hidrología subterránea

METODOS APLICADOS A HIDROLOGIA SUBTERRANEA

PROSPECCION GEOELECTRICA
ANALISIS MULTIVARIADO
GEOESTADISTICA

LA PLATA - DICIEMBRE DE 1990

CFI
CONSEJO FEDERAL DE INVERSIONES

Segundas Jornadas de Actualización
en Hidrología Subterránea

Introducción a la Prospección Geoeléctrica Aplicada en Geohidrología

CALVETTY AMBONI, Boris
RAPACCINI, Alicia

La Plata
1990

PROSPECCION GEOELECTRICA
Y SU APLICACION A LOS ESTUDIOS GEOHIDROLOGICOS DEL CFI

INTRODUCCION

En las páginas siguientes se exponen en breve detalle los principios y técnicas de cálculo a los que se ajusta la metodología aplicada en las mediciones geoeléctricas por los geofísicos del Consejo Federal de Inversiones.

Los métodos empleados, preponderantemente en exploraciones geohidrológicas, son los de Sondeo Electrico Vertical (SEV) y en menor medida el de Calicatas Eléctricas (CE), con aplicación excluyente de la configuración electródica de Schlumberger.

Por consiguiente la exposición se limita a una enunciación de conceptos generales, a la descripción de la práctica de los métodos SEV y CE y a las técnicas de interpretación adoptadas. Como complemento, en un anexo, se resumen los numerosos trabajos con aplicación de estos métodos realizados desde 1977 a la fecha, incluyendo los realizados por convenios de cooperación horizontal.

De tal manera, solo se pretende aqui introducir al estudio de los métodos de Prospección Geoeléctrica, dando los elementos que se consideran indispensables para su utilización, además de difundir las ventajas de la aplicación de tales técnicas geofísicas en la exploración geohidrológica.

1. CONCEPTOS GENERALES

PROSPECCION GEOFISICA

Sin intentar una definición estricta, puede decirse que Prospección Geofísica es la investigación de la parte superior de la corteza terrestre con fines utilitarios mediante la aplicación de métodos físicos y matemáticos.

Sus principios de aplicación se basan en la variabilidad de las propiedades físicas de las rocas como ser: conductividad, elasticidad, susceptibilidad magnética, radiactividad, etc. Cada una de ellas da lugar a un método de características diferentes, cuya clasificación exhaustiva se omite, indicándose tan solo que pueden ser agrupados en: Magnéticos, Gravimétricos, Sísmicos, Eléctricos, Radiactivos y Térmicos.

PROSPECCION ELECTRICA

Aunque el desarrollo de los métodos eléctricos es relativamente reciente, sus orígenes deben rastrearse hasta mediados del siglo XVIII, cuando Watson descubre la conductividad del suelo, o a comienzos del siglo XIX, cuando el inglés Fox descubre el fenómeno del Potencial Espontáneo y sugiere su aplicación en la prospección minera.

No obstante, su más importante avance y diversificación se da en el presente siglo. En gran medida, a partir de las investigaciones de Conrad Schlumberger, quien en 1913 realiza dos importantes experimentos: En el primero, mediante mediciones del potencial espontáneo, descubre un yacimientos de sulfuros en Bor (Servia, Yugoslavia). En el segundo, aplicando corriente artificial, estudia una cuenca silúrica en Calvados (Francia).

El rápido perfeccionamiento de estos y otros métodos estuvo ligado hasta la década del 30, a este ingeniero de minas alsa-

ciano quien, con su hermano Marcel, el matemático rumano Stefanescu y el físico francés Maillet, trabajó en su desenvolvimiento, no solo en su faz práctica sino que, investigando las condiciones de la propagación de la corriente eléctrica en medios estratificados, dotó a los mismos del indispensable sustentamiento teórico. Esta última cuestión fue descuidada por muchos investigadores que, especulando exclusivamente con eventuales logros experimentales, arribaron a conclusiones erróneas que cobraron amplia difusión merced a la sencillez de su aplicación, perjudicando notoriamente el desarrollo de estos métodos.

Siguiendo el camino de Schlumberger, identificado con la escuela francesa, son enormes los avances logrados con posterioridad, merced a la dedicación de numerosos investigadores y a los adelantos tecnológicos, los que permitieron su enriquecimiento y diversificación.

CLASIFICACION DE LOS METODOS ELECTRICOS

Entre los métodos de prospección eléctrica pueden diferenciarse, en primera instancia, aquellos que exploran campos naturales preexistentes de los que requieren una fuente de energía artificial, y entre éstos, los que utilizan campos constantes (corriente continua) de los que lo hacen con campos variables. Tal como lo muestra Orellana (1982) en el siguiente cuadro:

A. Metodos de campo natural:

1.- Potencial Espontáneo
2.- Corrientes Telúricas
3.- Magneto - telúrico
4.- AFMAG

B. Metodos de campo artificial:

B.1. de campo constante

1.- Líneas equipotenciales y del cuerpo cargado
2.- Sondeos Eléctricos
3.- Calicatas eléctricas

B.2. de campo variable

1.- de frecuencia
2.- por establecimiento de campo
3.- Calicatas electromagnéticas
4.- de radiografía hertziana

B.3. Polarización inducida

APLICACIONES

Los métodos eléctricos proporcionan información del subsuelo que puede ser utilizada con fines muy variados. Los más importantes son los siguientes:

- Investigaciones tectónicas para la búsqueda de petróleo
- Estudios para la localización de aguas subterráneas
- Estudio de cuencas carboníferas
- Detección de yacimientos de menas metálicas
- Investigaciones de basamento para cimentaciones
- Detección de zonas de fuga en embalses
- Investigaciones a profundidad muy reducida en estudios arqueológicos
- Estudios complementarios para cartografía de suelos
- Estudios de zonas muy profundas de la corteza terrestre.

RESISTIVIDAD DE LAS ROCAS

Dado un conductor cualquiera, se denomina corriente eléctrica a todo desplazamiento ordenado de cargas libres. Esta será continua si su sentido e intensidad no varían con el tiempo. En estas

condiciones, una corriente I que atraviesa una superficie S, establecerá una densidad de corriente dada por:

$$J = I.S^{-1} \qquad (1)$$

relacionada con el campo eléctrico E mediante la ley de Ohm:

$$J = \sigma.E = \rho^{-1}E \qquad (2)$$

donde σ y ρ representan la conductividad y la resistividad del conductor, respectivamente. Magnitudes físicas de gran amplitud de variación, debido a la diversidad de formas de conducción, que dependen de la naturaleza del cuerpo considerado.

Los minerales que constituyen las rocas son en general malos conductores de la electricidad, por lo que las propiedades conductivas de estas se deben a que presentan una gran proporción de vacíos o poros con algún contenido de agua intersticial, agua que a su vez tiene sales en solución que la hacen iónicamente conductora. En consecuencia, son muchos los factores que intervienen en la caracterización resistiva de una roca: forma y distribución de los poros, grado de saturación, naturaleza y concentración del electrolito, temperatura, etc. Esto hace imposible atribuir a cada variedad una determinada resistividad, aunque sí, un margen de variación que en casi todos los casos es bastante amplio, como puede apreciarse en la figura 1.

Suponiendo la saturación de una porción de roca, es decir, que todos sus poros se encuentran ocupados por algún fluido, su resistividad, ρ_s, será función de su porosidad, p, y de la resistividad del fluido, ρ_v, según la siguiente relación:

$$\rho_s = ap^{-m}\rho_v \qquad (3)$$

donde m depende principalmente del grado de cementación de la roca, variando entre 1.3 para rocas no cementadas y 2.3 para rocas

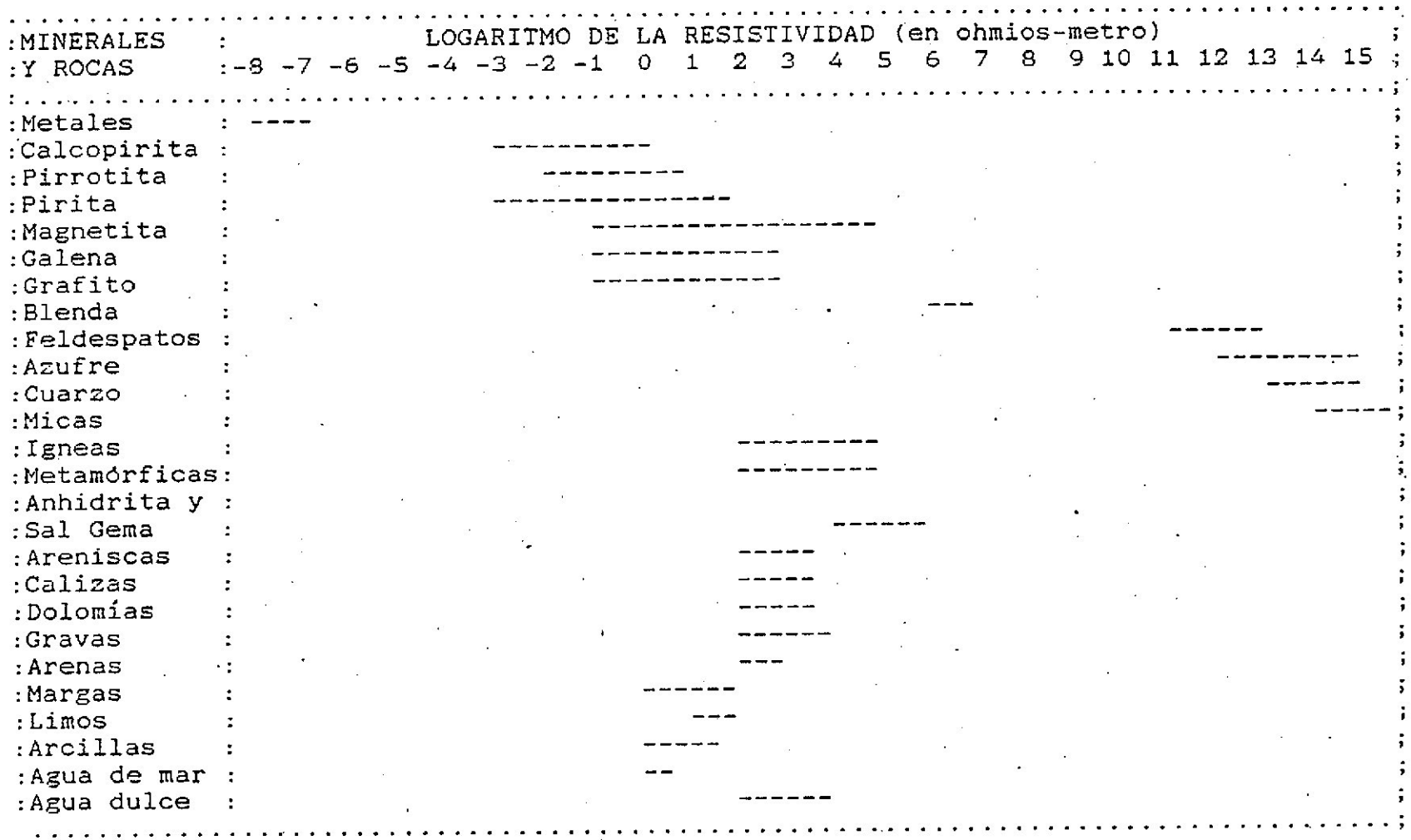


FIGURA 1- Resistividad de algunas rocas y minerales

Modificado de Orellana (1982)

porosas de elevada cementación. El valor de a varía entre 0.5 y 1.5 y depende de la textura de la roca. Para rocas sedimentarias y cuando no se dispone de datos previos, suele utilizarse en primera aproximación: a = 1, m = 2.

En general:

$$\rho_s = F\rho_w \qquad (4)$$

donde el coeficiente F se denomina Factor de formación.

Para rocas no saturadas, habrá que considerar un índice de saturación $\mathcal{I} = bs^{-n}$, donde b y n son constantes a determinar y s el grado de saturación. Como $b \cong 1$, $n \cong 2$, en el caso más general

$$\rho_r = (ps)^{-2}\rho_w \qquad (5)$$

proporcionará un valor aproximado.

FORMULAS Y DISPOSITIVOS EN LOS METODOS DE CORRIENTE CONTINUA

Los principios que rigen la práctica de estos métodos se basan en las leyes de la circulación de la corriente eléctrica en medios isótropos. Una idea suscinta de estos principios puede exponerse suponiendo que el terreno en el que se van a efectuar las mediciones es un medio homogéneo e isótropo, de resistividad ρ, separado por un plano horizontal de un semiespacio de resistividad infinita que representa a la atmósfera (figura 2).

Aplicando un potencial eléctrico entre dos puntos A y B de este plano, se produce, a través del medio isótropo, la circulación de corriente eléctrica, I, cuya densidad de corriente, J, en cada punto del medio es proporcional al campo eléctrico, E, provocado en dicho punto:

$$J = \rho^{-1}E \qquad (6)$$

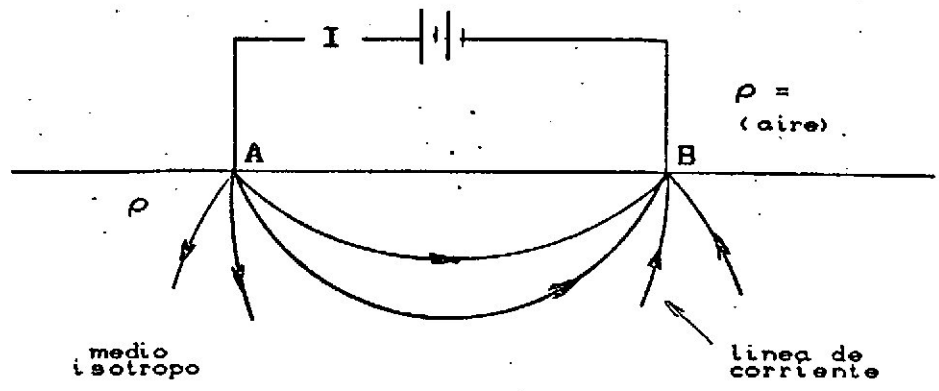


FIGURA 2- Circulación de la corriente en un medio isótropo de resistividad ρ

En caso de ubicar el electrodo A lo suficientemente alejado del B (electrodo B en infinito), las líneas de corriente en las proximidades de A serán radiales y divergentes (figura 3). En consecuencia, las superficies equipotenciales serán semiesféricas y el campo eléctrico **E** en cada una de ellas estará dado por:

$$E = \frac{\rho I}{2\pi} r^{-2} \qquad (7)$$

donde **I** es la corriente que penetra en el terreno por **A** y r representa el radio de la superficie semiesférica de referencia.

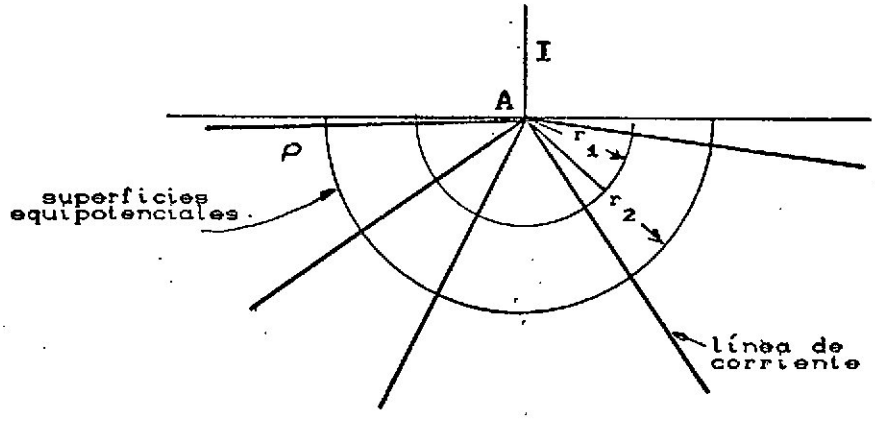


FIGURA 3- Líneas de corriente y superficies equipotenciales en las proximidades del electrodo A

Partiendo de la ec. 7, es posible calcular la diferencia de potencial, ΔV, entre dos de dichas superficies, de radios r_1 y r_2

$$\Delta V = V_1 - V_2 = \frac{\rho I}{2\pi}(r_1^{-1} - r_2^{-1}) \qquad (8)$$

donde V_1 es el potencial a la distancia r_1 de A, o lo que es equivalente, el potencial de la semiesfera de radio r_1.

De esta fórmula puede despejarse el valor de ρ, resistividad del medio isótropo, en función de I y ΔV que pueden ser medidos sobre la superficie plana:

$$\rho = 2\pi \frac{1}{r_1^{-1} - r_2^{-1}} \frac{\Delta V}{I} = K\frac{\Delta V}{I} \qquad (9)$$

Donde K se denomina constante geométrica por su dependencia estricta de las condiciones geométricas del dispositivo empleado, siendo sus dimensiones las de una distancia.

En condiciones reales, el subsuelo no es precisamente homogéneo ni isótropo, por lo que el valor obtenido aplicando la fórmula anterior, corresponderá a una integración de los valores reales implicados en la medición y dependerá también de las posiciones relativas de las tomas de tierra que posibilitan la energización del terreno, puntos A y B, y la medición de ΔV, puntos M y N. Por esta razón, lo que habitualmente se obtiene es una resistividad aparente, ρ_a, la que puede definirse como la resistividad que tendría un medio homogéneo en el que, al aplicar una corriente I entre electrodos de emisión, se observaría la misma diferencia de potencial entre los electrodos de medición que en el medio heterogéneo.

Para la obtención de este parámetro, se utilizan dispositivos cuya forma general se muestra en la figura 4:

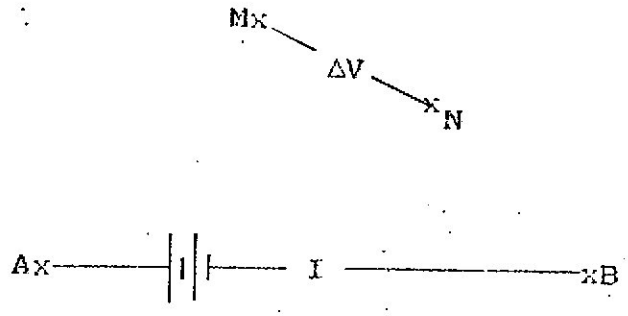


FIGURA 4- Dispositivo general para mediciones de la resistividad aparente (en planta)

En cuyo caso:

$$\rho_a = 2\pi\left(\frac{1}{AM} - \frac{1}{BM} - \frac{1}{AN} + \frac{1}{BN}\right)^{-1}\frac{\Delta V}{I} \qquad (10)$$

En la práctica, los dispositivos más usuales pueden diferenciarse entre lineales y dipolares.

DISPOSITIVOS LINEALES

Los electrodos de medición se disponen sobre una línea, siendo los mas difundidos los conocidos con los nombres de Schlumberger y Wenner:

a) Configuración Schlumberger

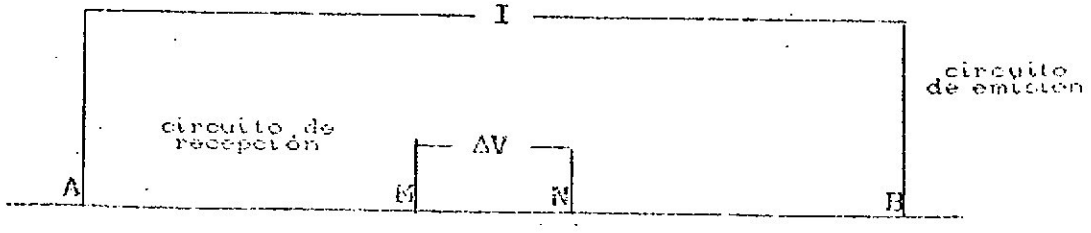


FIGURA 5

es un dispositivo simétrico que debe cumplir con la condición de que MN debe ser menor, a lo sumo igual, que AB/5. Los valores de resistividad aparente se grafican en función de AB/2. La constante geométrica es:

$$K = \frac{\pi}{4MN}(AB^2 - MN^2) \qquad (41)$$

b) Configuracion Wenner

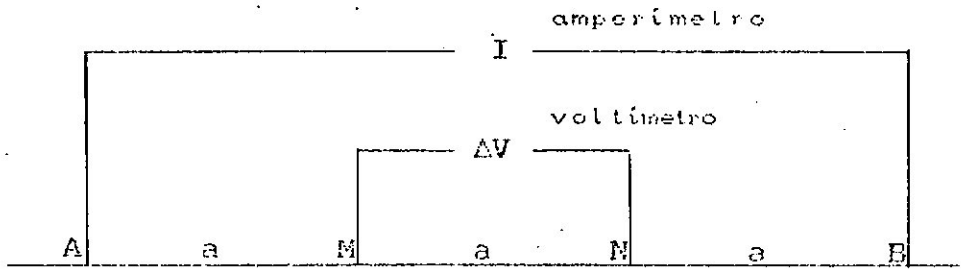


FIGURA 6

también simétrico y en el que los electrodos deben mantenerse equiespaciados. Los valores de resistividad aparente se grafican en función de a = AM = MN = NB. La constante geométrica es:

$$K = 2\pi a \qquad (42)$$

c) Configuraciones tricelectródicas

Son variaciones de las dos anteriores y resultan de colocar uno de los electrodos de corriente lo suficientemente alejado del centro del dispositivo como para que no influya sobre los valores del potencial (electrodo en infinito). La constante geométrica es la del dispositivo base multiplicada por 2.

