

# **Generación de valores de lluvia en matrices porosas. Una aplicación a la determinación de la avenida de diseño de la presa El Realito, Gto.**

Oscar David Santillán Hernández  
Instituto Mexicano de Tecnología del Agua, Jiutepec, Mor. MÉXICO.  
[osantill@tlaloc.imta.mx](mailto:osantill@tlaloc.imta.mx)

## **Resumen:**

Para garantizar que los resultados de los estudios hidrológicos tengan la máxima calidad en los procesos realizados y de los resultados obtenidos, es necesario contar con información climatológica e hidrométrica histórica, de características especiales como son la longitud de registros lo más completa posible. En particular en este trabajo se muestra una metodología para el caso en que falten datos diarios o mensuales de lluvia en estaciones climatológicas que estén ubicadas en cuencas de estudios hidrológicos. En este sentido, para mejorar la calidad de los datos utilizados, es necesario generar registros sintéticos con características estadísticas que permitan confiar en su generación. Se presenta en este trabajo una metodología para llevar a cabo la generación de este tipo de registros para rellenar con ellos los huecos en matrices porosas de datos reales de valores de lluvia diaria y mensual. Los resultados obtenidos se utilizan para llevar a cabo la revisión del estudio hidrológico río Santa María hasta la presa El Realito, en el estado de Guanajuato.

De los registros de la precipitación de las bases de datos ERIC II y CLICOM y luego de que se analizaron los registros en su contexto temporal y espacial de la lluvia registrada en las estaciones climatológicas ubicadas en la zona estudio tanto dentro como en sus alrededores, se estableció que los años de registro disponibles corresponden a una matriz de datos mensuales que puede considerarse como una “matriz porosa” sujeta a generar información para los meses sin dato. Por lo que como objetivo se plantea que se debe utilizar el máximo de registros de lluvia diaria y mensual, por lo que se realizó una depuración y ajuste con base en las longitudes de registro y años completos, además de tratar de encontrar un periodo común en los registros. El periodo de información obtenida fue de 1960-2005. Luego de un primer análisis, se seleccionaron sólo 18 estaciones para el estudio; además se analizó la correlación entre esas mismas estaciones. Una vez con la matriz de datos climatológicos completa, se determinó la lluvia media para cada una de las subcuencas que intervienen en el proceso del estudio hidrológico. La zona donde se encuentra la presa El Realito es una subcuenca de la región Bajo Pánuco, la cual pertenece al río Tampaón. La zona de estudio tiene una superficie total de escurrimiento de 5,038 kilómetros cuadrados. La cuenca de estudio se localiza entre los paralelos 21° 15' y 22° 15' y entre los meridianos 100° 00' y 101° 30'.

## **Localización de la zona de estudio.**

La zona de estudio se encuentra ubicada dentro de la Región Hidrológica Núm. 26, Pánuco, que recibe el nombre por la existencia del río más importante dentro de la región, el Pánuco. Es una de las regiones hidrológicas más importantes del país, tanto por la superficie que ocupa: 96,719 km<sup>2</sup> (que representa, aproximadamente, el 5% del territorio nacional), así como por su volumen de escurrimiento. En la Figura 1 se muestra la Región Hidrológica 26.

El área de estudio ocupa los estados de Guanajuato y San Luis Potosí en el ámbito estatal; en el ámbito municipal, los municipios directamente involucrados en el la zona son San Luis Potosí,

Santa María del Río, Zaragoza, Villa de Reyes, Tierra Nueva, Villa de Arriaga, San Felipe, San Diego de la Unión y San Luis de la Paz, mostrados en la Tabla 1 y Figura 2.

Para llevar a cabo la caracterización climática de la zona de estudio, especialmente de precipitación, se ubicaron las estaciones climatológicas dentro y cerca de la cuenca de estudio a 32 estaciones, mostradas en la Figura 3 y la Tabla 2.

