

Design of optimal rainfall networks by means of principal components analysis

Susana Herman de Bautista and Arnaldo Donoso M.

*Centro de Investigaciones del Agua y Escuela de Ingeniería Civil
Facultad de Ingeniería, Universidad del Zulia, Apdo. 526. Maracaibo, Venezuela*

Abstract

This paper presents a procedure for the design of optimal raingage networks to obtain the mean areal precipitation of monthly rain over a fixed area. It is based on the multivariate technique of Principal Components considering both density and location of the observations. In proportion to the increase of stations, the variance of the mean area precipitation estimated from the network observations toward the real variance. The variance is calculated as the first eigenvalue of the correlation matrix. The accuracy of the network is defined in terms of its capability to represent the real spatial variability of the events.

Key words: Rainfall, rainfall networks.

Optimización de redes pluviométricas mediante el uso del análisis en componentes principales

Resumen

El presente trabajo presenta un método para diseñar redes para medir la precipitación media mensual en un área, basada en el Análisis en Componentes Principales.

A medida que aumenta el número de estaciones en un área, la varianza de la precipitación media mensual calculada por medio de una red de estaciones converge a la varianza de la precipitación media real ocurrida en el área.

El método que se propone se basa en el análisis del mayor valor propio (λ_1) de la matriz de correlación de una red hipotética, que constituye la varianza del Primer Componente Principal que según estudios realizados puede considerarse como la Precipitación Media Mensual, para el área considerada.

Se diseñó una cuenca de 25 Km. x 40 Km. con alturas de 0 a 100 mts. con cuadrículas de 4 x 5 Km. con 20 estaciones distribuidas homogéneamente. Se calcularon las configuraciones óptimas de redes variando de 1 a 20 estaciones.

La precisión de la red se define en términos de su capacidad de representar la variabilidad espacial real de los eventos.

Palabras claves: Precipitación media, redes pluviométricas.

Introducción

La precipitación media en un área, caída durante un cierto intervalo de tiempo es un parámetro de primera importancia en Hidrolo-

gía. Las mediciones de este fenómeno se llevan a cabo en puntos fijos del espacio (estaciones pluviométricas). La estimación de la precipitación media en el área a partir de medidas puntuales implica un cierto grado de incertidumbre.

que dependerá del número de medidas de que se disponga así como de la ubicación espacial de las estaciones de medición.

El diseño de una red pluviométrica consiste en la determinación del número y ubicación de estaciones de medición sobre una región con la finalidad de obtener un registro histórico de datos que puedan caracterizar el fenómeno de precipitación en el espacio y en el tiempo.

El problema del diseño de redes pluviométricas ha sido dividido en varios niveles [1]. Los niveles 1 y 2 pueden ser clasificados como problemas de estimación regional [2]. El caso del cálculo de la precipitación total media en el área pertenece a estos niveles.

Algunos métodos de diseño de redes pluviométricas propuestos consideran como criterio el costo, la precisión o ambos. Rodríguez y Mejía [3] usaron el error cuadrático medio como criterio de decisión para definir la densidad de la red bajo condiciones de muestras aleatorias, luego Bras Rodríguez l. [2] introducen un modelo con costo y precisión, midiendo este término con el error cuadrático medio. Duband [4] presenta un método para racionalizar una red de medidas basado en Componentes Principales analizando el coeficiente de correlación múltiple entre las estaciones y los componentes principales significativos.

En este trabajo se propone un método para optimizar la distribución de estaciones de un red pluviométrica basado en la estructura de correlación de los datos obtenidos a partir de una red inicial.

Metodología propuesta

Dado un conjunto de eventos de precipitación sobre un área, ocurridos durante un periodo largo de tiempo, se dispone de una red de estaciones distribuidas sobre el área que miden características locales del fenómeno a lo largo del periodo de tiempo. El fenómeno global ocurrido, la precipitación total, es variable en el espacio [2], esto es, existe una variabilidad espacial propia del fenómeno, parte de la cual se reflejara en las mediciones de la red. En la medida en que la red se haga más densa, la variabilidad total

implícita en las mediciones tenderá a la variabilidad propia del fenómeno.