8.-DISPOSITIVOS DIPOLARES:

En este tipo de configuraciones, los pares AB y MN se hallan lo suficientemente separados como para suponer que cada uno de ellos constituye un dipolo. En principio, la posición del dipolo

de recepción MN respecto del de emisión AB, puede ser cualquiera.

En la práctica, se utilizan los de la figura siguiente:

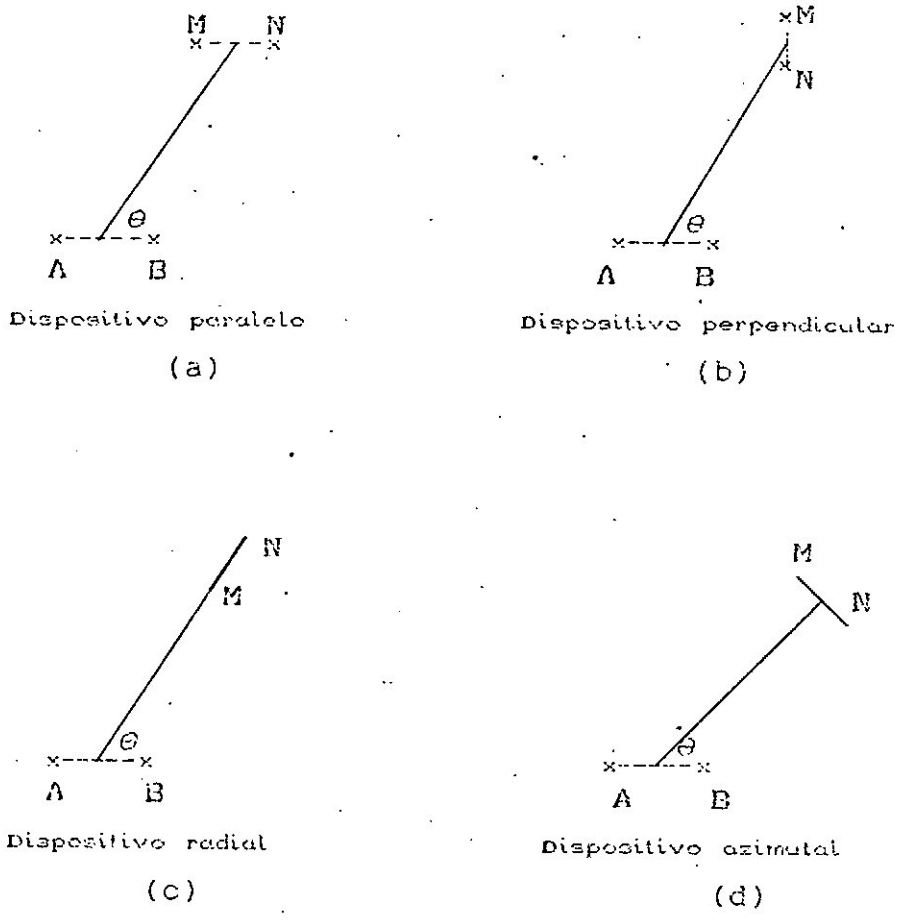


FIGURA 7- Disposición de electrodos en los dispositivos dipolares

2. EL METODO DE SONDEO ELECTRICO VERTICAL

Se denomina Sondeo Electrico Vertical (SEV), a una serie de mediciones de la resistividad aparente efectuadas con un dispositivo de azimut y centro fijos y separación creciente entre electrodos de emisión. Los dispositivos que cumplen con estas características son los de Wenner y Schlumberger.

Como en cualquier dispositivo electródico, la resistividad aparente se calcula mediante una fórmula del tipo

$$\rho_a = K(\Delta V/I) \qquad (13)$$

La finalidad del SEV es la determinación de un modelo de la variación en profundidad de la resistividad del subsuelo, a partir de mediciones en superficie.

La necesidad de energizar el terreno, medir la corriente que circula y el potencial provocado, requiere la materialización de dos circuitos eléctricos independientes entre sí, denominados de corriente o emisión y de potencial o recepción, respectivamente, tal como se muestra esquemáticamente en las figuras 5 y 6.

Circuito de emisión: Incluye una fuente de poder, un amperímetro, electrodos de corriente y cable.

En las mediciones rutinarias del CFI se utiliza como fuente de energía una batería de acumuladores de 12 V en serie con un convertidor de 250 W de potencia. La tensión de salida varía entre unos pocos voltios y algunos centenares de voltios.

Los amperímetros empleados permiten mediciones de hasta 10 A, con una precisión del 1 % y resolución de 0.1 mA en el menor alcance. La lectura es digital, con indicador numérico de $3\frac{1}{2}$ dígitos e indicador de polaridad. Se alimenta con una batería común de 9 V.

El cable, medio de transmisión de la corriente eléctrica desde la fuente hasta los electrodos, tiene un mm^2 de sección, y para su transporte y fácil extensión en las mediciones va arrollado en carretes con capacidad para 500 m.

Los electrodos de corriente son varillas de alrededor de 0.8 a 1 m de largo y 0.2 m de diámetro, de acero inoxidable.

Circuito de recepción: La medición de las diferencias de potencial se realiza mediante milivoltímetros electrónicos. En estas mediciones se debe considerar el efecto de los potenciales naturales, cuya influencia es necesario neutralizar. Ello se consigue con la inclusión de un circuito compensador en el instrumento de medición, gracias al cual es posible medir solamente el efecto de la corriente de energización.

Los utilizados en CFI tienen un alcance máximo de 1999 mV, precisión del 1 %, resolución de 10 µV en el menor alcance, impedancia de entrada de 10 MΩ, compensador de potencial espontáneo hasta 200 mV, lectura digital por cristal liquido de $3\frac{1}{2}$ dígitos, indicador de polaridad y alimentación con baterías comunes de 9 V o pilas comunes chicas de 1,5 V

El cable, necesario para conectar los electrodos de potencial con el voltímetro, tiene las mismas características del empleado en el circuito de corriente, y es transportado en carretes de no más de 250 metros.

En cuanto a los electrodos de potencial, se emplean los impolarizables, constituidos por un vaso de base porosa conteniendo una solución saturada de sulfato de cobre en agua, en el que se sumerge una varilla de cobre, la que se conecta al cable del circuito de medición.

Práctica del sondeo eléctrico vertical: Los SEV, de acuerdo a la longitud que alcanza la extensión de sus alas, pueden ser

clasificados en :

Cortos	cuando	AB < 300 m
Normales	"	300 m < AB < 3000 m
Largos	"	3000 m < AB < 30000 m
Muy largos	"	AB > 30000 m

Los sondeos cortos son empleados principalmente en Ingeniería Civil, los normales en Hidrogeología, los largos en prospección petrolera y los muy largos en investigaciones de la corteza terrestre, habiéndose alcanzado en este caso, una longitud superior a los 1000 km.

La medición de un sondeo "normal", requiere de un operador y tres o cuatro ayudantes. Pueden efectuarse entre 3 y 6 por día, dependiendo ello de: longitud de los SEV, distancias entre uno y otro y condiciones topográficas, principalmente.

Establecida la ubicación del sondeo y la dirección de sus alas, e instalado el instrumental en el lugar correspondiente al centro, se procede a colocar los electrodos en sus posiciones iniciales. Compensado el potencial natural, se energiza el terreno, se lee la corriente I que circula y la diferencia de potencial ΔV provocada, se anotan estos valores en la planilla correspondiente y se calcula el valor de ρ_a, el que es volcado en un gráfico bilogarítmico en función de AB/2 (si el dispositivo es el de Schlumberger) o de a (si es el de Wenner). Este procedimiento se repite para los sucesivos valores de AB/2 o de a hasta completar el sondeo. Terminado éste, se tendrán graficados puntos correspondientes a una Curva de Resistividad Aparente (CRA), figura 8, la que deberá ser construida en base a ellos para dar lugar al proceso de interpretación.

EL METODO DE CALICATAS ELECTRICAS

Las Calicatas Eléctricas son mediciones de la resistivid

CFI

OPERADOR: Calvetty Amboni

Provincia: Bs. As. S.E.V. Nº 8

Depto: SAAVEDRA Rumbo: N-S

Zona: Huanca A° Chasico Fecha: 23/0/90

Observaciones:

Rta 29

km 87,7

AB/2 (m)	MN (m)	I (mA)	V (mV)	ρ_a (Ωm)
2	1	46	234	59,9
3		38,4	111	79,5
4		24,8	66	93,8
5		42,9	57	103,3
6,5		29,1	30	101,2
8		40,7	203	90,9
10		41,2	12,0	91,3
13		84,4	12,6	79,1
16		145	11,4	63,2
20		344	13,1	47,8
25		454	8,4	36,3
32		705	6,0	27,4
40		313	1,50	24,1
50	1/20	422/211	1,22/13	22,9/23,2
65	1/20	820/543	1,85/12,9	21,4/22,3
80		608	12,3	21,6
100		772	10,3	20,8
125		710	5,8	19,9
160		630	2,8	17,8
200		450	2,1	17,6
250	20/100	477/480	078/3,8	16,0/14,9
320	20/100	730/740	0,71/3,4	15,6/14,4
400		1050	3,3	15,6
500		1010	2,85	16,1
650		920	1,6	23,0
800		810	1,13	27,9

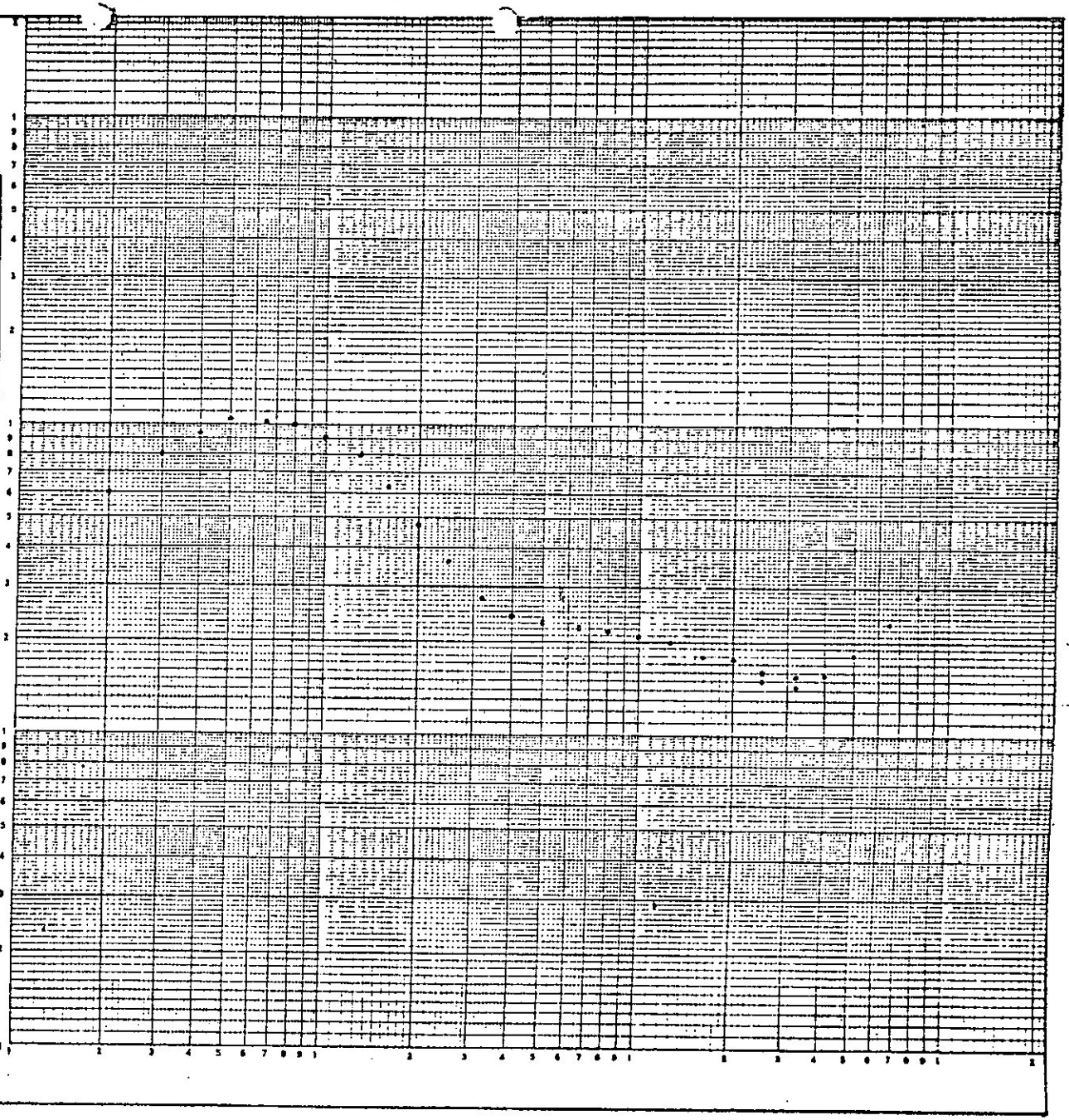


FIGURA 8- Planilla con datos de campo. Método SEV.

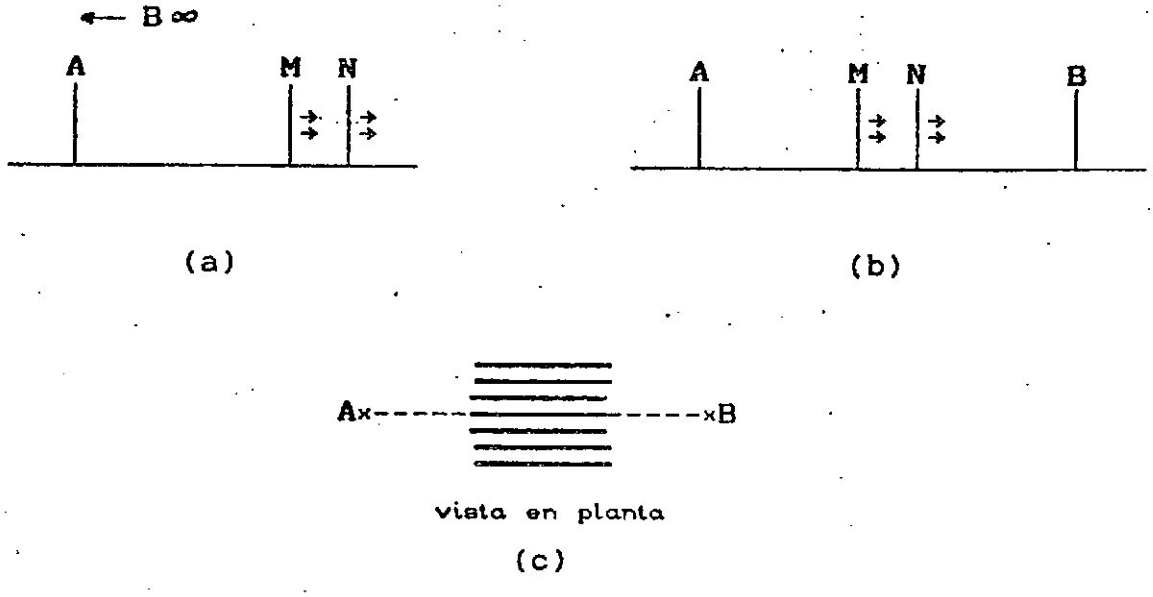
aparente efectuadas con el propósito de determinar variaciones laterales de la resistividad. Trátase, en síntesis, de obtener mapas o perfiles de las variaciones de la resistividad superficial, o subsuperficial, con la finalidad de poner en evidencia contrastes geológicos en sentido horizontal. Es decir, aunque son también investigaciones del subsuelo, no interesan las variaciones en profundidad.

Con tal finalidad, se emplean tanto dispositivos de campo fijo (el circuito de corriente no se desplaza durante la medición) como dispositivos de campo móvil (ambos circuitos se desplazan simultáneamente).

La conformación de los circuitos se ajusta a lo descrito en líneas anteriores, siendo las configuraciones más usuales las siguientes:

a) **Calicatas de campo fijo:**

de gradiente (fig 9a)
Schlumberger, método de bloques (fig 9b y 9c)



(a) (b) (c)

FIGURA 9- Calicatas de campo fijo

b) Calicatas de campo movil:

dipolares (fig 10a)
trielectródicas (fig 10b)
simétricas (fig 10c)
circulares

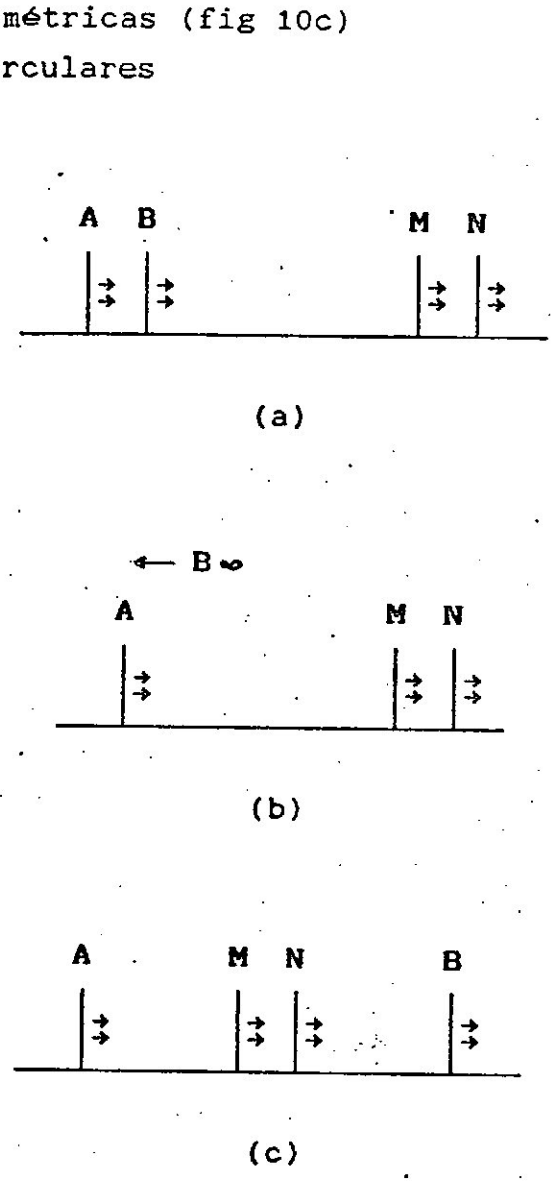


FIGURA 10- Calicatas de campo movil

3- INTERPRETACION DE SONDEOS ELECTRICOS VERTICALES

MEDIOS ESTRATIFICADOS

Sea un subsuelo heterogéneo constituido por estratos horizontales y paralelos, homogéneos e isótropos. Este medio puede caracterizarse mediante el espesor E y la resistividad ρ de cada capa. Cada uno de estos medios parciales se denomina capa geoeléctrica, y a un conjunto de capas geoeléctricas suprayaciendo a un medio de espesor infinito (sustrato) se lo denomina corte geoeléctrico.

Un corte geoeléctrico de n capas requiere para su especificación el conocimiento de n resistividades y (n-1) espesores o profundidades, es decir un total de (2n-1) parámetros (figura 11).

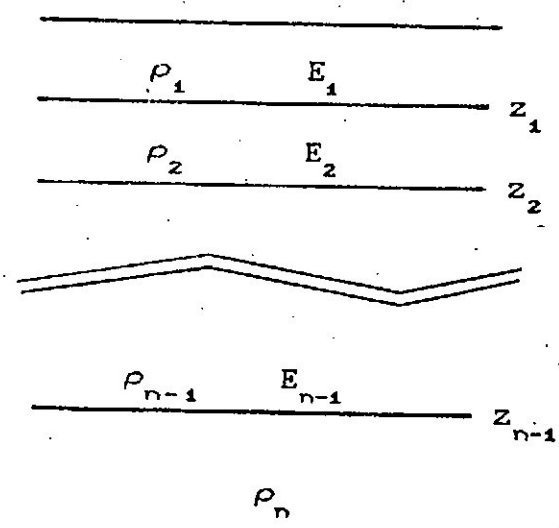


FIGURA 11 - Corte Geoeléctrico de n capas

Si se representan las resistividades de las capas en función de la profundidad en coordenadas logarítmicas se obtiene una representación gráfica del corte geoeléctrico que se denomina Curva de Resistividad Verdadera (CRV). Así, esta curva puede graficarse junto con la CRA.

Los cortes geoeléctricos se clasifican según relaciones de desigualdad establecidas entre sus resistividades, recibiendo de acuerdo a ellas distintos nombres. Por extensión, las CRV y las CRA reciben los mismos nombres que los cortes a los que corresponden.

a) cortes de dos capas

Existen dos tipos según sea la resistividad de la segunda capa mayor o menor que la de la primera. Se denominan ascendente si $\rho_1 < \rho_2$ (figura 12) y descendente si $\rho_1 > \rho_2$ (figura 13).

