

ESTIMACIÓN ESTADÍSTICA DE VALORES FALTANTES EN SERIES HISTÓRICAS DE LLUVIA

Rubén Darío Medina-Rivera*; Esther Cecilia Montoya-Restrepo*; Álvaro Jaramillo-Robledo**

RESUMEN

MEDINAR., R. D.; MONTOYAR., E. C.; JARAMILLO R., A. Estimación estadística de valores faltantes en series históricas de lluvia. *Cenicafé* 59 (3): 260-273.2008.

Las observaciones de lluvia que se registran en la red meteorológica de la Federación Nacional de Cafeteros de Colombia (FNC), son parte fundamental para la realización de estudios, investigaciones y actividades de seguimiento, que sirven para fundamentar la toma de decisiones en materia de recursos, costos de producción y política ambiental. En ocasiones, no se cuenta con información completa, debido a situaciones como equipos no calibrados, deterioro en unidades de medición, mediciones inoportunas o deficiencias en el procedimiento mismo de la observación y la medición, entre otros. Con el fin de tener series históricas completas de lluvia, se propuso un método de estimación de datos faltantes, de fácil aplicación, fundamentado y validado estadísticamente, basado en series de datos de precipitación acumulada por ciclos en períodos de tiempo mensuales, agrupados de acuerdo a la condición de El Niño, La Niña y Neutro. Con el método propuesto, el error de estimación relativo no superó el 30%, cuando se tienen hasta cuatro datos faltantes, independientemente de la condición, además, la estimación de datos faltantes, se ajusta a las características propias de cada estación y su estimación depende única y exclusivamente de su historia.

Palabras clave: Estimación datos faltantes, precipitación, regresión acumulada.

ABSTRACT

The observations of rainfall registered in the meteorological network of the National Federation of Coffee Growers of Colombia are a fundamental part to carry out studies, research and follow-up activities that are useful to take decisions regarding resources, production costs and environmental policy. Sometimes there is not complete information due to situations such as non-calibrated instruments, damage in measurement units, not timely measurements or deficiencies in the observation and measurement procedures, among others. In order to have complete rainfall historical series, a method to estimate missing data easy to implement was proposed, researched and statistically validated. That method was based on data series of accumulated rainfall by cycles in monthly periods of time grouped according to the condition of El Niño, La Niña and Neutral. Using this method, the relative estimation error was not higher than 30% when there are up to four missing data regardless the conditions. Besides, the missing data estimate fits the typical characteristics of each station and their estimate depends exclusively on its history.

Keywords: Missing Data Estimation, rainfall, cumulative regression.

* Investigador Científico I e Investigador Científico III, respectivamente. Biometría. Centro Nacional de Investigaciones de Café, Cenicafé. Chinchiná, Caldas, Colombia.

** Investigador Científico III. Agroclimatología. Centro Nacional de Investigaciones de Café, Cenicafé. Chinchiná, Caldas, Colombia.

El Centro Nacional de Investigaciones de Café (Cenicafé), cuenta con información meteorológica registrada continuamente desde 1950, en 232 estaciones climáticas que hacen parte de la red nacional de meteorología. Estas estaciones están clasificadas de acuerdo con la disponibilidad de equipos para registrar datos de precipitación, horas de sol, temperatura, humedad del aire y evaporación, entre otros. Dichas estaciones están localizadas en las laderas de las cordilleras, con un cubrimiento de las regiones cafeteras del país (2).

Las series históricas de clima completas y de períodos largos son un requisito para la elaboración de estudios confiables. En la medida que los datos sean el resultado de un proceso veraz y bien planeado, con registros periódicos que den continuidad a las observaciones, será factible obtener resultados ajustados a la realidad del fenómeno que se estudia. Sin embargo, existen situaciones que pueden afectar la calidad de los datos o registros de las series históricas de precipitación, tales como el cambio de lugar o movimiento de instrumentos, la transformación del espacio físico del entorno donde se encuentra la estación, o no registrar el dato en el momento adecuado, lo que conlleva a la inconsistencia o ausencia de datos.

El servicio meteorológico de la Federación Nacional de Cafeteros de Colombia (FNC), operado por Cenicafé, consciente de la necesidad de contar con información meteorológica de calidad como factor clave para los estudios climáticos en las áreas donde se cultiva el café, y ante la dificultad de no tener siempre información completa, específicamente en las series de lluvia, establece la prioridad de disponer de un método que sea de fácil aplicación, con fundamento estadístico, para completar las series históricas de lluvia.

Rubin (1987), citado por Gómez *et al.* (3), desarrolló una nueva metodología de

propósito general, flexible y fundamentada que denomina “imputación múltiple”, para el tratamiento de la información faltante, cuya técnica es completar un conjunto de datos, reemplazando cada valor faltante por un vector de posibles valores de dimensión m , que representan una distribución de valores probables.

López *et al.* (5), además de describir elementos como temperatura anual, humedad relativa, precipitación y distribución geográfica, caracterizaron la lluvia diaria, para diez estaciones de las cuencas del Río Negro (Argentina) y Santa Lucía (Uruguay). Uno de los propósitos, fue identificar y estimar datos faltantes de lluvia. Para ello, utilizaron una cadena de Markov estacionaria de primer orden, con la cual se determinó si el día correspondiente al dato faltante fue lluvioso o seco, y luego estimaron su valor por el método de la media, el cuál consiste en promediar los registros históricos de la variable de interés del día con el dato faltante. Los autores concluyeron que si bien existen algunos métodos “más elaborados” que mejoran la estimación de datos faltantes con respecto a otros “métodos simples”, la diferencia en las estimaciones no son contundentes, debido quizás a las dificultades inherentes al fenómeno físico considerado y la falta de precisión de las medidas de campo.