El análisis en componentes principales (C.P.) [5,4] proporciona una herramienta adecuada para cuantificar la variabilidad implícita en los datos registrados por una red mediante el estudio de su matriz de correlación. Esta medida consiste en el cálculo de la variación total y varianza del primer componente principal que representa la precipitación media en el área [5,6,7]. Por este procedimiento es posible obtener una buena estimación de la variabilidad propia de la precipitación si se considera una red de estaciones suficientemente densa. Una vez determinada la variabilidad espacial del fenómeno puede estudiarse la calidad de las estimaciones de la media para una red particular, comparando la variabilidad implícita en la estructura de correlación de los datos de la red con la variabilidad tomada como propia del fenómeno.

El primer componente principal, está dado por:

$$Z_{i,1} = \sum_{j=1}^p a_{j,1} x_{i,j} \quad (1)$$

donde $Z_{i,1}$ es la i -ésima ocurrencia del primer componente principal, $a_{j,1}$ es la j -ésima componente del autovector asociado al mayor autovalor de la matriz de correlación, $x_{i,j}$ es la i -ésima observación de precipitación en la j -ésima estación y p es el número de estaciones en la red. En estudios realizados con datos de precipitación en la cuenca del Lago de Maracaibo [5], en la región norte de Italia [7] y en la región norte de Francia y oeste de Bélgica [6] se observó que las componentes del autovector para el primer componente principal Z_1 diferían muy poco en magnitud, concluyendo que el primer componente principal es un índice de la precipitación media en el área. La varianza del primer componente principal es igual al mayor autovalor de la matriz de correlación [4,8], mientras que la variación total, que es una medida de la cantidad de información contenida en las mediciones hechas al fenómeno de precipitación, se define como la suma de los elementos de la diagonal de la matriz de correlación.

ción, suma que será igual al número de estaciones consideradas, p .

$$\sum_{i=1}^p \lambda_i = P \quad (2)$$

Es importante señalar dos aspectos relativos al significado de la fracción λ_1/p . En primer lugar, la variación total de las observaciones realizadas por una red de p estaciones estará distribuida jerárquicamente en el total de p componentes principales, según la magnitud de los autovalores correspondientes. Si bien la suma de los p autovalores de la matriz de correlación es igual a la variación total, la distribución de las magnitudes de los autovalores estará determinada por el grado de correlación de las observaciones en las p estaciones. Para un número fijo p de estaciones, λ_1/p será una medida del grado de correlación de las mediciones cuando se prueban distintas configuraciones de la red. El segundo aspecto es la variabilidad espacial de la precipitación que puede estimarse a partir de medidas locales en la región donde ocurre. La variación total p sólo depende del número de estaciones sin tomar en cuenta la distribución en el área. El efecto de la distribución espacial de las p estaciones es tomado en cuenta por λ_1/p .

La variabilidad espacial propia del fenómeno de precipitación ocurrido en la región en estudio determina la correlación que existirá entre las mediciones locales, reflejada por λ_1/p . En el diseño de una red se busca un equilibrio entre la información suministrada por el número de estaciones extendidas en el área, y que el grado de correlación entre ellas no indique una redundancia de información. Ambas tendencias se miden con la fracción λ_1/p .

La variabilidad propia del fenómeno de precipitación puede estimarse mediante un gráfico λ_1/p vs. p dado que λ_1/p tiende asintóticamente al valor de la varianza de la media real cuando la densidad de estaciones sobre el área crece. Esta medida puede ser utilizada para evaluar la calidad de las estimaciones de la media para una red cualquiera comparando el porcentaje de la

variación total explicado por el primer componente principal con la variabilidad real propia de la precipitación ocurrida, estimada como el valor a largo plazo de la fracción λ_1/p , el grado de incertidumbre en las estimaciones de la media por la red será la diferencia entre ambos valores.

