre los valores reales y los valores estimados a través de los diversos métodos de completación, comparándolos entre sí para detectar diferencias estadísticamente significativas entre ellos y se aplicaron las siguientes medidas de bondad de ajuste: Coeficiente de Determinación (R^2), Error Estándar de Estimación (EEE), Análisis de Varianza (ANOVA) y el Test de concordancia de Bland y Altman (ACBA). Este último modelo se basa en las diferencias promedio existentes entre el valor real y el estimado, permite apreciar la homogeneidad del error. El nivel de concordancia lo asignan las diferencias promedio (dp) entre el valor real y el estimado y la desviación estándar (SD) de dichas diferencias. Además, establece un límite de concordancia (LC) con un 95% del nivel de significación (Bland y Altman, 1999; Dunn y Roberts, 1999), definido por la siguiente expresión (8):

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

RESULTADOS

Análisis comparativo de los métodos

En la tabla 2 se presentan los valores del R^2 y el EEE para cada estación, según el método aplicado. Los Errores Estándar obtenidos en las estaciones, no presentaron grandes diferencias entre los distintos métodos de completación utilizados para la misma estación, a excepción de la estación Magdalena, donde se presenta una marcada diferencia entre los métodos relacionados con la distancia (RD) y disposición espacial (PV). Por otro lado, el VR presenta un EEE menor en todas las estaciones, por consiguiente, bajo este contexto estadístico los datos simulados son los más cercanos a los observados.

Tabla 2. Coeficientes de Determinación y Error Estándar de Estimación

METODO	Lives		Llapa		Magdalena		Contumaza		San Juan		Asunción	
	R^2	EEE	R^2	EEE	R^2	EEE	R^2	EEE	R^2	EEE	R^2	EEE
Regresión lineal 1estacion	0.84	21.50	0.91	26.51	0.92	9.16	0.90	19.58	0.98	14.99	0.98	11.71
Regresión lineal 2estacion	0.87	18.18	0.80	39.61	0.95	7.10	0.87	21.71	0.87	35.85	0.92	24.61
Regresión lineal 3estacion	0.81	18.05	0.92	22.53	0.90	13.20	0.87	22.25	0.95	22.98	0.88	33.18
Razones de distancia	0.84	31.20	0.86	38.64	0.94	49.64	0.90	21.45	0.99	11.04	0.98	18.12
Razones promedios	--	--	--	--	0.96	7.13	--	--	--	--	--	--
Promedios vecinales	--	--	--	--	0.96	39.79	--	--	--	--	--	--
Correlación con 2 estaciones vecinas	0.87	16.73	0.92	23.43	0.94	49.55	0.90	20.87	0.99	21.31	0.97	29.39
Correlación con 3 estaciones vecinas	0.87	23.62	0.92	23.42	0.95	47.24	0.91	19.90	0.98	19.21	0.97	28.16
Regresión múltiple 2 estaciones	0.85	21.09	0.92	19.91	0.93	7.94	0.90	20.40	0.99	13.88	0.98	12.27
Regresión múltiple 3 estaciones	0.85	20.93	0.92	19.86	0.94	8.11	0.90	18.86	0.98	12.68	0.98	12.50
Vector regional	0.91	16.55	0.94	15.56	0.96	5.75	0.91	18.03	0.98	11.00	0.98	11.22

En la Tabla 2 se muestran valores de R^2 superiores a 0.80, incluso en algunos métodos llegando a 0.99, pero en ellos el EEE es alto. Por lo tanto, el considerar al R^2 como un método válido de bondad de ajuste, es un error, ya que sólo sirve a modo de referencia y se debe aplicar con precaución, pues no mide qué tan apropiado es el método empleado (Montgomery y Runger, 1994, Pizarro et al., 2009).

Por otro lado, los valores más bajos del EEE con el método de RD se obtuvieron en las estaciones de una altitud casi igual como son Contumaza (2452 msnm), San Juan (2353 msnm)

y Asunción (2229 msnm). Además estas mismas estaciones cuentan con valores de EEE bajos de forma general en todos los métodos, debido a que se encuentran relativamente cerca y adicionalmente el factor de altura no es muy diferente entre ellas.

Para determinar la existencia de diferencias entre los distintos métodos utilizados para cada estación, se realizó un análisis de varianza (ANOVA), que mide el comportamiento de los efectos entre las medias de cada método (Tabla 3). La estación Magdalena es la única que manifestó diferencias significativas entre los distintos métodos utilizados, con un valor p muy bajo (inferior a 0,05).

