

EVALUACIÓN DE MÉTODOS HIDROLÓGICOS PARA LA COMPLETACIÓN DE DATOS FALTANTES DE PRECIPITACIÓN EN ESTACIONES DE LA REGIÓN DEL MAULE, CHILE.

EVALUATION OF HYDROLOGIC METHODS FOR COMPLETING RAINFALL MISSING VALUES IN MAULE REGION OF CHILE

Roberto Pizarro ¹, Paula Ausensi ², Dayanna Aravena ²,
Claudia Sangüesa ¹, Lastenia León ², Francisco Balocchi¹

Resumen

Este estudio evalúa la aplicabilidad de distintos métodos para la estimación de datos faltantes de precipitación puntual, en ocho estaciones pluviográficas de la Región del Maule, Chile. Estos métodos corresponden a correlación lineal, completación por razones de distancia, completación por promedios vecinales, completación por razones promedio y completación por correlación con estaciones vecinas, propuestos por UNESCO-ROSTLAC, en 1982, en el marco del balance hídrico para América del Sur. Además se agregó un sexto método, el de completación por regresiones múltiples. Los métodos se analizan a través del coeficiente de determinación (R^2), error estándar de estimación (EEE), test de concordancia de Bland y Altman y análisis de varianza, con los que se determina que método presenta mejor ajuste para la región. Según el análisis de Bland y Altman, el mejor método de completación fue el de regresión múltiple con 2 y 3 estaciones cercanas, lo que además se corrobora con los valores obtenidos del R^2 y el EEE. Asimismo, los errores estándar de los otros métodos estudiados fueron demasiado altos, lo que los hace no recomendables, excepto el de correlación con estaciones vecinas que tiende a acercarse en sus resultados al de regresiones múltiples.

Palabras clave: Completación datos faltantes, precipitación puntual.

Abstract

This study evaluates the applicability of different methods for the estimation of rainfall missing values, through eight raingauge stations in Maule Region of Chile. These methods were linear correlation, distance rate, local averages, mean rates and correlation with nearby stations, proposed by UNESCO-ROSTLAC in 1982, within the framework of the hydrologic balance for South America. In addition, was added another method named multiple regressions. These methods were analyzed through different statistical methods; these were determination coefficient (R^2), standard error of estimation (EEE), test of agreement of Bland and Altman and analysis of variance, in order to define the best goodness of fit method. According to the analysis of Bland and Altman, the best method of completing missing rain value, was multiple regression with 2 and 3 nearby stations. This result is also validated by R^2 values, and also validated by standard error of estimation. On the other hand, the other methods studied, showed important differences between real values and estimated values, so it was concluded that multiple regressions is the best method to complete missing rain value.

Keywords: Rainfall missing value; rainfall data base; rainfall estimation.

INTRODUCCIÓN

En la mayoría de los estudios relacionados con hidrología y en investigaciones de los recursos naturales o relacionados con el medio ambiente, el punto de partida es la estimación de las precipitaciones (Tapiador *et al*, 2003). Por ello, conocer el comportamiento y la forma de evaluación que tiene la precipitación es de gran importancia (Aparicio, 2001).

En la estimación de la precipitación, cuando hay carencia de datos, existen diversos métodos que van desde avanzadas tecnologías como el uso de satélites, programas estadísticos y modelaciones hidrológicas (Smith *et al*, 1997), que por su elevado

costo, no son de masiva utilización, hasta los métodos tradicionales, que son más factibles de utilizar. Estos últimos, 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. De esta forma, se pueden mencionar métodos estadísticos como promedios simples, regresiones lineales múltiples o modelos de series de tiempo (Kim *et al*, s.f.). Por otra parte, una de las modelaciones hidrológicas estadísticas que identificaron en décadas pasadas el problema de la falta de datos, el estudio realizado utilizó el modelo espacial normal de primer orden de Marcov; dicha investigación observó datos de sectores contiguos para estimar los valores

¹: Universidad de Talca, Facultad de Ingeniería Forestal, 2 Norte 285, Código Postal 3465548, Casilla Postal 747 y 721.
e-mail: rpizarro@utalca.cl

²: Dirección General de Aguas, . Ministerio de Obras Públicas de Chile

faltantes de precipitaciones en donde; los resultados arrojaron que es un buen modelo empírico para la estimación de los datos de precipitación para las ciudades de Kansas y Nebraska (Haining, 1984)..

Bajo el contexto de la metodología tradicional, la presente investigación compara cinco métodos de completación de datos para la estimación de precipitación puntual, cuando hay carencia de información en distintas estaciones pluviométricas de la Región del Maule, Chile, con el fin de evaluar la calidad de la predicción de dichos métodos, para su posterior recomendación.

MÉTODO

La investigación fue realizada en 8 estaciones de la Región de Maule, Chile (figura 1, tabla 1), la que se extiende entre los 34° 41' y 36° 33' de latitud Sur (IGM, 1985). La zona se encuentra dominada por condiciones anticiclónicas durante gran parte del año y por condiciones frontales intermitentes durante el invierno. Las precipitaciones son casi exclusivamente de origen frontal y se concentran en los meses de mayo y agosto, donde precipita entre un 70% ó 75% del total anual; entre octubre y marzo ocurre la estación seca, que se prolonga entre 4 y 6 meses.

Tabla 1. Coordenadas UTM estaciones pluviométricas utilizadas.

Estación con carencia de información	Coordenadas UTM	
	Este	Norte
Armerillo	309995	6047229
Colorado	294700	6056138
Gualleco	228487	6096938
Huapi	296125	6059871
El Guindo	286227	6094789
Pencahue	242669	6082024
San Javier	256902	6057037
Talca	260939	6075651



Figura 1. Ubicación estaciones pluviográficas

Descripción de los métodos de completación de información de datos faltantes

Para la determinación de los datos faltantes se utilizan variados métodos, los cuales normalmente se basan en fórmulas empíricas, donde se relacionan determinadas variables hidrológicas que, independientemente de su base teórica, responden a la estructura de un estudio estadístico (Mintegui y López, 1990).