Para el diseño de una red pluviométrica puede utilizarse la medida del grado de incertidumbre como criterio para la escogencia de la mejor alternativa en la medida en que se cambia la configuración de la red en cuanto a número y ubicación de estaciones. Se escoge un número inicial de estaciones suficientemente grande, según una cierta distribución espacial y luego se van eliminando estaciones, evaluando en cada caso el grado de incertidumbre y la combinación de estaciones que lo minimice.

Para ilustrar el método se realizaron simulaciones en una cuenca hipotética simple de forma rectangular de 40x25 Km. con un área de 1000 Km². La topografía de la cuenca es regular con cotas proporcionales a la distancia a uno de los vértices, tomado como origen de coordenadas. Las cotas varían desde cero en el origen hasta 100 mts. en el vértice opuesto (Fig.1). Rodríguez y Mejía [9] presentan una función de correlación en función de la distancia entre dos puntos y el tiempo. Se han identificado dos estructuras de correlación espacial para procesos superficiales: la función de decaimiento exponencial y la función de Bessel modificada de segundo orden [3]. Bras y Rodríguez I. [2] utilizan la función de Bessel e introducen el efecto de la diferencia de alturas a través de la matriz de errores de la varianza.

Para obtener la matriz de correlación de cualquier red de estaciones sobre esta cuenca se utilizó una función de correlación simplificada donde el coeficiente de correlación entre los datos de dos estaciones es función solamente de la distancia entre ellas y de su diferencia de cotas.

La función de correlación utilizada fue:

$$r_{i,j} = 0.908297 - 0.150324 \log(d_{i,j}) + 0.023743 \log(h_{i,j}) \quad (3)$$

donde $r_{i,j}$ es el coeficiente de correlación entre los datos de las estaciones i y j , $d_{i,j}$ es la distancia

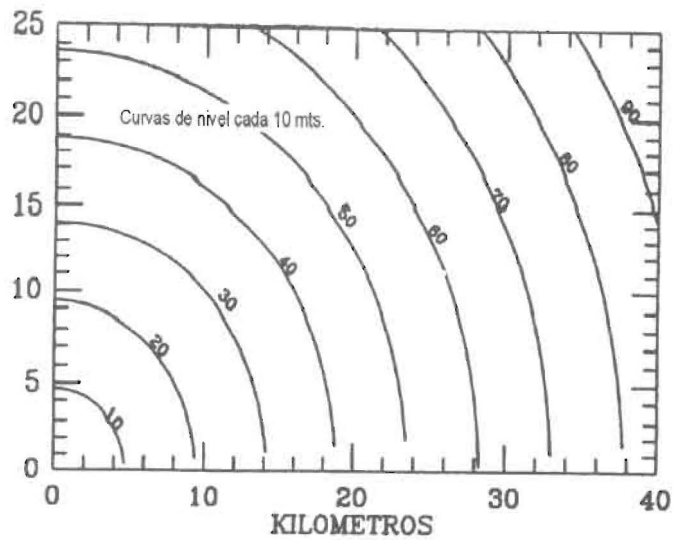


Figura 1. Características fisiográficas de la cuenca del ejemplo.

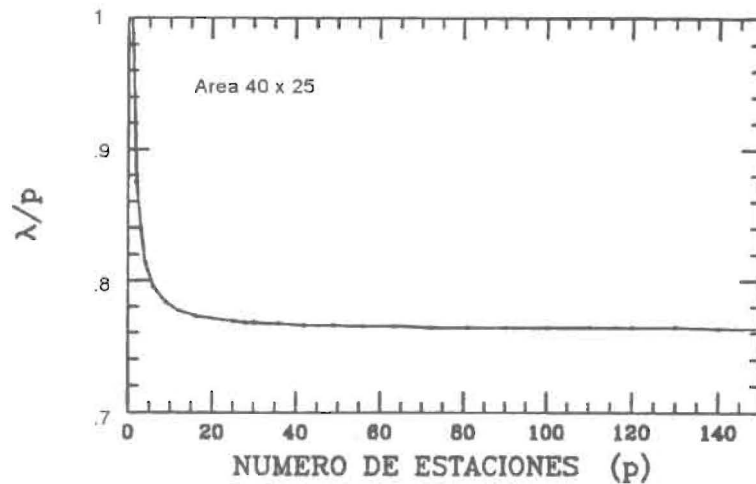


Figura 2. λ_1/p vs p para distribución homogénea de estaciones.