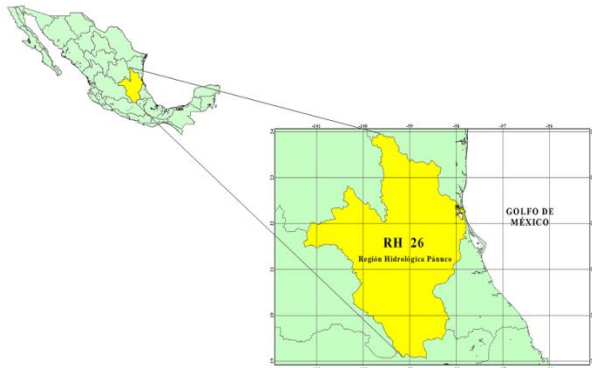


Figura 1 Región Hidrológica Núm. 26.

Tabla 1. Municipios involucrados en la zona de estudio

Estados	Municipios
San Luis Potosí	San Luis Potosí
	Santa María del Río
	Zaragoza
	Villa de Reyes
	Tierra Nueva
	Villa de Arriaga
Querétaro de Arteaga	San Felipe
	San Diego de la Unión
	San Luis de la Paz

### Metodología.

En la generación de información para completar registros, existen diferentes métodos, como el de proporción normal, que emplea las estaciones vecinas y su valor de lluvia media anual; otro es el método del inverso de la distancia, que también utiliza las estaciones vecinas y pondera el valor de la distancia entre ellas. En este trabajo, se muestra un método que considera también las estaciones vecinas, estableciendo la relación entre estas estaciones, a través de una correlación múltiple mediante la metodología de Componentes Principales, para lo cual establece criterios o hipótesis de aceptación y rechazo, tanto de manera paramétrica como visual.

En el desarrollo del trabajo se muestra primero una depuración de la información considerada, seleccionando aquellas estaciones que dispongan de la mejor información en su longitud histórica como en el contenido de esta en el tiempo. Para ello primero se muestra los análisis de selección de estaciones y luego se presenta la metodología de correlación. Finalmente se muestra un ejemplo de aplicación, además de los resultados finales de la aplicación hidrológica.

### Análisis de la información climatológica

Con la información de precipitación con las bases de datos ERIC II y CLICOM, se llevó a cabo el análisis temporal y espacial de la lluvia en la zona. Para el análisis temporal se ubicó a cada estación de manera georeferenciada; es decir, se ubicaron geográficamente cada una de las estaciones climatológicas ubicadas dentro y cerca de la cuenca en estudio (Tabla 2 y Figura 3). Además, se definió una longitud de registro que fuera común en todas las estaciones, o por lo menos lo mejor posible.

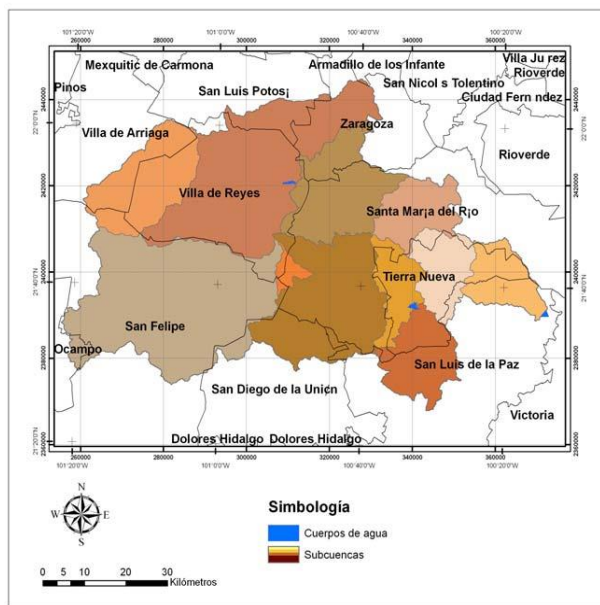


Figura 2. Municipios en la zona de estudio.