En Chile, se utilizan diversos procedimientos para la estimación de la precipitación puntual, cuando hay carencia de información, principalmente los métodos del “Balance Hídrico para América del Sur” (UNESCO-ROSTLAC, 1982).

Completación por regresión lineal

Este método es uno de los más utilizados; se recomienda para la estimación de datos mensuales y anuales de la estación en estudio y las de una estación 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 estación que tenga carencia de información, mediante una ecuación lineal de dos variables (1), del tipo:

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

Donde;

\hat{Y} = Valor estimado de la precipitación para la estación carente (mm).

x = Valor de precipitación registrado en la estación patrón (mm).

a, b = constantes de regresión.

Este método, a pesar de ser de muy fácil aplicación, no puede ser aplicado indiscriminadamente, dado que es necesario saber si la calidad del ajuste es buena o mala. Una mala calidad del ajuste, puede llevar a la generación de información sin consistencia, lo cual en lugar de mejorar la situación, la empeora por la agregación estadística de datos no representativos de la realidad que se pretende (UNESCO-ROSTLAC, 1982).

Por esta razón, es posible utilizar el coeficiente de correlación como una forma de establecer la calidad de los datos. Con la determinación del coeficiente de correlación (R), se puede estimar el grado de correlación lineal que existe entre las estaciones en estudio, y cuyo valor oscila entre $-1 \leq R \leq 1$, donde el valor 0 indica una correlación nula, en tanto los valores 1 y -1 , denotan una correlación total. En términos hidrológicos, se considera aceptable una regresión cuyo valor de R sea mayor a 0,8 ó menor que $-0,8$ (Pizarro *et al*, 1993). Además deben utilizarse distintos métodos estadísticos, que permitan evaluar la calidad de los ajustes obtenidos.

Completación por razones de distancias

Este método se utiliza para la estimación de datos menores de un año, en zonas planas no montañosas; las estaciones deben tener una disposición espacial lineal como la que se muestra en la Figura 2.

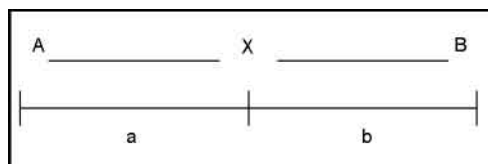


Figura 2. Disposición espacial para la completación, por razones de distancia. (Fuente: Pizarro et al, 1993).

Donde; X = Representa la posición de una estación con carencia de información; A y B = Señala la presencia de estaciones con información completa; a y b = Representa la distancia sobre un plano desde la estación X.

La estación con carencia de datos debe quedar entre dos estaciones que presenten una estadística completa; así, y utilizando la siguiente expresión (2), se puede estimar el dato faltante.

$$PX = PA + a * \left[\frac{(PB - PA)}{(a + b)} \right] \quad (2)$$

Donde;

PX, PA, PB, representan la precipitación para las estaciones X, A y B, respectivamente, para el período en estudio.

En este método se ocupan las precipitaciones y las distancias, por lo tanto se asume que existe una variación lineal de las precipitaciones, en función de la disposición espacial.

Estamethodologíaesunadelasutilizadasporhidrólogos e ingenieros en la estimación de precipitaciones para estudios realizados en Sudáfrica, cuyas proyecciones son de alta importancia especialmente para los países que carecen de agua (Lynch, 1998).

Completación 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 3).

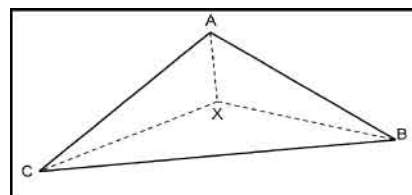


Figura 3. Disposición espacial para la completación de datos por promedios vecinales (Pizarro et al, 1993)

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.

De cumplir con esto, es posible la utilización de la expresión (3) para estimar las precipitaciones.

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

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

Este método, es básicamente una estimación que resulta del cálculo de un promedio aritmético en las n estaciones vecinas existentes.

Completación por razones promedio

Es complementario al método anterior, y se define por la siguiente base conceptual. Si en una zona cualquiera, la precipitación normal de la estación X, difiere en más de un 10% con alguna de las estaciones vecinas, entonces es necesario aplicar la ecuación (4) (UNESCO-ROSTLAC, 1982; Linsley et al, 1988; Pizarro et al, 1993):

$$P_x = \bar{P}_x * \left[\frac{PA}{\bar{P}_A} + \frac{PB}{\bar{P}_B} + \dots + \frac{PN}{\bar{P}_N} \right] \quad (4)$$

Donde; $\bar{P}_x, \bar{P}_A, \bar{P}_B, \dots, \bar{P}_N$ = promedio normal de las precipitaciones anuales registradas en un período común para las N estaciones y la estación X; PA, PB, ...PN = precipitación en las N estaciones durante el período que falta en X.

Lo que se logra al aplicar este método, es realizar una estimación para la estación faltante, en función de las relaciones entre las precipitaciones de un período en estudio y las precipitaciones normales (Pizarro et al, 1993).

Cabe señalar que se entiende por precipitación normal, el promedio de los últimos treinta años; de no contarse con los registros de esta duración, se puede establecer un periodo común a las estaciones en análisis, que se sugiere sea mayor o igual a 20 años en lo posible.

Completación por correlación con estaciones vecinas

Aquí se utilizan las precipitaciones estimadas a partir de correlaciones entre la estación con carencia de datos y cada una de las estaciones vecinas, a los que se asocian los coeficientes de correlación respectivos (UNESCO-ROSTLAC, 1982). 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 = Valor estimado de precipitación en X .

PXi = Valor estimado de precipitación en X , a partir de las regresiones con cada una de las i estaciones.

rXi = Coeficiente de correlación entre los registros de la estación X , y cada una de las i estaciones.

A, B, \dots, N = Estaciones consideradas.

Este método sirve para la completación de información de tipo anual, y su uso es sólo recomendable cuando el coeficiente de correlación del método de correlación lineal, no supera la barrera del valor $\pm 0,8$ (CAZALAC, 2005).