en kilómetros entre las estaciones y h_{ij} es la diferencia de cotas en metros. Esta función se obtuvo por regresión múltiple a partir de datos de precipitación mensual de una subcuenca del Lago de Maracaibo, con distancias máximas de 73 Km. y diferencia de altura máxima de 100 mts.

En primer lugar se determinó la variabilidad propia de la precipitación para lo cual se distribuyeron homogéneamente estaciones sobre la cuenca en una retícula de $M \times N$, variando M y N hasta un máximo de 150 estaciones sobre

el área. Para cada valor de $M \times N$ se construyó la matriz de correlación y se calculó el mayor autovector de la matriz por el algoritmo QL, previa reducción de la matriz a tridiagonal [10]. El comportamiento típico de la curva λ_1/p .vs. p es decreciente al aumentar p , tendiendo asintóticamente a un valor constante el cual se toma como la variabilidad propia de la precipitación ocurrida en el área (Fig. 2).

El programa elaborado para realizar las simulaciones lee la matriz de correlación correspondiente al total de estaciones a considerar

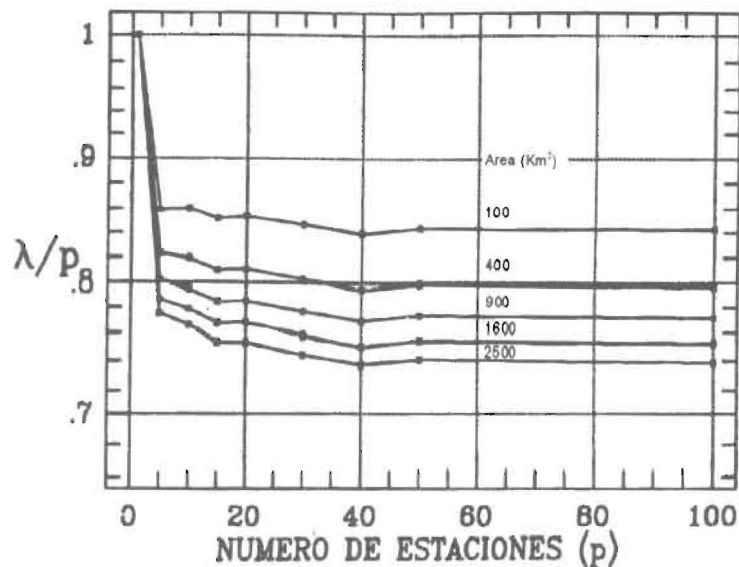


Figura 3. λ_1/p vs p para distribución aleatoria de estaciones y densidad de estaciones variable.

inicialmente y calcula el mayor autovalor de la matriz que resulta de descartar n estaciones. Numeradas las estaciones de 1 a p , el programa tiene un generador de combinaciones que elabora y muestra cada una de las C_p, n posibles combinaciones de estaciones a eliminar y selecciona aquella que hace mínimo el mayor autovalor λ de la matriz de correlación resultante, lo que corresponde a elegir la combinación de estaciones que registre la mayor variabilidad del fenómeno, pues ésta será la mejor muestra estadística. Dado el comportamiento decreciente de la varianza a medida que aumenta el número de estaciones (Fig.3), se observa que para un número fijo de estaciones, la red más confiable será la que produzca el menor pues ésta estará más cerca del comportamiento a largo plazo de λ_1/p . La red será óptima si minimiza el mayor autovalor de la matriz de correlación. El programa permite incrementar el número de estaciones a eliminar a fin de estudiar redes óptimas con distinto número de estaciones.

$$Z_{i,1} = \sum_{j=1}^p a_{ij}$$

Se eligió una red inicial de 20 estaciones dispuestas de manera homogénea en un retícula

de 5x4 (Fig.4) y se obtuvieron redes óptimas para un número de estaciones variable entre 2 y 19.