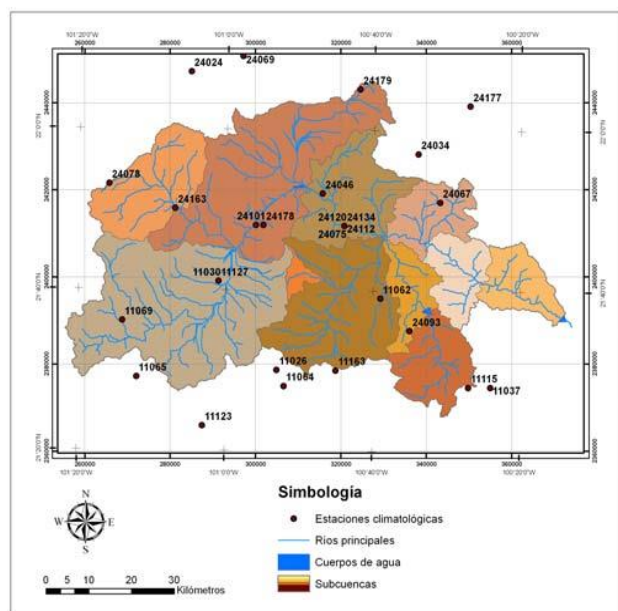


Figura 3. Estaciones climatológicas ubicadas dentro y cerca de la zona de estudio

Tabla 2.- Estaciones climatológicas dentro y fuera de la cuenca y su ubicación geográfica.

Clave	Nombre	Latitud	Longitud
11026	Hacienda San Juan	21.500	-100.8833
11030	Jaral de Berrios S.Fpe	21.6833	-101.0167
11037	Las Mesas, S. Luis Potosí	21.4667	-100.400
11046	Minas de Maguey, Vict.	21.5500	-100.1333
11062	San Aanton de los Mtz.	21.6500	-100.6500
11064	San Diego de la Unión	21.4667	-100.8667
11065	San Felipe, S. Felipe DGE	21.4833	-101.2000
11068	San Luís de la Paz	21.3000	-100.5167
11069	S. Pedro de los Almoloaya	21.600	-101.2333
11083	Xichu, Xichu (DGE)	21.3833	-100.0500
11115	El Refugio Pdte. Mateos	21.4667	-100.4500
11123	Buenavista, San Felipe	21.3833	-101.0500
11127	Jaral de Berrios, S. F.	21.6833	-101.0167
11163	La Jaula, San D. de la U.	21.5000	-100.7500
24024	El Peaje, S.L.P.	22.1167	-101.0833
24034	La Salitrera, Zaragoza	21.9500	-100.5667
24046	Ojo Caliente, V.de Reyes	21.8667	-100.7833
24047	Ojo de Agua, Alaquines	21.8500	-100.1000
24067	San José Albuquerque	21.8500	-100.5167
24069	San Luís Potosí (DGE)	22.1500	-100.9667
24075	Santa Maria del Río (DGE)	21.8000	-100.7333
24078	Santiago, Villa Arriaga	21.8833	-101.2667
24081	Soledad Diez Gutiérrez	22.1833	-100.9333
24093	Tierranueva, Tierranueva	21.5833	-100.5833
24101	Villa de Reyes	21.8000	-100.9333
24112	Santa Maria del Río	21.8000	-100.7333
24120	Santa Maria del Río	21.8000	-100.7333
24134	Santa Ma. del Río ETA.156	21.8000	-100.7333
24163	Bledos, Villa de Reyes	21.8333	-101.1167
24177	Santa Catarina	22.0500	-100.4500
24178	Villa de Reyes	21.8000	-100.9167
24179	Xoconoxtle, Zaragoza	22.0833	-100.7000

El periodo de información obtenida fue de 1960-2005. Algunas de las estaciones de la zona sólo se disponían de información hasta 2001. Se seleccionaron sólo 18 estaciones para el estudio, ya que presentaban mejores longitudes de registro y de mejor periodo común de información. En la Tabla 3, se muestran análisis de las estaciones consideradas y otras más, a las cuales se les determino el porcentaje de datos diarios disponibles. Mientras que en la Figura 4 se muestra en gráfica de barras la longitud de registro de las estaciones climatológicas.

Además de analizar si contenían meses completos; también se analizó a cada una de las estaciones año por año con datos anuales “completos” se aceptaron las que tenían 75% al menos.