Completación por regresiones múltiples

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

$$\hat{Y} = a + bX_1 + cX_2 + dX_3 + \dots + nX_i \quad (6)$$

Donde; \hat{Y} = Valor de precipitación estimada para la estación con carencia de Información; X_i = Valor de precipitación en estaciones con información completa.

a, b, c, n = Constantes de regresión.

Este método es utilizado en esta investigación, con el fin de investigar sus capacidades frente a los otros métodos tradicionales ya descritos, cuyos resultados presentan falencias importantes.

Por otra parte, existe otro método de amplia difusión en el mundo, denominado método Kriging, el cual ha sido ampliamente aplicado a varios tópicos de la investigación hidrológica, en particular la estimación espacial de la lluvia y las propiedades del suelo (Chang et al, 1999), lo que también se muestra en los trabajos de Yu y Chen (1996 y 1997) que estudiaron las fórmulas regionales de las curvas Intensidad Duración Frecuencia, IDF, y su extensión en zonas sin pluviómetros, y Yu et al (2004), que propusieron un acercamiento alternativo basado en las propiedades fractales de la lluvia, para construir las fórmulas de escalamiento de la IDF. Sin embargo, este método requeriría un estudio

por sí mismo, dado que involucra una serie de pasos estadísticos y contar con una amplia gama de datos, todo lo cual excede los objetivos de este estudio, que pretende analizar los métodos propuestos por la metodología de elaboración del balance hídrico de América del Sur, propuesta por Unesco en 1982 y que ha tenido sucesivas actualizaciones.

En relación a los datos requeridos para la aplicación de los métodos de completación de datos faltantes, éstos corresponden a los valores de precipitación mensual de las 8 estaciones consideradas para el estudio y para un período de 15 años. De este modo, se obtuvo un total de 180 datos mensuales por estación. Con esta información 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.

Una vez establecidos los vacíos de información mensual, los datos que se extrajeron se consideraron como la precipitación real, que fue la base para establecer las comparaciones entre los distintos métodos (Tabla 2).

Para la aplicación del método de completación por regresión lineal, se establecieron regresiones lineales simples, entre los valores de las precipitaciones que no fueron eliminados para la estación X , y los valores de la 1ª, 2ª y 3ª estación más cercana (Tabla 3).

Para comprobar la calidad de los ajustes, fue necesario aplicar los supuestos de normalidad, a través del Test de Kolmogorov – Smirnov y el gráfico de probabilidad normal de residuos; el supuesto de homocedasticidad, por medio del gráfico de residuos; y el supuesto de no autocorrelación o independencia, utilizando para ello el método d de Durbin-Watson, apoyado en el gráfico de residuos versus el tiempo.

En relación a los supuestos de regresión lineal, cabe destacar que si el objetivo del modelo que se quiere ajustar es únicamente la estimación puntual, (como lo es en el caso de los métodos de completación), el método de los mínimos cuadrados en conjunto con el coeficiente de determinación R^2 , son suficientes, y no es necesario que se cumplan los supuestos clásicos de regresión (Gujarati, 1992). Sin embargo, con el fin de tener una mejor herramienta de decisión y comparar los resultados del estudio con algunas extensiones de éste, de igual forma se obtuvieron los modelos con sus respectivos supuestos clásicos de regresión.

Por otro lado, para el método de completación por razones de distancias, la estación con carencia de información debe estar al centro de dos estaciones que tengan estadística completa. Por ello, en primer lugar fue necesario establecer qué estaciones podían

Tabla 2. Precipitaciones reales que fueron eliminadas en cada estación.

Año	meses	Estaciones							
		Armerillo	Colorado	El Guindo	Gualleco	Huapi	Pencahue	San Javier	Talca
1989	ene	5,5	3,5	3,0	1,5	1,5	6,3	5,8	3,0
	oct	50,0	38,0	23,0	11,0	24,0	8,0	8,5	10,6
	nov	6,0	2,5	7,0	0,0	2,0	0,0	2,5	1,7
1990	abr	108,9	50,2	41,0	41,1	50,0	25,3	41,5	22,3
	nov	82,5	47,8	21,5	14,7	30,5	17,6	29,0	19,5
1991	ene	11,5	3,5	2,0	20,1	6,0	24,0	21,0	8,7
	mar	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
	abr	150,0	70,2	45,5	16,4	55,9	20,9	33,7	24,7
1992	jun	438,1	286,0	117,5	114,5	170,5	149,1	171,1	155,3
	jul	159,3	134,3	46,5	36,5	62,0	22,1	41,3	28,2
1993	agost	199,6	152,5	66,5	117,7	99,0	121,0	106,7	95,6
	jul	223,9	159,0	69,5	89,9	138,5	60,1	95,4	85,3
	oct	65,4	52,5	10,0	5,9	0,0	9,2	34,5	14,7
1995	dic	113,4	37,5	9,0	9,3	0,0	2,5	10,6	9,6
	nov	1,0	0,8	0,0	0,0	1,0	0,0	0,0	0,0
1996	mar	41,9	9,2	1,0	0,0	5,0	4,0	5,7	4,8
	oct	33,0	8,5	6,5	0,0	10,0	5,6	3,5	5,9
1997	jun	1034,2	595,5	327,5	375,0	423,5	345,3	350,0	315,4
	oct	151,0	87,9	16,0	84,5	76,2	69,9	59,9	56,0
1998	jun	66,0	34,5	13,5	14,0	18,0	13,3	22,6	12,8
	agost	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
1999	agost	433,0	218,5	135,2	133,0	172,0	80,8	103,5	85,5
	sep	469,5	351,0	202,0	237,5	228,0	196,3	210,8	199,1
2000	abr	57,0	27,0	5,5	6,0	9,0	9,7	8,9	8,5
	may	110,9	71,4	45,0	40,5	57,0	24,4	49,0	23,1
	jun	1373,0	823,3	418,5	549,0	613,0	459,2	395,7	408,0
	sep	525,0	357,5	176,0	257,0	227,0	188,4	235,4	154,6
2001	mar	0,0	2,7	0,0	0,0	0,0	0,0	0,0	0,0
	jun	241,0	154,2	56,0	123,1	42,7	95,9	77,3	77,6
2002	ene	3,0	2,2	0,0	0,0	0,0	0,0	0,9	0,0
	jul	439,6	246,0	170,5	189,0	196,0	119,4	127,2	118,7
2003	ene	71,1	88,5	44,0	20,0		0,0	35,0	32,9
	may	204,0	126,5	89,5	139,0	94,0	112,0	75,3	58,2
	agost	44,0	53,0	42,0	36,0	47,5	34,5	46,6	54,5
	oct	146,0	75,0	29,0	28,5	66,0	17,2	24,2	15,0
	nov	124,0	80,0	48,0	64,0	51,0	53,5	75,2	46,4