Análisis de Resultados

En primer lugar se obtuvo el gráfico λ_1/p vs. p para una distribución homogénea de estaciones con un máximo de 150 estaciones, observándose el comportamiento asintótico de la fracción λ_1/p . (Fig.2). En la Figura 3 se observa como varía el valor estable de la fracción λ_1/p a medida que aumenta el área, lo que indica que en áreas pequeñas la media sobre el área representa un mayor porcentaje de la variabilidad total del fenómeno.

El comportamiento decreciente de λ_1/p indica que en la medida en que aumenta la densidad de estaciones se introduce mayor variabilidad en la estructura de correlación de los datos, hasta un punto a partir del cual sucesivos incrementos en el número de estaciones no aportan información nueva apreciable. El valor estable de la fracción puede interpretarse como una medida de la variabilidad propia del fenómeno de precipitación sobre el área puesto que es independiente a largo plazo del número de estaciones.

Luego se dispuso una red de 20 estaciones distribuidas homogéneamente en una cuadrícula

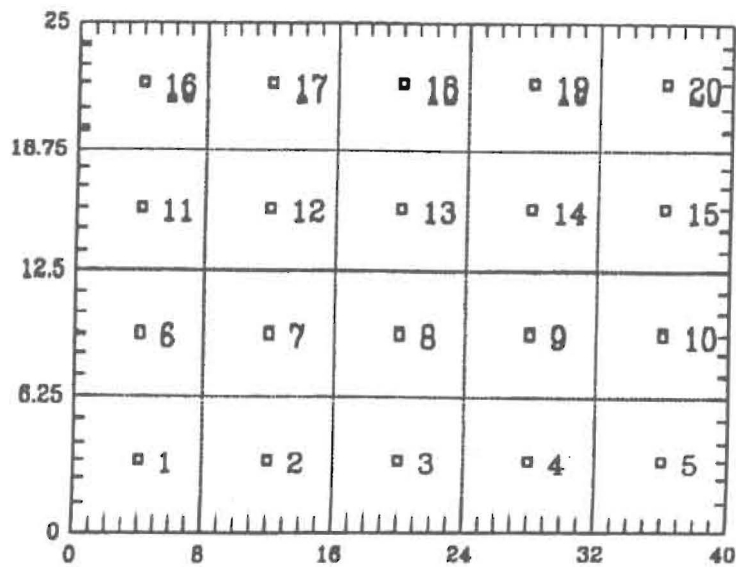


Figura 4. Distribución inicial de 20 estaciones.

