

UN MODELO ESTOCASTICO PARA EL ESTUDIO DE LA DENSIDAD DE DRENAJE
DE LA CUENCA SUPERIOR DEL RIO DARRO (GRANADA)

Gutierrez Jaimez, R.* ; Moreno Bas, E.*

ABSTRACT : The drainage pattern is controlled by the structure and it is a decisive factor in the infiltration rate. For this reason, the analysis of this pattern has a special interest in the underground works areas, for aiding the study of the preferential flow lines. In this paper we prove the accuracy of a lognormal model for describing the random variables "length of the link" and "areas of drainage" which we have observed. Starting with this bivariate model, we get a theoretical model for describing the random variable "density of drainage". After to study the adequate statistical test we see that we can accept the model to a high level of significance.

RESUME : Le réseau de drainage, facteur décisif dans l'infiltration est contrôlé par la structure, pour cela une analyse détaillée de celle-ci est d'un intérêt spécial dans les zones où des travaux souterrains vont être réalisés, pour aider l'étude des lignes d'écoulement préférentiel. Dans cette communication, nous démontrons l'exactitude d'un modèle lognormal pour décrire les variables fortuites "longueur du lit" et "surface de drainage" que nous avons observées. A partir de ce modèle de deux variables, nous obtenons un modèle théorique pour décrire la variable random "densité de drainage". Après avoir étudié l'analyse statistique adéquate, nous voyons que nous pouvons accepter le modèle à un haut niveau de signification.

RESUMEN : La red de drenaje, factor decisivo en la infiltración, está controlada por la estructura, por lo que un análisis detallado de la misma es de especial interés en las áreas donde se van a realizar trabajos subterráneos, para ayudar al estudio de las líneas de flujo preferencial. A partir de los datos medidos en la cuenca del río Darro (Granada), se ajusta un modelo estocástico bidimensional de tipo log-normal a las variables "longitud de cauce" y "área de drenaje", se constituye toda la inferencia estadística posible para dicho modelo, y se deduce una ley de distribución para la densidad de drenaje. Se prueba, además, el alto grado de ajuste de dicha distribución, y por tanto la aplicabilidad del modelo estudiado para la descripción de la red de sistemas de tercer orden analizada, y para la predicción estadística de variables aleatorias hidrogeológicas implicadas en dicho Modelo.

* Departamento de Estadística Matemática
Facultad de Ciencias - Granada (España)

1. INTRODUCCION

La finalidad del trabajo que presentamos es describir bajo el punto de vista probabilístico la variable, δ , densidad de drenaje que obtenemos como cociente de las variables "longitud de cauces", L , y "áreas de drenaje", A , de una cierta zona de la provincia de Granada. La información muestral de que disponemos consta de 83 datos de la variable (L,A) , de la zona antes dicha, en base a los cuales se estima y contrasta un modelo logarítmico-normal.

A partir de este, una vez aceptado como válido, se deduce una ley teórica ya estimada con la que se describe las probabilidades con que δ está en intervalos (δ_0, δ_1) arbitrarios y pertenecientes al cambio de variación de δ . Dicha ley se contrasta con la información que se deduce de δ a partir de la muestra, concluyéndose que, en este caso, el modelo logarítmico-normal es aceptable a un alto grado de confianza.

Por otra parte es interesante resaltar que en la bibliografía en que se utilizan métodos matemáticos en relación con el problema de drenaje, hay dos tendencias bien definidas. Una que analiza estadísticamente los datos |por ejemplo en la línea descrita en DOORN KAMP & KING, 1971.cap.I| sin suponer un modelo probabilístico básico, con lo cual el conocimiento de las leyes teóricas que siguen - las variables hidrogeológicas queda sin resolver. Y otra línea más reciente |de la cual el modelo más desarrollado es el de DAVID-FIX, 1961 y su perfeccionamiento por el modelo de SMART, 1976| que trata de analizar el problema citado, a partir del establecimiento de un modelo teórico de carácter aleatorio, el cual relacione las variables hidrogeológicas en juego, y del que pueda extraerse información probabilística sobre el fenómeno físico en cuestión.

En este trabajo, y entre otros fines más concretos que luego se detallan, lo que hemos pretendido en un método mixto probabilístico-estadístico, con el que creemos que se obtiene el máximo de información posible de los datos observados en la realidad, al tener en consideración las dos tendencias antes citadas.

2. LA DISTRIBUCION LOGARITMICO-NORMAL BIVARIANTE COMO DISTRIBUCION DE "L" Y DE "A".

Supuesto que se adopta un punto de vista probabilístico en el estudio de las variables L , "longitud de los cauces de la zona en estudio" y A , "áreas de drenaje hacia esos cauces de esa zona", - el conocimiento más completo a que se puede aspirar es, desde luego, conocer las distribuciones univariantes y, en su caso, la bivalente que gobiernan dichas variables. El primer paso será pues de

SIAMOS-78. Granada (Españo)

cidir si dichas variables son independientes, y en este caso proponer las distribuciones univariantes que las gobiernan, o si dichas variables están ligadas entre sí y por tanto la descripción probabilística debe hacerse con una distribución bidimensional. La primera hipótesis es plenamente rechazada si se considera la información muestral dada en las Tablas I y II y en el estudio sobre el coeficiente de correlación que se hace en el apartado 3, estudio apoyado a su vez en las consideraciones que hacemos en el párrafo 2.1.. Supuesto aceptado estos cálculos, el problema que inmediatamente se plantea es la elección de la distribución bidimensional teórica de la variable (L,A) . Por estudios precedentes, (véase DAVID-FIX) y por razones que se desprenden, de manera evidente, de la información muestral, la familia de distribución que se adopta como posibles leyes de (L,A) son del tipo exponencial, y de entre estas, del tipo logarítmico-normal por dos bloques de razones fundamentalmente. El primer grupo de razones que justifica la elección de este tipo de leyes como primera aproximación probabilística, las damos en el párrafo 2.1, y, bajo nuestro punto de vista, son suficiente para emprender la construcción de los test de hipótesis estadísticos correspondientes. El segundo grupo de razones son menos rigurosas bajo el punto de vista de la Estadística Matemática pero complementan las de 2.1 con algunas consideraciones de tipo intuitivo.

2.1. En general, dadas dos variables (X,Y) no independientes sabemos que las transformaciones $(f_1(.), f_2(.))$ tales que hagan que las variables $(f_1(L), f_2(A))$ se distribuyan según una normal bivalente, cumple que el valor absoluto del coeficiente de correlación ρ tiene el mayor valor posible, para toda elección de funciones $(g_1(.), g_2(.))$ (véase Jhonson-Kotz, 1972). Si se considera las transformaciones $(\log L, \log A)$, puede verse que el valor del coeficiente de correlación muestral $\underline{r} = 0,8716$ que es el mayor de los obtenidos con otros tipos de transformaciones que se han considerado (por ejemplo, $(\sqrt{L}, \sqrt{A}), (L^2, A^2)$, etc.).