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

Estación con falta de datos (Y)	Estaciones utilizadas (X), para la estimación de la precipitación (Y)		
	1ª estación más cercana	2ª estación más cercana	3ª estación más cercana
Armerillo	Colorado	Huapi	San Javier
Colorado	Armerillo	Huapi	San Javier
Huapi	Colorado	El Guindo	Talca
El Guindo	Talca	Huapi	San Javier
Gualleco	Pencahue	Talca	San Javier
Pencahue	Gualleco	Talca	San Javier
San Javier	Talca	Pencahue	Huapi
Talca	San Javier	Pencahue	El Guindo

Tabla 4. Esquema para la aplicación del método de razones de distancias.

Estación con carencia de información	Estación A	Estación B
Armerillo	Colorado	Huapi
Colorado	Armerillo	Gualleco
Gualleco	El Guindo	Colorado
Huapi	El Guindo	Pencahue
El Guindo	Huapi	Gualleco
Pencahue	Huapi	Colorado
San Javier	Talca	Pencahue
Talca	Huapi	Pencahue

estimar a las otras (Tabla 4), para posteriormente calcular las distancias entre las estaciones seleccionadas. Así, para determinar las distancias entre las estaciones, éstas fueron ubicadas en una carta topográfica con una escala 1:500.000; se eligió esta escala, ya que permite visualizar todas las estaciones en la misma carta.

El método de completación por promedios vecinales considera una disposición espacial de tipo triangular, y requiere de tres estaciones para estimar la precipitación en la estación que tiene información faltante. Dada esta condición, sólo fue posible aplicarlo a 5 de las 8 estaciones (Tabla 5), ya que

las estaciones de Armerillo, Colorado y Gualleco, no cumplieron con la disposición espacial necesaria.

En relación al método de completación por razones promedio, éste es complementario al método de promedios vecinales, por lo que se ocupó la misma conformación espacial de las estaciones. Además, fue necesario calcular la precipitación normal para cada estación, que para este caso se hizo a través del cálculo del promedio aritmético anual de los últimos 15 años y no de los 30 como se sugirió anteriormente, y por que sólo se pretende comparar el comportamiento de los distintos métodos de completación.

Tabla 5. Esquema para la aplicación del método de promedios vecinales.

Estación con carencia de información	Estación A	Estación B	Estación C
Talca	Pencahue	San Javier	El Guindo
San Javier	Colorado	Huapi	Talca
Pencahue	Gualleco	Talca	San Javier
El Guindo	Huapi	Talca	Pencahue
Huapi	Armerillo	El Guindo	San Javier

Para la aplicación del método de completación por correlación con estaciones vecinas, se utilizaron las estimaciones hechas con la completación por regresión lineal y sus respectivos coeficientes de correlación (R), para cada una de las estaciones. 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 (Tabla 6 y 7).

Tabla 6. Configuración del método de completación por correlación con estaciones vecinas, con dos estaciones.

Estación con carencia de información.	Estación A	Estación B
Armerillo	Huapi	Colorado
Colorado	Huapi	Armerillo
Gualleco	Talca	Pencahue
Huapi	Colorado	El Guindo
El Guindo	Talca	Huapi
Pencahue	Talca	San Javier
San Javier	Talca	Pencahue
Talca	Pencahue	San Javier

Tabla 7. Configuración del método de completación por correlación con estaciones vecinas, con tres estaciones.

Estación con carencia de información	Estación A	Estación B	Estación C
Armerillo	Huapi	Colorado	Pencahue
Colorado	Huapi	Armerillo	Pencahue
Gualleco	Talca	Pencahue	San Javier
Huapi	Colorado	El Guindo	Talca
El Guindo	Talca	Huapi	Colorado
Pencahue	Talca	San Javier	Gualleco
San Javier	Talca	Pencahue	Gualleco
Talca	Pencahue	San Javier	Gualleco

Después de la aplicación de los métodos de completación, se observó que los valores de los errores eran demasiados altos para la calidad de R^2 que se observaba, por lo que se decidió incluir un sexto método de completación, el de regresiones múltiples. Este método lo que busca es establecer si la cantidad y calidad de las estimaciones hechas con el método de correlación lineal, se ve afectada por la cantidad de estaciones incluidas. A modo de poder realizar comparaciones, se ocupó la misma conformación de estaciones utilizadas para el método de correlación con estaciones vecinas.

Análisis comparativo de la calidad de la estimación

Una vez obtenida la información faltante, se hizo una comparación entre los valores reales y los valores estimados a través de los diversos métodos de completación. Así, se analizaron los resultados generados por cada uno, comparándolos entre sí para detectar diferencias estadísticamente significativas entre ellos.

Para detectar estas diferencias, se aplicaron las siguientes medidas de bondad de ajuste:

Coefficiente de Determinación (R^2): que expresa el porcentaje de la variación total de las precipitaciones reales, que son explicadas por el método de completación.