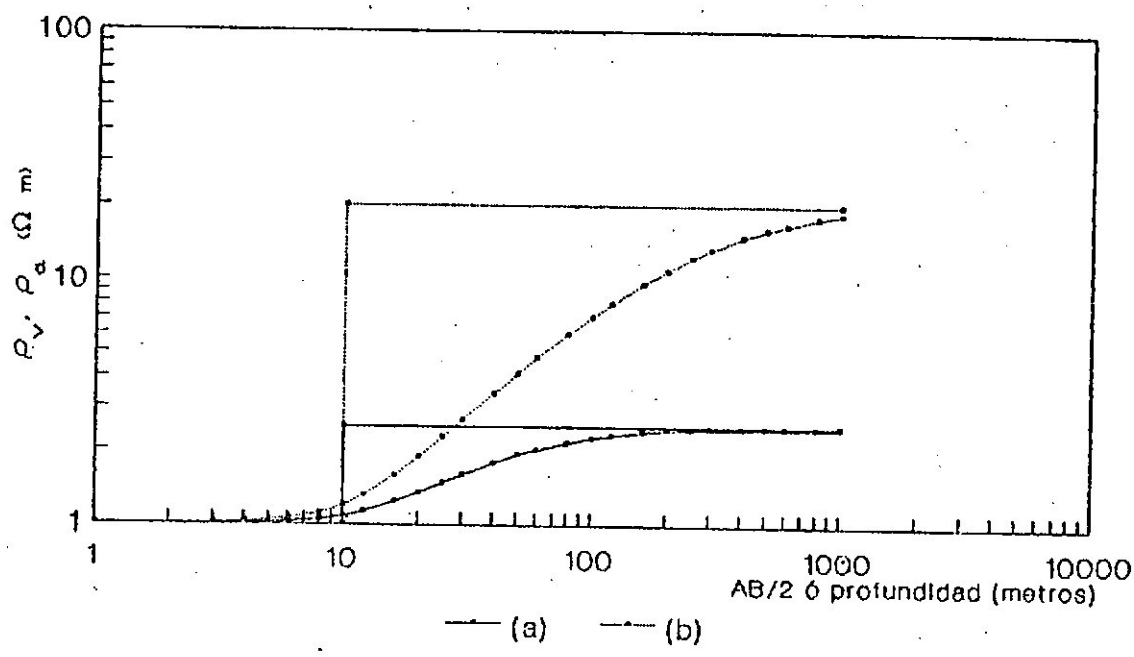


FIGURA 12 - Cortes geoeléctricos ascendentes
$E_1 = 10$, $\rho_1 = 1$, $\rho_2 = 2.5$ (a) y 20 (b)

b) cortes de tres capas

Hay cuatro tipos posibles (figuras 14 a 17)

Tipo H si $\rho_1 > \rho_2 < \rho_3$

Tipo K si $\rho_1 < \rho_2 > \rho_3$

Tipo Q si $\rho_1 > \rho_2 > \rho_3$

Tipo A si $\rho_1 < \rho_2 < \rho_3$

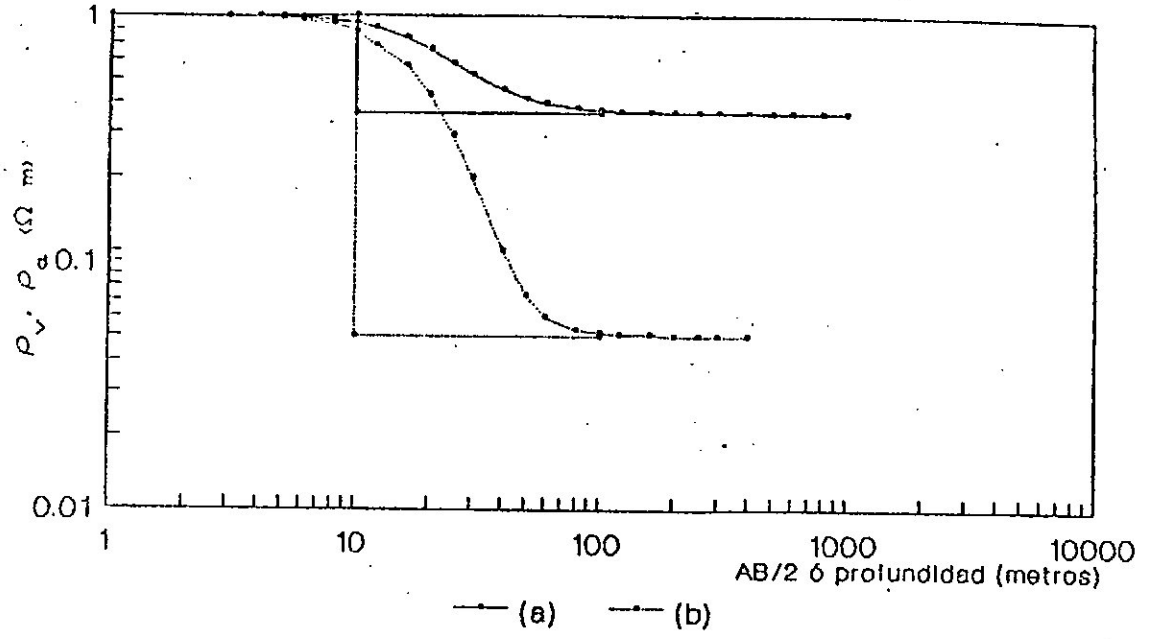


FIGURA 13 - Cortes Geoeléctricos descendentes
$E_1 = 10$, $\rho_1 = 1$, $\rho_2 = 0.4$ (a) y 0.05 (b)

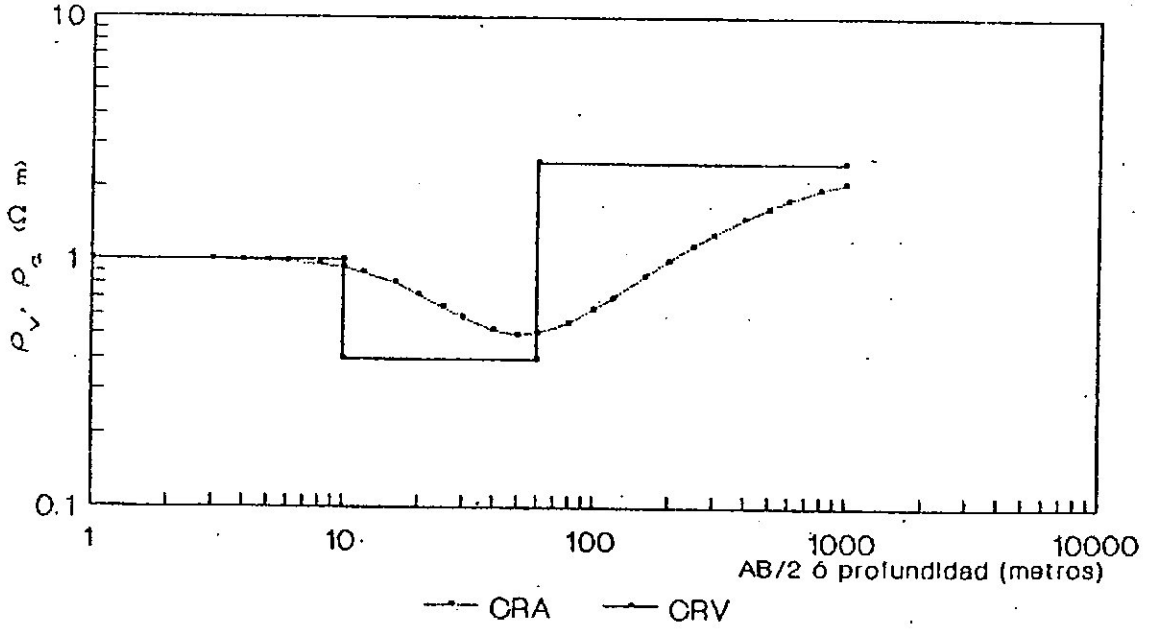


FIGURA 14 - Corte Geoeléctrico Tipo H
$E_1 = 10$, $E_2 = 50$, $\rho_1 = 1$, $\rho_2 = 0.4$, $\rho_3 = 2.5$

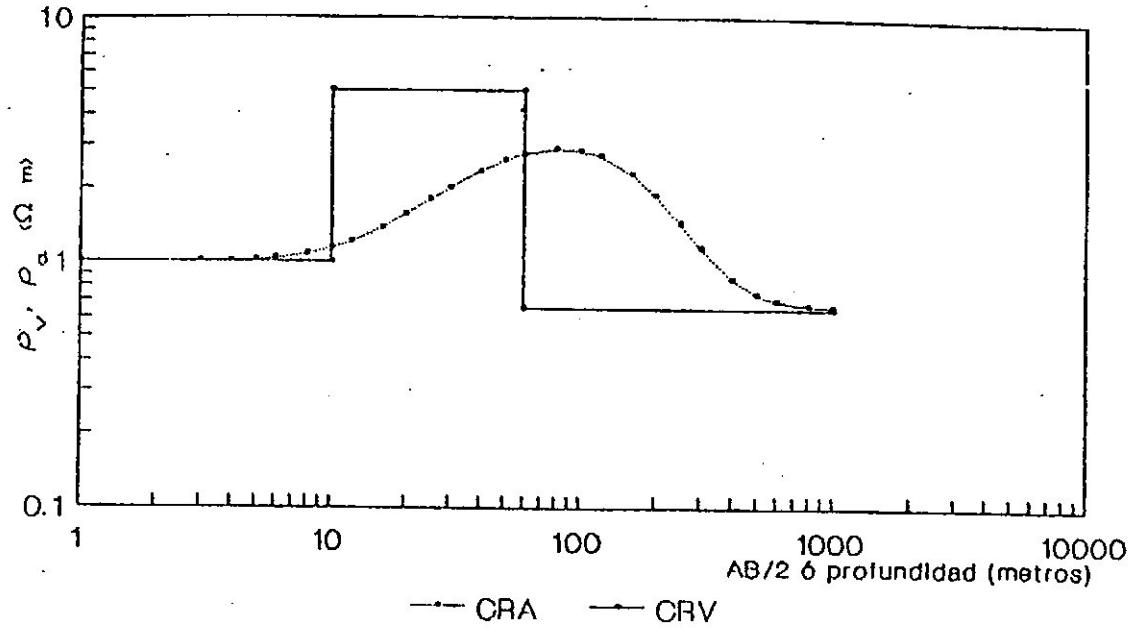


FIGURA 15 - Corte Geoeléctrico Tipo K
$E_1 = 10$, $E_2 = 50$, $\rho_1 = 1$, $\rho_2 = 5$, $\rho_3 = 0.65$

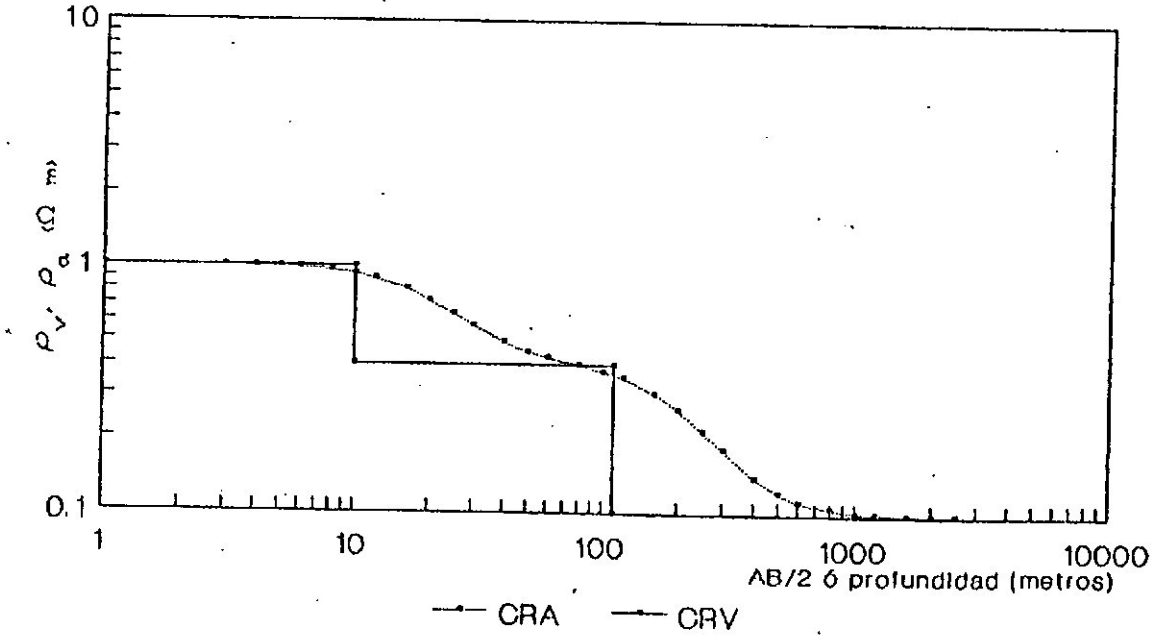


FIGURA 16 - Corte Geoeléctrico Tipo Q
$E_1 = 10$, $E_2 = 100$, $\rho_1 = 1$, $\rho_2 = 0.4$, $\rho_3 = 0.1$

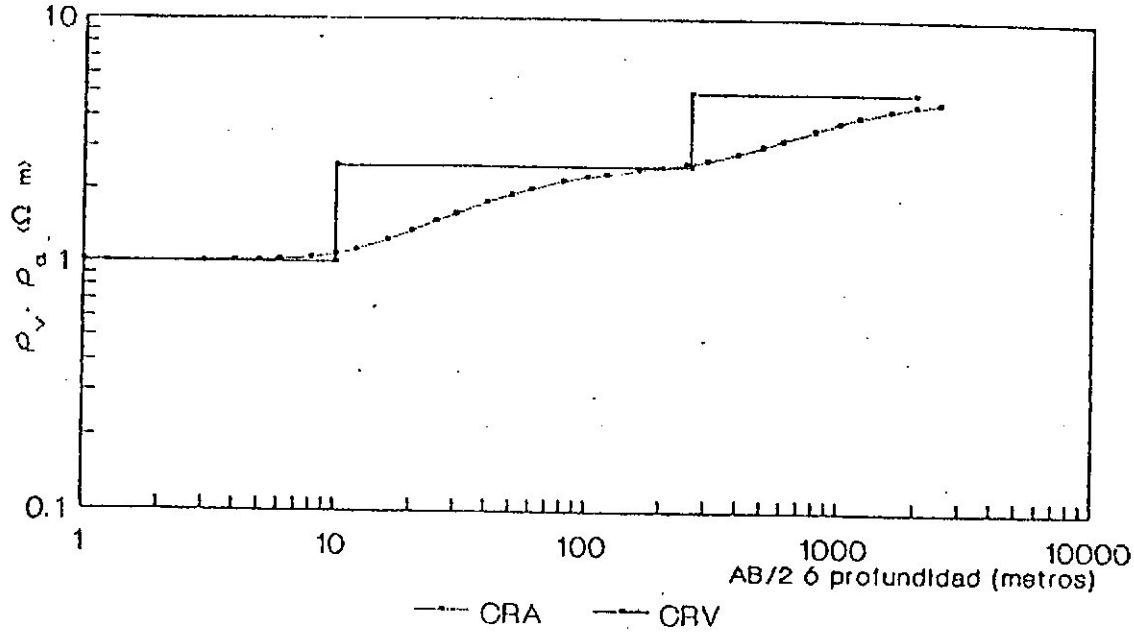


FIGURA 17 - Corte Geoeléctrico Tipo A
$E_1 = 10$, $E_2 = 250$, $\rho_1 = 1$, $\rho_2 = 2.5$, $\rho_3 = 5$

c)cortes de más de tres capas

Se designan mediante una combinación de los nombres utilizados para cortes y curvas de tres capas. Para ello se consideran las tres primeras capas y se le asigna la letra correspondiente, luego se toman la segunda, tercera y cuarta capa se clasifica y se agrega la letra correspondiente a la anterior. Se repite el proceso hasta agregar la letra correspondiente a las tres últimas capas.

Por ejemplo un corte de cinco capas con resistividades tales que $\rho_1 > \rho_2 < \rho_3 > \rho_4 > \rho_5$ será Tipo HKQ; Tipo H por ser $\rho_1 > \rho_2 < \rho_3$, Tipo K por ser $\rho_2 < \rho_3 > \rho_4$ y Tipo Q por ser $\rho_3 > \rho_4 > \rho_5$. (figura 18)

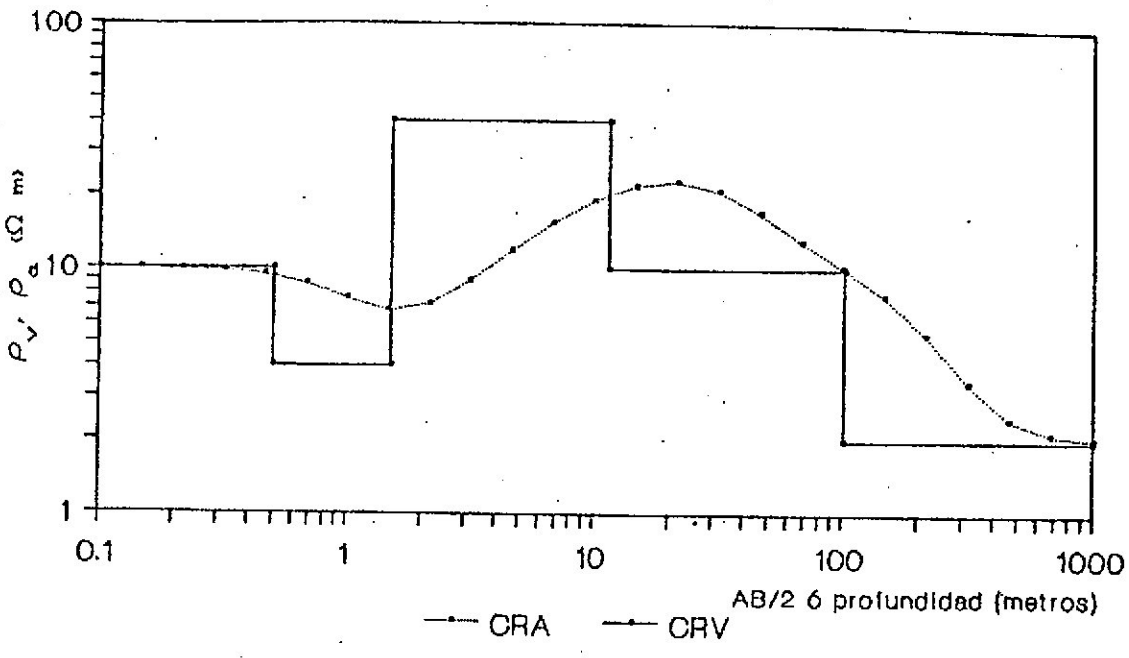


FIGURA 18 - Corte Geoeléctrico Tipo HKQ
$E_1 = 0.5$, $E_2 = 1$, $E_3 = 10$, $E_4 = 90$
$\rho_1 = 10$, $\rho_2 = 4$, $\rho_3 = 40$, $\rho_4 = 10$, $\rho_5 = 2$

CURVAS DE RESISTIVIDAD APARENTE

Como ya se mencionó, la CRA surge de representar los datos de resistividad aparente ρ_{ai} originados en la medición de un SEV en función de AB/2 en el caso de SEV Schlumberger o de a en curvas Wenner, en coordenadas logarítmicas. Esto no significa que la profundidad de investigación sea igual a AB/2 (ó a), o a una fracción constante de ella. Esta profundidad depende, no sólo de la separación entre electrodos sino fundamentalmente de la distribución de resistividades en el subsuelo.

La adopción de coordenadas logarítmicas está justificada por varias razones:

a) Las curvas presentan una forma más regular, de manera que las variaciones de la resistividad aparente tienden a mantener el orden de magnitud sobre toda la longitud de la curva.

b) Se pone de manifiesto que la detectabilidad de una capa disminuye con su profundidad de yacencia, es decir que su espesor debe ser tanto mayor cuanto más profunda se encuentre para que su presencia se refleje en la CRA.

c) Permite dar la misma importancia a las variaciones de resistividad de las capas conductivas que de las resistivas, es decir, variaciones de pequeña magnitud que pueden ser importantes en capas de baja resistividad, son despreciables en capas muy resistivas.

d) Esta representación es muy ventajosa para la interpretación por métodos gráficos, ya que en coordenadas logarítmicas la multiplicación por un factor cualquiera de todas las resistividades o de todos los espesores de las capas sólo provoca una traslación de la CRA, paralela a los ejes coordenados.

RESOLUCION DEL PROBLEMA DIRECTO

El Problema Directo en prospección geoeléctrica es aquél cuya resolución permite determinar la curva de resistividad aparente que le corresponde a un corte geoeléctrico determinado; a diferencia del Problema Inverso que consiste en encontrar un corte correspondiente a una CRA y que constituye el objetivo de la interpretación geoeléctrica.

Desde el punto de vista físico-matemático la resolución del problema directo implica conocer la distribución del potencial eléctrico en la superficie de un medio, en general heterogéneo, cuando por él se hace circular una corriente eléctrica.

Para su resolución, se considerará sólo el caso de un medio estratificado, compuesto por varias capas homogéneas e isótropas, de extensión lateral infinita y limitadas entre sí por planos horizontales y paralelos. Sobre éste, se ubica un medio de resistividad infinita que representa la atmósfera.

Dado que la corriente aplicada es estacionaria, entonces el campo eléctrico **E** será conservativo o irrotacional, por lo tanto deriva de un potencial V. Es decir que:

$$\mathbf{E} = -\nabla V \qquad (14)$$

donde ∇ es el operador gradiente.