Por tanto, si no hemos encontrado otras transformaciones que hagan que el coeficiente de correlación sea mayor que este es porque la distribución normal bivalente constituye una buena aproximación a la distribución de $(\log L, \log A)$.

En consecuencia, tomando transformadas inversas se concluye que (L,A) siguen una distribución aproximada logarítmico-normal (véase Aitchinson-Brown, 1966).

La inferencia sobre dicho coeficiente \underline{r} se realiza en el apartado 3, con la adopción de normalidad para $(\log L, \log A)$.

SIAMOS-78. Granada (Español)

2.2. Argumentos para considerar esa aproximación, menos rigurosos - que los dados pero que corroboran lo expuesto son los siguientes: las marginales teóricas de la lognormal bivalente son lognormal - univariantes, (véase Jhonson-Kotz, 1972), y como puede verse en las figuras 1 y 2 y más tarde demostramos por métodos analíticos, la información muestral asegura este hecho. Ciertamente el recíproco no es cierto, es decir pudieran L y A distribuirse según lognormales - univariantes y no ser (L,A) variables distribuidas según una lognormal bivalente.

Por otro lado, considerar esta aproximación proporciona no pocas ventajas de tipo técnico como por ejemplo simplicidad en el tratamiento de los estimadores de los parámetros que intervienen, propiedades óptimas de estos, tabulaciones, etc..

Finalmente diremos que, en una primera aproximación considerar las distribuciones lognormales como modelos del fenómeno que nos -- ocupa puede ser adecuado, aunque evidentemente hablamos solo a nivel de hipótesis de trabajo que será contrastada con la información muestral que disponemos.

3. INFERENCIA SOBRE EL COEFICIENTE DE CORRELACION LINEAL.

Como antes se dijo en el apartado 2.1 el coeficiente de correlación lineal resulta valer 0,8716, calculado sobre los datos logaríticos. Aquí se completa la información sobre dicho coeficiente mediante la inferencia que es posible efectuar sobre él, en la hipótesis de normalidad. Fundamentalmente son dos los resultados teóricos básicos al respecto.

- a). El comportamiento asintótico de la distribución del citado coeficiente en el muestreo, cuando tomara el valor cero, es decir cuando la independencia entre las dos variables fuese total. En este caso

$$|r| / \frac{1-r^2}{n-2} \xrightarrow{d} t_{n-2; 2\alpha}$$

Para la deducción de este resultado ver por ejemplo Anderson, -- 1958, pág.61.

- b). El comportamiento asintótico de la citada distribución en el caso de un valor teórico no nulo ρ cualquiera. En este caso, se -- sabe que: el coeficiente r se distribuye con una densidad dada por:

SIAMOS-78. Granada (España)

$$n=N-1; \left| 2^{n-2} (1-\rho^2)^{n/2} \cdot (1-r^2)^{\frac{n-3}{2}} / (n-2) \pi \right| \cdot \sum_{\alpha=0}^{\infty} \frac{(2\rho r)^{\alpha}}{\alpha} \Gamma^2\left(\frac{n+\alpha}{2}\right)$$

en una Normal bidimensional con coeficiente teórico ρ .
(véase Anderson, 1958, pág. 69).

Utilizando la transformación de Fisher se aproxima dicho resultado por este otro asintótico:

Llamando $Z = \frac{1}{2} \ln \frac{1+r}{1-r}$ ("transformación de Fisher") y

$Z = \frac{1}{2} \ln \frac{1+\rho}{1-\rho}$ se verifica que $N-1(Z-Z) \xrightarrow{d} N(0,1)$.

En función de este resultado se obtienen los contrastes (de diferencias de coeficientes de correlación, significativas, etc.) y los intervalos de confianza de un coeficiente de correlación muestral r .

En el caso presente, teniendo en cuenta el valor calculado del coeficiente y el tamaño muestral puede llegarse a estos resultados concretos, basados en los resultados de a) y b) antes citados.

- (1) El intervalo de confianza al 95% correspondiente al valor $r = 0,8716$ calculado, con $N = 83$, resulta ser este:

$$\begin{aligned} \text{Prob. } |\text{tg h } (Z - 0,22336) \leq \rho \leq \text{tg h } (Z + 0,22336)| &= \\ = \text{Prob. } |\text{tg h } (1,11385) \leq \rho \leq \text{tg h } (1,56057)| &= \\ = \text{Prob. } |0,80157 \leq \rho \leq 0,9155| &= 0,95 \end{aligned}$$

- (2) Por otro lado es descartable la independendencia total, puesto que:

$$\frac{|r|}{\frac{1-r^2}{n-2}} \longrightarrow \frac{0,8716}{0,0544} = 16,022 \quad \text{que es claramente significativo -}$$

frente a $t_{81g.l.; 2\alpha} = 2,719$. ($\alpha = 0,05$)

- (3) Utilizando además el contraste frente a $\rho = 1$ es descartable, - igualmente al 95%, dicho valor teórico para la población, a par-

tir del dato muestral $r = 0,8716$ ya que el valor $|Z-Z| \cdot n-3$ es significativo frente al valor de la distribución teórica.

Podemos decir en resumen, que dada la muestra en estudio, el coeficiente de correlación lineal de la población correspondiente toma un valor entre 0,80157 y 0,9155, con una confianza del 95%, -- siendo descartables tanto la independencia total, como el caso $\rho = 1$ de dependencia funcional máxima (coincidencia de las dos rectas de regresión).

4. ESTIMACION Y CONTRASTE DE LA LEY LOGARITMICO-NORMAL A LA "LONGITUD DE LAS CUENCAS (L)" Y "AREAS DE DRENAJE (A)".

Por el análisis realizado, en el apartado 3, del coeficiente de correlación de los datos empíricos que aparecen en las Tablas I y II, y las razones dadas en el apartado 2, vamos a considerar la hipótesis siguiente: la longitud de las cuencas, L , es una variable aleatoria de distribución logarítmico normal que representaremos -- por $\Lambda(L|\mu_1, \sigma_1^2)$; el área de drenaje, A , es otra variable aleatoria, no independiente de L , de distribución logarítmico-normal que representaremos por $\Lambda(A|\mu_2, \sigma_2^2)$; y la variable aleatoria bidimensional -- (L, A) supondremos se distribuye según una logarítmico-normal bidimensional que representaremos por $\Lambda(\bar{\mu}, \Sigma)$, en donde $\bar{\mu}$ representa el vector bidimensional y Σ la matriz de 2×2 , asociadas a la ley.