Error Estándar de Estimación (EEE): 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, que hace el método aplicado (Caro, 2001). Dado que los valores que resultan de esta prueba, son inherentes al tipo de información utilizada, no son directamente comparables con los resultados de otro conjunto de datos (Sánchez, 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. Esta última prueba permite además, apreciar la homogeneidad del error en la misma unidad de medida que los datos (Bland y Altman, 1999; Dunn y Roberts, 1999). En

términos estadísticos, 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, definido por la expresión (7):

$$LC = dp \pm 1.96 * SD \quad (7)$$

Análisis de Varianza (ANOVA): permite definir si existen o no diferencias significativas entre los modelos. Corresponde a un procedimiento utilizado para descomponer la variabilidad total de un experimento en componentes independientes que pueden asignarse a causas distintas, como la variación entre y dentro de los modelos (Peña, 1995; Marabolí, 2000). El diseño experimental aplicado lleva a cabo pruebas de hipótesis de efectos fijos de tratamientos (modelos) cuyas hipótesis son:

Ho: Los efectos de los métodos sobre la media son iguales

Ha: Los efectos de los métodos son distintos, es decir, al menos uno de los métodos difiere significativamente de los demás.

El criterio utilizado para determinar la significancia estadística, establece que si el valor p es menor o igual que el nivel de significación $p \leq 0,05$, entonces existe evidencia significativa en contra de Ho (Cid et al, 1990).

RESULTADOS

En el tabla 8 se entregan los valores del coeficiente de determinación (R^2) y el error estándar de estimación (EEE) para cada estación, según el método aplicado.

Los Errores Estándar obtenidos en la mayoría de las estaciones, presentaron grandes diferencias entre los distintos métodos de completación utilizados para la misma estación. Asimismo, los errores fueron más altos de lo que se esperaba. Se observó que las mayores diferencias se presentan para la estación Armerillo, la cual a pesar de presentar valores de R^2

Tabla 8. Coeficiente de Determinación (R²) y Error Estándar de Estimación (EEE).

Método	Armerillo		Colorado		Huapi		El Guindo		Gualleco		Pencahue		San Javier		Talca	
	R ²	EEE	R ²	EEE	R ²	EEE	R ²	EEE	R ²	EEE	R ²	EEE	R ²	EEE	R ²	EEE
Razones de Distancia	0,68	164,9	0,94	43,9	0,94	31,7	0,95	21,4	0,93	32,1	0,98	13,2	0,96	17,8	0,97	15,4
Promedios Vecinales	-	-	-	-	0,87	46,7	0,95	22,1	-	-	0,98	13,2	0,77	46,1	0,98	13,7
Razones Promedios	-	-	-	-	0,98	20,4	0,96	19,0	-	-	0,93	25,7	0,96	18,5	0,98	12,0
Correlación con Estaciones Vecinas (2 estaciones)	0,0	333,6	0,99	7,2	0,99	5,7	0,98	14,9	0,84	14,9	0,98	2,5	0,99	6,3	0,98	11,9
Correlación Con Estaciones Vecinas (3 estaciones)	0,0	474,1	0,05	173,4	0,95	27,9	0,96	18,3	0,89	40,8	0,99	11,1	0,99	4,3	0,96	18,5
Regresiones Múltiples (2 estaciones)	0,92	62,3	0,95	29,2	0,93	23,9	0,89	22,3	0,98	13,9	0,96	13,3	0,96	14,9	0,97	10,5
Regresiones Múltiples (3 estaciones)	0,93	61,1	0,97	20,9	0,94	22,7	0,91	21,8	0,98	13,7	0,98	9,8	0,96	14,6	0,97	10,4
Regresión Lineal simple (1ª más cercana)	0,92	64,1	0,93	33,9	0,93	23,9	0,89	24,4	0,98	14,5	0,98	11,0	0,96	15,2	0,96	13,3
Regresión Lineal simple (2ª más cercana)	0,89	72,2	0,92	38,3	0,84	37,5	0,83	27,6	0,96	19,5	0,96	13,7	0,94	18,1	0,96	12,7
Regresión Lineal simple (3ª más cercana)	0,87	82,9	0,91	42,1	0,85	36,4	0,89	24,4	0,94	21,9	0,94	16,9	0,93	27,9	0,89	21,9

- Método no aplicado

superiores a 0,8 (salvo en el método por correlaciones con estaciones vecinas con dos y tres estaciones que tuvo valores 0), presentó errores superiores a 60 mm residuales para todos los métodos. Por ello, 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).

Por otro lado, el EEE más bajo se obtuvo en la estación Pencahue con el método de correlación con estaciones vecinas con 2 estaciones que presentó un valor de 2,5 mm, con un R^2 de 0,98. En general en esta estación se presentaron bajos errores y elevados R^2 , los que fueron en todos los casos superiores a 0,93.

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 9).

El resultado de esta prueba entregó que la única estación que manifestó diferencias significativas

entre los distintos métodos utilizados, con un valor p muy bajo (inferior a 0,05), fue Armerillo.

Lo señalado anteriormente muestra que ocurre un caso especial con esta estación, ya que por lo general en las otras estaciones, el método de correlación con estaciones vecinas y regresiones múltiples presentan resultados muy similares, en cambio en la estación de Armerillo, ello no ocurre.

DISCUSIÓN

Para determinar qué método fue el que estimó mejor los valores reales de precipitación media, se aplicó el test de concordancia de Bland y Altman (ACBA), para cada estación en particular; los datos que se ajustan mejor a los valores reales, son aquellos donde las diferencias promedio son menores. En la figura 4, se presentan las gráficas de las diferencias promedio entre los métodos utilizados para cada estación en donde, (**RD**): Razones de Distancia; **PV**: Promedios Vecinales; **RP**: Razones Promedio; **CEV**, **i**: Correlación con estaciones vecinas: i : cantidad de estaciones incluidas; **RM**, **p**: Regresiones Múltiples con p : Cantidad de estaciones Incluidas; **R1**, **j**:

Tabla 9. ANOVA de métodos de completación de datos aplicados por estación.