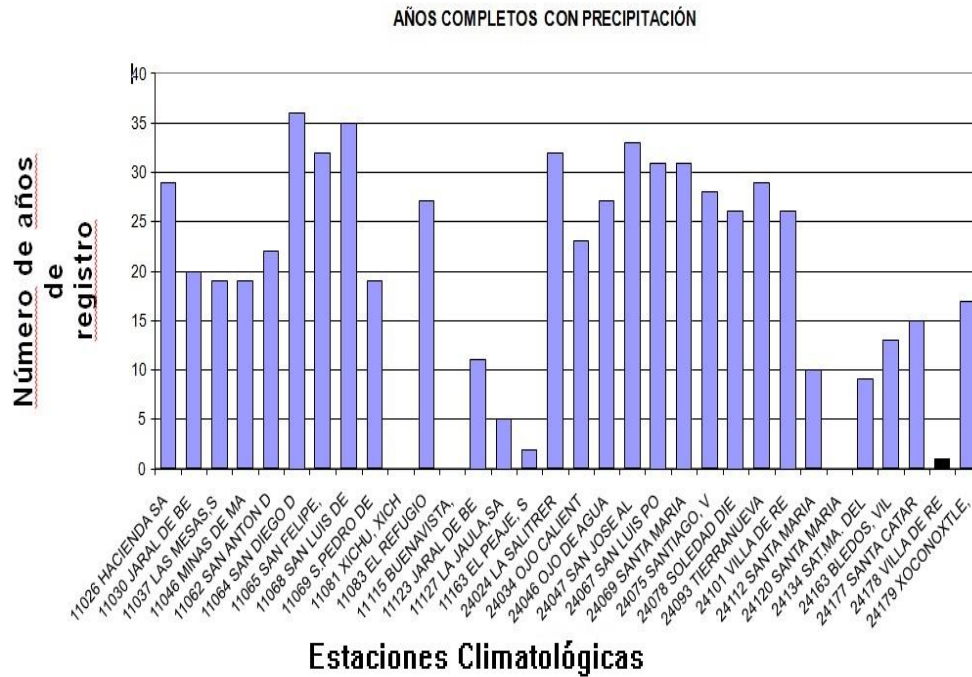


Figura 4. Longitudes de registro de las estaciones climatológicas.

Tabla 3. Estaciones climatológicas consideradas (dentro y alrededor de la cuenca):

ClaveNombre	Años completos	% de años completos
11037 LAS MESAS, S. LUIS DE LA PAZ	19	41.304
11046 MINAS DE MAGUEY, VICTORIA	19	41.304
11062 SAN ANTÓN DE LOS MTZ.	22	47.826
11064 SAN DIEGO DE LA UNIÓN	36	78.261
11065 SAN FELIPE, S. FELIPE DGE	32	69.565
11068 SAN LUIS DE LA PAZ	35	76.087
11069 S. PEDRO DE LOS ALMOLOYA	19	41.304
11083 XICHU, XICHU (DGE)	0	0.000
24034 LA SALITRERA, ZARAGOZA	32	69.565
24046 OJO CALIENTE, V.DE REYES	23	50.000
24047 OJO DE AGUA, ALAQUINES	27	58.696
24067 SAN JOSÉ ALBURQUERQUE	33	71.739
24069 SAN LUIS POTOSÍ (DGE)	31	67.391
24075 SANTA MARÍA DEL RÍO (DGE)	31	67.391
24078 SANTIAGO, VILLA ARRIAGA	28	60.870
24081 SOLEDAD DIEZ GUTIÉRREZ	26	56.522
24093 TIERRANUEVA, TIERRANUEVA	29	63.043
24101 VILLA DE REYES	26	56.522

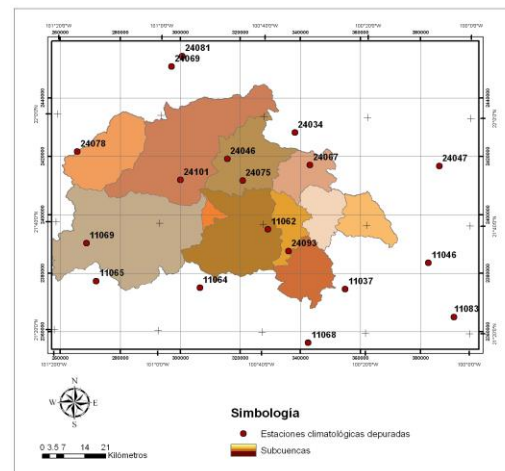


Figura 3. Estaciones seleccionadas en la zona de estudio.