Alfaro y Pacheco (1), para estimar datos faltantes en series de precipitación anual para estaciones meteorológicas localizadas en distintas regiones de Costa Rica, aplicaron los métodos de la razón, el del cociente normal, el análisis de regresión simple y el múltiple. Para cada método se compararon los valores estimados con los valores observados, obteniendo diferencias máximas hasta de un 95% del valor real. También, a partir de la prueba no paramétrica de suma de rangos de Wilcoxon, determinaron que la distribución de probabilidad de series de datos estimados

era igual a la serie de datos observados. Concluyeron que el mejor método es el de la regresión múltiple y que las diferencias máximas más altas se dieron con el método de la razón. Estos autores sugieren la utilización de métodos “más complejos” que los presentados en el estudio, si se quieren hacer estimaciones más precisas de precipitación.

En general, de acuerdo con los antecedentes y la revisión bibliográfica, se resume lo siguiente:

- En los métodos para estimar datos faltantes que utilizan estaciones de referencia, la teoría establece que estas estaciones estén lo más cercanas; sin embargo, dicho criterio no aplica para las estaciones situadas en regiones montañosas, como la red meteorológica de la FNC, dado que la cercanía de estaciones no implica igual condición topoclimática.
- La aplicación de estos métodos se hace sobre registros de precipitación simulada y no con series de datos registrados.
- Ningún método tiene en cuenta los eventos de El Niño y La Niña, de gran influencia en la variabilidad climática de Colombia.
- Los métodos referenciados no han sido adoptados para la estimación de datos faltantes en la red climática de la FNC.

Por lo anterior, se realizó esta investigación, con el objetivo de disponer de una herramienta estadística para estimar datos faltantes en las series históricas de lluvia de la red de estaciones climáticas de la FNC, teniendo en cuenta las condiciones topoclimáticas y los eventos de El Niño, La Niña y las condiciones neutras.

MATERIALES Y MÉTODOS

La investigación fue realizada en Cenicafé y se tuvieron en cuenta los registros de lluvia diaria de diez estaciones climáticas de la red meteorológica de la FNC, distribuidas en toda la geografía cafetera, operada por la Disciplina de Agroclimatología de Cenicafé, con más de 25 años de información de lluvia diaria (Tabla 1).

Se identificaron los registros faltantes de lluvia diaria por estación y mes, y se descartaron los meses con al menos un dato faltante, con el fin de obtener series completas de precipitación. En cada estación y para cada mes, se establecieron los episodios de El Niño, de La Niña y condición Neutra (grupos), de acuerdo con los Índices Oceánicos de El Niño-ONI (Tabla 2), establecidos internacionalmente y publicados en la página del National Weather Service, Climate Prediction Center-NOAA (4). Estos valores ONI, indican las desviaciones de la temperatura del agua superficial del océano Pacífico ecuatorial, donde temperaturas iguales o mayores a los $0,5^{\circ}\text{C}$ por encima de un valor medio normal (media del año 1971 al 2000) durante tres meses consecutivos, indica la ocurrencia de El Niño; si por el contrario, la temperatura es de $-0,5^{\circ}\text{C}$ o menos, es un período de La Niña, si no se reúnen estas condiciones se considera un período Neutro (normal).

Esta clasificación en grupos constituye el primero de los cuatro componentes, que hacen parte de la construcción del método. Es de resaltar que los meses de aquellos últimos años en los que se presentó cada condición de El Niño, La Niña y Neutro, no se incluyeron, con el fin de evaluar el método.

Tabla 1. Localización geográfica de las estaciones climáticas analizadas y número de años de observación.

Estación	Municipio / Depto	Latitud Norte		Longitud Oeste		Altitud (m)	Nº años observados de lluvia	Periodo	
		°	'	°	'			Inicio	Final
Pueblo Bello	Valledupar Cesar	10	25	73	34	1.134	49	1958	2006
Bertha	Moniquirá Boyacá	05	53	73	34	1.677	54	1953	2006
Chapetón	Ibagué Tolima	04	28	75	16	1.353	52	1955	2006
Jorge Villamil	Gigante Huila	02	20	75	31	1.420	52	1955	2006
Julio Fernández	Restrepo Valle	03	49	76	32	1.381	53	1954	2006
Manuel Mejía	El Tambo Cauca	02	24	76	44	1.735	54	1953	2006
Santa Helena	Marquetalia Caldas	05	19	75	00	1.395	26	1981	2006
Naranjal	Chinchiná Caldas	04	58	75	39	1.381	51	1956	2006
Paraguaicito	Buenavista Quindío	04	24	75	44	1.203	45	1963	2007
Ospina Pérez	Consacá Cauca	01	15	77	29	1.603	54	1953	2006

El segundo componente del método lo constituyen los “Intervalos de confianza”, para el promedio de la lluvia acumulada, es decir, el acumulado desde el primero hasta el final de cada mes. Estos intervalos, se construyeron para cada estación y grupo con un coeficiente de confianza del 95%. Además, para caracterizar cada una de las estaciones por mes y grupo, se obtuvieron los percentiles de precipitación diaria.

A partir de las distribuciones de frecuencia observadas para la variable cantidad de lluvia diaria a través de los años, se identificó la probabilidad de que llueva en cada uno de los días por estación, grupo y mes. Esta información constituye el tercer componente del método, “Probabilidad”.