Estación	vF	SC	GI	CM	FC	VALOR P
Armerillo	Dentro del grupo	9,4059e6	8	1,1757e6	5,09	0,0000*
	Entre grupos	7,2793e7	315	231092,0		
	TOTAL	8,2199e7	323			
Colorado	Dentro del grupo	808978,0	8	101122,0	1,96	0,0512
	Entre grupos	1,6262e7	315	51628,1		
	TOTAL	1,7071e7	323			
Huapi	Dentro del grupo	172904,0	10	17290,4	1,00	0,4388
	Entre grupos	6,6269e6	385	17212,9		
	TOTAL	6,79986e6	395			
El Guindo	Dentro del grupo	66736,9	10	6673,69	0,81	0,6240
	Entre grupos	3,1916e6	385	8289,99		
	TOTAL	3,2583e6	395			
Gualleco	Dentro del grupo	85091,4	8	10636,4	0,61	0,7709
	Entre grupos	5,5088e6	315	17488,3		
	TOTAL	5,5088e6	323			
Pencahue	Dentro del grupo	11691,1	10	1169,11	0,13	0,9993
	Entre grupos	3,3742e6	385	8764,16		
	TOTAL	3,3858	395			
San Javier	Dentro del grupo	77694,3	10	7769,43	0,85	0,5763
	Entre grupos	3,5004e6	385	9092,07		
	TOTAL	3,5781e6	395			
Talca	Dentro del grupo	74354,4	10	7435,44	0,84	0,5880
	Entre grupos	3,3984e6	385	8827,09		
	TOTAL	3,4727e6	395			

* Existe diferencias significativas entre los métodos de completación; por lo menos 1 es distinto de los demás

Regresión Simple: j : grado de cercanía desde la estación con falta de información hasta la estación que está estimando. 1: la más cercana, 2: la segunda más cercana; 3 : la tercera más cercana)

En la estación Armerillo (Figura 4a), el método de regresión múltiple, con 2 y 3 estaciones es el que más se acerca a los valores reales con una diferencia promedio de 0,27 mm y 6,01 mm respectivamente. Luego se puede observar un alza en la diferencia media hasta llegar a los 90 mm con el método de razones de distancia, pasando a superar los 100 mm de diferencia promedio entre datos reales y datos estimados, con los otros métodos. El método de regresión lineal con la estación San Javier, es el que arrojó los resultados más desfavorables, lo que se puede explicar por la ubicación de la estación San Javier, porque si bien ésta es una de las más cercanas, (en distancia lineal), no es representativa de lo que ocurre en la estación Armerillo, ya que se ubican en zonas climáticas distintas, donde Armerillo está en la precordillera andina y San Javier en la depresión intermedia del país.

Para la estación Colorado, el método que mejor ajusta a los valores reales también es el de regresiones múltiples, con 2 y 3 estaciones, y los que más difieren son correlación con estaciones vecinas, con 3 estaciones, y el método de regresión lineal con la 3ª estación más cercana. Gráficamente se refleja que existe una gran diferencia entre el método que estima mejor, que en este caso es el de regresión múltiple con dos estaciones (Huapi y Armerillo), y el que estima peor, el método de correlación con estaciones vecinas, que contempló 3 estaciones (Huapi, Talca y Pencoahue) (Figura 4b).

En el caso de la estación Huapi, el método que obtuvo las menores diferencias promedio, fue el método de correlación con estaciones vecinas con 2 estaciones seguido del método de regresión múltiple con 3 estaciones, (Colorado, Armerillo y Talca). Los métodos restantes se comportan muy parecidos; la diferencia promedio más alta es de 35,49 mm, que se obtuvo con el método de regresión simple con la estación Talca. (Figura 4c).

En la estación el Guindo, el análisis de concordancia de Bland y Altman arroja que las diferencias promedio entre los métodos de completación, son muy parecidas entre sí, es decir, que no hay diferencias importantes entre los métodos de completación utilizados (Figura 4d). El método que obtuvo menor valor de diferencia promedio fue el método de razones de distancia, y el valor más alto se alcanzó con el método de regresión lineal con la estación Colorado, que es la 3ª más cercana.

Para la estación Gualleco se pudo establecer que las menores diferencias promedio entre los métodos de completación, se obtuvieron con el método de regresión múltiple con 2 estaciones

(Talca y Pencoahue), que denotó un valor de 3,04 mm y, en el caso del método de regresión múltiple con 3 estaciones, (Pencoahue, San Javier y Talca), se alcanzó los 4,78 mm. Con los otros métodos, las diferencias promedio fueron más altas variando entre los 11,81 mm y los 28,02 mm (Figura 4e).

En la estación Pencoahue, los métodos de completación no difieren mucho de la precipitación real, ya que el mayor valor de diferencia promedio fue de 12,77 mm, (con el método de regresión lineal simple con la estación Gualleco) y, el más bajo 1,8 mm, con el método de regresión lineal simple con la estación Talca (Figura 4g).

Esto se puede explicar, como en las otras estaciones, por la relación que existe entre la ubicación espacial de la estación con carencia de información y las distancias de la estaciones que sirven para estimar la precipitación de la misma. En la depresión intermedia ocurre que las estaciones están más homogéneamente distribuidas, y las distancias entre ellas son más cortas.

El análisis de concordancia de Bland y Altman en la estación San Javier indica que las diferencias promedio de los métodos de completación son relativamente bajas. Van desde los 0,32 mm (con el método correlación con estaciones vecinas con 3 estaciones, Talca, Pencoahue y Gualleco) y los 16,38 mm, con el método de regresión lineal con la estación Huapi (Figura 4f).

Para la estación Talca este test indica que de los métodos analizados, los que presentan menor diferencia promedio son los métodos de regresión múltiple con 2 estaciones (Pencoahue y San Javier) y regresión múltiple con 3 estaciones (Pencoahue, San Javier y Gualleco) con 1,38 mm y 1,28 mm, respectivamente. El método con el valor más alto es el de correlación con estaciones vecinas (Figura 4h).

En general se puede decir que las diferencias promedio entre los métodos de completación, fueron bajas y ningún método superó los 10 mm, lo que indica que los métodos aplicados, estiman bien a esta estación.

Con los resultados obtenidos después de la aplicación de los métodos de completación de datos faltantes, y para cada una de las estaciones utilizadas en este estudio, en el tabla 8 se puede observar las veces que los distintos métodos, fueron elegidos dentro de los tres mejores estimadores, en orden descendente, según el análisis de concordancia de Bland y Altman.

En general y a partir de los resultados entregados por los diferentes métodos, y principalmente a partir de lo que señala el test de Bland y Altman, los métodos más eficientes son los de regresión lineal, simple y múltiple, a lo cual se debe poner especial atención en el valor del EEE.