Al ser los medios homogéneos este potencial verificará la ecuación de Laplace (ec. 15) en todo el semiespacio salvo en los electrodos y en las superficies de discontinuidad (contacto entre capas).

$$\nabla^2 V = 0 \qquad (15)$$

donde ∇^2 representa el laplaciano.

Por lo tanto el problema directo puede resolverse por integración de esta ecuación diferencial.

A partir de la expresión del potencial, podrá obtenerse el campo **E** por derivación y, conocido éste, para un dispositivo Schlumberger, la resistividad aparente se calcula mediante:

$$\rho_a = \pi r^2 \frac{\mathbf{E}}{I} \qquad (16)$$

No se pretende dar aquí un detalle de las formulaciones matemáticas necesarias para resolver el problema, por lo tanto se obviarán los pasos intermedios que conducen a la expresión de la resistividad aparente. Para el caso planteado, de un medio estratificado de n capas tiene la forma:

$$\rho_a(r) = \rho_1 r^2 \int_0^\infty N_n(\lambda)\, J_1(\lambda r)\, \lambda\, d\lambda \qquad (17)$$

donde λ es la variable de integración, J_1 es la función de Bessel de primera especie y orden 1, y $\mathbf{N}_n(\lambda)$ se denomina **Función Característica (FC)** y está determinada por las condiciones de

contorno del problema. Por lo tanto, N_n depende de las resistividades y espesores de las capas y representa al corte geoeléctrico considerado.

La (17) permite obtener la CRA que corresponde a un corte dado una vez determinada la función característica; este paso se realiza mediante fórmulas de recurrencia que van calculando su valor en el techo de cada capa desde el sustrato y hasta llegar a la superficie. El algoritmo correspondiente a Sunde es uno de los más utilizados y se lo puede encontrar en cualquier texto de Geoeléctrica.

La FC suele representarse en coordenadas logarítmicas en función de $1/\lambda$, un ejemplo para un corte de tres capas se muestra en la figura 19.

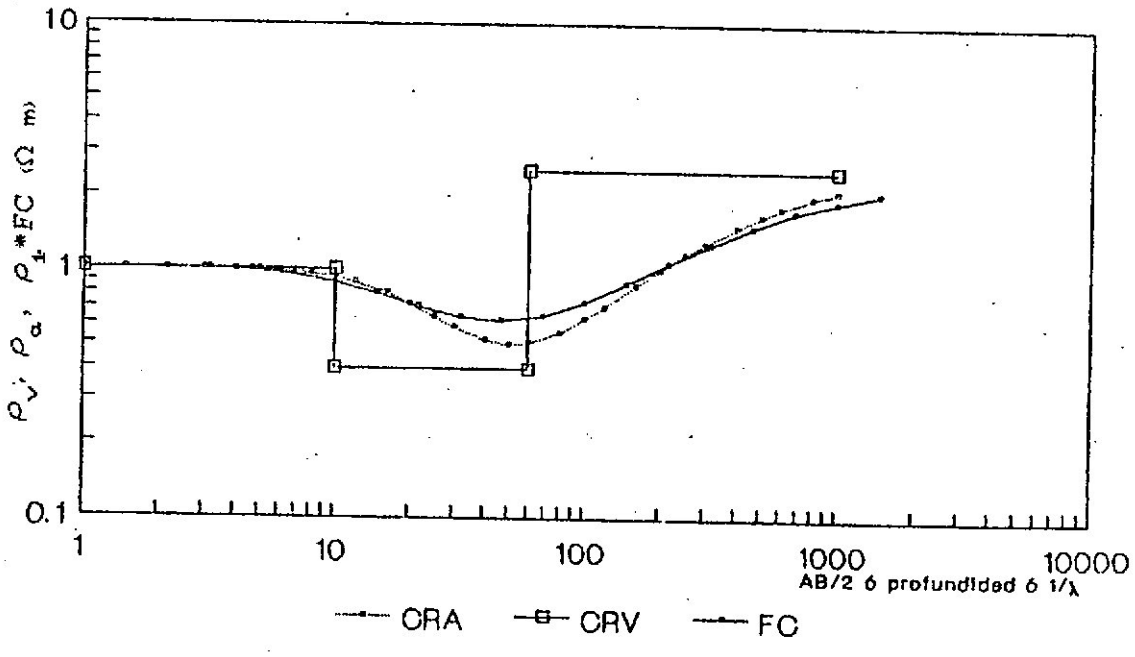


FIGURA 19 - CRV, CRA y Función Característica para un corte de tres capas Tipo H.
$E_1 = 10$, $E_2 = 50$, $\rho_1 = 1$, $\rho_2 = 0.4$, $\rho_3 = 2.5$

En cuanto al cálculo de la CRA, la integral de (17) no es fácil de resolver en forma analítica y se han propuesto varios métodos para hacerlo numéricamente. El de mayor uso en la

actualidad aplica la teoría de filtros digitales, está basado en una idea de Kunetz (1966) y fué desarrollado posteriormente por Ghosh (1971).

Si en la expresión integral se definen nuevas variables x e y de manera que $x = \ln(1/\lambda)$ e $y = \ln(r)$ la ecuación (17) se reduce a:

$$\rho_a(y) = \int_0^{\infty} f_1(x)\ f_2(y-x)\ dx \qquad (18)$$

donde $f_1(x) = \rho_1\ N_n(e^{-x})$ y $f_2(y-x) = J_1(e^{y-x})\ e^{2(y-x)}$

Una integral de este tipo se denomina de **convolución** y puede aproximarse mediante un **operador lineal** o **filtro**, siempre que la función considerada sea "regular", es decir curvas suaves y con variaciones de largo período. Las CRA cumplen con esta condición cuando se utilizan coordenadas logarítmicas.

Así, el valor de la resistividad aparente correspondiente a cada abscisa estará dado por:

$$(\rho_a)_m = \sum_{j=-\alpha}^{\beta} b_j\ f_1(m-j) \qquad (19)$$

donde m indica la absisa en la que se está evaluando la resistividad aparente, b_j representa los coeficientes del filtro digital, α y β dependen del diseño del filtro y f_1 es la función característica del corte geoeléctrico.

Desde la publicación de Ghosh hasta hoy, se encuentra en la literatura una gran cantidad de filtros obtenidos por distintos métodos que difieren entre si en el número de coeficientes y el intervalo de muestreo aplicado, lo que les confiere diferentes precisiones para el cálculo de la CRA.

PARAMETROS Y CURVAS DE DAR ZARROUK

Como se vió, una capa geoeléctrica queda perfectamente determinada si se dan su espesor y su resistividad. Otra manera de caracterizarla es a través de los denominados **parámetros de Dar Zarrouk** que se definen mediante las siguientes relaciones:

$$T = E\,\rho$$
$$S = E/\rho \qquad (20)$$

T recibe el nombre de **resistencia transversal unitaria** y S se denomina **conductancia longitudinal unitaria**.

Los parámetros T y S son aditivos por lo tanto a un conjunto de n capas le corresponde la suma de sus parámetros individuales, es decir:

$$T = \sum_{1}^{n} T_i$$
$$S = \sum_{1}^{n} S_i \qquad (21)$$

La principal utilidad de estos parámetros es que, dado un conjunto de capas geoeléctricas, se lo puede reemplazar por una sola capa homogénea e isótropa de resistividad ρ_m y espesor E_m tales que los parámetros de Dar Zarrouk del conjunto se mantengan constantes, es decir que se deben verificar las siguientes ecuaciones:

$$T = \rho_m E_m$$
$$S = \frac{Em}{\rho_m} \qquad (22)$$

E_m y ρ_m pueden denominarse pseudoespesor y resistividad media, respectivamente. De las anteriores se deduce que:

$$\rho_m = \sqrt{T/S}$$

$$E_m = \sqrt{TS} \qquad (23)$$

Hasta aquí se ha considerado que cada capa es una unidad indivisible, pero T y S pueden determinarse también para profundidades intermedias de manera que se pueden considerar funciones de la profundidad z.

De esta manera, la resistividad media ρ_m puede representarse como una función de la profundidad media o pseudoprofundidad z_m en coordenadas logarítmicas, dando como resultado una curva que se denomina Curva de Dar Zarrouk (CDZ) (figura 20).

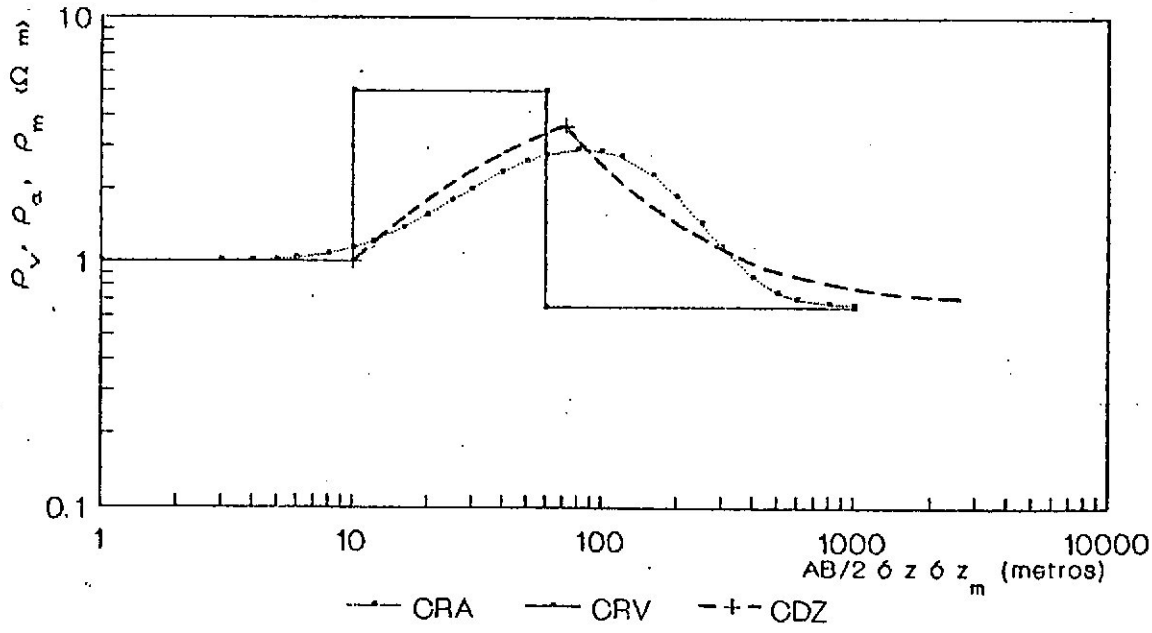


FIGURA 20 - CRV, CRA y Curva de Dar Zarrouk para un corte de tres capas Tipo K
$E_1 = 10$, $E_2 = 50$, $\rho_1 = 1$, $\rho_2 = 5$, $\rho_3 = 0.65$

Sus características principales son:

a) La CDZ se compone de un número de arcos igual al número de capas del corte geoeléctrico. Cada uno de ellos se puede obtener a partir de dos arcos fundamentales.

b) El primer arco de la curva coincide con el primer tramo de la CRV (semi-recta horizontal cuyo extremo derecho es el punto de coordenadas $z_m = E_1$, $\rho_m = \rho_1$).

c) El arco correspondiente a cada capa tiende asintóticamente a la resistividad de esa capa.

d) Los arcos ascendentes son cóncavos hacia abajo y los descendentes, hacia arriba.

e) Si una capa es perfectamente aislante ($\rho=\infty$) o conductora ($\rho=0$) el arco correspondiente es una semirecta de pendiente +1 ó -1, respectivamente.

CORTES EQUIVALENTES

Si bien, el método SEV no es ambiguo en teoría, es decir que a cada curva de resistividad aparente le corresponde un único corte geoeléctrico, esto no ocurre en la práctica.

Las curvas de campo están determinadas por una serie de puntos medidos y, por lo tanto, afectados de errores de observación, de manera que en lugar de una curva, lo que se obtiene es una serie de segmentos de error por los que pasan infinitas curvas figura 21. Un ejemplo de dos curvas de campo muy similares correspondientes a distintos cortes geoeléctricos se muestra en la figura 22.

Lo anterior conduce al concepto de **equivalencia**: "Se llaman **cortes equivalentes**" a aquellos que aunque difieran en los parámetros de sus capas e incluso en el número de éstas, tienen curvas de campo que difieren entre sí en menos del límite del error experimental." (Orellana y Hernandez, 1979)

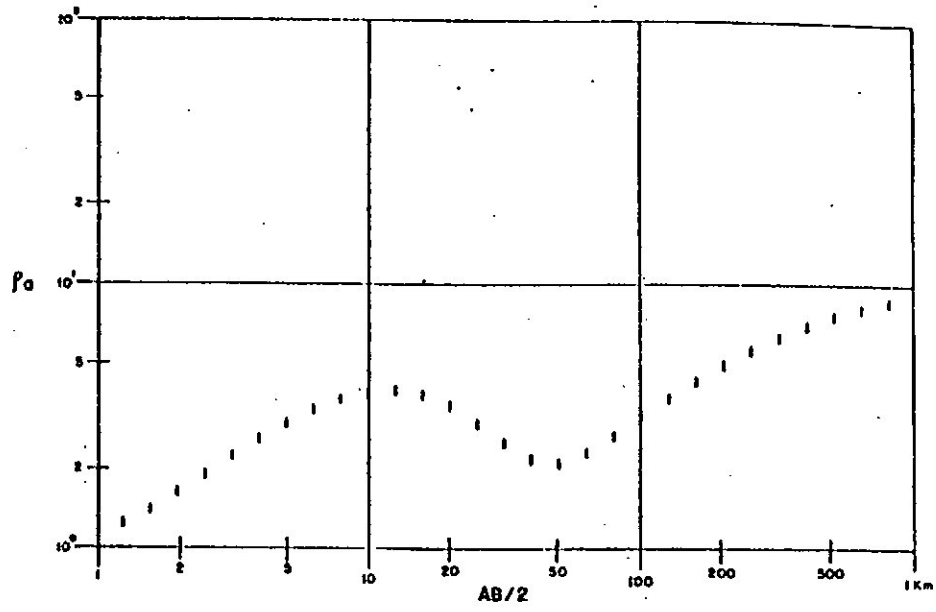


FIGURA 21 - Datos reales de observación en un SEV (tomado de Orellana, 1982)

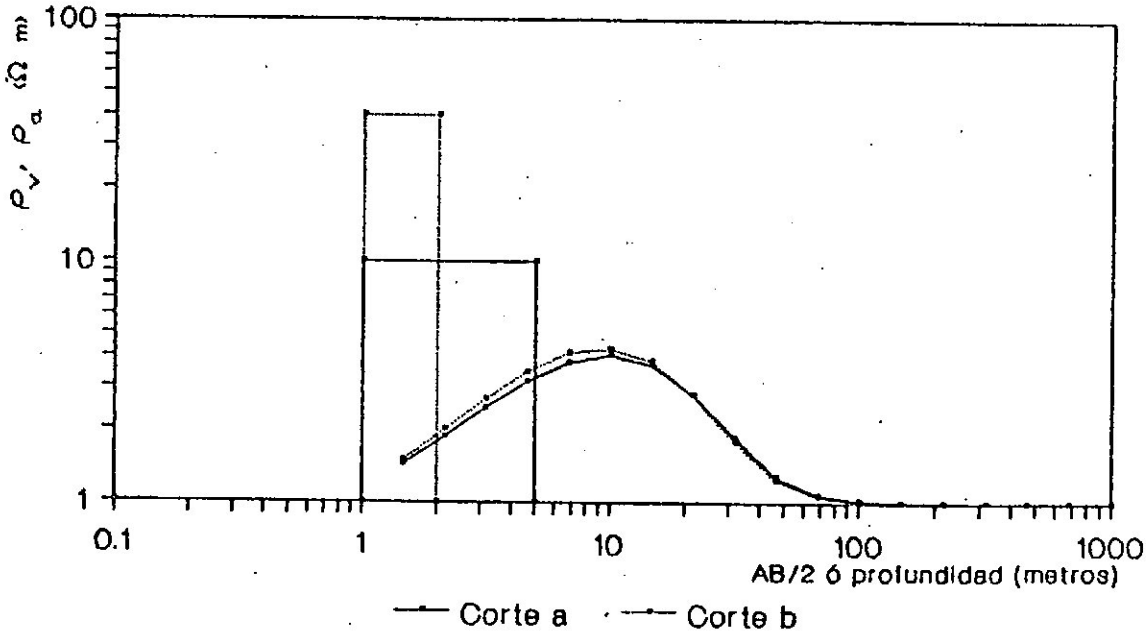


FIGURA 22 - CRV y CRA de dos cortes geoeléctricos Tipo K
(a) $E_1 = 1$, $E_2 = 4$, $\rho_1 = 1$, $\rho_2 = 10$, $\rho_3 = 1$
(b) $E_1 = 1$, $E_2 = 1$, $\rho_1 = 1$, $\rho_2 = 40$, $\rho_3 = 1$

Dos cortes equivalentes pueden tener CRV muy distintas, sin embargo sus CDZ serán muy similares, por lo tanto el dominio más adecuado para encontrar cortes equivalentes a uno dado es el de Dar Zarrouk.

METODOS DE INTERPRETACION DE CURVAS DE SEV

La finalidad de la interpretación de un SEV es determinar un modelo de la distribución vertical de resistividades en el subsuelo. Dada la "ambigüedad práctica" del método, dicho modelo deberá ser tal que, no sólo se ajuste a la curva de campo sino que también sea coherente con los SEV contiguos y con la información geológica disponible.

Existen varios métodos de interpretación que pueden clasificarse en métodos gráficos y numéricos, existiendo dentro de los segundos una amplia gama de posibilidades. Se describirán sólo algunos de ellos por ser los utilizados en los trabajos del CFI.

a) Métodos gráficos de superposición y reducción

El método de superposición se basa en la comparación de la curva que se desea interpretar con curvas teóricas de un catálogo. Su principal limitación consiste en que se requiere una colección de curvas patrón bastante amplia y, dado que la cantidad de parámetros variables aumenta con el número de capas es difícil contar con suficientes curvas teóricas como para realizar la interpretación.

En esos casos se utiliza el método de reducción, que consiste en sustituir varias capas por una sola capa ficticia, equivalente a ellas, permitiendo así el uso de curvas patrón de menor número de capas.

Entre los métodos propuestos, el más empleado es el de Ebert-Kalenov por ser el de mayor exactitud y eficacia.

El método gráfico tiene dos importantes restricciones, la primera se refiere al número de capas que pueden interpretarse ya que, si bien en teoría éste es ilimitado, prácticamente no tiene validez para casos de más de seis o siete capas. La segunda es que los cortes con capas muy delgadas no se prestan al uso de este método a menos que los contrastes de resistividad entre ellas sean importantes.

b) Métodos de aproximaciones sucesivas

Estos métodos requieren de un modelo inicial aproximado que puede obtenerse gráficamente o por estimación a partir de la información geológica.

El procedimiento más difundido es el propuesto por Johansen (1975). Se calcula la CRA correspondiente al modelo inicial y se la compara con la curva de campo, si no se obtiene un ajuste adecuado se modifica uno o más parámetros del modelo, se calcula una nueva CRA teórica y se vuelve a realizar la comparación. El proceso se repite hasta conseguir un corte geoeléctrico solución cuya CRA ajuste satisfactoriamente. El número de iteraciones necesarias para encontrar dicha solución depende de lo acertado del modelo inicial y de la experiencia del interpretador. Una limitación importante es que su aplicación en casos de muchas capas es lenta y dificultosa. La figura 23 muestra un diagrama del método de Johansen.

c) Métodos en los que se obtiene la CDZ

El más difundido es el propuesto por Zohdy (1975) que se basa en la similitud que, en muchos casos, presenta la CDZ con la CRA. Su principal ventaja es que permite obtener un corte geoeléctrico directamente a partir de la CRA sin necesidad

de contar con un modelo inicial.

La CRA de campo se ingresa muestreada a intervalos logarítmicos constantes, es decir que se deben interpolar entre los valores observados los correspondientes a ciertas abscisas (AB/2) predeterminadas.

Estos valores de resistividad aparente ($\rho_{a,o}$) se consideran como una primera aproximación de la CDZ es decir, se toman como si fuesen resistividades medias ($\rho_{m,1}$). Esto implica suponer que cada intervalo entre puntos muestreados consecutivos representa una capa cuyos parámetros pueden determinarse a partir de:

$$E_i\rho_i = T_i - T_{i-1} \qquad y \qquad E_i/\rho_i = S_i - S_{i-1} \tag{24}$$

donde el subíndice i representa la capa considerada.

Una vez determinados los parámetros del corte de esta manera, se calcula la CRA teórica ($\rho_{t,1}$) y se compara con la de campo.

La siguiente etapa corresponde al ajuste del modelo para lo cual se supone que un pequeño cambio de la CDZ en un punto producirá un cambio de la CRA de la misma magnitud, en el punto considerado.

Por lo tanto se calcula otra CDZ ($\rho_{m,2}$) de manera que:

$$\frac{\rho_{m,2}}{\rho_{m,1}} = \frac{\rho_{a,o}}{\rho_{t,1}} \tag{25}$$

Así, con la nueva CDZ se obtiene un nuevo corte geoeléctrico cuya CRA se vuelve a comparar con la observada. Si no hay ajuste entre éstas, se repite el proceso tantas veces como sea necesario.