Se desprende con este análisis que se tiene que la matriz de datos mensuales faltantes, la cual

puede considerarse como una “matriz porosa” sujeta a generar información para los meses sin dato

Dentro de los análisis estadísticos que se deben realizar para caracterizar la información climatológica, es la de identificar los grados de correlación espacial y temporal de las estaciones dentro de la zona de estudio. Con este fin se muestran las matrices de correlación múltiple a nivel mensual (Tabla 4). Como observa en matriz de correlación múltiple mensual, existe una buena dependencia estadística entre las estaciones elegidas De este análisis se desprende que existen 14 estaciones prioritarias (Tabla 5) por su ubicación y peso en la estimación de la precipitación media, estas estaciones tienen en promedio 77% del registro completo mensual. Por esta razón se decidió completar estas series como proceso inicial de la inferencia de las series mensuales, iniciando con las estaciones 24069, 24046, 24093, 24075 y 24067.

**Tabla 4. Matriz de correlación múltiple mensual entre estaciones climatológicas.**

Núm Est	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Estac/Estac	11037	11046	11062	11064	11065	11068	11069	11083	24034	24046	24047	24067	24069	24075	24078	24081	24093	24101
11037	1	0.93	0.89	0.79	0.77	0.45	0.79	0.84	0.92	0.75	0.78	0.78	0.76	0.66	0.72	0.42	0.69	0.33
11046	0.93	1	0.85	0.77	0.74	0.5	0.77	0.88	0.92	0.69	0.67	0.85	0.79	0.63	0.7	0.45	0.69	0.35
11062	0.89	0.85	1	0.79	0.82	0.46	0.84	0.81	0.87	0.77	0.76	0.85	0.72	0.72	0.76	0.36	0.8	0.36
11064	0.79	0.77	0.79	1	0.91	0.66	0.93	0.82	0.74	0.88	0.71	0.73	0.83	0.87	0.87	0.47	0.79	0.33
11065	0.77	0.74	0.82	0.91	1	0.48	0.96	0.77	0.71	0.85	0.78	0.71	0.72	0.77	0.89	0.45	0.78	0.3
11068	0.45	0.5	0.46	0.66	0.48	1	0.53	0.53	0.43	0.54	0.34	0.47	0.61	0.63	0.45	0.42	0.55	0.44
11069	0.79	0.77	0.84	0.93	0.96	0.53	1	0.79	0.77	0.87	0.8	0.78	0.76	0.82	0.84	0.41	0.82	0.31
11083	0.84	0.88	0.81	0.82	0.77	0.53	0.79	1	0.79	0.71	0.66	0.77	0.82	0.69	0.77	0.43	0.68	0.31
24034	0.92	0.92	0.87	0.74	0.71	0.43	0.77	0.79	1	0.75	0.73	0.88	0.8	0.64	0.65	0.37	0.74	0.4
24046	0.75	0.69	0.77	0.88	0.85	0.54	0.87	0.71	0.75	1	0.68	0.72	0.86	0.86	0.74	0.4	0.82	0.4
24047	0.78	0.67	0.76	0.71	0.78	0.34	0.8	0.66	0.73	0.68	1	0.58	0.6	0.56	0.64	0.44	0.63	0.34
24067	0.78	0.85	0.85	0.73	0.71	0.47	0.78	0.77	0.88	0.72	0.58	1	0.76	0.72	0.64	0.33	0.84	0.46
24069	0.76	0.79	0.72	0.83	0.72	0.61	0.76	0.82	0.8	0.86	0.6	0.76	1	0.81	0.74	0.51	0.77	0.5
24075	0.66	0.63	0.72	0.87	0.77	0.63	0.82	0.69	0.64	0.86	0.56	0.72	0.81	1	0.71	0.38	0.8	0.36
24078	0.72	0.7	0.76	0.87	0.89	0.45	0.84	0.77	0.65	0.74	0.64	0.64	0.74	0.71	1	0.5	0.67	0.21
24081	0.42	0.45	0.36	0.47	0.45	0.42	0.41	0.43	0.37	0.4	0.44	0.33	0.51	0.38	0.5	1	0.32	0.27
24093	0.69	0.69	0.8	0.79	0.78	0.55	0.82	0.68	0.74	0.82	0.63	0.84	0.77	0.8	0.67	0.32	1	0.41
24101	0.33	0.35	0.36	0.33	0.3	0.44	0.31	0.31	0.4	0.4	0.34	0.46	0.5	0.36	0.21	0.27	0.41	1