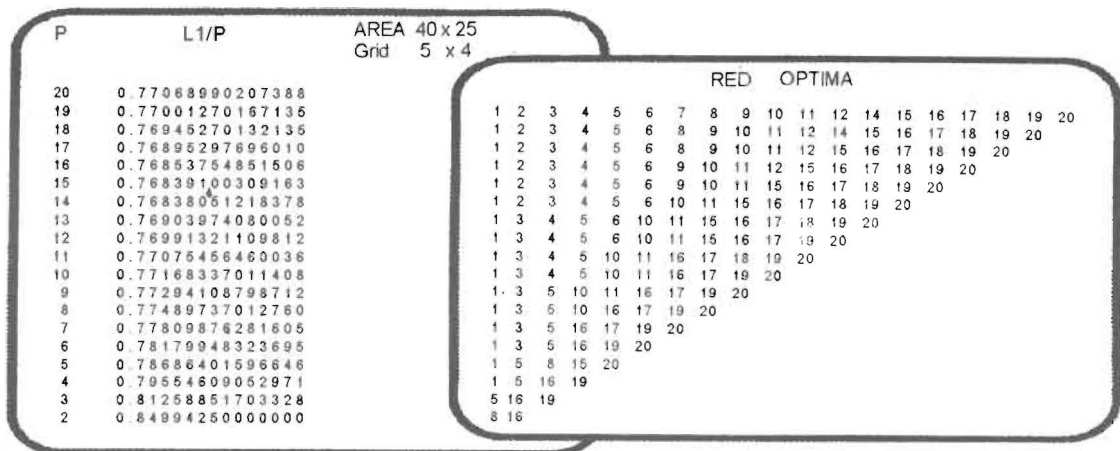


Figura 5. Salida típica del programa para redes óptimas con número de estaciones de 2 a 19

la de 5x4 (Fig.4) obteniéndose la matriz de correlación correspondiente como dato de entrada para el programa. Este generó la combinación de N estaciones que produce el menor valor de λ_1 , esto es, la red de N estaciones cuyos datos están menos correlacionados y por lo tanto suministran la mayor cantidad posible de información, lo cual minimiza la incertidumbre en las estimaciones de la media. Se varió N desde 2 a 19. La Figura 5 muestra una salida típica del programa, las configuraciones de redes correspondientes a esta salida se muestran en la Figura 6.

Para obtener una medida del grado de incertidumbre correspondiente a cada configuración de red, se tomó como 100% el intervalo (1-VR) donde VR es el valor estable de λ_1/p , expresándose el grado de incertidumbre en porcentaje:

$$\%I = \frac{\frac{\lambda_1}{p} - V.R.}{1 - V.R.} \tag{4}$$

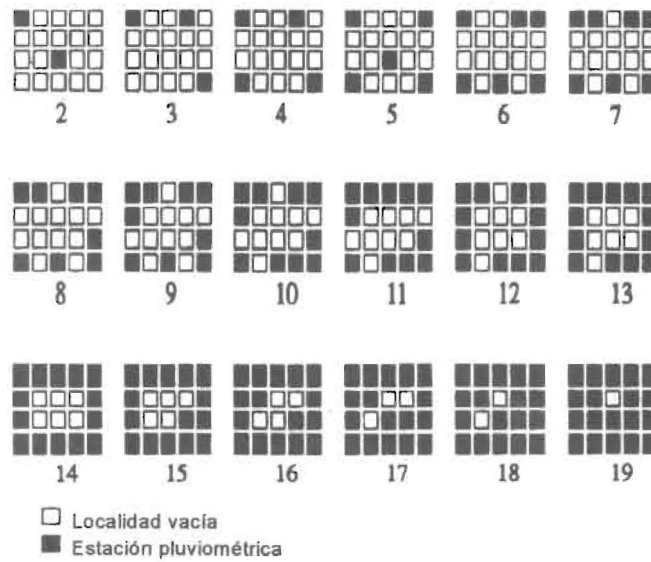


Figura 6. Configuraciones de redes óptimas encontradas por el programa.

En las simulaciones realizadas para 150 estaciones se encontró un valor para VR de 0.7634 de manera que el grado de Incertidumbre en las estimaciones de la media por las redes analizadas es:

$$\% \hat{I} = \frac{\frac{\lambda_1}{P} - 0.7634}{0.2366} \quad (5)$$

como se muestra en la Figura 7 y la Tabla 1.

En la Fig. 6 se observa que se obtuvo una gráfica decreciente que presenta un mínimo para 14 estaciones.

Conclusiones

Los resultados obtenidos mediante las simulaciones concuerdan con el análisis teórico desarrollado. Las configuraciones de redes encontradas en las simulaciones son óptimas para la función de correlación utilizada, que considera solamente la distancia y diferencia de altura entre estaciones. Con el uso de funciones de correlación más complejas, se podrán obtener redes óptimas más adecuadas a las características fisiográficas de cada área en particular.