Nótese que dadas las características del método el modelo resultante tendrá un número de capas igual al número de puntos de la CRA ingresados. Por lo tanto habrá que reducirlo para obtener un corte geoeléctrico con sentido geológico. Esta reducción se efectúa en el dominio de Dar Zarrouk.

La descripción anterior es muy simplificada y se han obviado algunos detalles tales como los casos en que la CRA no puede asimilarse a una CDZ; para un estudio más detallado se recomienda consultar la publicación de Zohdy ya citada. Las figura 24 muestra un diagrama del método.

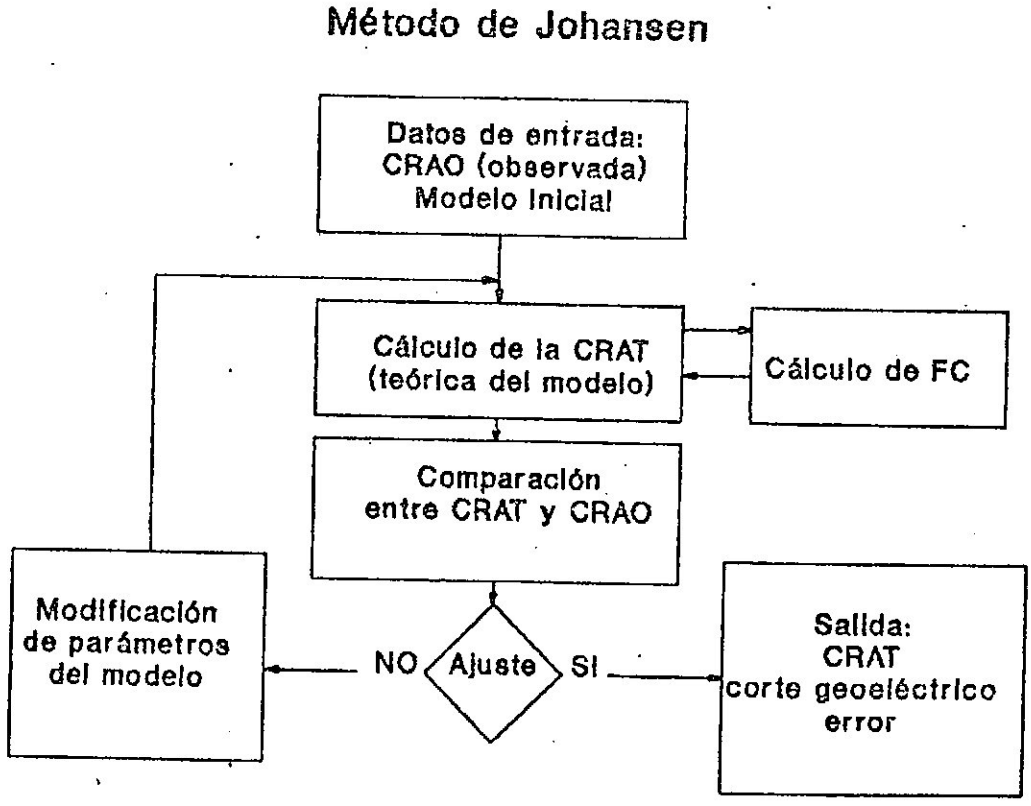


FIGURA 23 - Método de Johansen. Diagrama.

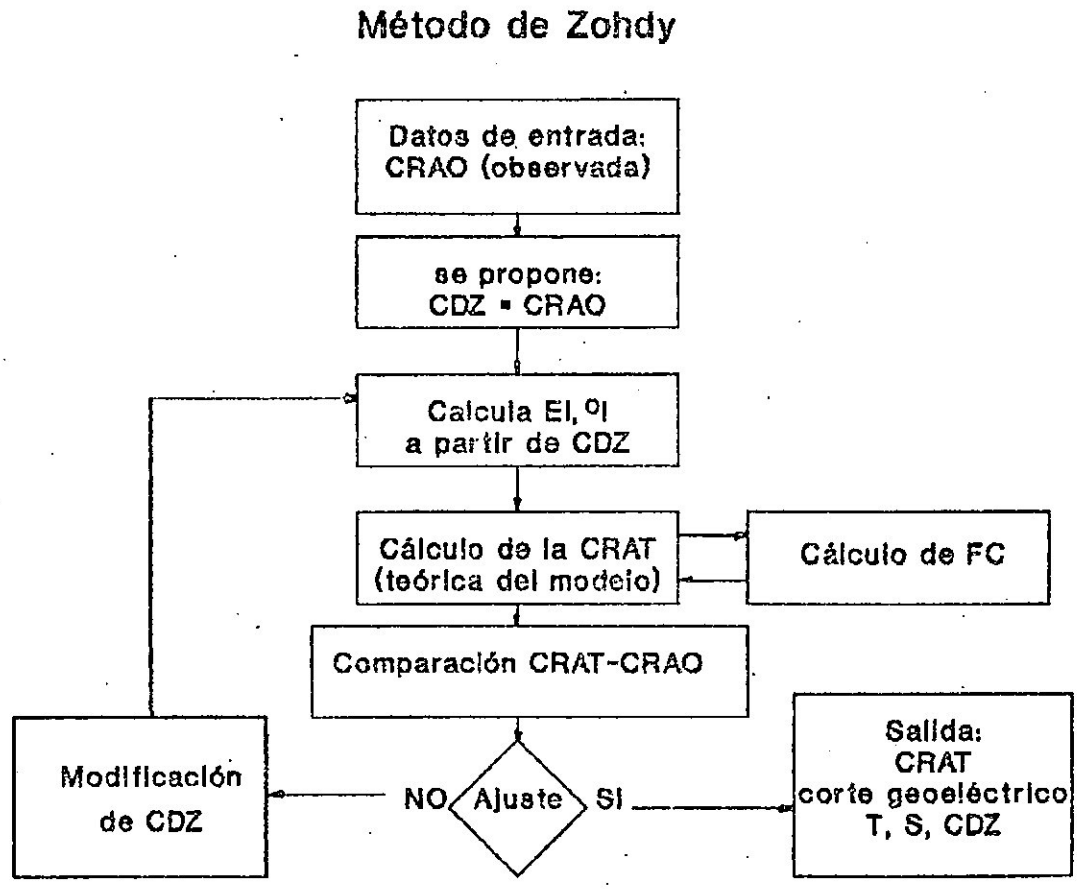


FIGURA 24 - Método de Zohdy. Diagrama.

4.- BIBLIOGRAFIA DE REFERENCIA

ASTIER, J.L.; 1975. Geofísica aplicada a la Hidrogeología. Paraninfo. Madrid.

BHATTACHARYA, P.K. y PATRA, H.P.; 1968. Direct Current Geoelectric Sounding. Elsevier. Amsterdam.

CANTOS FIGUEROLA, J.; 1973. Tratado de Geofísica Aplicada. Litoprint. Madrid.

DIAZ UCHA, E.L.; 1988. Interpretación Automática de Sondeos Eléctricos Verticales, Base de Datos y Aplicaciones. Tesis Doctoral. Universidad de Granada.

GHOSH, D.P.; 1971a. The Application of Linear Filter Theory to the Direct Interpretation of Geoelectrical Resistivity Sounding Measurements. Geophysical Prospecting, 19, 192-217.

GHOSH, D.P.; 1971b. Inverse Filter Coefficients for the Computation of Apparent Resistivity Standard Curves for a Horizontally Stratified Earth. Geophysical Prospecting, 19, 769-775.

JOHANSEN, H.K.; 1975. An Interactive Computer / Graphic-Display-Terminal System for Interpretation of Resistivity Soundings. Geophysical Prospecting, 23, 449-458.

KELLER, G.V. y FRISCHKNECHT, F.C.; 1966. Electrical Methods in Geophysical Prospecting. Pergamon Press. Oxford. Londres.

KOEFOED, O.; 1979. Geosounding Principles. 1.- Resistivity sounding measurements in Methods in Geochemestry and Geophysics. Vol. 14a. Elsevier. Amsterdam.

ORELLANA, E. y MOONEY, H.; 1966. Tablas y Curvas Patrón para Sondeos Eléctricos Verticales sobre terrenos estratificados. Interciencia. Madrid.

ORELLANA, E.; 1982. Prospección Geoeléctrica en corriente continua. Paraninfo. Madrid.

PARASNIS, D.S.; 1970. Principios de Geofísica Aplicada. Paraninfo. Madrid

PARASNIS, D.S.; 1971. Geofísica Minera. Paraninfo. Madrid.

ZOHDY, A.; 1974. Use of Dar Zarrouk Curves in the Interpretation of Vertical Electrical Sounding Data. Geological Survey Bulletin 1313-D.

ZOHDY, A.; 1975. Automatic Interpretation of Schlumberger Sounding Curves, Using Modified Dar Zarrouk Functions. Geological Survey Bulletin 1313-E.

ANEXO

Trabajos de Prospección Geoeléctrica realizados entre 1976 y 1990 por el Consejo Federal de Inversiones

Hace 14 años que la Prospección Geoeléctrica ocupa un espacio entre los multidisciplinarios equipos del CFI. En este lapso se realizaron numerosos estudios de asistencia a equipos provinciales, tanto en estudios hidrogeológicos como en estudios del subsuelo para la proyección de obras de ingeniería hidráulica y electricista, siendo la modalidad relevante su inserción en los estudios hidrogeológicos programados y ejecutados por CFI.

El método aplicado es preponderantemente el de Sondeo Eléctrico Vertical (SEV), aunque esporádicamente se ha empleado el de Calicata Eléctrica (CE).

Las tareas de programación, medición e interpretación son ejecutadas por los dos geofísicos del CFI, con participación de personal de planta en las tareas de dibujo, dactilografía y administración. En las mediciones de campo, los auxiliares necesarios son contratados en la zona de trabajo.

El instrumental utilizado, se adapta al modo de operación que impone su rápido traslado de un punto a otro del pais y en las zonas de trabajo, en cualquier clase de vehículo, incluyendo el transporte "a hombro" en terrenos inaccesibles por otros medios.

Entre las actividades realizadas deben contabilizarse también las de perfeccionamiento, que involucran la asistencia a cursos de post grado sobre temas relacionados con la actividad desarrollada y las de asesoramiento y divulgación, entre las que pueden mencionarse el dictado de cursos (Santiago del Estero, 1977 y La Pampa, 1979) y la presentación de trabajos en Reuniones Científicas.

En este anexo se enumeran los estudios realizados, con un breve resumen de cada uno. Cabe destacar que tres de ellos fueron se realizaron por Convenio con la Universidad Nacional de La Plata, con participación del Departamento de Geofísica Aplicada de la Facultad de Ciencias Astronómicas y Geofísicas.

RESUMEN DE LOS ESTUDIOS REALIZADOS

A. APLICADOS A LA GEOHIDROLOGIA

ESTUDIO GEOELECTRICO EN AREAS DE TINOGASTA Y FIAMBALA, PROVINCIA DE CATAMARCA. 1977.

Para el "Proyecto de Desarrollo de los recursos hídricos del Noroeste Argentino" (NOA-HIDRICO)

Se midieron 60 SEV Schlumberger de reconocimiento, de hasta 2000 m de longitud, con la finalidad de obtener un primer modelo de la geometría y las características litológicas de la cubierta cuartaria.

INFORME DE LA INTERPRETACION DE LOS SONDEOS GEOELECTRICOS MEDIDOS EN EL AREA DE SAN CARLOS, PROVINCIA DE SALTA. 1977.

Para el Proyecto NOA-HIDRICO

Interpretación e informe sobre 58 SEV Schlumberger, de 400 a 800 m de longitud, medidos por técnicos de la Universidad Nacional de Salta (UNSA), con la finalidad de evaluar el uso del agua subterránea en obras de riego.

La interpretación se realizó por el método de comparación con curvas patrón, presentándose los resultados mediante un mapa de resistividad y cinco secciones geoeléctricas.

INFORME DE LA INTERPRETACION DE LOS SONDEOS GEOELECTRICOS MEDIDOS EN AREA DE LA QUEBRADA DE HUMAHUACA. PROVINCIA DE JUJUY. 1977.

Para el Proyecto NOA-HIDRICO

Por el método de comparación con curvas patrón se interpretaron 42 SEV medidos por la Cátedra de Hidrogeología de la UNSA a lo largo del lecho del río Grande, entre Rodero y el Angosto de Perchel y algunas quebradas laterales, con la finalidad de aproximar el espesor de los sedimentos aluviales para la evaluación de las posibilidades de riego de las distintas áreas involucradas.

Los resultados se presentaron en secciones geoeléctricas distribuidas por áreas.

APLICACION DE SEV EN LA PROSPECCION HIDROGEOLOGICA DEL ACUIFERO TOAY - ANGUIL. 1979.

Para la Dirección de Recursos Hídricos de La Pampa.

Entre las tareas de prospección hidrogeológica realizadas por la provincia con la finalidad de detectar y cuantificar los

sectores de mayor aptitud del acuífero subterráneo para la provisión de agua potable a la ciudad de Santa Rosa, se midieron 106 SEV, en un area de 350 km^2.

La interpretación se realizó por el programa de Zodhy, presentándose los resultados en secciones geoeléctricas correlacionadas con la información geológica proveniente de parforaciones existentes. Estas secciones proporcionan información sobre: profundidad del basamento cristalino (del orden de los 200 m), profundidad de la zona de transición agua dulce - agua salada y espesor de la cubierta arenosa que cubre prácticamente toda el área.

ESTUDIO GEOELECTRICO CHACHARRAMENDI-LIMAY MAHUIDA, PROVINCIA DE LA PAMPA. 1980.

Para la Dirección de Recursos Hídricos de La Pampa

En base a 172 SEV Schlumberger, de longitud variable entre 100 y 1000 m, con un promedio de 527 m, interpretados por el programa de Zohdy, se obtuvieron cuatro secciones geoeléctricas en el sector comprendido entre Chacharramendi, Limay Mahuida y La Reforma, poblaciones del centro de la Provincia de La Pampa, que definen un área triangular de aproximadamente 1700 km^2.

Las secciones permiten estimar la profundidad de un basamento resistivo, cónstituido por rocas metamórficas y graníticas, el que aflora en diversos puntos del borde de la zona en cuyo interior se mantiene, aparentemente, a una profundidad del orden de los 50 m.

Los valores de resistividad obtenidos para la zona saturada, conformada principalmente por sedimentos limo-arenosos a limo-arcillosos, desalientan espectativas sobre la existencia de agua con menos de 2000 ppm, mínimo de salinidad en las cercanías de Chacharramendi. Por el contrario, los mínimos resistivos, con $\rho < 1$ Ω.m, correspondiente a capas suprabasamentales, obedecerían a aguas con mas de 30000 ppm.

MEDICIONES GEOELECTRICAS PARA EL ESTUDIO HIDROGEOLOGICO EN LA ZONA DE TRANCAS, PROVINCIA DE TUCUMAN. 1982.

Para la Dirección Provincial del Agua.

Dada la escasa información de subsuelo, se midieron en el área de interés 79 SEV Schlumberger, en base a los que se elaboró un modelo geoeléctrico simple, de a lo sumo cuatro capas, fácilmente correlacionable con los perfiles litológicos disponibles.

Este modelo permitió definir la geometría del complejo acuífero regional, identificado con las dos capas superiores (con $\rho > 20$ Ω.m), apoyado sobre un acuitardo acuícludo constituido por facies terciarias predominantemente samíticas y pelíticas, aflorantes en la zona de estudio, e identificado con resistividades inferiores a 20 Ω.m.

PROSPECCION GEOELECTRICA EN EL VALLE DE CATAMARCA. 1980-1983

Para la Dirección de Aguas Subterráneas de la Provincia de Catamarca.

Mediante 362 SEV Schlumberger de 1546 m de longitud promedio, se obtuvo un modelo del Valle de Catamarca basado en la distribución vertical de la resistividad, el que es utilizado por la Dirección de Hidráulica de la Provincia en prospección hidrogeológica orientada a satisfacer demandas de agua para uso humano, agrícola e industrial.

El modelo se presenta mediante secciones geoeléctricas, mapa de isopacas del relleno sedimentario, mapa de isolíneas de la resistividad transversal de los espesores resistivos (con $\rho > 20$ Ω.m) y mapa de secciones reducidas.

De acuerdo con él, las condiciones hidrogeológicas más favorables se dan en las áreas que son dominio de los principales cursos de agua, las que conforme se avanza hacia el sur, quedan prácticamente restringidas a los cauces actuales y antiguos del principal colector del Valle.

ESTUDIO GEOELECTRICO EN COLONIA HUACO, DTO. ANDALGALA, PROVINCIA DE CATAMARCA. 1983.

Para las Direcciones de Hidráulica y de Colonización de la Provincia de Catamarca.

Dirigido a la evaluación y caracterización del recurso hídrico subterráneo en el área de emplazamiento futuro de la Colonia Huaco, 20 km al SO de la ciudad de Andalgalá, se midieron 16 SEV de entre 1000 y 2000 m de longitud a lo largo de tres picadas de orientación NO-SE.

La interpretación se efectuó por el programa de Zohdy y sus resultados se presentaron en base a cinco secciones geoeléctricas y un mapa de isobatas del piso conductor (predominantemente arcilloso).

INFORME PRELIMINAR DE LAS MEDICIONES GEOELECTRICAS EN EL CAMPO DEL PUCARA, PROV. DE CATAMARCA. 1981-1983.

Para la Dirección de Hidráulica de la Provincia de Catamarca

En una región sobre la que hasta ese momento se carecía de datos de subsuelo, se midieron 49 SEV cuya finalidad fue la de proporcionar un modelo geológico preliminar tendiente a facilitar las subsiguientes tareas de exploración.

En la interpretación se utilizó el programa de Zohdy, expresándose los resultados en secciones geoeléctricas y un mapa de la Resistencia Transversal de los horizontes con resistividad mayor que 10 Ω.m.

DETERMINACION DE LA RESISTIVIDAD Y EL ESPESOR, CON FINES HIDROGEOLOGICOS, DEL RELLENO SEDIMENTARIO EN EL AREA RURAL DE BELEN, PROVINCIA DE CATAMARCA. 1984.

Para la Dirección de Aguas Subterráneas de Catamarca.

Mediante 28 SEV Schlumberger se determinaron las variaciones de resistividad y espesor del relleno sedimentario en la zona rural de Belén, a lo largo de siete perfiles. En base al modelo geoeléctrico obtenido, se calcularon los valores más probables del espesor de la columna sedimentaria (profundidad del basamento), proponiéndose hipótesis sobre su constitución granulométrica y la profundidad a que se encontraría el nivel freático, desalentadoras las últimas respecto al posible uso para riego del agua subterránea.

Este trabajo complementa la escasa información hidrogeológica disponible por la Dirección de Hidráulica de la provincia, basada en un muy reducido número de perforaciones poco profundas, en un ámbito donde éstas son difíciles de realizar por la existencia de bloques y rodados grandes en la masa sedimentaria del valle.

ESTUDIO DEL SUBALVEO DEL VALLE DEL RIO DESEADO. SECTOR PICO TRUNCADO, PROVINCIA DE SANTA CRUZ. 1984.

Para "Provisión de agua a Pico Truncado". CFI-SPSE.

Con el propósito de definir los espesores del depósito aluvional del valle y sus características resistivas, se midieron 44 SEV distribuidos en cuatro perfiles transversales al eje del valle. El dispositivo fue el de Schlumberger con separación máxima entre electrodos de corriente de 250 a 500 m.

Se definieron zonas de interés a partir de mapas de T y secciones geoeléctricas que permitieron obtener la geometría del reservorio y contribuyeron al desarrollo del modelo hidroestratigráfico.

MEDICIONES GEOELECTRICAS EN LA CUENCA SUPERIOR DEL RIO TURBIO, PROVINCIA DE SANTA CRUZ. 1986.

Para "Estudio Geohidrológico en la cuenca superior del rio Turbio". CFI-SPSE.

Realizado con el apoyo logístico de Servicios Públicos S.E. de la provincia, consiste en la medición de 19 SEV Schlumberger de 1000 m de longitud, cuya finalidad es la de calcular los espesores del subálveo del valle del Río Turbio entre Ea. La Primavera y 28 de Noviembre.

Los resultados se expresan en dos perfiles longitudinales y tres transversales al valle, que muestran valores máximos de 60 m en el sector norte, de 100 m en el sector de la toma del acueducto a 28 de Noviembre y de 50 m entre Julia Duffour y 28 de Noviembre.

PROSPECCION GEOELECTRICA DEL VALLE DE SANTA MARIA. PROVINCIA DE CATAMARCA, 1985-1987.

Para "Remodelación de las obras de riego de Santa María". CFI - Dirección de Hidráulica provincial.

Ante la dificultad de ampliar el conocimiento del subsuelo del valle por métodos directos y para extender la información obtenida de perforaciones existentes se midieron 95 SEV Schlumberger según 12 perfiles transversales al valle, entre Fuerte Quemado y Punta Balasto.

Las secciones geoeléctricas permiten obtener un modelo estimativo de la geometría del valle, pese a que son pocos los SEV cuya penetración permite el cálculo de la profundidad del basamento.

Por otra parte, las correlaciones posibles con los datos existentes son descritas con detalle para cada uno de los perfiles.