Las expresiones de estas leyes son las siguientes. Para la longitud L .

$$d\Lambda(L|\mu_1, \sigma_1^2) \begin{cases} \frac{1}{L\sigma_1\sqrt{2\pi}} \cdot \exp\left\{-\frac{1}{2\sigma_1^2}(\log L - \mu_1)^2\right\}dL, & \text{si } L > 0 \\ 0 & , \text{ si } L \leq 0 \end{cases}$$

y análogamente para A , sin más que sustituir los parámetros (μ_1, σ_1^2) por (μ_2, σ_2^2) . La bidimensional viene dada por la expresión

$$d\Lambda(\bar{\mu}, \Sigma) = \begin{cases} \frac{1}{L \cdot A \cdot |\Sigma| \cdot 2\pi} \cdot \exp\left\{-\frac{1}{2} \left| \begin{pmatrix} (L-\mu_1) & (A-\mu_2) \end{pmatrix} \cdot \Sigma^{-1} \cdot \begin{pmatrix} (L-\mu_1) \\ (A-\mu_2) \end{pmatrix} \right|\right\} \cdot dL \cdot dA & , \text{ si } L > 0, A > 0 \\ 0 & , \text{ en otro caso} \end{cases}$$

SIAMOS-78. Granada (Españo)

Evidentemente, la aceptación o el rechazo de estas hipótesis tendremos que hacerlos en base a tests estadísticos apropiados al problema. La estimación y el contraste de estas leyes, con su consiguiente aceptación o rechazo, se lleva a cabo en las subsecciones 4.1, 4.2 y 4.3 .

4.1. ESTIMACIONES Y ERRORES DE LAS ESTIMACIONES PARA LOS PARAMETROS DE LAS DISTRIBUCIONES DE L Y A.

Los estimadores de los parámetros que aparecen en cada una de estas leyes que se van a considerar se obtienen por el método de la máxima-verosimilitud, y las estimaciones que daremos están apoyadas en los datos empíricos que aparecen en la Tabla I para la variable L, y en la Tabla II para la variable A. La expresión de los estimadores y sus varianzas es la siguiente: (véase KENDALL Y STUART, vol. 2, 1967)

$$\hat{\mu}_1 = \frac{1}{n} \sum_{j=1}^n \log L_j, \quad \hat{\sigma}_1^2 = \frac{1}{n} \sum_{j=1}^n (\log L_j - \hat{\mu}_1)^2 \quad (1)$$

$$\text{Var}(\hat{\mu}_1) = \frac{\hat{\sigma}_1^2}{n}, \quad \text{Var}(\hat{\sigma}_1^2) = \frac{2\hat{\sigma}_1^4}{n} \quad (2)$$

$$\hat{\mu}_2 = \frac{1}{n} \sum_{j=1}^n \log A_j, \quad \hat{\sigma}_2^2 = \frac{1}{n} \sum_{j=1}^n (\log A_j - \hat{\mu}_2)^2 \quad (3)$$

$$\text{Var}(\hat{\mu}_2) = \frac{\hat{\sigma}_2^2}{n}, \quad \text{Var}(\hat{\sigma}_2^2) = \frac{2\hat{\sigma}_2^4}{n} \quad (4)$$

habiéndose determinado las expresiones (3) y (4) a partir de la información de Fisher en virtud de la eficiencia de los estimadores. Además, por ser las distribuciones de $(\hat{\mu}_j, \hat{\sigma}_j^2)$, $j = 1, 2$, del tipo exponencial podemos asegurar la suficiencia de dichos estimadores (véase S. ZACKS, 1971). Si consideramos las expresiones (1) y (2), y los datos de la Tabla I, las estimaciones que resultan, junto con su varianza, son las siguientes:

$$\hat{\mu}_1 = 0,4190, \quad \hat{\sigma}_1^2 = 0,4945 \quad (5)$$

$$\text{Var}(\hat{\mu}_1) = 0,0060, \quad \text{Var}(\hat{\sigma}_1^2) = 0,0059 \quad (6)$$

Así mismo, si tenemos en cuenta las expresiones (3) y (4) y los datos de la Tabla II, las estimaciones que se obtienen, y sus varianzas, vienen dadas por:

$$\hat{\mu}_2 = -1,8369 \quad , \quad \hat{\sigma}_2^2 = 1,0148 \quad (7)$$

$$\text{Var}(\hat{\mu}_2) = 0,0122 \quad , \quad \text{Var}(\hat{\sigma}_2^2) = 0,0248 \quad (8)$$

Si determinamos la distribución logarítmico-normal para las estimaciones dadas en (5), esto es, la distribución teórica de la variable "longitud de las cuencas", podemos calcular la probabilidad de que una cuenca tenga una longitud comprendida entre dos valores fijados. Para visualizar los resultados hemos listado las frecuencias teóricas junto con las obtenidas empíricamente de la variable $-\log L$, (el hecho de trabajar con $\log L$ es simplemente por comodidad en la integración de la ley), para la partición del eje real que aparece en la Tabla III.

Análogamente, considerando la estimación dada en (7), los valores teóricos obtenidos con el modelo logarítmico-normal para la variable "áreas de drenaje", aparecen en la Tabla IV listadas junto a los valores empíricos. (realmente se han listado las frecuencias para la variable $\log A$, por la misma razón que en la variable L).

4.2. CONTRASTE DE LA HIPOTESIS DE QUE L Y A SON VARIABLES DISTRIBUIDAS SEGUN LOGARITMICO-NORMALES.

El test estadístico que vamos a utilizar para medir la posible discrepancia entre los valores teóricos y empíricos que aparecen en la Tabla III para el caso de L y en la Tabla IV en el caso de A , es el conocido "test χ^2 ". Aunque los resultados que este test establece son asintóticos, en nuestro caso puede utilizarse sin problemas debido a que la cantidad de datos es "suficientemente grande" --- ($n = 83$), de acuerdo con los órdenes de aproximación del estadístico χ^2 que se considera, a la distribución límite.

En el caso de la variable L , el valor experimental del estadístico χ^2 es 3,6496, lo que significa que la hipótesis de que L sigue una distribución logarítmico-normal de parámetros los dados en (5) puede aceptarse a un nivel de confianza del 93%. En consecuencia, dicha hipótesis será aceptada.

Repitiendo los cálculos para la variable A , el valor experimental del estadístico χ^2 es 2,5439, lo que nos dice que la ley lognormal de parámetros los dados en (7) puede aceptarse como distribu---

ción de A con un nivel de confianza del 92%.