En la proceso de la inferencia de las series se establecieron una serie de criterios para determinar cuáles serían las estaciones auxiliares.

Las estaciones auxiliares, son aquellas que resultan complementarias en los periodos en donde no se cuenta con registros en la estación analizada y que además exista una dependencia estadística y espacial.

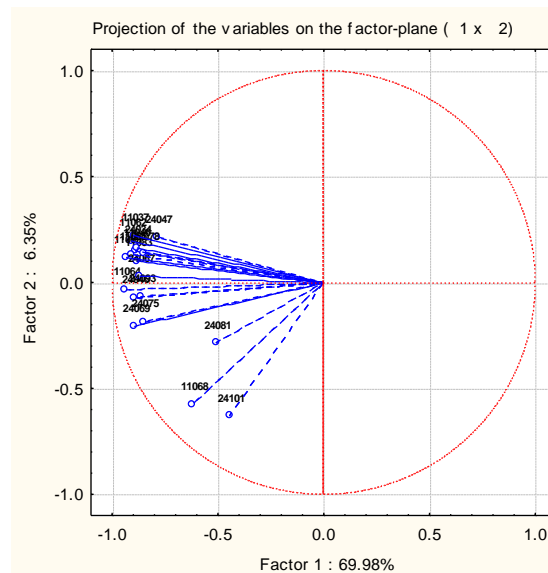
En el proceso de selección de las estaciones se utilizó el análisis multivariado de datos con la finalidad de discriminar aquellas estaciones que presenten un comportamiento estadístico diferente; es decir, identificar estaciones atípicas dentro de la zona de estudio.

En el proceso de estimación de la ecuación múltiple es necesario realizar un análisis de componentes principales con la finalidad de garantizar la existencia de una dependencia estadística.

**Tabla 5. Estaciones con mayor correlación y mayor registro.**

Estación	% de años completos	Estación	% de años completos
11037	61.0	24034	75.0
11046	53.7	24046	78.7
11062	63.8	24067	90.4
11064	88.2	24069	94.1
11065	82.3	24075	73.2
11068	86.6	24078	83.7
11069	57.3	24093	88.6

La Figura 4, muestra las estaciones atípicas dentro de la zona de estudio (24081, 11068 y 24101), con un 76% de significancia para las componentes principales 1 y 2.