ESTUDIO DEL SUBALVEO DEL VALLE DEL RIO DESEADO, SECTOR LAS HERAS, PROVINCIA DE SANTA CRUZ. 1988.

Para "Provisión de agua potable a Las Heras". CFI-SPSE.

Se midieron en total 36 SEV Schlumberger de entre 250 y 500 m de extensión que contribuyeron a definir las condiciones geohidrológicas generales del subálveo del valle, facilitando la delimitación de áreas propicias para su explotación y la elaboración del anteproyecto de captación.

PROSPECCION GEOELECTRICA EN EL QUIMILO, PROVINCIA DE CATAMARCA. 1988.

Para la Secretaría de Ciencia y Técnica de la Provincia de Catamarca (SECYTCA).

Para evaluar la posibilidad de satisfacer la demanda de agua potable y para ganadería de la población El Quimilo y su zona de influencia ubicada al sur del Departamento La Paz, al borde de las Salinas Grandes, se desarrolló un programa de medición de la resistividad del subsuelo en base a 15 SEV distribuidos en el área.

Los valores obtenidos son menores que 1 Ωm, salvo en las reducidas áreas medanosas, cuestión que desalienta la posibilidad de existencia de agua subterránea explotable con los fines previstos.

PROSPECCION GEOELECTRICA EN EL AREA DE RIEGO DE MICHIHUAO, PROVINCIA DE NEUQUEN. 1988

Departamento de Geofísica Aplicada de la Facultad de Ciencias Astronómicas y Geofísicas.
Convenio de Cooperación Horizontal CFI-UNLP

Para "Anteproyecto preliminar para el desarrollo del área de riego de Michihuao"

Se midieron 154 SEV en un área de 500 km^2 y se re-interpretaron 80 SEV del Centro Regional de Aguas Subterráneas (CRAS).

Su análisis mediante un programa interactivo, semiautomático, permitió determinar el techo de los sedimentos cretácicos y las variaciones del relleno cuaternario, con la finalidad de proporcionar información sobre las condiciones de drenaje ante una eventual incorporación de riego .

Los resultados se expresan mediante 26 perfiles geoeléctrico y 6 mapas de distribución de resistividades a distintas profundidades, incluida la superficie del terreno.

DETERMINACION DEL ESPESOR DE LA CAPA PREPONDERANTEMENTE SEFITICA EN EL CENTRO URBANO DE CALETA OLIVIA, PROV DE SANTA CRUZ. 1988.

Para "Anegamiento urbano, manejo del acuífero freático en Caleta Olivia" (en elaboración) CFI-SPSE.

En base a 22 SEV cortos, al perfil litológico de dos perforaciones de reconocimiento y a una red freatimétrica, se realizó el cálculo del volumen de agua mínimo que provoca situaciones de anegamiento en el centro urbano de Caleta Olivia.

Ello implica una estimación de la profundidad de la base del freático, asi como de los límites de variación de su superficie.

ESTUDIO GEOELECTRICO DEL ACUIFERO COSTERO COMPRENDIDO ENTRE FARO PUNTA MEDANOS Y FARO QUERANDI. 1988.

Para "Evaluación del Recurso Hídrico Subterráneo de la Región Costera Atlántica Bonaerense. Región II. CFI - Provincia de Buenos Aires.

Departamento de Geofísica Aplicada de la Facultad de Ciencias Astronómicas y Geofísicas.
Convenio de Cooperación Horizontal CFI-UNLP

Como parte de las tareas previstas en el Convenio de Cooperación suscripto entre el CFI y la UNLP, se realizó el estudio Geoeléctrico de una franja costera de 75 km de largo y 4 km de ancho, que se extiende entre Faro Punta Médanos y Faro Querandí.

El mismo comprendió la medición de 91 SEV distribuidos en 16 perfiles transversales a la costa y tuvo como objetivo la

determinación de la geometría del acuífero costero mediante la definición del basamento hidrogeológico. Los resultados se presentan en secciones geoeléctricas y un mapa de isobatas del sustrato conductor.

CALCULO DE LA PROFUNDIDAD DEL BASAMENTO PORFIRICO EN PAMPA DE LA COMPAÑIA. PROVINCIA DE SANTA CRUZ. 1989

Para "Provisión de Agua a San Julián". CFI-SPSE

Entre las tareas de exploración hidrogeológica que se realizan para la provisión de agua a la ciudad de Puerto San Julián, se efectuaron determinaciones de la resistividad del terreno en 36 puntos mediante SEV Schlumberger de hasta 3000 m de longitud.

Los resultados, expresión de un modelo apoyado en perfiles litológicos y eléctricos de perforaciones que atraviesan parcialmente los horizontes sedimentarios, se presentaron mediante secciones geoeléctricas y mapas de espesores de los horizontes explorados, apoyados sobre rocas volcánicas y piroclásticas del Grupo Bahía Laura, las que constituyen el Basamento Hidrogeológico de la región.

PROSPECCION GEOELECTRICA DEL VALLE DEL FARIÑANGO, PROVINCIA DE CATAMARCA. 1989.

Para la Dirección de Hidráulica de la Provincia de Catamarca

Con la finalidad de evaluar las reservas hídricas subterráneas de este Valle, situado inmediatamente al N de la capital provincial, y planificar su explotación para consumo humano, se midieron 49 SEV distribuidos en seis picadas transversales al valle, en el marco de un programa de exploración geohidrológica de la Dirección de Hidráulica de la Provincia.

Se obtuvo asi una estimación de los espesores sedimentarios cuyo piso está constituido por un basamento cristalino precámbrico-paleozoico.

MEDICIONES GEOELECTRICAS EN EL SUDOESTE DE LA PROVINCIA DEL CHACO. 1987 - 1989.

Para "Estudio de fuentes para la provisión de agua potable a once localidades del sur de la Provincia del Chaco". CFI-Dirección de Hidráulica del Chaco.

Entre junio de 1987 y noviembre de 1988 se midieron 190 SEV Schlumberger distribuidos entre las localidades de Haumonia, Horquilla, La Sabana, Charadai, Cote Lai, Hermoso Campo, Gancedo y Santa Silvina. En estas últimas se midieron además calicatas eléctricas.

El objetivo de tales mediciones fue el de complementar los estudios geohidrológicos realizados en todas estas localidades con la finalidad de evaluar las posibilidades de explotación del agua subterránea para su uso en la dotación de agua potable a sus

poblaciones.

El análisis de los SEV se realizó aplicando el programa de Zohdy, obteniéndose secciones geoeléctricas, apoyadas algunas en perforaciones de reconocimiento.

Las resistividades encontradas son muy bajas en las localidades del departamento Tapenagá, en las que sólo capas muy superficiales superan los 10 Ω.m. La resistividad disminuye con la profundidad por efecto del incremento en la proporción de sedimentos pelíticos y el aumento de la salinidad del agua.

Hacia el oeste, en Hermoso Campo, Santa Sylvina y Gancedo, especialmente en esta última, situada cerca del límite con Santiago del Estero, las condiciones son algo más promisorias. No obstante, tampoco en ellas se puede dejar de lado la proyección de zonas de recarga por represas poco profundas, adecuadamente proyectadas (su excavación en exceso puede llevar a su deterioro por invasión de aguas salinizadas). En estos casos, resultaron de utilidad los mapeos eléctricos, de sectores previamente seleccionados, mediante calicatas eléctricas con AB = 10, 20 y 40m.

PROSPECCION GEOELECTRICA EN PAMPA ALTA, PROVINCIA DE SANTA CRUZ. 1989.

Para "Provisión de agua a Puerto Deseado". CFI-SPSE

Integrando los estudios hidrogeológicos en el norte de la provincia, emprendidos por CFI y SPSE, se realizó prospección geoeléctrica mediante la medición de 106 SEV de hasta 2000 m de longitud en una zona aledaña a Puerto Deseado y a lo largo de 100 km sobre la ruta 281, entre Fitz Roy y Tellier.

El objetivo fue determinar las características resistivas del espesor sedimentario y definir la profundidad del techo del Grupo Bahía Laura, base del sistema hidrogeológico.

La interpretación, basada en un modelo inicial obtenido mediante el método de Zohdy, conduce a secciones geoeléctricas que permiten identificar capas que se correlacionan con las gravas arenosas de los sedimentos terciarios de la Fm Patagonia y la Fm Sarmiento, y con el basamento porfírico jurásico.

MEDICIONES GEOELECTRICAS DE RECONOCIMIENTO PARA EL ESTUDIO DE FUENTES EN JUNIN DE LOS ANDES, PROVINCIA DE NEUQUEN. 1990

Para el Ente Provincial de Agua y Saneamiento (EPAS)

Como parte del estudio geohidrológico del valle del río Chimehuín, aguas arriba de Junín de Los Andes, se midieron 27 SEV Schlumberger de 500 m de longitud máxima, con la finalidad de obtener información hidrogeológica del subsuelo.

La interpretación se efectuó por el programa de Zohdy, presentándose los resultados en ocho secciones geoeléctricas esquemáticas y un mapa se isopacas de las capas con resistividad mayor que 100 Ω.m, en cuya base se elaboró un modelo esquemático de la cuenca subterránea.

PROSPECCION GEOELECTRICA EN EL SUDESTE DE LA PROVINCIA DE LA PAMPA. 1990.

Para la Dirección de Hidrología de la Pampa.

Con la finalidad de estimar las variaciones de profundidad del basamento hidrogeológico en el sudeste pampeano, se midieron 77 SEV de longitudes variables entre 500 y 2000 m, interpretados por el programa de Zohdy. Los cortes geoeléctricos obtenidos muestran :

.Una capa resistiva superior relacionada con la cubierta de arena y bancos de tosca presentes en el área.

.Una capa de resistividad intermedia que representa sedimentos más finos que los anteriores así como una posible variación en la salinidad del agua.

.Un sustrato resistivo asociado al basamento cristalino, aflorante al oeste del área, que se profundiza a partir de la ruta 154, presentando una estructura de bloques de distintas profundidades.

PROSPECCION GEOELECTRICA EN LA COSTA ATLANTICA DE LA PROVINCIA DE BUENOS AIRES ENTRE PUNTA RASA Y PUNTA MEDANOS. 1986-1990

Departamento de Geofísica Aplicada de la Facultad de Ciencias Astronómicas y Geofísicas.
Convenio de Cooperación Horizontal CFI-UNLP

Para "Evaluación del Recurso Hídrico Subterráneo de la Región Costera Atlántica Bonaerense. Región I. CFI - Provincia de Buenos Aires.

Efectuado en virtud del convenio de cooperación con la U.N.L.P,constituye un capítulo del estudio del acuífero costero. Se basa en 114 SEV distribuidos en perfiles transversales a la costa, un perfil longitudinal y dos mapas de isolíneas de resistencia transversal del acuífero frático y del semiconfinado.

El proceso de interpretación empleado fue el de Ebert-Kalenov, controlado mediante un programa interactivo que resuelve el problema directo por convolución con un filtro de 29 coeficientes. En el ajuste final se aprovechó toda la información obtenida en el estudio geohidrológico lográndose asi un buen modelo del complejo sistema acuífero.

PROSPECCION GEOELECTRICA EN LA COSTA ATLANTICA DE LA PROVINCIA DE BUENOS AIRES ENTRE FARO QUERANDI Y MAR DE COBO. INFORME PRELIMINAR.1990

Para "Evaluación del Recurso Hídrico Subterráneo de la Región Costera Atlántica Bonaerense. Región III. CFI-Provincia de Buenos Aires.

Se midieron 61 SEV Schlumberger con máxima separación entre

electrodos de corriente variable entre 100 y 1000 metros, a lo largo de 9 perfiles transversales a la linea de costa y distribuidos en una franja litoral de aproximadamente 46 km de longitud.

El objetivo del trabajo fue aportar elementos de juicio para los estudios Geológico-geomorfológico e Hidrogeológico que, en el marco de un convenio de cooperación, desarrolla el Centro de Geología de Costas de la Universidad Nacional de Mar del Plata.

La interpretación se realizó por el método de aproximaciones sucesivas, a partir de un modelo inicial obtenido gráficamente. Los resultados se presentan en secciones geoeléctricas en las que se diferencia un espesor superior resistivo asociado a las arenas que conforman el cordón medanoso costero, superpuesto a un sustrato conductivo que representa el basamento hidrogeológico en el área.

B. APLICADOS A PROYECTOS DE INGENIERIA HIDRAULICA

APLICACION DE TECNICAS GEOELECTRICAS EN EL ESTUDIO DE PREFACTIBILIDAD PARA LA CONSTRUCCION DE UN DIQUE EN EL ANGOSTO DE ANDALUCA, PROVINCIA DE CATAMARCA. 1981.

Para el Comité de Cuenca Hídrica del rio Abaucán - Colorado-Salado (CCHACS)

La estrechez del Angosto de Andaluca (la distancia mínima entre bloques opuestos es de 600 metros) indujo al CCHACS a la realización de estudios de prefactibilidad para la construcción de un dique de embalse, entre los que se encuentra el de geoeléctrica realizado por CFI, cuyo objetivo fue calcular el espesor del relleno sedimentario en la zona del proyecto.

Se midieron 11 SEV Schlumberger de 160 a 800 metros de longitud, dos de ellos paramétricos, y una doble calicata eléctrica simétrica con AB = 32 y 100 metros.

Interpretados los SEV por Zohdy se obtuvo una sección geoeléctrica aproximadamente perpendicular al eje del valle mostrando la profundidad calculada para el basamento granítico y perfiles de ρ_a por calicata, congruentes con la sección, obteniéndose una profundidad máxima de 65 metros.

MEDICIONES GEOELECTRICAS EN EL ARROYO COLORADO, DEPARTAMENTOS YAVI Y COCHINOCA, PROVINCIA DE JUJUY. 1986.

Para la Dirección de Hidráulica de Jujuy.

Se midieron cinco SEV Schlumberger en la zona de ubicación de un muro de afloramiento existente en el arroyo Colorado, donde la Dirección de Hidráulica de Jujuy necesitaba un mejor conocimiento del subálveo para realizar modificaciones en un proyecto de ingeniería conducente a la construcción de un dique de embalse elaborado por el Proyecto NOA-HIDRICO en 1980 y en el que ya existia un estudio similar, del que las mediciones realizadas fueron simplemente complementarias.

INFORME SOBRE UNA SECCION GEOELECTRICA EN EL RIO SANTA RITA, DTO. SANTA BARBARA, PROV. DE JUJUY. 1986

Para la Dirección de Hidráulica de Jujuy

La sección se obtuvo en base a 5 SEV medidos según una línea transversal al río Santa Rita, cerca de Palma Sola, con la finalidad de evaluar el espesor del subálveo en un sector donde la Dirección Provincial de Hidráulica de Jujuy proyectó la construcción de un dique derivador.

RESULTADO DE LAS MEDICIONES GEOELECTRICAS EN LA SALIDA DE LA QUEBRADA DE RIO GRANDE, DTO. TINOGASTA, PROV. DE CATAMARCA. 1986.

Para la Dirección de Hidráulica de Catamarca.

En diciembre de 1986, se midieron tres SEV en la salida de la quebrada de Rio Grande, con la finalidad de calcular el espesor del relleno sedimentario en una sección transversal al eje de la quebrada. El estudio se integró al anteproyecto de dique nivelador destinado al aprovechamiento de las aguas del río en la irrigación de aproximadamente 1200 Has.

Pese a lo angosto de la quebrada (poco menos de 100 m), se aplicó la metodología de capas horizontales, obteniéndose una estimación de la profundidad del basamento constituido por esquistos micáceos (muy resistivo).

C. APLICADOS A INGENIERIA ELECTRICISTA

DETERMINACION DE LA RESISTIVIDAD DEL SUELO EN LAS ESTACIONES TRANSFORMADORAS DE PASO DE LA PATRIA, LA CRUZ, GOYA Y SALADAS, PROVINCIA DE CORRIENTES. 1989.

Para la Dirección Provincial de la Energía de la Provincia de Corrientes.

Comprende este trabajo el análisis de los "Estudios de resistividad eléctrica" y de los "Estudios de suelos", efectuados por un contratista, para los Anteproyectos Definitivos de Ingeniería para la construcción de las estaciones transformadoras del sector electrico de la provincia de Corrientes.

Al efecto, se efectuaron mediciones de resistividad en los puntos mencionados en el encabezamiento, las que sirvieron de referencia para aconsejar un cambio en la metodología utilizada.

ANALISIS DE LA VARIANZA

Jorge Luis Fasano

ANALISIS DE VARIANZA

En casos sencillos, tales como establecer la probabilidad de que una muestra dada pertenezca a una población con características específicas o para contrastar hipótesis acerca de las equivalencias de dos muestras, se utilizan los test basados en la distribución t o de Student. En muchos otros casos se deben , en cambio, tomar decisiones en problemas que involucran más de dos poblaciones.

Por ejemplo, para el primer caso se puede utilizar el test t para comprobar la hipótesis que un conjunto de 10 muestras de agua provienen de una población que tiene una salinidad media de 1,1 g/l, Considerando que la muestras fueron extraídas de una población normal, con media $\bar{X}$= 1,15 g/l, desvío estandard s =0,3027 g/l y varianza s^2= 0,0916 $(g/l)^2$, el estadístico t se calcula según

$$t = [\bar{X} - \mu] \sqrt{n}/s \qquad [1,15-1,1] \sqrt{10}/0,3 = 0,52 \qquad (1)$$

con μ= media poblacional

Formalmente se contrastan la hipótesis nula (H_0) y la alternativa (H_1) siguientes:

H_0: $\mu_1 = \mu$

H_1: $\mu_1 \neq \mu$

Si se quiere una probabilidad de rechazar la hipótesis nula cuando es verdadera, el valor calculado de t debe exceder el valor tabulado correspondiente para 9 grados de libertad y nivel de significación α de 0,05.Si como en este caso la prueba es bilateral (la muetra puede provenir de una población con media mayor o menor que μ) el valor de t se debe buscar para $\alpha/2$= 0,025. El nivel de significación corresponde a la probabilidad de cometer un error de tipo 1, es decir de rechazar una hipótesis correcta. La combinación de aceptar o rechazar una hipótesis, que a su vez puede ser falsa o verdadera, resulta en cuatro posibilidades, dos de ellas correctas y dos incorrectas.

	Hipótesis correcta	Hipótesis incorrecta
Hipótesis aceptada	Decisión correcta	Error tipo II
Hipótesis rechazada	Error tipo I	Decisión correcta

Supongamos ahora que en otra área se extraen otras 10 muetras del mismo acuífero con media X_2= 1,26 , s_2= 0,576 y s_2^2= 0,332, y se quiere comparar las medias respectivas más que confrontarlas con parámetros estadísticos propuestos. Se plantean ahora las siguientes hipótesis

H_0= $\mu_1 = \mu_2$ $\qquad$ H_1= $\mu_1 \neq \mu_2$

que establecen la igualdad o no de las medias poblacionales de las que fueron extraídas los grupos de muestras. Estas pueden

presentar tamaños iguales ($n_1=n_2$) o diferentes y las varianzas de las poblaciones pueden ser iguales ($\sigma_1^2=\sigma_2^2$) o diferentes. La combinación de estas situaciones produce cuatro casos distintos Por ello el primer paso consiste en determina si hay o no homogeneidad entre las varianzas poblacionales de donde proceden las muestras utilizando la distribución F. Esta es una distribución teórica de los valores que se esperarían si se procede a un muestreo aleatorio de una población normal y se calcula, para todos los pares posibles de varianzas muestrales, el cociente de éstas. Las hipótesis planteadas son:

H_0: $\sigma_1^2 = \sigma_2^2$ $\quad$ H_1: σ^2 distintas

Se calcula el valor F

$$F = s_1^2/s_2^2 \quad ; \; s_1 > s_2 \qquad (2)$$

que en nuestro caso es F= 0,269/0,0916= 2,94. Las tablas de distribución F tienen tres entradas: nivel de significación α, grados de libertad de la muestra de mayor varianza ϕ_1 y de la de menor ϕ_2.

Para un nivel de significación del 5 %, es decir, se quiere correr un riesgo de concluir que las varianzas de salinidades son diferentes cuando en realidad son iguales una vez cada 20, y grados de libertad $\phi_1 = \phi_2 = 9$, el valor es F= 3,18. Como el valor calculado es mayor, se acepta la homogeneidad de varianzas y es posible entonces, calcular una varianza ponderada común (varianza combineda) de acuerdo a la expresión

$$s^2 = \frac{(n_1-1)\, s_1^2 + (n_2-1)\, s_2^2}{n_1 + n_2 - 2} \qquad (3)$$

y puesto que en este caso $n_1=n_2$, la expresión (3) se reduce a

$$s^2 = [s_1^2 + s_2^2] / 2 \qquad (4)$$

El valor de t se calcula

$$t = [\bar{x}_1 - \bar{x}_2] / s_d \qquad (5)$$

con s_d, la desviación típica de las diferencias

$$s_d = \sqrt{\frac{s^2}{n} + \frac{s^2}{n}} = \sqrt{\frac{2\, s^2}{n}} = \sqrt{\frac{s_1^2 + s_2^2}{n}} \qquad (6)$$

El valor de t $\alpha/2$ para 18 grados de libertad es de 2,10 y el calculado es -0,579 por lo que se concluye que no existen diferencias entre los contenidos de salinidad de los sectores analizados.