4.3. ESTIMACION Y CONTRASTE DEL MODELO LOGNORMAL BIVARIANTE.

Los estimadores que proponemos para los parámetros del modelo bivariente son también máximo-verosímiles, y sus expresiones vienen dadas por: (véase KENDALL y STUART, 1967,)

$$(\hat{\mu}_1, \hat{\mu}_2) = \left(\frac{1}{n} \sum_{j=1}^n \log L_j, \frac{1}{n} \sum_{j=1}^n \log A_j \right) \quad (9)$$

$$\hat{\Sigma} \equiv \begin{pmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12} \\ \hat{\sigma}_{21} & \hat{\sigma}_2^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{n} \sum_{j=1}^n (\log L_j - \hat{\mu}_1)^2 & \frac{1}{n} \sum_j (\log L_j - \hat{\mu}_1) (\log A_j - \hat{\mu}_2) \\ & \frac{1}{n} \sum_{j=1}^n (\log A_j - \hat{\mu}_2)^2 \end{pmatrix} \quad (10)$$

estimadores cuyas propiedades son bien conocidas y óptimas cuando el tamaño de la muestra es suficientemente grande. En nuestro caso, el tamaño muestral es $n = 83$, con lo que puede asegurarse que sus varianzas están próximas al menor valor posible, que son centrados y, por supuesto, suficientes para la distribución de dichos vectores es exponencial. Un análisis detallado de las propiedades dadas y de las distribuciones de estos estimadores pueden encontrarse en la obra de KENDALL y STUART, 1967, o en la de ZACKS, 1971. El estimador del coeficiente de correlación y la dispersión de los estimadores (9) y (10) son:

SIAMOS-78. Granada (España)

$$(\text{con } \hat{\sigma}_{21} = \hat{\sigma}_{12}) \quad \hat{\rho} = \frac{\hat{\sigma}_{12}}{\hat{\sigma}_1 \hat{\sigma}_2} \quad (11)$$

y la matriz de dispersión por:

$$V_1(\hat{\mu}_1, \hat{\mu}_2) = \begin{pmatrix} \frac{1}{\hat{\sigma}_1^2} & -\frac{\hat{\rho}}{\hat{\sigma}_1 \hat{\sigma}_2} \\ -\frac{\hat{\rho}}{\hat{\sigma}_1 \hat{\sigma}_2} & \frac{1}{\hat{\sigma}_2^2} \end{pmatrix} \frac{n}{(1-\hat{\rho}^2)} \quad (12)$$

$$V_2(\hat{\sigma}_1, \hat{\sigma}_2, \hat{\rho}) = \begin{pmatrix} \frac{2-\hat{\rho}^2}{4\hat{\sigma}_1^2} & -\frac{\hat{\rho}^2}{4\hat{\sigma}_1^2\hat{\sigma}_2^2} & -\frac{\hat{\rho}}{2\hat{\sigma}_1^2} \\ -\frac{\hat{\rho}^2}{4\hat{\sigma}_1^2\hat{\sigma}_2^2} & \frac{2-\hat{\rho}^2}{4\hat{\sigma}_2^4} & -\frac{\hat{\rho}}{2\hat{\sigma}_2^2} \\ -\frac{\hat{\rho}}{2\hat{\sigma}_1^2} & -\frac{\hat{\rho}}{2\hat{\sigma}_2^2} & \frac{1+\hat{\rho}^2}{1-\hat{\rho}^2} \end{pmatrix} \frac{n}{1-\hat{\rho}^2} \quad (13)$$

con lo que las estimaciones que se obtienen considerando los datos empíricos, (aquellos cuyas marginales son las Tablas I y II), son los siguientes:

$$(\hat{\mu}_1, \hat{\mu}_2) = (0,4190 \quad -1,8369) , \quad \hat{\Sigma} = \begin{pmatrix} 0,4945 & 0,6174 \\ 0,6174 & 1,0148 \end{pmatrix} \quad (14)$$

con varianzas para $(\hat{\mu}_1, \hat{\mu}_2)$, las dadas por

$$V_1^{-1} = \begin{pmatrix} 0,0008134 & 0,0010156 \\ 0,0010156 & 0,0016693 \end{pmatrix} \quad (15)$$

y para $(\hat{\sigma}_1, \hat{\sigma}_2, \hat{\rho})$ las dadas por

$$V_2^{-1} = \begin{pmatrix} 0,00589 & 0,0091862 & 0,0012476 \\ 0,0091862 & 0,024814 & 0,0025604 \\ 0,0012476 & 0,0025604 & 0,000695 \end{pmatrix} \quad (16)$$

SIAMOS-78. Granada (España)

Si este modelo bivalente puede ser aceptado como bueno, las líneas de regresión de L sobre A y de A sobre L darían una expresión funcional de como varia la variable L cuando lo hace la A y recíprocamente. Además los coeficientes de la función quedarían automáticamente determinados por los valores dados en (14) y (11).

Desde luego, el admitir el modelo como válido para describir el fenómeno llevaría consigo construir un test estadístico apropiado al caso, no obstante, ya sabemos algo al respecto, pues las marginales de esta distribución serían en efecto logarítmico-normales que han sido aceptadas a un alto nivel de significación. Sin embargo, sabemos que esto no es suficiente, es decir, las marginales pueden ser las adecuadas y sin embargo la conjunta no está determinada en general, unívocamente por estas. Un posible test sería el siguiente: ya que una de las finalidades básicas del trabajo es encontrar la distribución de probabilidad para la densidad de drenaje, vamos a determinar la distribución de dicha densidad de drenaje a partir de la lognormal bivalente aquí obtenida. Hecho esto, contrastaremos con la distribución empírica de dicha densidad. Nótese que esto es conceptualmente distinto a los contrastes realizados para L y para A, pues en estos casos las leyes teóricas fueron estimadas a partir de los datos empíricos.

La obtención de la ley de la densidad de drenaje a partir del modelo bivalente y el contraste con los datos empíricos de dicha densidad se dan en el apartado 6., y como veremos los resultados son positivos. Por esta razón, en el apartado siguiente vamos a dar líneas de regresión que se deducen con el modelo lognormal-bivalente.