**Figura 4. Análisis de componentes principales de las 18 estaciones seleccionadas.**

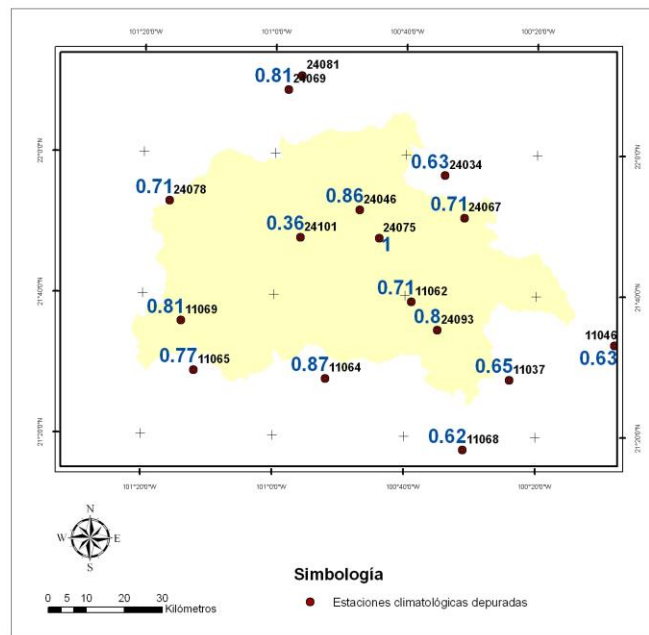
### **Criterios y proceso de la inferencia de series.**

1. Depurar el número de estaciones que cuentan con el número de años de periodo de registro seleccionado.
2. Identificar las estaciones auxiliares para la inferencia de datos.

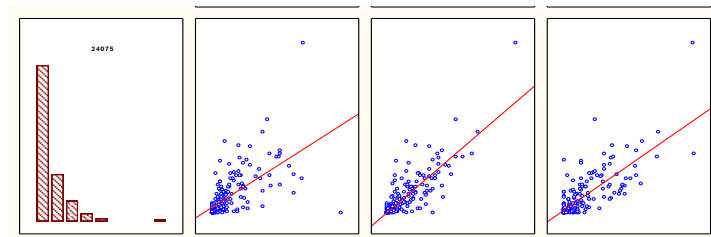
3. Analizar la matriz de correlación múltiple, para establecer que estaciones auxiliares se pueden utilizar en la inferencia de las series mensuales.
4. Con la ayuda del sistema de información geográfica, definir cuales estaciones auxiliares seleccionadas presentan una adecuada cercanía con la estación en estudio.
5. Con un análisis de regresión múltiple lineal, definir la o las ecuaciones que se utilizaran en la inferencia de la serie.
6. Evaluar la ecuación, utilizando el coeficiente de regresión múltiple.

**Ejemplo de aplicación de la metodología.**

La estación 24075 cuenta con el 73% de la información y a su correlación con las estaciones 24069, 24046 y 24093 es de 0.81, 0.86 y 0.8 (Figura 5 y 6); además de que los periodos de datos faltantes resultan ser complementarias con la estación de interés, adicionalmente se revisó que exista una cercanía geográfica.



**Figura 5. Correlación entre la estación 24075 y las estaciones 24069, 24046 y 24093.**



**Figura 6. Correlación entre la estación 24075 y las estaciones 24069, 24046 y 24093.**



Teniendo las estaciones auxiliares identificadas, se procede a la determinación de la ecuación lineal múltiple, para inferir la precipitación de la estación 24075.

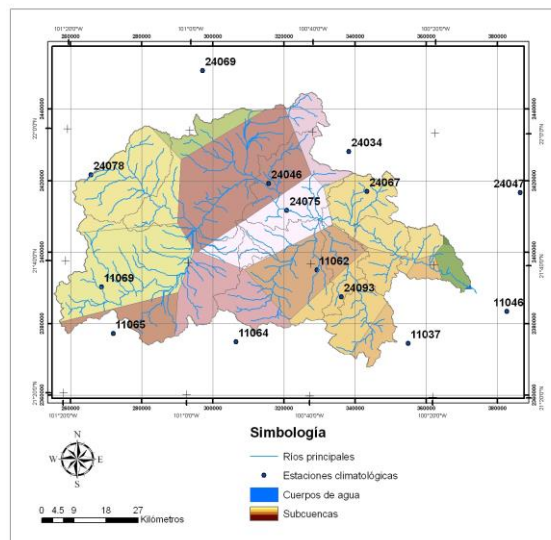
$$Hp_{24075} = Hp_{11046} * 0.011216 + Hp_{11069} * 0.482577 * + Hp_{24093} * 0.348124 + 1.4233$$