En el caso que $n_1 \neq n_2$, s_d se calcula con la primera expresión de (6) pero reemplazando n

$$s_d = \sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}} = \sqrt{\frac{s^2\,(n_1 + n_2)}{n_1\, n_2}} \qquad (7)$$

Si $\sigma_1 \neq \sigma_2$ no se halla una varianza común y

$$s_d = \sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n}} \qquad (8)$$

y si $n_1 \neq n_2$ se reemplaza n por el valor correspondiente en cada término

$$s_d = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \qquad (8')$$

Si existe un tercer grupo de muestras, o muchos más, se pueden tomar éstas de a pares y comprobar estadísticamente sus medias o tratar el problema dentro de la rama de la estadística denominada **análisis de la varianza.** Esta técnica involucra la separación de la varianza total de las mediciones efectuadas en varias componentes o fuentes. Además considera en forma simultánea las diferencias en las medias y en las varianzas.

Supongamos ahora que incorporamos un tercer grupo de muestras (k=3), de tamaño n_3= 10, media 1,28, desvío estándard 0,514 y varianza 0,264. Consideramos un valor x_{ij} como el correspondiente al contenido salino de una muestra j ($j=1,2,\ldots,n_j$) de la i-ésima área ($i= 1,2,\ldots,k$). De esta forma los datos se hallan divididos en k clases con n_i valores en la i-ésima clase. El número total de valores considerando todas las clases es

$$N = \sum_{i}^{k} n_i \qquad (9)$$

la media para cada clase es

$$\bar{x}_i = \frac{1}{n_i} \sum_{j}^{n_i} x_{ij} \qquad (10)$$

y la de todas las clases

$$\bar{x} = \frac{1}{N} \sum_{i}^{k} \sum_{j}^{n_i} x_{ij} \qquad (11)$$

La suma total de los cuadrados de los desvíos de las x_{ij} respecto a la media $\bar{x}$ se divide en dos partes, Una primera parte es debida a la variabilidad que no puede ser explicada por diferencias de los efectos regionales entre las clases. La segunda se debe a la variabilidad entre clases promediada sobre las determinaciones individuales en cada clase. Es decir que estas partes corresponden a una fuente de variación **dentro de las muestras** (también llamado **error experimental**) y otra **entre las muestras** que corresponden a las diferencias existentes entre las medias de la clase respecto a la media general.

La suma total de las desviaciones al cuadrado queda definida por

$$SCT = \sum_{i}^{k} \sum_{j}^{n_i} (x_{ij} - \bar{x})^2 \qquad (12)$$

la suma de cuadrados dentro de los grupos o clases es

$$SCE = \sum_{i}^{k} \sum_{j}^{n_i} (x_{ij} - \bar{x}_i)^2 \qquad (13)$$

y la suma de cuadrados entre los grupos es

$$SCA = \sum_{i}^{k} n_i (\bar{x}_i - \bar{\bar{x}})^2 \qquad (14)$$

La identidad fundamental

$$SCT = SCA + SCE \qquad (15)$$

es válida independientemente de si se cumple o no la hipótesis de igualdad de las medias poblacionales.

Se puede considerar a los tres grupos de muestras como tres poblaciones de varianza constante σ^2 y si suponemos que $\mu_1 = \mu_2 = \mu_3 = \mu$ se puede considerar a su vez como tres grupos de muestras de una única población. Bajo estas consideraciones se pueden definir 3 estimadores insesgados de la varianza poblacional:

1- Mediante combinación de las varianzas muestrales en forma análoga a lo realizado anteriormente

$$\hat{\sigma}^2 = \frac{n_1 s_1^2 + n_2 s_2^2 + n_3 s_3^2}{n_1 + n_2 + n_3 - 3} = \frac{1}{N-3} \sum_{i}^{k} \sum^{n_i} (x_{ij} - \bar{x}_i)^2 = \frac{SCE}{N-3} = CME \qquad (16)$$

CM : cuadrado medio

2- A partir de la relación:

$$\sigma_{\bar{x}}^2 = \sigma^2 / n \qquad \sigma^2 = n \sigma_{\bar{x}}^2 \qquad (17)$$

sumando para todos los i conjuntos de muestras

$$k\sigma^2_{\bar{x}} = \sum_{i}^{k} n_i (\bar{x}_i - \bar{\bar{x}})^2 \qquad o \qquad \sigma^2 = \frac{1}{k} \sum_{i}^{k} n_i (\bar{x}_i - \bar{\bar{x}})^2 \qquad (18)$$

Para una estimación insesgada se utilizan k-1 grados de libertad, por lo tanto

$$\hat{\sigma}^2 = \frac{SCA}{k-1} = CMA$$

3- Considerando a las tres muestras juntas como una gran muestra de tamaño $N = n_1 + n_2 + \ldots + n_k$, el estimador insesgado de σ^2 es

$$\sigma^2 = \frac{1}{N-1} \sum \sum (x_{ij} - \bar{\bar{x}})^2 = \frac{SCT}{N-1} = CMT \qquad (19)$$

Puesto que los CMA y CME son estimadores de la varianza poblacional, su relación debe ser cercana a la unidad y presentar una distribución F

$$F = \frac{CMA}{CME} = \frac{\text{varianza estimada "entre"}}{\text{varianza estimada "dentro"}} \qquad (20)$$

Se puede observar que el estimador del denominador no está afectado cuando la hipótesis no es cierta ya que se obtiene combinando las varianzas muestrales y sólo depende de las desviaciones dentro de cada muestra o grupo de muestras. Por el contrario, el estimador del numerador será mayor ya que las medias muestrales $\bar{x}_i$ pueden diferir entre sí y de $\bar{\bar{x}}$ en una proporción mayor que la que cabría esperar por azar. La varianza CMA estima la σ^2 sólo cuando las medias poblacionales de las distintas poblaciones son iguales. Caso contrario, la varianza estimada a

partir de la SC "entre" es igual a $\sigma^2 + c$, con $c > 0$, debido a la desigualdad de las medias de la población.

La tabla de análisis de la varianza siguiente resume las consideraciones anteriores

Fuente variación	G. L.	Σ Cuadrados	Cuadrados Medios	F
Entre muestras	k-1	SCA	(SCA/k-1)= CMA	CMA/CME
Dentro muestras	N-k	SCE	(SCE/N-k)= CME	
Total	N-1	SCT	(SCT/N-1)= CMT	

También se cumple N-1 = (N-k) + (k-1)

Si el F calculado excede el valor del F tabulado para un intervalo de confianza dado, con k-1 y N-k grados de libertad, se rechaza la hipótesis de homogeneidad. Para determinar en cual de las medias hay diferencia significativa se puede utilizar el procedimiento de comparación múltiple desarrollado por Tukey. Para este test se selecciona un nivel de significación α para determinar la significación de una o más de las hipótesis nulas (H_0: $\mu_1=\mu_2$, H_0: $\mu_1=\mu_3$, H_0: $\mu_2=\mu_3$, etc). El test de Tukey calcula un valor único, la **diferencia significativa honesta** (DSH) con el cual se comparan todas las diferencias de medias.

$$DSH = ASS_{(\alpha,k,N-k)} (CME/n)^{\frac{1}{2}} \qquad (21)$$

donde ASS es la Amplitud Studiantizada Significativa que se obtiene de la tabla correspondiente a partir de α grados de libertad, k número de medias (= número de clases) y N-k, con N número total de observaciones. CME es el cuadrado medio del error y n el número de observaciones de las clases.

Se calculan para todos los pares posibles de diferencias de medias y aquellas que den un valor mayor a la DSH se considera significativa. Si las muestras no son todas del mismo tamaño n no se considera y el segundo factor en (21) corresponde por lo tanto, al desvío estándard del error.

Otro test que se realiza es el de **Barlett** que permite comprobar la homogeneidad de varianzas de más de dos muestras. Los pasos a seguir son:

1- Se calcula la varianza de cada muestra (clase) i (i:1,2,...,k) cada una de tamaño n_i

$$s_i^2 = \frac{(SC)_i}{(GL)_i} \qquad (22)$$

2-Se calcula el log s_i^2

3-Se halla la varianza estimada acumulada

$$s^2 = \frac{\Sigma\ SC}{\Sigma\ GL} \qquad (23)$$

4- Se define

$$B = \log s^2 \cdot \Sigma\ (n_i-1) \qquad (24)$$

5- Se calcula un valor chi cuadrado e^2

$$e^2 = \ln 10 \cdot [B - \Sigma(n_j - 1) \cdot \log s^2] \quad (25)$$

siendo ln 10 = 2,3026

Se compara el valor obtenido de e^2 con el tabulado para α y grados de libertad $\emptyset = k-1$. Si el valor calculado es menor que el tabulado se acepta la hipótesis nula de igualdad de varianzas. Caso contrario el análisis de varianza debe aplicarse con ciertas reservas ya que uno de los supuestos básicos para realizar este tipo de análisis es la homogeneidad de varianzas.

Hasta ahora se ha visto lo que se denomina Análisis de la Varianza de una vía o de un factor (One way analysis of variance) en donde los valores en cada clase (por ej. área de muestreo) se consideran réplicas sujetas a variaciones aleatorias. Sin embargo puede haber variaciones significativas entre los valores individuales en cada una de las k clases. Mediante el análisis de la varianza de dos factores o vías se incorpora esta segunda fuente de variabilidad. Supongamos que tenemos k = 3 clases que corresponden a estaciones meteorológicas en cada una de las cuales se realizaron observaciones de la precipitación durante un lapso dado (j años con $j = 1,2,\ldots,n$). Un valor x_{ij} representa entonces la lluvia caída durante el año j en las estación i. Las columnas j se referirán a grupos o generalizando al factor II y las filas a las clases o factor I. En cada una de las k clases hay un valor de cada grupo y en cada uno de los n grupos hay un valor por clase. Se define una media de grupo

$$\bar{x}_j = \frac{1}{k} \sum_i^k x_{ij} \quad (26)$$

y la suma total de cuadrados se divide ahora en tres partes

$$SCT = SCA + SCG + SCE \quad (27)$$

donde SCA y SCE representan la suma de cuadrados entre clases y el error o variabilidad residual respectivamente; SCG la suma de cuadrados entre grupos.

$$\sum_i^k \sum_j^n (x_{ij} - \bar{x})^2 = \sum_i^k n_j (\bar{x}_i - \bar{x})^2 + \sum_j^n k (\bar{x}_j - \bar{x})^2 + \sum_i^k \sum_j^n (x_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2 \quad (28)$$

y los respectivos grados de libertad son:

$$nk-1 = N-1 = (k-1) + (n-1) + (k-1)(n-1) \quad (29)$$

En forma análoga los cuatro términos de (28) divididos por sus grados de libertad (29) proveen los estimadores insesgados de la varianza.

Para el análisis de una vía el modelo matemático es

$$x_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad (30)$$

siendo μ la media poblacional, $\alpha_i = (\mu_i - \mu)$ los efectos o desviaciones y $\varepsilon_{ij} = (x_{ij} - \mu_i)$ una variable aleatoria con media 0 y varianza desconocida que corresponde a los errores al azar.

ANALISIS DE REGRESION

Jorge Luis Fasano

ANALISIS DE REGRESION

En muchas investigaciones resulta de interés obtener una expresión cuantitativa que relacione variables que están ligadas funcionalmente. De esta forma no sólo se puede conocer la relación sino que además es posible estimar la variable dependiente a partir de la independiente. Por ejemplo la conductividad de las aguas dependen de su salinidad y es posible obtener una expresión de la variable dependiente Y (conductividad) en función de la salinidad X. Graficando los pares de valores X-Y se puede tener una idea de la existencia o no de una relación, y si ésta es lineal o curvilínea. Para nuestro caso la relación es lineal y la ecuación es

$$Y'_i = b_o + b_1 X \qquad (1)$$

siendo Y'_i el valor estimado de Y_i para un valor dado de X_i. Estas Y' calculadas deben ser tales que los desvíos respecto a los valores originales sean mínimos, es decir

$$\Sigma\, e_i^2 = \Sigma\, (Y_i - Y'_i)^2 = \Sigma\, (Y_i - b_o - b_1 X_i)^2 = \text{mínimo} \qquad (2)$$

donde e es el error residual o desvíos de Y respecto a ls Y'.

Partiendo de las ecuaciones normales se pueden calcular los coeficientes b_o y b_1 que definen la recta con las características deseadas

$$\Sigma\, Y_i = b_o + b_1\, \Sigma X_i \qquad (3a)$$

$$\Sigma\, X_i Y_i = b_o\, \Sigma X_i + b_1\, \Sigma X_i^2 \qquad (3b)$$

que en forma matricial se expresa como

$$\begin{bmatrix} \Sigma\, Y_i \\ \Sigma X_i Y_i \end{bmatrix} = \begin{bmatrix} n & \Sigma X_i \\ \Sigma X_i & \Sigma X_i^2 \end{bmatrix} \begin{bmatrix} b_o \\ b_1 \end{bmatrix} \qquad (4)$$

siendo la solución

$$b_1 = \frac{n\Sigma XY - \Sigma Y\, \Sigma X}{n\Sigma X^2 - (\Sigma X)^2} \qquad (5a)$$

$$b_o = \frac{\Sigma Y\, \Sigma X^2 - \Sigma X\, \Sigma XY}{n\Sigma X^2 - (\Sigma X)^2} \qquad (5b)$$

Se puede demostrar que (Apéndice I) la expresión (5a) es equivalente a

$$b_1 = \frac{\Sigma (X - \bar{X})\, (Y - \bar{Y})}{\Sigma (X - \bar{X})^2} \qquad \frac{SP_{XY}}{SC_X} \qquad (6a)$$

donde SP_{XY} es la suma de productos corregida y SC suma de cuadrados corregida. El coeficiente b_o, ordenada al origen, también se puede expresar como

$$b_o = \frac{\Sigma\, Y_i}{n} - b_1 \frac{\Sigma\, X_i}{n} = \bar{Y} - b_1 \bar{X} \qquad (6b)$$

Si suponemos a los Y_j como variables aleatorias se puede considerar que los datos siguen un modelo teórico poblacional

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \qquad (7)$$

es decir que un valor observado de Y es igual a la suma de una constante relacionada a la media (β_0), más una función de X_i más un desvío o error aleatorio ϵ_i. Por lo tanto, lo que se trata a partir de una muestra, es estimar los coeficientes de regresión de la población (β) y se desea para ello que ($b_0 + b_1X$) sea un estimador de varianza mínima (el mejor) de ($\beta_0 + \beta_1X$) y que además sea un estimador lineal insesgado. Esto equivale a

$$E(b_0 + b_1X) = \beta_0 + \beta_1X = \mu_{yx} = E(Y/X) \qquad (8)$$

donde E() es el valor esperado y por lo tanto E(Y/X) es la media de las Y en cada subpoblación para una dada X. Esto es cierto cuando

$$E(b_0) = \beta_0 \quad ; \quad E(b_1) = \beta_1 \qquad (9)\ (10)$$

Es insesgado porque el valor esperado del estadistico empleado como estimador es igual al parámetro de la población que se va a estimar (como $E(X) = \mu$). Es de mínima varianza porque si consideramos i grupos de muestras se pueden calcular mediante el método de cuadrados mínimos los coeficientes b_{0i} y b_{1i} y si por otro método se obtiene otro grupo de estimaciones c_{0i} y c_{1i}, de las rectas con los coeficientes b serán las que presenten la menor dispersión alrededor de la ecuación de regresión de la población.

En la ecuación (7) ϵ, el término de error o de perturbacio estadistica, es una variable independiente aleatoria, de media = 0 y cuya distribución puede o no estar especificada (poblaciones tipo II-IV y I-III respectivamente). De (7) y de las igualdades de (8) surge que

$$\epsilon = Y - (\beta_0 - \beta_1X) = Y - E(Y/X) \qquad (11)$$

La varianza poblacional es

$$\sigma^2 = E[Y - E(Y/X)] \qquad (12)$$

y se considera que todas las subpoblaciones tiene la misma varianza. Por lo tanto

$$\sigma^2 = E(\epsilon)^2 = E[\epsilon - E(\epsilon)]^2 = Var\ \epsilon = VAR\ Y \qquad (13)$$

Volviendo a la ecuación (1) vemos que la Y' calculada es una estimación de μ_{yx}, es decir es un valor esperado de Y y no de sus valores individuales. En forma análoga a (11) se puede definir las desviaciones de Y respecto a las Y'

$$e = Y - Y' = Y - b_0 - b_1X \qquad (14)$$

Resulta entonces que e es un stimador de ϵ y nos permite estimar la **varianza de la regresión de la población o varianza residual** σ_{yx} ya que cuanto más concentrados estén los puntos Y_i alrededor de la línea de regresión mejor será la estimación. La medida de la dispersión de observaciones en torno a la línea de regresión es la desviación tipica de la recta.

Se supone que para cada conjunto de valores Y correspondientes a X_i las varianzas son iguales es decir

$$\sigma^2 = E[Y - E(Y/X)] \qquad (15)$$

$$\sigma_i = \frac{\Sigma (Y - \mu_{Y/x_i})^2}{N_i} \qquad (16)$$

la varianza de los valores Y alrededor de la recta de regresión para valores de X es

$$\sigma^2_{y/x} = \frac{\Sigma (Y - \mu_{Y/X})^2}{N}$$

Y $\sigma_{y/x}$ es la **desviación típica de la regresión de la población.** Como $\mu_{Y/X}$ no se conoce, pero Y' es su estimación, se define un estimador insesgado de $\sigma^2_{y/x}$

$$s^2_{y/x} = \frac{\Sigma (y - Y')^2}{n-2} = \frac{\Sigma (Y - b_0 - b_1 X)^2}{n-2} = \frac{\Sigma e^2}{n-2} \qquad (18)$$

donde la sustracción de dos grados de libertad corresponde al número de coeficientes de regresión utilizados (b_0 y b_1). $s_{y/x}$ se denomina **error típico de estimación o desviación típica de la regresión estimada.** Aquí es importante aclarar el hecho que si la variable aleatoria Y no tiene una distribución especificada no se puede medir la importancia de $s_{y/x}$ en términos de una distribución tal como la normal o en función de la tabla t. Si por el contrario las poblaciones de Y se distribuyen normalmente con una varianza común $s^2_{y/x}$, resulta que el 95 % de los valores de Y caen dentro del intervalo $\pm$ 2 $s_{y/x}$ de la línea de regresión. La interpretación de este error típico está ligada al concepto de coeficiente de correlación r.

El desvío o error total de los valores individuales de Y está dado por la diferencia entre éstos y la media aritmética $\bar{Y}$, que es el estimador de las Y_i cuando no se usa la recta de regresión. Se lo puede descomponer en:

$$Y - \bar{Y} = (Y - Y') + (Y' - \bar{Y}) \qquad (19)$$

$(Y - Y')$ = e = error no explicado. Es el error que todavía persiste luego del ajuste de la recta de regresión

$(Y' - \bar{Y})$ = error explicado, cantidad de error que se elimina cuando se ajusta la recta de regresión

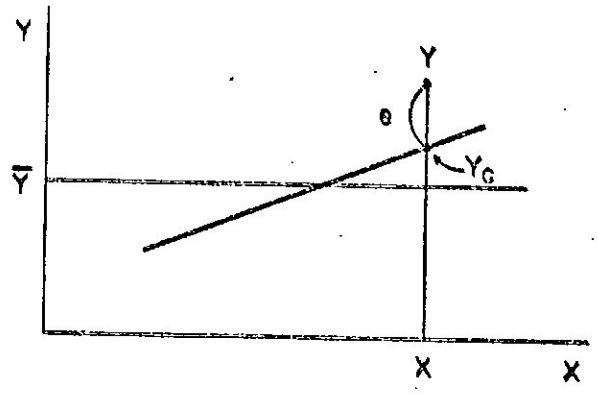


Fig. 1

Elevando al cuadrado cada error y sumando todos los valores de las muestras se llega a la relación fundamental del análisis de la

regresión (ver deducción en Apéndice II)

$$\underset{\text{SCT (totales)}}{\Sigma (Y - \bar{Y})^2} = \underset{\text{SCI (inexplicado)}}{\Sigma (Y - Y')^2} + \underset{\text{SCR (explicado)}}{\Sigma (Y' - \bar{Y})^2} \quad (20)$$

y dividiendo por $\Sigma (Y - \bar{Y})^2$ queda

$$1 = \frac{\Sigma (Y - Y')^2}{\Sigma (Y - \bar{Y})^2} + \frac{\Sigma (Y' - \bar{Y})^2}{\Sigma (Y - \bar{Y})^2} \quad (21)$$

El segundo término de la derecha se define como el **coeficiente de determinación** de la muestra, r^2

$$r^2 = \frac{\text{SC explicada}}{\text{SC total}} \quad (22)$$

Su raíz cuadrada se denomina **coeficiente de correlación** r y tiene el mismo signo que b_1. De la definición de r^2 y de la relación fundamental surge que si el error no explicado = 0, $r^2 = 1$ ya que $Y = Y'$; si el error explicado = 0, $Y' = \bar{Y}$ y $r^2 = 0$. Por lo tanto r varía entre ± 1. Así el coeficiente de determinación brinda tres tipos de información:

1- Mide la cantidad de mejoramiento obtenido a partir de la línea de regresión en términos de la reducción del error total. Un r^2 = 0,8 indica que la suma total de los cuadrados se redujo en un 80 %. Para un $r^2 = 1$ la reducción es del 100 % , el error $e = (Y-Y') = 0$ y todos los puntos están sobre la recta de regresión.