Tabla 3. Análisis de varianza de los métodos de completación de datos

Estación	VF	SC	GI	CM	FC	Valor P	Valor Crítico F
Lives	Dentro del grupo	15631.72	9	1736.86			
	Entre grupos	2869523.08	310	9256.53	0.1876	0.9953	1.9101
	TOTAL	2885154.80	319				
Llapa	Dentro del grupo	56424.30	9	6269.37			
	Entre grupos	2576377.12	310	8310.89	0.7544	0.6588	1.9101
	TOTAL	2632801.42	319				
Contumaza	Dentro del grupo	3896.91	9	432.99			
	Entre grupos	3338073.64	310	10767.98	0.0402	0.9999	1.9101
	TOTAL	3341970.56	319				
Magdalena	Dentro del grupo	178004.26	11	16182.21			
	Entre grupos	2891901.79	372	7773.93	2.0816	0.0209*	1.8144
	TOTAL	3069906.05	383				
Asunción	Dentro del grupo	35179.46	9	3908.83			
	Entre grupos	5013056.69	310	16171.15	0.2417	0.9880	1.9101
	TOTAL	5048236.15	319				
San Juan	Dentro del grupo	23518.54	9	2613.17			
	Entre grupos	4176641.96	310	13473.04	0.1940	0.9947	1.9101
	TOTAL	4200160.50	319				

Con el test de concordancia de Bland y Altman (ACBA), se determinó los datos que se ajustan mejor a los valores reales, a través de las diferencias promedio. En la Figura 5 se muestran las diferencias promedios (dp) de cada estación entre los valores simulados y los observados.

El vector regional se muestra en las estaciones Contumaza (CON), San Juan (SAJ) y Asunción (ASC) (Figura 5a, 5e, 5f) como el mejor método que se ajusta a los valores observados. Teniendo un dp de 1.64 mm, 1.39 mm y 1.87 mm, respectivamente.

La regresión lineal de forma general se ubica como mejor método en las estaciones Levis (LEV) y Magdalena (MAG), dp de 3.63 mm y 0.45 mm, respectivamente.

En la estación Llapa (LLA) se presenta como mejor método según el ACBA el de regresión múltiple con dos estaciones con un dp de 2.90 mm.

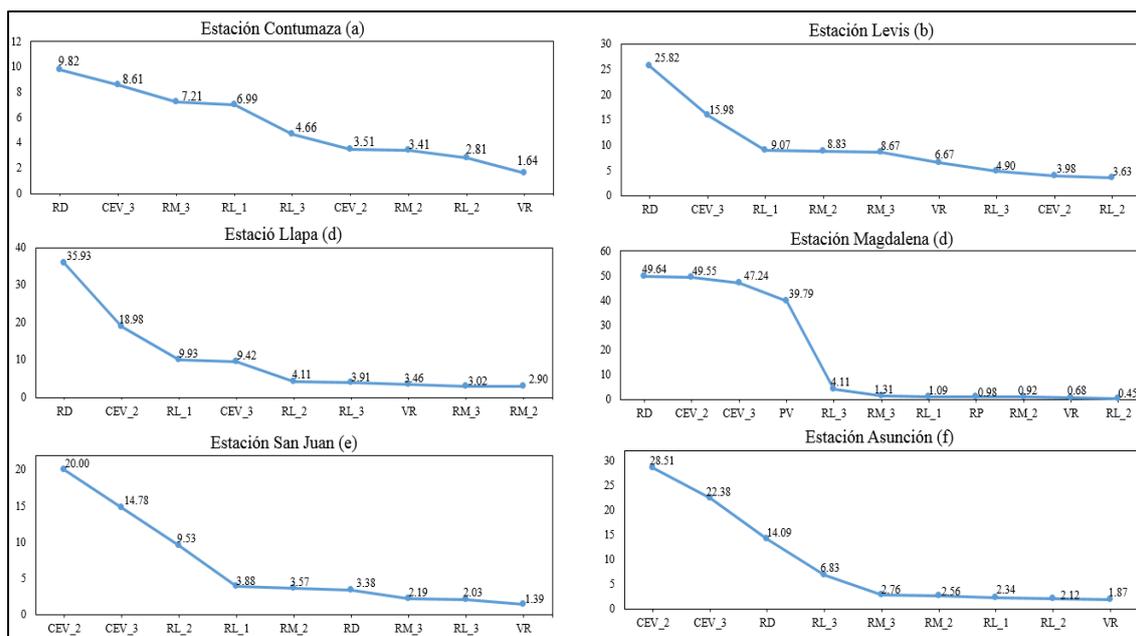


Figura 5. Diferencia promedio del análisis de concordancia de Bland y Altman (ACBA)

CONCLUSIONES

En el análisis realizado a los diferentes métodos de completación de datos en la cuenca del río Jequetepeque, se concluye con los resultados obtenidos por el test ACBA, que el vector regional es el mejor método en las estaciones SAJ, CON y ASC, la regresión lineal se presenta como mejor método en MAG y LEV, en tanto la regresión múltiple con 2 estaciones se muestra como mejor método en la estación LLA.

El resultado del vector regional obtenido con el EEE coinciden con las estaciones del ACBA, las cuales se encuentran cerca entre ellas y a una altitud similar, presentando precipitaciones con buena correlación en cuanto a la distribución espacial.