5. LINEAS DE REGRESION DE L SOBRE A Y DE A SOBRE L.

5.1 Supuesto el modelo lognormal aceptado, esto es la variable aleatoria bidimensional (L, A) se distribuye según $\Lambda(\mu, \sigma)$, puede verse (véase por ejemplo Johnson y Kotz "Distributions in Statistics" Vol.4, John Wiley, 1972, pág.15) que la variable condicionada (L|A) se distribuye según una lognormal de parámetros

$$\Lambda\left(\hat{\mu}_1 + \frac{\hat{\mu}_1 \hat{\sigma}_1}{\hat{\sigma}_2} (\log A - \hat{\mu}_2), (1 - \hat{\rho}^2) \hat{\sigma}_1^2\right)$$

y que (A|L) sigue otra lognormal de parámetros

$$\Lambda\left(\hat{\mu}_2 + \frac{\hat{\rho} \hat{\sigma}_2}{\hat{\sigma}_1} (\log L - \hat{\mu}_1), (1 - \hat{\rho}^2) \hat{\sigma}_2^2\right).$$

Por tanto, las líneas de regresión vendrán dadas por las esperanzas matemáticas de cada una de estas leyes, esto es:

SIAMOS-78. Granada (Españo)

Línea de regresión de (L|A)

$$E |L|A| = A \frac{\hat{\sigma}_1}{\hat{\sigma}_2} \cdot \exp\left\{\frac{1}{2}(1-\hat{\rho}^2)\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} + \hat{\mu}_1 - \hat{\rho} \hat{\sigma}_1 \frac{\hat{\mu}_2}{\hat{\sigma}_2}\right\}$$

Línea de regresión de (A|L)

$$E |A|L| = L \frac{\hat{\sigma}_2}{\hat{\sigma}_1} \cdot \exp\left\{\frac{1}{2}(1-\hat{\rho}^2)\frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} + \hat{\mu}_2 - \hat{\rho} \hat{\sigma}_2 \frac{\hat{\mu}_1}{\hat{\sigma}_1}\right\}.$$

cuyas estimaciones son inmediatas considerando (14) y (11). Así --
pués los resultados numéricos son:

$$E |L|A| = 2,1340 \cdot A^{0,4247} \quad (17)$$

$$E |A|L| = 4,2022 \cdot L^{1,7887} \quad (18)$$

5.2 REGRESIONES DE PRIMERA Y SEGUNDA CLASE.

Todo problema de ajuste de distribuciones probabilísticas, como el presente, pretende en el fondo predecir con el mayor grado de precisión posible los valores de ciertas variables implicadas en el modelo, a partir del conocimiento de otros datos, de otras variables.

La primera aproximación al respecto, y en un modelo bidimensional como el que aquí se propone, es utilizar las líneas de regresión correspondientes. A continuación, e independientemente de las predicciones dadas para el modelo log-normal bidimensional [ver 5.1] vamos a construir la inferencia sobre las rectas de regresión basadas en la variable bidimensional normal, asociada a la lognormal citada, ya que en este caso es posible utilizar resultados exactos para los contrastes, bandas de confianza y tolerancia etc. de las citadas rectas de regresión normales.

Los datos numéricos básicos son los siguientes:

$$\bar{L} = 0,419 \quad \bar{A} = -1,8369$$

$$S_L^2 = 0,4945 \quad S_A^2 = 1,0148$$

$$S_L = 0,7032 \quad S_A = 1,0073$$

$$S_{LA} = 0,61738$$

$$r = 0,8716 \quad r^2 = 0,75968 \quad 1-r^2 = 0,24032$$

SIAMOS-78. Granada (España)

5.3 En primer lugar se construyen las rectas de regresión de primera clase, con datos no agrupados (tamaño muestral $N=83$), y las correspondientes hipérbolas que dan los límites de confianza -- así como los límites de tolerancia correspondientes. Estas hipérbolas, así como las elipses que más tarde aparecen en la regresión de segunda clase, constituyen "bandas de confianza" --- realmente, midiéndonos en cada valor predicho de una variable a partir del valor dado de la otra, el error cometido en la predicción. En ambos casos vamos pues a construir las bandas dejando variable el valor a que se condiciona la predicción, por lo que tiene sentido hablar de hipérbola y elipse pues sus ecuaciones son las que nos dan los límites de las bandas en donde está contenida la predicción más o menos el error correspondiente.

En efecto sucesivamente calculamos:

(a) Recta de regresión de A|L.

$$\begin{aligned} A &= -1,8369 + 1,2485 (L - 0,419) \\ &= -2,36002 + 1,2485 L \end{aligned}$$

(a.1) Significación frente a cero del coeficiente angular b_{AL} .

$$\frac{b_{AL} - b_{AL}}{S_{b_{AL}}} = \frac{b_{AL}}{S_{b_{AL}}} = \frac{1,24849}{6,0886 \times 10^{-3}} = 205,053$$

claramente significativo frente a $t_{2\alpha; n-2}$ g.l.

(a.2) Límites de confianza para el valor b_{yx} técnico.

$$b_{AL} \pm t_{2\alpha; S_{b_{AL}}} \quad \text{con } v = n-2 \text{ g.l. si } n-2 < 200.$$

En este caso:

$$\begin{aligned} &1,24849 \pm 1,664, 6,0886 \times 10^{-3} = \\ &= 1,24849 \pm 0,01013. \end{aligned}$$

SIAMOS-78. Granada (Españo)

(a.3) Límites de confianza para el valor a_{yx} teórico.

En este caso $a_{AL} = -2,36002$ con unos límites dados -- por

$$-2,36002 \pm t_{2\alpha; n-2} \cdot S_{a_{yx}} \quad (\text{ya que } n < 200)$$

$$\text{donde } S_{a_{AL}}^2 = S_{b_{AL}}^2 \cdot \frac{\Sigma x^2}{n}$$

(b) Predicción de A dado L. Límites de confianza.

$\bar{y} + b_{AL}(L - \bar{L}) \pm t_{2\alpha; n-2} \cdot S_{AL}$ que toma la forma:

$$-2,36002 + 1,2485 L \pm t_{2\alpha; 81} \cdot 4,367 \cdot 10^{-5} (L - 0,419)^2 + 2,996 \cdot 10^{-5}$$

que es la hipérbola de confianza correspondiente.

(c) Predicción de L cuando se da A, según la recta de regresión calculada.

$$L A = \bar{L} + \frac{A - \bar{A}}{b_{AL}} = 0,419 + \frac{A - 1,81369}{1,24849}$$

(c.1) Límites de confianza de esta predicción.