**Resultados de la aplicación de la regresión múltiple de las 13 estaciones analizadas.**

**Tabla 5. Estaciones auxiliares y coeficientes de la regresión múltiple.**

Estación	estación	coef.	Estación	hef.	Estación	coef.	Termino	coeficiente de
24069	11062	0.538					11.7	0.53
24069	11046	0.421					13.3	0.50
24046	11069	0.068					4.1	0.39
24046	11069	0.696	24075	0.274			8.3	0.51
24093	11046	0.185	11069	0.482			3.3	0.69
24075	11046	0.011	11069	0.483	24093	0.348	1.4	0.72
24067	11046	0.144	11069	0.384	24093	0.515	1.2	0.75
11062	24067	0.469	24075	0.051	24093	0.576	2.9	0.78
24034	11046	0.488	24067	0.461	24075	0.098	3.9	0.82
11037	11062	0.847	24034	0.438	24093	0.135	7.5	0.77
24078	24046	0.310	24069	0.593	24075	0.139	2.6	0.61
11069	24069	0.294	24075	0.388	24078	0.352	4.3	0.71
11064	11062	0.087	11069	0.383	24046	0.517	2.0	0.80
11065	11064	0.419	11069	0.278	24078	0.286	3.6	0.78
11046	11062	0.149	24034	0.763	24067	0.025	4.4	0.64
11047	11037	0.658	24034	0.146	24069	0.760	21.8	0.57

En la obtención de los polígonos de Thiessen, únicamente se consideraron las 13 estaciones de la Tabla 5. En la Figura 7 se muestran gráficamente los polígonos de Thiessen.



**Figura 7. Polígonos de Thiessen.**



**Tabla 6. Precipitación media para el periodo 1961-2001.**

ESTACIÓN	PRECIPITACIÓN ANUAL	PRECIPITACIÓN ANUAL
	SERIE CON MESES FALTANTES	SERIE COMPLETA
11037	702.4	735.1
11046	551.5	575.7
11062	452.9	461.3
11064	429.4	416.3
11065	471.5	457.7
11069	454.3	371.4
24034	624.1	577.7
24046	357.2	374.1
24067	459.2	453.6
24069	376.1	450.2
24075	377.0	363.5
24078	394.1	422.5

**Tabla 7. Distribución mensual de la precipitación.**

Precipitación mensual	
enero	13.7
febrero	8.5
marzo	10.8
abril	22.8
mayo	41.8
junio	70.6
julio	69
agosto	61.1
septiembre	67.6
octubre	29.4
noviembre	11
diciembre	12.5
<b>Total</b>	<b>426.8</b>

Finalmente, luego de haber generado los datos faltantes a través de las ecuaciones recursivas obtenidas con el análisis Multivariado, se muestran los valores de la precipitación media con series de datos con información faltante y con la serie de datos completada con la metodología. Así mismo, en la Tabla 7 se presentan la distribución de datos de lluvia media, deducida con la información ya completaría.

## Conclusiones

Hair et al. (1999), define al análisis multivariado como cualquier análisis de dos o más variables que midan, expliquen y pronostiquen el comportamiento de las relaciones existente entre ellas. En muchas ocasiones las técnicas multivariadas son un medio de representar en un análisis simple aquello que requirió varios análisis utilizando técnicas univariadas. En el Análisis de Componentes Principales (ACP) se usa un procedimiento matemático que transforma un conjunto de variables correlacionadas en un nuevo conjunto de variables no correlacionadas conocidas como "Componentes Principales". La principal aplicación del ACP es la de generar subregiones de datos de muestreo con la característica de estar fuertemente correlacionadas.

Por otro lado, al existir plataformas computacionales y software especializado, estas técnicas permiten una aplicación que puede considerar el mayor número de estaciones y obtener también resultados de una manera relativamente fácil.

Se considera que con esta metodología se da una mayor certidumbre a los resultados de la aplicación de los métodos hidrológicos e hidrometeorológicos.

## Referencias bibliográficas.

Hair, J.F., Anderson R.E., Tatham, R.L., Blanck W.C. (1999). "Análisis Multivariante". Prentice Hall. p 4.