En cada estación, para cada grupo y cada mes, se obtuvo el acumulado de cada ciclo involucrado. Con esta información y con el tiempo transcurrido en días, para cada mes, se obtuvo la tasa diaria de precipitación acumulada (Z), a partir de una regresión lineal simple. La estimación de esta tasa es el cuarto componente del método.

Para la evaluación del método, en cada estación y en cada grupo, se tomó aleatoriamente un mes de los que no se incluyeron para su construcción, y de él, a su vez, aleatoriamente se seleccionó un día o tiempo t (caso 1), dos días (caso 2), tres días (caso 3), cuatro días (caso 4) hasta completar diez días (caso 10), los cuales se consideraron, como datos faltantes. Una vez determinados los días con datos faltantes, se aplicó el método. Con el

Tabla 2. Índice Oceánico de El Niño(ONI) National Weather Service, Climate Prediction Center NOAA. (4)

Year	DJF	JFM	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND	NDJ
1950	-1,8	-1,5	-1,4	-1,4	-1,4	-1,2	-0,9	-0,8	-0,8	-0,8	-0,9	-1,0
1951	-1,0	-0,8	-0,6	-0,4	-0,2	0,1	0,4	0,5	0,6	0,7	0,7	0,6
1952	0,3	0,1	0,1	0,1	0,0	-0,2	-0,3	-0,3	-0,1	-0,2	-0,2	-0,1
1953	0,1	0,3	0,4	0,5	0,5	0,4	0,4	0,4	0,4	0,4	0,3	0,3
1954	0,3	0,2	-0,1	-0,5	-0,7	-0,7	-0,8	-1,0	-1,1	-1,1	-1,0	-1,0
1955	-1,0	-0,9	-0,9	-1,0	-1,1	-1,0	-1,0	-1,0	-1,5	-1,8	-2,1	-1,7
1956	-1,2	-0,8	-0,7	-0,6	-0,6	-0,6	-0,7	-0,8	-0,9	-0,9	-0,9	-0,8
1957	-0,5	-0,1	0,2	0,6	0,7	0,8	0,9	0,9	0,8	0,9	1,2	1,5
1958	1,6	1,5	1,1	0,7	0,5	0,5	0,4	0,1	0,0	0,0	0,1	0,3
1959	0,4	0,4	0,3	0,2	0,0	-0,3	-0,4	-0,5	-0,4	-0,4	-0,3	-0,3
1960	-0,3	-0,3	-0,3	-0,2	-0,1	-0,1	0,0	0,0	-0,1	-0,2	-0,3	-0,2
1961	-0,2	-0,2	-0,2	-0,1	0,1	0,1	0,0	-0,3	-0,6	-0,6	-0,5	-0,5
1962	-0,5	-0,5	-0,5	-0,5	-0,4	-0,3	-0,2	-0,3	-0,4	-0,6	-0,7	-0,7
1963	-0,6	-0,3	0,0	0,1	0,1	0,3	0,6	0,8	0,8	0,9	1,0	1,0
1964	0,8	0,4	-0,1	-0,5	-0,7	-0,7	-0,8	-0,9	-1,0	-1,1	-1,1	-1,0
1965	-0,8	-0,5	-0,3	0,0	0,2	0,6	1,0	1,2	1,4	1,5	1,6	1,5
1966	1,2	1,1	0,8	0,5	0,2	0,1	0,1	0,0	-0,2	-0,3	-0,3	-0,4
1967	-0,4	-0,5	-0,6	-0,5	-0,3	0,0	0,0	-0,2	-0,4	-0,5	-0,5	-0,6
1968	-0,7	-0,9	-0,8	-0,8	-0,4	0,0	0,3	0,3	0,2	0,4	0,6	0,9
1969	1,0	1,0	0,9	0,7	0,6	0,4	0,4	0,4	0,6	0,7	0,7	0,6
1970	0,5	0,3	0,2	0,1	-0,1	-0,4	-0,6	-0,8	-0,8	-0,8	-0,9	-1,2
1971	-1,4	-1,4	-1,2	-1,0	-0,8	-0,8	-0,8	-0,8	-0,9	-0,9	-1,0	-0,9
1972	-0,7	-0,3	0,0	0,3	0,5	0,8	1,1	1,3	1,5	1,8	2,0	2,1
1973	1,8	1,2	0,5	-0,1	-0,5	-0,8	-1,1	-1,3	-1,4	-1,7	-1,9	-2,0
1974	-1,8	-1,6	-1,2	-1,1	-0,9	-0,7	-0,5	-0,4	-0,5	-0,7	-0,8	-0,7
1975	-0,6	-0,6	-0,7	-0,8	-1,0	-1,1	-1,3	-1,4	-1,6	-1,6	-1,7	-1,8
1976	-1,6	-1,2	-0,9	-0,7	-0,5	-0,2	0,1	0,3	0,5	0,7	0,8	0,8
1977	0,6	0,5	0,2	0,1	0,2	0,3	0,3	0,4	0,5	0,7	0,8	0,8

Valores en rojo: El Niño; en Azul: La Niña; en Negro: Neutro

acumulado real (dato real), es decir, incluyendo el dato que fue retirado para suponer el dato faltante, se estableció la diferencia absoluta y relativa entre él y el dato estimado, es decir, los errores de estimación absoluto y relativo, respectivamente.

Con los resultados de esta investigación se evaluaron las hipótesis: “el error de estimación no depende del número de datos faltantes en la serie” y “el error relativo promedio de estimación es menor del 30%”.