La estación MAG se encuentra cerca de SAJ, CON y ASC, pero con una diferencia de altitud de alrededor de 900 m. LEV y LLA se encuentran más alejadas de este grupo de estaciones, debido a la distribución espacial de la lluvia en estas estaciones se presenta como mejor método la regresión, lineal y múltiple, respectivamente.

La ventaja de utilizar el VR, es la extensión de datos, pero para ello se debe tener un grupo de estaciones que guarde una buena correlación y además la DED debe ser menor a 0.2; y bajo estos lineamientos el VR realizará una correcta completación de datos, tal como lo revela el ACBA. Por otro lado, las estaciones que se encuentren espacialmente muy alejadas y donde exista una variabilidad de la precipitación, lo más idóneo será utilizar los métodos de regresión.

BIBLIOGRAFIA

- Bland, J, Altman, D. 1999. Statistical methods in medical research. Measuring agreement in methods comparative studies. Vol. 8 (2) pp. 35-160.
- Brunet-Moret, Y. 1979. Homogénéisation des précipitations. Cahiers ORSTOM, Série Hydrologie, 16:3-4.
- Centro del Agua para Zonas Áridas de América Latina y el Caribe (CAZALAC). 2005 Guía metodológica para la elaboración del mapa de zonas áridas, semiáridas y subhúmedas secas de América Latina y El Caribe. Chile. 66p.
- Collier, C.G., 2007. Flash flood forecasting: what are the limits of predictability? Q. J. R. Meteorol. Soc. 133 (622), 3–23.

- Dunn, G., Roberts, C. 1999. Statistical methods in medical research. Biostatistics group. The medical school. Modelling method comparison data. Vol. 8 (2) pp. 161-179.
- Espinoza, J. 2005. Utilización del vector regional con HYDRACCES. SENAMHI- PERÚ.
- Hiez, G. 1977. L'homogénéité des données pluviométriques. Cahiers ORSTOM, série Hydrologie, 14:129-172.
- Hou, A.Y., kakar, R., Neeck, S., Azabarzin, A., Kummerow, C., Kojima, M., Oki, R., Nakamura, K., Iguchi, T., 2013. The Global Precipitation Measurement (GPM) Mission. Bull. Am. Meteorol. Soc. <http://dx.doi.org/10.1175/BAMS-D-13-00164>. 1 (e-View).
- Kidd, C., Levizzani, V., 2009. Status of satellite precipitation retrievals. Hydrol. Earth Syst. Sci. 15, 1109–1116.
- Waldo Sven Lavado Casimiro , Josyane Ronchail , David Labat , Jhan Carlo Espinoza & Jean Loup Guyot. 2012. Basin-scale analysis of rainfall and runoff in Peru (1969–2004): Pacific, Titicaca and Amazonas drainages, Hydrological Sciences Journal, 57:4, 625-642, DOI: 10.1080/02626667.2012.672985.
- Montgomery, D., Runger, G. 1994. Applied statistics and Probability for engineers. New York: Editorial Mc Graw-Hill.
- 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: Universidad de Talca. pp 135.
- Pizarro, R., Ausensi, P., Aravena, D., Sangüesa, C., León, L., Balocchi, F. 2009. Evaluación de métodos Hidrológicos para la completación de datos faltantes de precipitación en estaciones de la región de Maule, Chile. Chile. Revista Aqua-LAC. Vol. 1 - Nº 2 - Sep. 2009. pp. 172-185.
- Ramesh S. V. Teegavarapu. 2014. Missing precipitation data estimation using optimal proximity metricbased imputation, nearest-neighbour classification and cluster-based interpolation methods, Hydrological Sciences Journal, 59:11, 2009-2026, DOI: 10.1080/02626667.2013.862334.
- Scheel, M.L.M., Rohrer, M., Huggel, C., Santos Villar, D., Silvestre, E., Huffman, G.J., 2011. Evaluation of TRMM Multi-Satellite Precipitation Analysis (TMPA) performance in the Central Andes region and its dependency on spatial and temporal resolution. Hydrol. Earth Syst. Sci. 15, 2649–2663.
- Schuermans, J.M., Bierkens, M.F.P., 2007. Effect of spatial distribution of daily rainfall on interior catchment response of a distributed hydrological model. Hydrol. Earth Syst. Sci. 11, 677–693.
- Scofield, R.A., Kuligowski, R.J., 2003. Status and outlook of operational satellite precipitation algorithms for extreme-precipitation events. Weather Forecast. 18, 1037–1051.
- Tapiador, F., 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, Vol. 43 (4) pp. 576–594.
- Tapiador, F., Turk, F.J., Petersen, W., Hou, A., García-Ortega, E., Machado, L., Angelia, C., Salio, P., Kidd, C., Huffman, G., de Castro, M., 2012. Global precipitation measurement: Methods, datasets and applications. Atmos. Res. 104–105, 70–97.
- 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 p - España. 745p.
- Vauchel, P., 2005. Hydraccess: Logiciel de gestion et traitement de données hydro météorologiques, version 2.1.4. Downloadable at <http://www.mpl.ird.fr/hybam.01/01/2011>.