Al ser $k^2 = t_{2\alpha}^2 \cdot S_{b_{LA}}^2 / b_{LA}^2 = 7,7577 \cdot 10^{-5} \ll 0,05$ se

puede tomar $1 - k^2 \approx 1$ con lo que las expresiones correspondientes a los límites indicados se simplifican quedando de esta forma:

$$0,419 + \frac{A - 1,81369}{1,24849} \pm 0,00705 \cdot 1,0696 + (A - 1,81369)^2$$

5.4 Regresión de segunda clase e Inferencia asociada.

Como es sabido, en los problemas estadísticos con dos variables x e y y pueden darse, en la práctica, dos casos bien definidos de relación funcional lineal:

SIAMOS-78. Granada (España)

- (a) Que una variable produzca un efecto en la otra (efecto casual)
- (b) Que ambas variables sean "casuales" (regresión de segunda clase).

Se caracterizan porque en el primer caso existe una sola recta de regresión (A|L), que se estudió antes para nuestro problema concreto, la cual se emplea para cálculo en ambos sentidos -por eso antes en el punto (C de 5.3) la utilizamos de A a L en lugar de L a A-.

Y en el caso (b) anterior hay dos rectas de regresión distintas, por consiguiente la inferencia de A a L es distinta de la de L a A.

Aquí desarrollamos los cálculos e inferencia de segunda clase para perfeccionar los de primera antes estudiados, ya que en el fenómeno hidrogeológico en cuestión suponemos que esta segunda posibilidad puede ser más realista en general (la fluctuación estadística casual de ambas variables L y A).

En efecto se encuentran sucesivamente los siguientes resultados, para nuestro problema:

(a) Rectas de regresión.

Toman la forma:

$$A|L = -1,8369 + 1,24849 (L - 0,419)$$

$$L|A = 0,419 + 0,60837 (A + 1,8369)$$

(b) Los límites de confianza son de la forma

$$\bar{A} + b_{AL} (L - \bar{L}) \pm S_{b_{AL}} \sqrt{2F \cdot S_L / n - (n-2) (L - \bar{L})^2}$$

$$\bar{L} + b_{LA} (A - \bar{A}) \pm S_{b_{LA}} \sqrt{2F \cdot S_A / n - (n-2) (A - \bar{A})^2}$$

que nos proporciona respectivamente las elipses de confianza de las regresiones A|L y L|A respectivamente:

$$A|L = -1,8369 + 1,24849 (L - 0,419) \pm 0,078 \cdot \sqrt{2 F_{2;81} \cdot 40,549 / 83 - 81 (L - 0,419)^2}$$

SIAMOS-78. Granada (España)

$$L|A = 0,419 + 0,60837(A+1,8369) \pm 0,038 \quad 2 F_{2;81}$$

$$83,2136/83-81(A+1,8369)^2$$

(c) En forma análoga obtenemos los límites de tolerancia que vienen dados por sendas elipses de tolerancia, que resultan ser estos:

$$A|L = -1,8369 + 1,24849(L-0,419) \pm 0,078 \quad 2,85 \cdot F_{2;81}$$

$$40,549/83-81(L-0,419)^2$$

$$L|A = 0,419 + 0,60837(A+1,8369) \pm 0,038 \quad 2 F_{2;81}$$

$$83,2136/83-81(A+1,8369)^2$$

Como es sabido es interesante representar gráficamente tanto - estas elipses así obtenidas como las hipérbolas de la regresión de primera clase vista en (5.3 (b)). Para ello son de interés las pendientes de los ejes principales y los semiejes de las elipses, que son los siguientes respectivamente:

$$b_0 - \frac{1}{b_0} = \frac{S_A - S_L}{2S_{AL}} \pm 1 + \left(\frac{S_A - S_L}{2S_{AL}} \right)^2 = 0,42137 \pm 1,08515$$

$$l_1 = \frac{F_{2;81}}{83,81} \quad 123,7626 \pm 15137,181 - 3245,3089 =$$

$$= 0,0121 \quad F_{2;81} \cdot 15,28511 = 0,41177$$

$$l_2 = 0,0121 \quad F_{2;81} \cdot 3,7270 = 0,09940$$

$$l_1; l_2 = k \quad S_L + S_A \pm (S_L + S_A)^2 - 4(S_L S_A - S_{LA}^2)$$

con $k = \frac{F_{2;n-2}}{n(n-2)}$ (para elipse de confianza) y $k = \frac{F_{2;n-2} \cdot (n+1)}{n(n-2)}$ (para

la elipse de tolerancia).

Por ejemplo en el caso de la elipse de confianza resultan estos valores calculados para los datos hidrogeológicos básicos de este trabajo.

Y análogamente se calculan los semiejes de la elipse de tolerancia.

Para completar los cálculos y las representaciones gráficas antes aludidas -y de las que aquí prescindimos- son de interés los datos sobre las tangentes a las elipses anteriores y sobre las ordenadas en el origen, fácilmente calculables conociendo los datos ya empleados en la determinación de las elipses básicas. Con todos estos datos, se determina fácilmente el "rectángulo" de confianza y de tolerancia mediante las tangentes horizontales y verticales a las citadas cónicas. En estos rectángulos nos movemos siempre, a la hora de utilizar las rectas de regresión de segunda clase como base de predicción de L dado A o bien de A dado L, en nuestro problema concreto de tipo hidrogeológico, y al 95%.

6. DISTRIBUCION DE LA DENSIDAD DE DRENAJE $\delta = \frac{L}{A}$

Supuesto que la variable bidimensional (L, A) se distribuye según el modelo logarítmico-normal bivalente dado en el apartado 4, la densidad de drenaje δ , definida como el cociente L/A, tiene la distribución resultante de determinar la conjunta (L/A, Z) y determinar a partir de aquí la marginal mediante la integración en Z. Ahora bien, si (L, A) sigue una logarítmico-normal, esto equivale a que -- (log L, Log A) se distribuya según una normal bivalente de parámetros (μ, Σ) . Con esto la transformación, de Jacobiano la unidad, dada por

$$\begin{pmatrix} \log \delta \\ z \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \log L \\ \log A \end{pmatrix},$$

transforma dicha normal bivalente en la normal $N(\bar{C}\bar{\mu}^1, C\Sigma C^1)$ siendo C la matriz de la transformación y habiendo utilizado la "prima" para indicar la operación de transponer. Operando resulta que $(\log \delta, z)^1$ se distribuye según una normal de parámetros.

$$\bar{\mu}^* = \begin{pmatrix} \hat{\mu}_1 & -\hat{\mu}_2 \\ \hat{\mu}_2 \end{pmatrix} \quad \text{y} \quad \Sigma^* = \begin{pmatrix} \hat{\sigma}_1^2 & -2\hat{\sigma}_{12} + \hat{\sigma}_2^2 & \hat{\sigma}_{12} - \hat{\sigma}_2^2 \\ \hat{\sigma}_{12} - \hat{\sigma}_2^2 & \hat{\sigma}_2^2 \end{pmatrix}$$

con lo que inmediatamente se deduce que δ se distribuye según una -- lognormal de parámetros

$$\Lambda(\hat{\mu}_1 - \hat{\mu}_2, \hat{\sigma}_1^2 - 2\hat{\sigma}_{12} + \hat{\sigma}_2^2)$$

siendo los parámetros que aquí aparecen los que se determinan a partir de los datos en el apartado 4, expresión (14).