RESULTADOS Y DISCUSIÓN

Teniendo en cuenta el período de estudio de las series de datos para cada estación (Tabla 1), se ilustra el número de meses o ciclos estudiados en cada grupo, sin contar con los meses de los últimos años en los que se presentó cada condición de El Niño, La Niña y Neutro (Tabla 3). La estación Santa Helena, presentó el menor número de ciclos mensuales para los tres grupos, debido no

Tabla 3. Número de ciclos mensuales por estación y grupo, utilizados para la estimación de datos faltantes de lluvia.

Estación	Grupo		
	El Niño	La Niña	Neutro
Bertha	169	167	293
Chapetón	122	147	221
Jorge Villamil	160	167	277
Julio Fernández	162	176	283
Manuel Mejía	162	176	296
Naranjal	162	155	281
Ospina Pérez	160	167	277
Paraguaicito	150	136	238
Pueblo Bello	148	142	274
Santa Helena	95	60	143

sólo a la cantidad de años analizados en la serie, sino al número de meses descartados por tener al menos un dato faltante. El mayor número de ciclos en los grupos de El Niño, La Niña y Neutro, lo presentan las estaciones Bertha (169), Julio Fernández (176) y Manuel Mejía (296), respectivamente.

Al analizar la precipitación acumulada mensual por grupo, en ocho de las diez estaciones, La Niña es el grupo de mayor precipitación (Figura 1), de tal manera que las estaciones Chapetón, Manuel Mejía, Naranjal y Paraguaicito, presentaron en promedio precipitaciones entre 204 y 256 mm, con errores de estimación entre 7 y 9 mm para este mismo grupo. Las estaciones Bertha y Pueblo Bello presentaron en promedio la misma precipitación acumulada mensual para los tres grupos, según prueba t al 5% (Figura 1). Para las estaciones Naranjal y Santa Helena los promedios de precipitación acumulada difieren en los tres grupos. Las demás estaciones tienen igual promedio para los grupos de El Niño y Neutro. La estación Julio Fernández tuvo los menores promedios de precipitación en cada uno de los grupos con respecto a las estaciones

Bertha, Chapetón, Manuel Mejía, Naranjal, Ospina Pérez, Paraguaicito, Pueblo Bello y Santa Helena. Esta última estación presentó el mayor promedio de precipitación en los grupos El Niño, La Niña y Neutro, con respecto a las demás estaciones.

Al analizar por grupo y mes el promedio de las precipitaciones acumuladas mensuales (Tabla 4), para la estación Bertha, específicamente en el grupo El Niño, se aprecia la menor precipitación acumulada mensual promedio en los meses de enero, febrero y diciembre, mientras que en los meses de abril, mayo y octubre, se observa la mayor precipitación. Similar comportamiento se presentó para el grupo La Niña y Neutro. Así mismo, para las nueve estaciones restantes se construyó el promedio de la precipitación acumulada y su intervalo, con un coeficiente de confianza del 95%.

Al analizar los percentiles de precipitación diaria se encontró que independientemente del grupo, para las estaciones Bertha, Chapetón, Manuel Mejía, Naranjal y Paraguaicito, el 75% de los días de los meses de abril, mayo, octubre y noviembre, presentaron

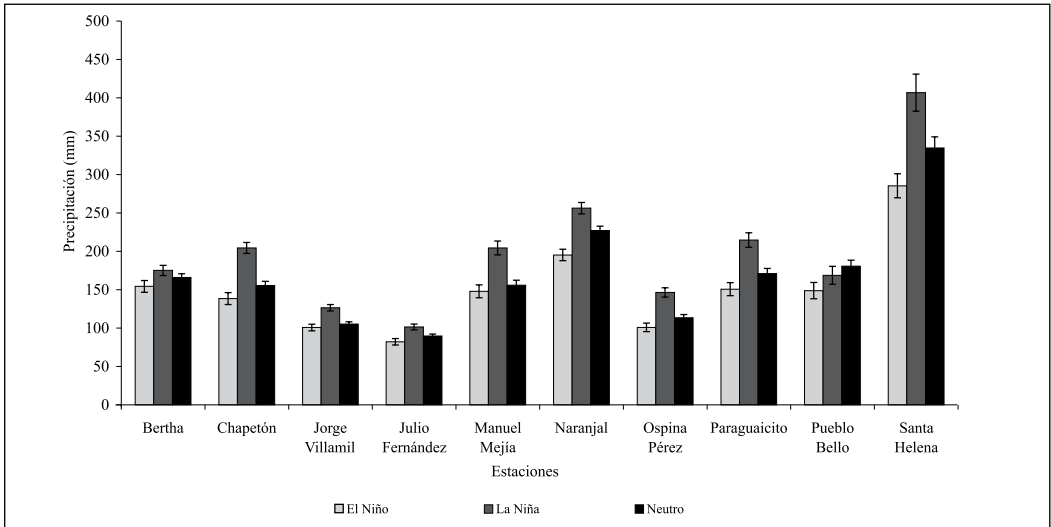


Figura 1. Intervalo para el promedio de precipitación acumulada mensual por grupo, para cada estación objeto de estudio.

precipitaciones menores a los 15,2 mm. Para estos mismos meses, el 25% de las precipitaciones más altas se encuentran entre 15 y 125 mm, lo que explica la variabilidad en la estimación del promedio (Figura 1). La precipitación máxima que se ha registrado en un día, para las estaciones Jorge Villamil y Julio Fernández, no ha superado los 103 mm de lluvia, además, en el 75% de los días para cualquier mes y grupo, no alcanza los 7 mm de lluvia.