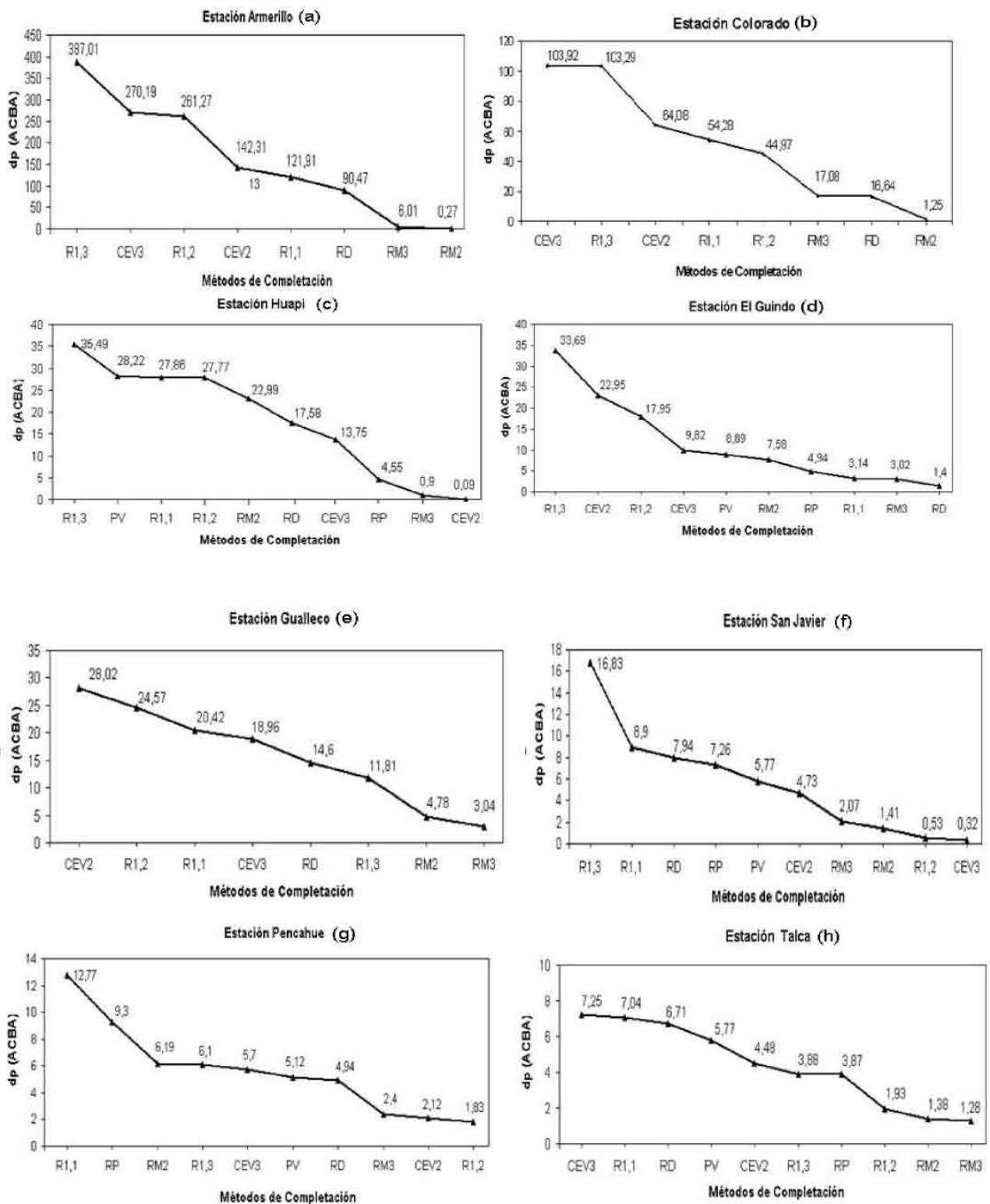


Figura 4. Diferencias promedio del Test de Concordancia de Bland y Altman, para las estaciones en estudio en donde, a) Estación Armerillo; b) Estación Colorado; c) Estación Huapi; d) Estación El Guindo; e) Estación Gualleco; f) Estación San Javier; g) Estación Péncahue y h) Estación Talca.

(RD: Razones de Distancia; PV: Promedios Vecinales; RP: Razones Promedio; CEV, i: Correlación con estaciones vecinas: i: cantidad de estaciones incluidas; RM, p: Regresiones Múltiples con p : Cantidad de estaciones Incluidas; R1, j : Regresión Simple: j : grado de cercanía desde la estación con falta de información hasta la estación que está estimando. 1: la más cercana, 2: la segunda más cercana, 3 : la tercera más cercana).

CONCLUSIONES

Sobre la base de los análisis desarrollados y considerando los objetivos planteados para el estudio, se concluye que para la completación de datos faltantes en la Región del Maule, el método que obtuvo mejores resultados, fue el método de Regresiones Múltiples, con 3 y 2 estaciones respectivamente.

Con respecto al método de regresión lineal simple, éste es uno de los más utilizados en Chile, por su fácil aplicación, dado que no requiere mucha información previa, sólo conocer el monto de las precipitaciones de las estaciones cercanas. Sin embargo, a pesar de ser el más recomendado en la literatura, en la práctica para las estaciones utilizadas en este estudio, no obtuvo los mejores resultados.

El método de razones de distancia también es de fácil aplicación, pero requiere de información adicional, como las distancias lineales, lo que hace preciso contar con una cartografía adecuada e instrumentos de medición, entre otros. Sin embargo, este método tampoco tuvo buenos resultados en el estudio.

El principal problema de aplicación del método de promedios vecinales, es la distribución espacial triangular que deben cumplir las estaciones, donde la estación con carencia de información se sitúa en el centro. Esto dejó estaciones en las cuales no fue posible estimar los datos faltantes. Además sus resultados también fueron deficientes.