SIAMOS-78. Granada (Español)

Nótese que el resultado será aceptable si la hipótesis de partida es válida, es decir si (L, A) es lognormal bivalente. Bajo esta hipótesis la lognormal obtenida es la $\Lambda(2,256, 0,2744)$.

Para comprobar si esto es cierto, consideremos los datos empíricos de δ y determinemos sus análogos teóricos con la ley obtenida.

Los resultados que se obtienen aparecen en la Tabla VI bajo el epígrafe "frecuencias teóricas". Estas han sido listadas junto a los datos empíricos que aparecen en esa misma Tabla bajo el epígrafe -- "frecuencias empíricas", (como en los casos anteriores listamos --- log δ en vez de δ por comodidad en los cálculos).

Como en los casos anteriores el grado de discrepancia entre las distribuciones teórica teórica y empírica lo medimos con un test χ^2 . La χ^2 experimental que se obtiene vale 2,5566 lo que equivale a decir que la ley obtenida es aceptable a un nivel de significación superior al 95%. Por este motivo no solo aceptamos dicha ley para la distribución de δ , sino que la logarítmico-normal bivalente puede también tomarse como válida lo que a su vez lleva consigo la aceptación de las líneas de regresión dadas en el apartado 5.

AGRADECIMIENTOS. - Agradecemos las orientaciones que en el campo de la Hidrogeología nos ha proporcionado el Profesor Fernández-Rubio. Damos las gracias al Profesor J.L. Díaz que nos facilitó los datos básicos de la cuenca analizada. Por último, nuestro agradecimiento a la Dra. Vila Miranda por el procesamiento de los datos utilizados en este trabajo, y a D^a Soledad Cantos Hoces por su cuidadosa mecanografía.

BIBLIOGRAFIA:

1. AITCHINSON, J. & J.A.C. BROWN, 1966 "The lognormal distribution". Cambridge at University Press.
2. ANDERSON, 1958, "An Introduction to Multivariate Statistical Analysis" Ed. Wiley.
3. DAVID, F.N. y FIX, E., 1961, "Rank correlation and regression in a nonnormal surface, en Proc. of Fourth Berkeley Sym. vol.1, Univ. California Press, p.177-197.
4. DIAZ, J.L., 1977, "Análisis Hidrogeológico del Alto-Medio Darro - (Granada). Tesis Licenciatura. Cátedra Hidrogeología. Universidad de Granada.

SIAMOS-78. Granada (Español)

5. DOOKNKAMP, J.C. y KING, C.A.M., 1971, "Numerical Analysis in Geomorphology: An Introduction". Ed. Arnold.
6. JHONSON, & KOTZ, 1972, "Distributions in Statistics: Continuous - Multivariate Distributions". Jhon Wiley.
7. KENDALL y STUART, 1967, "Advanced Theory of Statistics", 2ª edición. Griffin.
8. SMART, 1976, "Joint Distribution Functions for Link Length and - Drainage Areas" en Random Processes in Geology, edited by D.F. - Merrian. Springer-Verlag.
9. SMART, J.S., 1972, "Quantitative characterization of channel network structure". Water Resources Res. vol.8, n°6, p.1487-1496.
10. ZACKS, S., 1971, "The Theory of Statistical Inference" Wiley and Son.

TABLA I

DATOS EMPIRICOS DE LAS LONGITUDES DE LOS CAUCES DE LA RED
IMPLANTADA EN LOS SISTEMAS DE TERCER ORDEN (L_j)

(Cuenca Superior del Río Darro)

1,06	0,43	2,75	1,67	2,67	0,95	1,03	1,70	1,46	1,35
3,88	0,58	7,22	1,67	6,68	3,49	2,48	1,48	2,58	0,80
4,47	4,39	2,13	2,12	2,17	2,38	4,13	1,60	1,10	0,95
1,13	0,62	1,41	0,93	1,60	1,24	1,40	1,73	0,86	0,88
1,36	2,59	0,66	2,62	1,17	0,59	0,47	1,24	0,66	0,67
0,50	0,92	1,20	3,35	1,52	0,55	1,57	0,58	1,28	2,85
0,95	2,81	2,03	2,65	0,70	1,02	4,51	4,72	5,54	1,28
0,17	5,91	2,16	1,40	1,82	0,98	1,34	2,37	1,78	3,27
1,80	1,25	0,65							

Fuente: D. José Luis Díaz Hernández, del Grupo de Trabajo de Hidrogeología de la Universidad de Granada (Dominio Bético S. Str)

SIAMOS-78. Granada (Españo)

TABLA II

DATOS EMPIRICOS DE LAS SUPERFICIES DE DRENAJE (A_j) DE LOS -
CAUCES DE LA TABLA I.

0,221	0,020	0,212	0,140	0,176	0,062	0,069	0,156	0,154
0,181	0,304	0,052	1,189	0,156	1,988	1,296	0,558	0,397
0,462	0,078	1,247	0,833	0,384	0,351	0,239	0,341	0,477
0,243	0,085	0,059	0,063	0,035	0,129	0,071	0,177	0,078
0,145	0,239	0,110	0,137	0,206	0,838	0,060	0,418	0,060
0,030	0,035	0,089	0,078	0,060	0,035	0,054	0,134	0,363
0,100	0,054	0,134	0,020	0,081	0,265	0,101	0,336	0,528
0,347	0,062	0,053	0,355	0,610	1,137	0,109	0,079	0,531
0,104	0,074	0,134	0,127	0,156	0,154	0,127	0,414	0,226
0,254	0,034							

Fuente: D. José Luis Díaz Hernandez, del Grupo de Trabajo de
Hidrogeología de la Universidad de Granada (Dominio
Bético S. Str)

TABLA III

FRECUENCIAS EMPIRICAS Y TEORICAS DE LA VARIABLE LOG L

INTERVALOS		FR. EMPIR.	FR. TEOR.
-∞	-0,987	1	1,89
-0,987	-0,706	2	2,66
-0,706	-0,425	7	5,00
-0,425	-0,143	6	8,03
-0,143	0,138	12	11,02
0,138	0,419	16	12,90
0,419	0,700	10	12,90
0,700	0,982	12	11,02
0,982	1,26	7	8,03
1,26	1,544	5	5,00
1,544	1,825	3	2,65
1,825	∞	2	1,89