Pueblo Bello es una estación que para el mes de junio del grupo El Niño, llegó a registrar un máximo de 139 mm. No obstante, en los meses de mayo, septiembre y octubre, se registran de 13 a 126 mm para el 25% de los días de mayor precipitación. En general, independiente del grupo, en el 50% de los días de cada mes, para todas las estaciones, las precipitaciones no superan los 10 mm.

A partir de las frecuencias observadas para el número de días con precipitaciones mayores a cero, por estación y grupo, se

generaron las Tablas de probabilidad de que llueva en cada uno de los días de los meses del año. Por ejemplo, para la estación Bertha, en el grupo El Niño (Tabla 5), los días 22, 23, 25 y 29 del mes de abril, históricamente siempre ha llovido (valores de probabilidad igual al 100%); contrario a los días 1 del mes de febrero y 27 del mes de diciembre, donde la probabilidad de lluvia (precipitaciones mayores a cero) es sólo del 8,3 y 6,3%, respectivamente.

Dado que Pueblo Bello fue una de las estaciones que presentó mayor número de días sin precipitación, se tomaron 96 días de los meses de aquellos últimos años en los que se presentó cada condición de El Niño, La Niña y Neutro. En estos 96 días, históricamente la probabilidad de que lloviera fue menor o igual al 20%, y en éstos se observó que efectivamente en el 89% de los días no hubo precipitaciones. De allí, se determinó el criterio que, cuando la probabilidad de que llueva en un día dado sea menor o igual al 20%, la estimación del dato faltante es cero.

Tabla 4. Número de meses, intervalo para el promedio mensual acumulado, por grupo y mes. Estación Bertha.

Grupo	Mes	Número de meses	Intervalo		Promedio
			Li	Ls	
El Niño	Ene	15	20,4	55,3	37,9
	Feb	12	45,1	78,6	61,8
	Mar	10	86,3	162,6	124,4
	Abr	10	207,0	316,6	261,8
	May	13	275,3	341,2	308,2
	Jun	13	135,9	221,4	178,6
	Jul	12	86,0	172,3	129,2
	Ago	12	86,2	146,0	116,1
	Sep	15	142,8	201,6	172,2
	Oct	15	198,7	292,1	245,4
	Nov	16	146,6	199,1	172,9
	Dic	16	49,1	84,3	66,7
La Niña	Ene	17	52,4	106,9	79,6
	Feb	14	59,0	132,2	95,6
	Mar	13	112,3	184,5	148,4
	Abr	14	189,4	279,4	234,4
	May	13	220,5	335,2	277,8
	Jun	11	143,0	243,0	193,0
	Jul	11	155,7	204,5	180,1
	Ago	10	160,8	234,5	197,7
	Sep	14	176,1	221,3	198,7
	Oct	17	223,5	291,3	257,4
	Nov	17	130,3	184,8	157,5
	Dic	16	83,2	143,6	113,4
Neutro	Ene	19	35,3	65,9	50,6
	Feb	25	64,8	95,4	80,1
	Mar	29	120,0	171,7	145,8
	Abr	28	222,5	267,7	245,1
	May	26	223,8	281,6	252,7
	Jun	28	161,1	227,5	194,3
	Jul	28	130,7	171,7	151,2
	Ago	30	127,0	172,5	149,7
	Sep	23	171,4	219,6	195,5
	Oct	20	217,7	266,9	242,3
	Nov	18	139,5	194,6	167,1
	Dic	19	54,4	99,1	76,8

Li y Ls : Límites inferior y superior para el promedio con una confiabilidad del 95%

Tabla 5. Probabilidad de que llueva en cada día del mes. Estación Bertha, grupo El Niño.