Tabla 1
Grado de Incertidumbre (I)

No. de estac.	λ_1/p	I
2	0.8499425	0.3657756
3	0.8125885	0.2078973
4	0.795546	0.1358664
5	0.786864	0.09917152
6	0.7817994	0.07776575
7	0.7780987	0.06212469
8	0.7748974	0.04859424
9	0.7729411	0.04032592
10	0.7716833	0.03500962
11	0.7707546	0.03108443
12	0.7699132	0.0275283
13	0.7690397	0.02383639
14	0.7683805	0.02105039
15	0.768391	0.02109473
16	0.7685375	0.2171395
17	0.768953	0.0234701
18	0.7694527	0.02558196
19	0.7700127	0.02794876
20	0.7706899	0.03081109

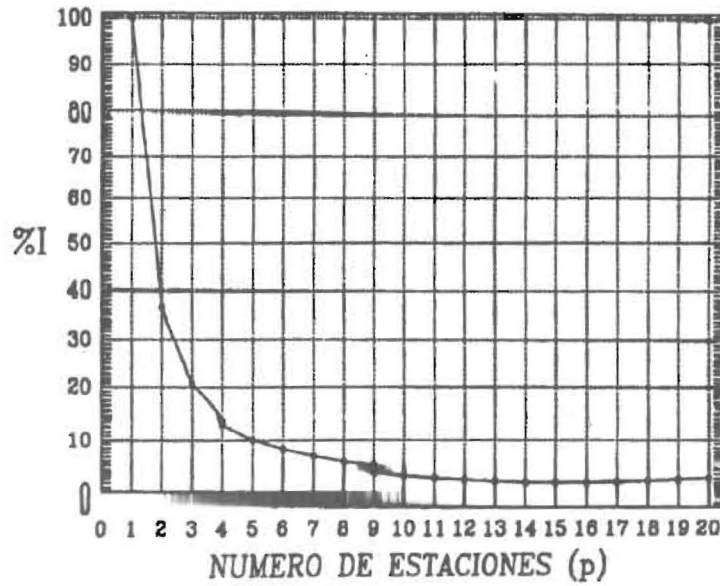


Figura 7. Estimación del grado de incertidumbre (%) para cada configuración de red.

Referencias Bibliográficas

1. Rodda, S.C. Hydrologic Network Design. Needs problems and approaches, Rep. 12, World Meteorology Organization, Geneva 1969.
2. Bras, R.L. and I. Rodriguez Iturbe, Network Design for the Estimation of Areal Mean of Rainfall Events. Water Resources. Res., 12(6), pag. 1185-1195, 1976.
3. Rodriguez Iturbe, I. y J.M. Mejia. The transformation of point rainfall to areal rainfall. Water Resour. Res., 10 No.4, 1974
4. Duband D. Hydrologie. Statistique Approfondie. Institut National Polytechnique de Grenoble. 1982.
5. Bautista de, S.; Villasmil, J.J.; "Estudio de la Distribución Espacial de la Precipitación Mensual en la Cuenca del Lago de Maracalbo" Revista Técnica. Universidad del Zulia, Vol 15, No.2, p. 87-94, 1992.
6. Mallants D. and Feyen J.. Defining Homogeneous Precipitation Regions by Means of Principal Components Analysis. Journal of Applied Meteorology. September 1990. Volume 29. p. 892-901
7. Bonelli P., Bacci P., Molteni F.. Precipitation over Northern Italy: A Description by Means of Principal Components Analysis. Journal of Climate and Applied Meteorology. October 1983. Volume 22. p. 1738-1752
8. Mardia, K.V.; Kent, J.T. and Bibby, J.M., Multivariate Analysis, Academic Press, 1979
9. Rodriguez Iturbe, I. y J.M. Mejia. The design of rainfall networks in time and space. Water Resour.Res., 10, 1974.
10. Press, W; Flannery,B.; Teukolsky, S. and Vetterling, W. Numerical Recipes. The Art of Scientific Computing. Cambridge University Press, 1987.

Recibido el 12 de Marzo de 1993

En forma revisada el 2 de Noviembre de 1993