SIAMOS-78. Granada (España)

TABLA IV

FRECUENCIAS EMPIRICAS Y TEORICAS DE LA VARIABLE LOG A

INTERVALOS		FR. EMPIR.	FR. TEOR.
-∞	-3,852	2	1,89
-3,852	-3,046	7	7,65
-3,046	-2,643	8	8,03
-2,643	-2,240	11	11,02
-2,240	-1,837	16	12,90
-1,837	-1,434	10	12,90
-1,434	-1,031	11	11,02
-1,031	-0,628	9	8,03
-0,628	0,178	6	7,65
0,178	∞	3	1,89

TABLA V

VALORES EMPIRICOS DE LA DENSIDAD DE DRENAJE $\phi_j = \frac{L_j}{A_j}$; $j=1, \dots, 83$

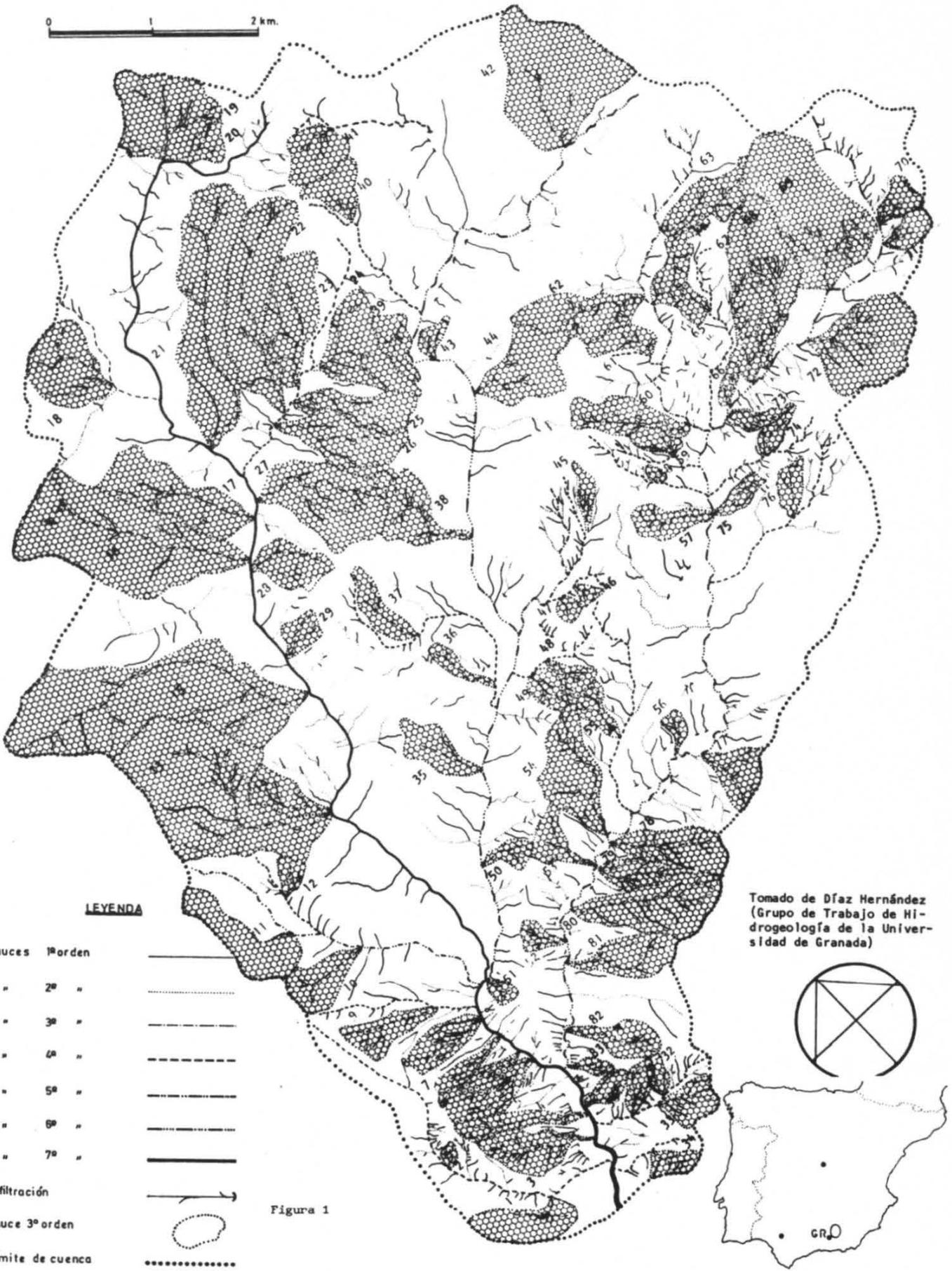
4,796	21,500	12,972	11,929	15,170	15,323	14,928	10,897
9,481	7,459	12,763	11,154	6,072	10,705	3,360	2,693
4,444	3,728	5,584	10,256	3,585	5,270	5,547	6,040
9,079	6,979	8,658	6,584	12,791	16,102	17,937	17,714
10,930	13,099	9,040	15,897	9,655	7,238	7,818	6,423
6,602	3,091	11,000	6,268	19,500	19,667	13,429	13,933
8,462	11,167	14,286	17,037	8,955	9,229	15,200	10,185
12,463	29,000	15,802	10,755	9,406	8,363	3,845	7,637
11,290	19,245	12,986	7,738	4,872	11,743	2,152	11,130
20,769	18,919	13,582	7,717	8,590	15,390	14,016	7,899
7,965	4,921	19,118					

TABLA VI

FRECUENCIAS EMPIRICAS Y TEORICAS DE LA VARIABLE LOG δ

INTERVALOS		FR. EMPIR.	FR. TEOR.
$-\infty$	1,208	3	1,89
1,208	1,627	8	7,66
1,627	1,837	6	8,03
1,837	2,046	9	11,02
2,046	2,256	13	12,90
2,256	2,465	13	12,90
2,465	2,675	12	11,02
2,675	2,885	10	8,03
2,885	3,304	8	7,65
3,304	∞	1	1,89

0 1 2 km.



LEYENDA

Cauces	1º orden	
"	2º "	
"	3º "	
"	4º "	
"	5º "	
"	6º "	
"	7º "	
Infiltración		
Cauce 3º orden		
Límite de cuenca		

Tomado de Díaz Hernández
(Grupo de Trabajo de Hidrogeología de la Universidad de Granada)

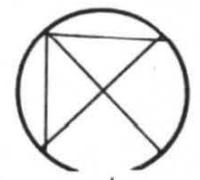


Figura 1