Día	MES											
	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
1	26,7	8,3	90,0	60,0	76,9	69,2	66,7	41,7	53,3	73,3	81,3	50,0
2	26,7	41,7	60,0	90,0	84,6	69,2	50,0	66,7	53,3	66,7	81,3	56,3
3	13,3	16,7	40,0	80,0	84,6	76,9	66,7	58,3	66,7	66,7	75,0	37,5
4	26,7	41,7	70,0	60,0	76,9	92,3	66,7	66,7	60,0	80,0	75,0	37,5
5	46,7	41,7	50,0	90,0	84,6	76,9	66,7	58,3	73,3	66,7	81,3	62,5
6	53,3	50,0	50,0	70,0	92,3	69,2	91,7	50,0	60,0	73,3	62,5	56,3
7	46,7	58,3	30,0	80,0	84,6	76,9	66,7	58,3	26,7	86,7	68,8	62,5
8	40,0	50,0	40,0	70,0	76,9	69,2	58,3	66,7	66,7	86,7	68,8	50,0
9	40,0	66,7	30,0	70,0	69,2	61,5	75,0	66,7	73,3	86,7	68,8	56,3
10	26,7	75,0	60,0	80,0	92,3	69,2	41,7	58,3	60,0	80,0	81,3	50,0
11	20,0	41,7	50,0	80,0	69,2	69,2	66,7	41,7	46,7	80,0	56,3	50,0
12	26,7	58,3	50,0	60,0	76,9	76,9	58,3	33,3	33,3	80,0	50,0	43,8
13	46,7	41,7	90,0	70,0	69,2	53,8	58,3	50,0	60,0	86,7	62,5	43,8
14	33,3	25,0	50,0	80,0	100,0	61,5	66,7	41,7	60,0	66,7	62,5	50,0
15	40,0	41,7	40,0	90,0	76,9	76,9	33,3	50,0	60,0	73,3	68,8	43,8
16	26,7	33,3	50,0	80,0	84,6	30,8	50,0	58,3	46,7	86,7	75,0	31,3
17	33,3	33,3	40,0	90,0	100,0	61,5	58,3	58,3	66,7	86,7	81,3	43,8
18	26,7	50,0	30,0	80,0	84,6	69,2	41,7	58,3	73,3	53,3	87,5	56,3
19	26,7	33,3	70,0	80,0	92,3	38,5	66,7	75,0	80,0	40,0	75,0	31,3
20	26,7	25,0	60,0	90,0	92,3	61,5	75,0	83,3	93,3	66,7	56,3	50,0
21	26,7	50,0	80,0	90,0	76,9	46,2	50,0	50,0	73,3	86,7	56,3	68,8
22	40,0	41,7	70,0	100,0	92,3	53,8	33,3	66,7	86,7	86,7	75,0	50,0
23	33,3	58,3	50,0	100,0	100,0	38,5	58,3	50,0	93,3	73,3	87,5	43,8
24	20,0	50,0	80,0	90,0	76,9	46,2	58,3	50,0	86,7	73,3	81,3	12,5
25	33,3	25,0	60,0	100,0	76,9	69,2	50,0	50,0	80,0	80,0	93,8	25,0
26	13,3	33,3	80,0	90,0	84,6	53,8	50,0	50,0	80,0	73,3	62,5	25,0
27	40,0	58,3	60,0	80,0	61,5	46,2	41,7	50,0	80,0	73,3	68,8	6,3
28	33,3	50,0	50,0	80,0	69,2	69,2	50,0	75,0	73,3	73,3	50,0	25,0
29	33,3	100,0	80,0	100,0	76,9	53,8	58,3	66,7	80,0	66,7	37,5	25,0
30	26,7		60,0	90,0	76,9	69,2	66,7	75,0	73,3	80,0	50,0	37,5
31	40,0		50,0		84,6		58,3	91,7		53,3		37,5

En todos los casos (número de días con datos faltantes), estación, grupo y mes, el coeficiente de regresión lineal fue diferente estadísticamente de cero, según la prueba *t* al 1%. Las estimaciones de los coeficientes de regresión con sus errores de estimación para cada estación y mes, constituyen la componente “tasa de precipitación acumulada”. Estas estimaciones (tasa diaria de precipitación) a su vez, constituyen la estimación del dato faltante, para el día de un mes y grupo determinado, siempre y cuando la probabilidad de que llueva para ese día específico sea mayor al 20% y el acumulado del mes donde falta el dato, no supere el límite superior para el promedio histórico de precipitación del mes en cuestión. La regresión para estimar la tasa diaria se hizo pasando por el origen, dado que de 36 regresiones por estación, en el 86% de ellos, el intercepto fue igual a cero estadísticamente, según la prueba *t* al 1%.

En general, los pasos del método para estimar datos faltantes de lluvia son los siguientes:

- a. Identificar el grupo al cual pertenece el mes con el día del dato faltante: El Niño, La Niña o Neutro.
- b. Obtener el acumulado de lluvia (*X*) sin tener en cuenta el dato faltante.
- c. Si la probabilidad de que llueva en dicho día es menor o igual a 20%, la estimación del dato faltante es cero.
- d. Si la probabilidad de que llueva es mayor del 20% y *X* es mayor descriptivamente, que el límite superior para el intervalo del acumulado histórico, del mes y grupo, respectivamente, la estimación del dato faltante es cero.
- e. Si la probabilidad de que llueva es mayor del 20% y *X* no es mayor descriptivamente,

que el límite superior para el intervalo del acumulado histórico, del mes y grupo, respectivamente, la estimación del dato faltante es *Z*.

Con los siguientes ejemplos, se ilustra la aplicación del método para estimar datos faltantes de lluvia:

Ejemplo 1. No se tiene la precipitación del día 29 del mes de octubre, del grupo El Niño de la estación Bertha. La precipitación mensual acumulada sin el día 29, es de 230 mm y la probabilidad de que llueva en dicho día es del 66,7% (Tabla 5); dado que esta probabilidad es mayor al 20% y el acumulado mensual es menor al límite superior del acumulado histórico (292,1), según la Tabla 4, entonces, la estimación para dicho día será de 7,8 mm con un error de estimación de 0,037 mm (Tabla 6).

Ejemplo 2. Para la misma estación, se tiene como dato faltante el día 8 del mes de enero clasificado en el grupo El Niño. La estimación de lluvia para este día sería cero, debido a que el acumulado de este mes con el dato faltante es de 113,7 mm, el cual es mayor a 55,3, que es el límite superior del promedio histórico para el mes de enero (Tabla 4).

En la evaluación del método, independientemente del grupo, los promedios de los errores relativos son estadísticamente menores del 30%, cuando se estiman hasta cuatro datos faltantes. Para los demás casos, los promedios son estadísticamente iguales al 30% (Tabla 7). Para el caso de tres datos faltantes, el 95% de los errores de estimación relativos estuvieron por debajo de 25,3%. En el caso de 8 datos faltantes, el 75% de las estimaciones tiene un promedio del error relativo hasta de 29,9%. Así mismo, el 50% de los errores de estimación relativos obtenidos para cada caso (uno, dos hasta

Tabla 6. Tasa diaria para el acumulado de lluvia diaria esperada, por grupo y mes. Estación Bertha.