El método de razones promedio complementa al anterior; en lo que respecta a su aplicación es más complicado, ya que a la disposición espacial triangular, se le agrega el promedio normal de las precipitaciones en las estaciones.

En relación al método de correlación con estaciones vecinas, éste generalmente se utiliza cuando los coeficientes de determinación R^2 del método de regresión lineal no superan $\pm 0,8$. Pero para este estudio y a modo de comparación, se decidió probar qué sucedía, inclusive obteniendo buenos coeficientes de correlación entre las estaciones. De este modo, los resultados obtenidos fueron muy favorables, ya que en 3 oportunidades se ubicó dentro de los tres mejores estimadores.

En relación al R^2 , este puede ser un buen indicador en una primera aproximación pero no lo suficiente para señalar calidad de ajuste, como es el caso de lo ocurrido con la estación Armerillo, que en la mayoría de los métodos aplicados presentó valores de R^2 superiores a 0,8, pero los EEE superaron los 60 mm.

En cuanto a los análisis realizados a los resultados, se puede señalar que la forma más certera de determinar cuál de los métodos se ajustó mejor a los datos reales, fue el análisis de concordancia de Bland y Altman, ya que éste relaciona directamente el dato real con el dato estimado en distintas dimensiones

ligadas a las desviaciones promedio, los límites y la desviación estándar.

Por otro lado, es recomendable al momento de estimar datos faltantes, contar con estaciones cercanas y confiables, ya que los mejores resultados se obtuvieron con las estaciones más cercanas, que no necesariamente son las más completas.

Al mismo tiempo, sería interesante extender este estudio a zonas áridas y semiáridas y a regiones húmedas, para poder establecer comparaciones, y determinar cuáles métodos se comportan de mejor manera dependiendo de la zona de aplicación.

Finalmente, se recomienda la aplicación del método de regresiones múltiples con 2 y 3 estaciones, para la completación de datos faltantes, y considerar de manera especial la estimación del EEE, porque un nivel bajo de esta medida de bondad de ajuste, asegura que la predicción se situará en marcos técnicos confiables.

AGRADECIMIENTOS

Los autores agradecen a la Dirección General de Aguas de Chile, institución que facilitó la entrega de la información pluviométrica de la Región del Maule.

REFERENCIAS

- Aparicio, F. 2001. Fundamentos de hidrología de superficie. México. Editorial Limusa S. A.. 303 p
- Bland, J, Altman, D. 1999. Statistical methods in medical research. Measuring agreement in methods comparative studies. Vol. 8 (2) pp. 35-160
- Caro, J. 2001. Modelación de caudales recesivos para la cuenca del Río Purapel, Estación Nirivilo. Tesis de Pregrado. Talca. Universidad de Talca. 113p
- 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
- Cid, L., Mora, C., Valenzuela, M. 1990. Estadística matemática. Probabilidades e Inferencia Estadística. Departamento de Matemáticas. Facultad de Ciencias. Concepción: Universidad de Concepción. Chile. 319p
- Chang, T., Shyu, G., Lin, Y., Chang, N. 1999. Geostatistical Analysis of Soil Arsenic Content in Taiwan. Journal of Environmental Science and Health, Part A 34(7) pp. 1485-1501
- 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
- Gujarati, D. 1992. Econometría. Segunda edición. México: Editorial Mc Graw-Hill Latinoamericana. 597p
- Haining, R., Griffith, D., Bennett, R. 1984. A statistical approach to the problem of missing spatial data

- using a first-order markov model. *The Professional Geographer*. Vol. 36 (3), pp. 338–345
- Instituto Geográfico Militar (IGM). 1985. Atlas geográfico de Chile. Talleres gráficos del Instituto Geográfico Militar de Santiago. Chile. 140 p
- Kim, T-W., Ahn, H., Ahn, Jae-H., Byeon, S-H., Wi, S-W., Kim, M. (s.f.) Estimating missing values of daily rainfall using classification techniques. Dept. of Civil and Environmental Engineering, Hanyang University, Ansan, Korea. 8 p
- Linsley, R., MKholer. 1988. Hidrología para Ingenieros. 2ª Edición. Editorial Mc Graw-Hill Latinoamericana. D.F. México. 386 p
- Lynch, S. 1998. Converting Point Estimates of Daily Rainfall onto a Rectangular Grid. Department of Agricultural Engineering, University of Natal, South Africa
- Marabolí, F. 2000. Evaluación de tres modelos precipitación- escorrentía (Budyco, Turc-Pike, Pizarro), en la cuenca del río Achibueno, VII Región, Chile. Tesis de pregrado. Universidad de Talca. Talca-Chile. 79 p
- Mintegui, J., López, F. 1990 La ordenación agrohidrológica en la planificación. Servicio central de publicaciones del gobierno Vasco. pp 306 – 1990
- Montgomery, D., Runger, G. 1994. Applied statistics and Probability for engineers. New York: Editorial Mc Graw-Hill
- Peña, D. 1995. Estadística. Modelos y métodos. Modelos lineales y series temporales. Editorial Alianza. Madrid
- 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
- Sánchez, F. 2001. Análisis y evaluación comparativa de siete modelos precipitación-escorrentía (Budyko, Coutange, Grunsky, Peñuelas, Pizarro, Turc, Turc-Pike), en la cuenca del Río Malleco, IX Región. Tesis de pregrado. Talca: Universidad de Talca. 75p
- Smith, D., Kniveton, D., Barrett, E. 1997. Statistical modeling approach to passive microwave rainfall retrieval. *Journal of Applied Meteorology*. Vol. 37 (2) pp. 135–154
- 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
- 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
- Yu, P., Chen, C. 1996. Extending the rainfall intensity-duration-frequency relationship to non-recording raingauges, *J. Chin. Inst. Eng.* (4) pp. 523-532
- Yu, P., Chen, C. 1997. Potencial of Extending the rainfall intensity-duration-frequency relationship to non-recording raingauges, *Hydrol. Process.*, (11) pp. 377-390
- Yu, P., Yang, T., Lin, C. 2004. Regional rainfall intensity formulas base on scaling property of rainfall, *Journal of Hydrology*, 295 (1-4) pp. 108-123