Grupo	Mes	Tasa diaria	
		Estimación (Z)	Error de estimación
El Niño	Enero	1,3	0,018
	Febrero	2,2	0,080
	Marzo	3,9	0,033
	Abril	8,5	0,073
	Mayo	10,0	0,080
	Junio	6,5	0,090
	Julio	4,0	0,033
	Agosto	3,5	0,029
	Septiembre	5,0	0,134
	Octubre	7,8	0,037
	Noviembre	6,1	0,055
	Diciembre	2,4	0,044
La Niña	Enero	2,7	0,025
	Febrero	3,1	0,042
	Marzo	4,7	0,019
	Abril	7,8	0,040
	Mayo	9,1	0,024
	Junio	6,6	0,053
	Julio	5,8	0,069
	Agosto	6,1	0,095
	Septiembre	6,5	0,035
	Octubre	8,1	0,056
	Noviembre	5,7	0,064
	Diciembre	3,9	0,056
Neutro	Enero	1,5	0,020
	Febrero	2,8	0,012
	Marzo	4,2	0,080
	Abril	7,8	0,041
	Mayo	8,2	0,031
	Junio	7,0	0,099
	Julio	5,2	0,050
	Agosto	4,5	0,038
	Septiembre	6,4	0,028
	Octubre	8,1	0,039
	Noviembre	6,0	0,074
	Diciembre	2,9	0,095

diez datos faltantes), no superaron el 17,5%. Además, los errores de estimación, asociados al promedio de los errores relativos están entre 1,1 y 10,7, para uno y diez datos faltantes, respectivamente.

La distribución de los errores de estimación relativos para diferente número de datos faltantes, muestra que para el grupo El Niño, el 75% de éstos no superaron el 18% (Tabla 8), mientras que para el grupo La Niña, a

partir de tres datos faltantes (caso 3) el error relativo es mayor del 30%. Para el grupo Neutro, con 6 datos faltantes, el 75% de los errores son inferiores al 26% (Tabla 8).

En la Figura 2, se ilustra el comportamiento del error relativo promedio para cada grupo y, en general, en función del número de datos faltantes (casos), donde se observa descriptivamente, que a mayor número de datos faltantes mayor error de estimación.

La primera hipótesis de investigación: **“El error de estimación no depende del número de datos faltantes en la serie”**, no fue corroborada, dado que el coeficiente de regresión fue diferente de cero estadísticamente, para la relación entre el número de datos faltantes y el error de estimación relativo, tanto para cada grupo, como en general (Tabla 9).

En el grupo EL Niño, por cada dato faltante a partir del primero (caso 1), se incurre en un error de estimación de 1,1%; en el grupo La Niña de 5,7% y para el

grupo Neutro 2,4%. Así, el menor error de estimación para varios datos faltantes en un mismo mes, se da en el grupo El Niño y el mayor en el grupo La Niña, de acuerdo con el estadístico de prueba t al 1%.

La segunda hipótesis de investigación: **“El error relativo promedio de estimación es menor del 30%”**, fue corroborada en cuatro de los diez casos, de acuerdo con la Tabla 7, es decir sólo se cumple hasta cuatro datos faltantes, independientemente del grupo. Para el grupo El Niño, se cumple la segunda hipótesis de investigación en todos los casos evaluados del número de datos faltantes; para el grupo la Niña, sólo para dos datos faltantes y para el grupo Neutro, para seis datos faltantes (Tabla 8)

Se puede concluir que el método propuesto para la estimación de datos faltantes de lluvia, tiene en cuenta las características propias de cada estación y grupo y su estimación depende única y exclusivamente de su historia.

Tabla 7. Percentiles 95, 75 y 50, promedio y error estándar para errores de estimación relativos según el número de datos faltantes, independientemente del grupo.

Número de datos faltantes	Percentil			Promedio	Error estándar	
	95	75	50			
1	13,6	8,3	4,4	6,0	*	1,1
2	26,6	11,8	6,0	9,1	*	1,7
3	25,3	19,8	12,8	14,6	*	2,9
4	53,2	20,7	14,4	19,0	*	4,1
5	51,4	20,9	12,9	19,6	**	5,0
6	54,1	26,0	10,8	20,8	**	6,0
7	93,1	28,2	12,4	26,6	**	7,7
8	97,6	29,9	14,1	28,0	**	8,4
9	115,0	38,8	17,5	32,4	**	6,9
10	133,0	31,3	14,4	34,6	**	10,7

* Promedio < 30%, según prueba t al 1%

** Promedio = 30%, según prueba t al 1%

Tabla 8. Percentiles 95, 75 y 50, promedio y error estándar para errores de estimación relativos según el número de datos faltantes por grupo.

Grupo	Número de datos faltantes	Percentil			Promedio	Error estándar
		95	75	50		
El Niño	1	9,2	5,8	2,8	3,4 *	1,0
	2	11,8	6,3	3,7	4,5 *	1,0
	3	21,7	16,8	9,6	10,2 *	2,5
	4	17,9	16,7	11,7	10,4 *	2,1
	5	20,8	16,6	10,4	11,4 *	1,7
	6	43,8	16,0	10,3	13,3 *	3,9
	7	38,8	16,1	13,4	14,0 *	3,2
	8	29,5	15,4	10,3	11,2 *	2,7
	9	37,4	17,8	15,5	16,2 *	2,7
	10	33,2	17,2	10,2	12,6 *	2,8
La Niña	1	30,0	9,9	6,6	8,8 *	2,6
	2	40,0	20,2	11,4	14,6 *	3,9
	3	90,0	23,5	16,0	22,3 **	7,9
	4	120,0	33,2	21,1	30,8 **	10,9
	5	150,0	25,4	21,5	32,8 **	13,3
	6	180,0	33,9	21,1	35,6 **	16,5
	7	210,0	59,2	37,8	50,7 **	20,1
	8	240,0	49,5	28,4	47,9 **	22,5
	9	170,0	72,8	43,2	55,3 **	16,7
	10	300,0	74,7	16,5	61,8 **	29,3
Neutro	1	13,6	12,4	3,4	5,9 *	1,5
	2	23,5	9,0	6,6	8,2 *	2,4
	3	19,9	19,1	12,8	11,4 *	2,3
	4	49,2	18,5	14,0	15,8 *	4,3
	5	51,4	20,6	7,6	14,7 *	5,2
	6	54,1	25,8	5,4	13,4 *	5,4
	7	84,9	14,4	6,2	15,1 **	8,0
	8	97,6	43,0	13,1	24,9 **	9,7
	9	96,5	36,7	19,6	25,8 **	8,9
	10	90,0	66,1	11,6	29,4 **	10,4

* Promedio < 30%, según prueba t al 1%.

** Promedio = 30%, según prueba t al 1%.

Para obtener errores mínimos de estimación con la aplicación del método, se requiere de series de lluvia diaria superiores a 25 años de registro.

El error relativo promedio, estadísticamente no es superior al 30%, cuando se estiman hasta cuatro días de un mismo mes, independientemente del grupo.

El error relativo promedio, estadísticamente es menor del 30%, cuando se estiman hasta 10 datos faltantes para un mismo mes del grupo El Niño, con errores promedio entre 3,4 y 16,2%; hasta dos datos faltantes, cuando el grupo es de La Niña, con errores promedio de 8,8 y 14,6%; y hasta seis datos faltantes cuando el grupo es Neutro, con errores promedio entre 5,9 y 15,8.

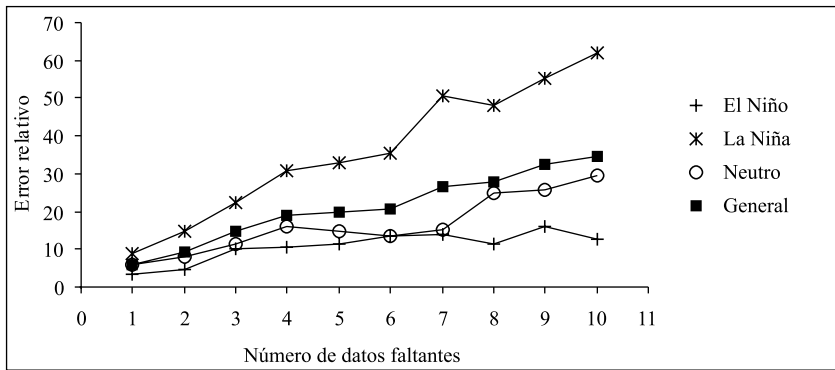


Figura 2. Error relativo promedio en función del número de datos faltantes, por grupo y en general.

Tabla 9. Estimación de los coeficientes de regresión, con sus respectivos errores de estimación (E.E.) y coeficiente de determinación (general).

Grupo	Intercepto		Coeficiente de regresión		R ²
	Estimación	E.E.	Estimación	E.E.	
El Niño	4,6618	1,5925	1,10461 *	0,2567	69,84%
La Niña	4,2933	2,0145	5,77312 *	0,3247	97,53%
Neutro	3,1728	1,9371	2,4145 *	0,3122	88,20%
General	4,0426	0,9726	3,09742 *	0,1568	97,99%

* Coeficiente diferente de cero estadísticamente, según prueba t al 1%

El alcance del método no sólo es útil para la estimación de datos faltantes de lluvia de las estaciones analizadas, sino que es aplicable para cualquier estación meteorológica dada, supeditado a la calidad y cantidad de registros históricos de dicha estación.

LITERATURA CITADA

1. ALFARO R; PACHECO, R. Aplicación de algunos métodos de relleno a series anuales de lluvia de diferentes regiones de Costa Rica. *In: Revista Tópicos Meteorológicos*, Vol. 7, No.1, 2000. p. 42-45. [On line Internet]. Disponible en: <http://www.imn.ac.cr/publicaciones/JUL00.html> (Consultado en abril de 2007)
2. CENTRO NACIONAL DE INVESTIGACIONES DE CAFÉ –Cenicafé. El clima en la zona cafetera. 2007. [On line Internet]. Disponible en: http://www.cenicafe.org/modules.php?name=Estado_del_tiempo_en_la_Zona_Cafetera (Consultado en marzo de 2007).
3. GÓMEZ, G. J.; PALAREA, A. J.; MARTIN, F. J. Métodos de inferencia estadística con datos faltantes, Estudio de simulación sobre los efectos en las estimaciones. *In: Estadística Española* (online) vol 48, no 162, p. 241 – 270, 2006. . [On line Internet] Disponible en: http://www.ine.es/revistas/estaespa/162_2.pdf (Consultado en diciembre de 2006)
4. NOAA, NATIONAL WEATHER SERVICE, Climate Prediction Center. [On line Internet] Disponible en: http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml (Consultado en marzo de 2007)
5. LÓPEZ, C.; BLANCO, J.; BIDEGAIN, M. Algoritmos de control de calidad de datos y de imputación de valores ausentes *In* parámetros meteorológicos. Proyecto CONICYT-BID 51/94 1995-1998. Informe final, 1999. [On line Internet] Disponible en: <http://www.geo.upm.es/postgrado/CarlosLopez/papers/ProyectoCONICYT51-94.pdf> (Consultado en enero de 2007)