2- Mide la perfección del ajuste de la línea de regresión a los puntos. Si $r^2 = 1$ los puntos caen sobre la recta. Si la recta de regresión es horizontal y los puntos Y dispersos, $r^2 = 0$.

3- r^2 mide la linealidad de los puntos. Si $r^2 = 1$ la dispersión de los puntos se acerca a una recta.

El coeficiente de determinación es una medida útil de la dispersión de Y respecto a los Y' cuando, como se mencionó anteriormente, $s^2_{y.x}$ no se puede valorar por desconocerse la distribución de Y.

De (21) y (22) surge que

$$\Sigma (Y - Y')^2 = (1 - r^2) \Sigma (Y - \bar{Y})^2 \quad (23)$$

introduciendo el estimador insesgado de σ_Y

$$s^2_y = \frac{\Sigma (Y - \bar{Y})^2}{n-1} \quad (24)$$

$$\sigma^2_y = \frac{\Sigma (Y - \mu_Y)^2}{N} \quad (25)$$

(23) puede escribirse como

$$(n-2)\ s^2_{y.x} = (1-r^2)\ (n-1)\ s^2_y \quad (26a)$$

o $$s^2_{y.x} = (1-r^2)\ s^2_y \frac{n-1}{n-2} \quad (26b)$$

lo que significa que la varianza de Y, s^2_y se ha reducido en un porcentaje igual a $(r^2 . 100)$ y hay una parte residual, $(1-r^2)$ por ciento no explicada de $s^2_{y.x}$ después de ajustada la línea de

regresión. Es decir cuando $r^2 = 1$ la s^2_{y} está totalmente explicada, eliminada o reducida y $s^2_{y/x}$ es igual a cero (---> Y = Y'). Si $r^2 = 0$ no se ha explicado nada al ajustar la recta de regresión.

Supongamos ahora que en una región dada, existen 50 puntos donde es posible tomar muestras de agua pero se desean recolectar sólo 20 muestras. Si m personas diferentes realizaron el muestreo seleccionando cada una de ellas 20 muestras se tendrán m rectas de regresión de salinidad--conductividad de los $4,7 \cdot 10^{13}$ (!) conjuntos de muestras posibles que oscilarán alrededor de la recta de regresión poblacional. Si se seleccionan muestras de tamaño n = 40 habrá $1 \cdot 10^{10}$ conjuntos posibles. Si aplicamos ahora el análisis de la regresión a cada conjunto de tamaño 40, las m rectas obtenidas se acerrarán más a la recta poblacional y por lo tanto la variación de los b_1 será más pequeña. Como conclusión las rectas resultantes serán mejores estimadores de la línea de regresión poblacional. La varianza de b, que indica su variación, se define como

$$\sigma^2_b = \frac{\Sigma (b_1 - \beta_1)^2}{M} \qquad (27)$$

con M la cantidad de muestras posibles de tamaño n elegidas de una población de tamaño N. Cuando σ^2_b es pequeña, la recta de regresión de la muestra será un buen estimador de la recta de regresión de la población. Pero como no se conoce β y M es muy grande, se puede calcular VAR b por medio de una expresión equivalente

$$\sigma^2_b = \frac{\sigma^2_{y/x}}{\Sigma (X - \bar{X})^2} \qquad (28)$$

y su estimador insesgado es

$$s^2_b = \frac{s^2_{y/x}}{\Sigma (X - \bar{X})^2} \qquad (29)$$

s^2_b será pequeño si $s^2_{y/x}$ lo es también y/o $\Sigma(X-\bar{X})^2$ es grande. Como se supone que $\sigma^2_{y/x}$ es un parámetro de la población constante, se espera que $s^2_{y/x}$ no variará demasiado con el tamaño n de la muestra. En cambio sí lo hará el denominador y por lo tanto s^2_b será menor cuando el tamaño de la muestra es mayor. Si como en el caso de la población tipo I la distribución de las Y no está especificado, la de los b_1 tampoco lo estarán; pero si el tamaño de la muestra es suficientemente grande, en general n >30, se puede asignar un intervalo de confianza a β_1.

De la fórmula de r^2 y de s^2_b se puede deducir la relación entre ambos. Si r^2 es grande, $s^2_{y/x}$ es pequeño y luego s^2_b será pequeño; pero un r^2 pequeño no significa un s^2_b grande, ya que si n es suficientemente grande s^2_b será pequeña. Por lo tanto, el ajuste no es perfecto pero b_1 será un estimador confiable de β_1 y la recta de regresión de la muestra es un buen estimador de la recta de regresión de la población.

También, si bien menos importante, se puede estimar la varianza poblacional de b_0 a través de

$$s^2_{b0} = \frac{s^2_{y/x} (\Sigma X^2)}{n \Sigma (X - \bar{X})^2} \qquad (30)$$

que será más pequeña a medida que aumente el tamaño de la muestra.

Cuando nos referimos a una población tipo II, en donde se conoce la distribución de frecuencias de las variables dependientes, por hipótesis de normalidad se pueden encontrar las distribuciones de b_0, b_1, Y' e Y, y construir intervalos de confianza para β_0, β_1, μ_{yx} e Y utilizando la distribución t.

Como el coeficiente β_1 es el de mayor interés se puede construir una prueba estableciendo la hipótesis nula y alternativa siguientes

$$H_0 = \beta_1 = 0 \quad ; \; H_1 = \beta_1 \neq 0 \quad \text{(prueba bilateral)}$$

La implicancia de β_1 es la independencia de X e Y y la recta de regresión es horizontal. De aquí surge la importancia de demostrar, a partir de b_1 si se cumple la hipótesis nula. Para ello se construye el estadístico t

$$t = \frac{b_1 - \beta_1}{s_{b1}} = \frac{b_1}{s_{b1}} \qquad (31)$$

Si el t calculado es mayor que el tabulado para $\alpha/2$ se rechaza la hipótesis nula. Si la probabilidad es mayor que $\alpha/2$ se considera entonces $\beta_1 = 0$. Si $\alpha = 5\ \%$, entonces

$$P\left[-t_{\alpha/2} < \frac{b_1 - \beta_1}{s_{b1}} > t_{\alpha/2}\right] = 0{,}95 \qquad (32)$$

o

$$P\left[b_1 - t_{\alpha/2} \cdot s_{b1} < \beta_1 > b_1 + t_{\alpha/2} \cdot s_{b1}\right] = 0{,}95 \qquad (33)$$

$$\beta_1 = b_1 \pm t_{\alpha/2} \cdot s_{b1}$$

lo que significa que 95 de cada 100 muestras seleccionadas contendrán al verdadero parámetro poblacional β_1. Este test o variantes de él son importantes en el estudio de series de tiempo, ya que estos procedimientos se basan en el supuesto que no hay tendencia en los datos (la regresión en el tiempo -o espacio- es cero). Si efectivamente la hubiera, debe ser removida. Las series sin tendencia se dicen estacionarias.

En forma análoga se puede calcular los intervalos de confianza para β_0.

Puesto que Y' es un estimador de $\mu_{yx} = E(Y/X)$ se puede desear conocer la confiabilidad o seguridad que ofrece como estimador. Para ello se define el **error típico de la estimación de la regresión** para una dada X_p

$$s_{y'} = s_{yx} \cdot \sqrt{\frac{1}{n} + \frac{(X_p - \bar{X})^2}{\Sigma (X - \bar{X})^2}} \qquad (34)$$

y como surge de la ecuación este error es mínimo cuando $X_p = \bar{X}$ y crece a medida que su diferencia aumenta. Para $X_p = \bar{X}$

$$s_{y'} = s_{yx}/\sqrt{n} \qquad (35)$$

Se construye un intervalo de confianza sabiendo que Y' tiene una distribución t con n-2 grados de libertad

$$\mu_{yx} = Y' \pm t_{\alpha/2} \cdot s_{y'} \qquad (36)$$

Graficando los intervalos de confianza para cada subpoblación se obtendrá la **banda de confianza**, que de acuerdo a lo mencionado

será más ancha en los extremos que en el medio.

En otros casos nos puede interesar conocer o predecir valores *individuales* de Y -no el valor medio para una X determinada-, y para ello necesitamos conocer el intervalo de confianza para Y. Esto se halla a partir de la varianza de $(Y - Y')$

$$Var\ (Y - Y') = Var\ Y + Var\ Y' \qquad (37)$$

De (34) y sabiendo que VAR Y = $\sigma^2_{y/x}$ se estima por $s^2_{y/x}$, (37) se puede escribir como

$$VAR\ (Y - Y') = s^2_{y/x} \cdot \left(1 + \frac{1}{n} + \frac{(X_p - \bar{X})^2}{\Sigma(X - \bar{X})^2}\right) \qquad (38)$$

El estadístico t se construye

$$t \qquad \frac{(Y - Y') - E(Y - Y')}{\left[s^2_{y/x} \cdot \left(1 + \frac{1}{n} + \frac{(X_p - \bar{X})^2}{\Sigma(X - \bar{X})^2}\right)\right]^{1/2}} \qquad (39)$$

y como $E(Y - Y') = E(Y) - E(Y') = \mu_{y/x} - \mu_{y/x} = 0$ el intervalo de confianza para una valor de X_p dado es

$$Y = Y' \pm t_{\alpha/2} \cdot \sqrt{s^2_{y/x}\left(1 + \frac{1}{n} + \frac{(X_p - \bar{X})^2}{\Sigma(X - \bar{X})^2}\right)} \qquad (38)$$

Cuando n aumenta el intervalo disminuye y aumenta cuando mayor sea la diferencia X_p y $\bar{X}$.

ANALISIS DE CORRELACION

El análisis de correlación sirve como medida de *covariabilidad* de dos variables X e Y y como medida de la bondad de ajuste de una recta de regresión a la distribución de las observaciones. Mediante este análisi se mide el grado de asociación entre dos atributos de una distribución conjunta de las variables o distribución bivariada. En lugar de tener una curva de frecuencia se tendrá una superficie de frecuencia (Fig. 2)

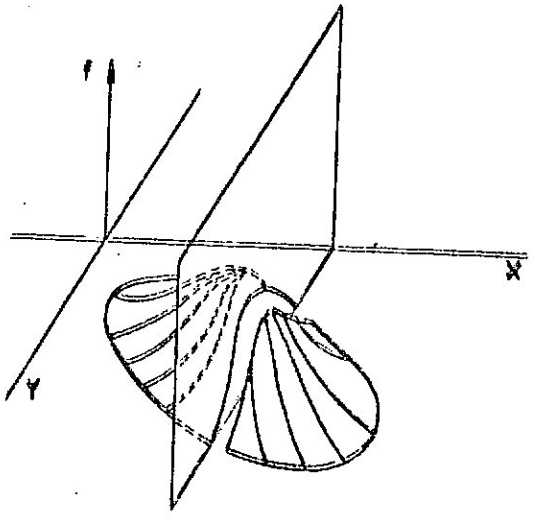


Fig. 2. Distribución bivariada

Por ejemplo, si en un medio reductor subterráneo la oxidación de la materia orgánica reduce los compuestos de Fe y Mn, ambos variarán en forma conjunta sin que por ello exista una relación causa-efecto, sino que obedecen a una causa común. La medida que indica la covariabilidad entre el Fe y Mn es el **coeficiente de correlación.** Para una distribución bivariada

$$E(X) = \mu_x \quad ; \quad E(Y) = \mu_y \qquad (41\ a\ y\ b)$$

$$Var(X) = \sigma^2_x \quad VAR\ (Y) = \sigma^2_y \qquad (42\ a\ y\ b)$$

y la covarianza de X e Y resulta

$$COV(X,Y) = E[X_i - E(X_i)]\,[Y_j - E(Y_j)] = E\,(X_i - \mu_x)\,(Y_j - \mu_j) \qquad (43)$$

y el **coeficiente de correlación poblacional** δ es

$$\delta = \frac{COV\ (X,Y)}{\sigma_x\ \sigma_y} = \frac{E(X - \mu_x)\ (Y - \mu_y)}{[E(X - \mu_x)^2]^{1/2}\,[E(Y - \mu_y)^2]^{1/2}} \qquad (44)$$

Se observa que:
1- para una relación lineal entre X e Y, si ambas covarian en la misma dirección COV (X,Y) > 0; si lo hacen en dirección opuesta resulta < 0.
2- δ es una medida del grado de covariación entre X e Y y varía entre ± 1. El signo de δ tiene la misma implicancia que el de la COV (X,Y).
3-Mide el grado de dependencia lineal entre X e Y. Si δ= 1 o δ= -1 la dependencia lineal es perfecta.

Como el tamaño de las poblaciones es grande, δ se estima a partir de una muestra estableciéndose la hipótesis que la distribución bivariada es normal (población tipo IV). Su estimador es el **coeficiente de correlación** r definido anteriormente

$$\hat{\delta} = r = \frac{\Sigma\ (X - \bar{X})\ (Y - \bar{Y})\ /\ n-1}{[\Sigma(X-\bar{X})^2/n-1]^{1/2}\,[\Sigma(Y-\bar{Y})^2/n-1]^{1/2}} = \frac{COV\ (X,Y)}{s_x\ \ s_y} \qquad (45)$$

$$= \frac{SP_{xy}}{[SC_x\ .SC_y]} = \frac{\Sigma\ (X - \bar{X})\ (Y - \bar{Y})}{[\Sigma(X-\bar{X})^2\ \Sigma(Y-\bar{Y})^2]^{1/2}}$$

$$= \frac{\Sigma\ XY\ -\ n\ \bar{X}\bar{Y}}{[(\Sigma X^2 - n\bar{X}^2)\ (\Sigma Y^2 - n\bar{Y}^2)]^{1/2}} = \frac{n\ \Sigma\ XY - (\Sigma X)\ (\Sigma Y)}{([n\Sigma X^2 - (\Sigma X)^2]\ [n\Sigma Y^2 - (\Sigma Y)^2])^{1/2}} \qquad (46)$$

SP_{xy} = suma de productos X-Y
SC_x \ SC_y = suma de cuadrados de X \ Y

Cuando n --> N, r --> δ; por lo tanto cuando la distribución es bivariada r se puede interpretar de dos maneras
1- la ya vista que mide la exactitud del ajuste de la recta de la regresión
2- como estimador de δ que mide la covariación.

De 1- se desprende que la definición de r^2 como cociente de sumas al cuadrado explicado/total es equivalente a (46) (ver Apéndice III)

$$r^2 = \frac{\Sigma\ (Y' - \bar{Y})^2}{\Sigma\ (Y - \bar{Y})^2} = \frac{[\Sigma(X - \bar{X})(Y - \bar{Y})]^2}{\Sigma(X - \bar{X})^2\ \Sigma(Y - \bar{Y})^2} \qquad (47)$$

Si la distribución es bivariada es posible hallar la regresión

de Y sobre X o de X sobre Y

$$Y' = b_0 + b_1 X \qquad ; \quad X' = b_0^* + b_1^* Y$$

$$b_1 = \frac{\Sigma (X - \bar{X})\ (Y - \bar{Y})}{\Sigma\ (X - \bar{X})^2} \qquad b_1^* = \frac{\Sigma\ (X - \bar{X})\ (Y - \bar{Y})}{\Sigma\ (Y - \bar{Y})^2}$$

Multiplicando b_1 por b_1^* se llega a (47)

$$b_1 \cdot b_1^* = \frac{[\Sigma (X - \bar{X})(Y - \bar{Y})]^2}{\Sigma (X - \bar{X})^2\ \Sigma (Y - \bar{Y})^2} = r^2 \qquad (48)$$

$$y\ r = (b_1 \cdot b_1^*)^{1/2} \qquad (49)$$

por lo que r es la media geométrica de los coeficientes de regresión. Su signo es igual a los b_1.

El coeficiente de correlación r puede someterse a una prueba para establecer su significación. Si el valor de r calculado es mayor que el valor de la tabla para un determinado nivel de significación y n-2 grados de libertad con número total de variables igual a 2, se dice que la correlación es significativa.

REGRESION LINEAL MULTIPLE

Cuando se mencionó el modelo de regresión poblacional se introdujo el término de perturbación ϵ que tiene en cuenta una serie de factores uno de los cuales puede ser la dependencia de Y de otras variables no consideradas. El valor de Y puede estimarse mejor si se incluyen estas variables independientes. Tal puede ser el caso en el estudio de caudales no sólo incluir la precipitación sino también las condiciones de humedad o en el estudio para determinar el espesor de un lente de agua dulce en islas de barrera considerar la precipitación, ancho de la zona medanosa y su altura entre otras variables. Para k variables independientes el valor estimado de Y' se expresa por la ecuación de regresión de la muestra

$$Y' = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_k X_k \qquad (50)$$

y para un valor de Y

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_k X_k + e \qquad (51)$$

Los coeficientes se estiman por mínimos cuadrados y han de ser tales que

$$\Sigma e^2 = \Sigma (Y - b_0 - b_1 X_1 - b_2 X_2 - \ldots - b_k X_k)^2 = \text{mínimo} \qquad (52)$$

Para k= 3 las ecuaciones normales son

$$n\, b_0 + b_1 \Sigma X_1 + b_2 \Sigma X_2 = \Sigma Y \qquad (53)$$

$$b_0 \Sigma X_1 + b_1 \Sigma X_1^2 + b_2 \Sigma X_1 \Sigma X_2 = \Sigma X_1 Y$$

$$b_0 \Sigma X_2 + b_1 \Sigma X_2 X_1 + b_2 \Sigma X_2^2 = \Sigma X_2 Y$$

o en forma matricial

$$\begin{bmatrix} n & \Sigma X_1 & \Sigma X_2 \\ \Sigma X_1 & \Sigma X_1^2 & \Sigma X_1 \Sigma X_2 \\ \Sigma X_2 & \Sigma X_2 X_1 & \Sigma X_2^2 \end{bmatrix} \begin{bmatrix} b_o \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \Sigma Y \\ \Sigma X_1 Y \\ \Sigma X_2 Y \end{bmatrix} \quad (54)$$

Las ecuaciones normales (53) se pueden simplificar llevándolas del origen (0,0,0) a $(\bar{Y}, \bar{X}_1, \bar{X}_2)$. De esta forma tenemos que

$\Sigma(X_1 - \bar{X}_1) = \Sigma X_1 - n\bar{X}_1 = \Sigma X_1 - \Sigma X_1 = 0$ y lo mismo para ΣX_2 y ΣY. Si

$$x_1 = X_1 - \bar{X}_1 \quad ; \quad x_2 = X_2 - \bar{X}_2 \quad ; \quad y = Y - \bar{Y}$$ las ecuaciones (53) quedan

$$b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2 = \Sigma x_1 y \quad (55)$$

$$b_1 \Sigma x_2 x_1 + b_2 \Sigma x_2^2 = \Sigma x_2 y$$

con $\Sigma x_1^2 = \Sigma (X_1 - \bar{X}_1)^2 = \Sigma X_1^2 - n(\bar{X}_1)^2$

$$\Sigma x_1 x_2 = \Sigma (X_1 - \bar{X}_1)(x_2 - \bar{X}_2) = \Sigma X_1 X_2 - n(\bar{X}_1)(\bar{X}_2)$$

y su resolución da

$$y' = b_1 x_1 + b_2 x_2 \quad (56)$$

o

$$(Y' - \bar{Y}) = b_1 (X_1 - \bar{X}_1) + b_2 (X_2 - \bar{X}_2) \quad (57)$$

de donde

$$Y' = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2 + b_1 X_1 + b_2 X_2 \quad (58)$$

por lo tanto

$$b_o = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2 \quad (59)$$

y (58) se transforma

$$Y' = b_o + b_1 X_1 + b_2 X_2 \quad (60)$$

Los b_1 y b_2 se denominan **coeficientes de regresión parcial** y representan el cambio promedio de Y como consecuencia de un cambio unitario en X cuando el resto de las X se mantienen constantes e indican la pendiente del plano de regresión. b_o corresponde a la altura o elevación del plano.

Una vez hallada la ecuación (60) interesa saber la bondad del ajuste del plano de regresión a los puntos dados y la significación de los coeficientes de regresión parcial. Para el primer punto, un plano ajustado a los puntos dados y que pase por la media $(\bar{Y}, \bar{X}_1, \bar{X}_2)$ se puede considerar el plano básico con respecto al cual se medirán las mejoras introducidas con la regresión. Las desviaciones están relacionadas por la identidad fundamental

$$\underset{\text{SCT}}{\underset{\text{Total}}{\Sigma (Y - \bar{Y})^2}} = \underset{\text{SCI}}{\underset{\text{Inexplicada}}{\Sigma (Y - Y')^2}} + \underset{\text{SCR}}{\underset{\text{Explicada}}{\Sigma (Y' - \bar{Y})^2}} \quad (61)$$

siendo la relación entre SCR y SCT el **coeficiente de determinación múltiple**

$$R^2_{y.12} = \frac{\text{SCR}}{\text{SCT}} = \frac{\Sigma (Y' - \bar{Y})^2}{\Sigma (Y - \bar{Y})^2} \quad (62)$$

y se puede interpretar como el índice de mejoramiento del ajuste del plano de regresión a los puntos reales u observaciones con respecto al ajuste que suponía el plano que pasa por la media $(\bar{Y}, \bar{X}_1, \bar{X}_2)$. Es decir indica la fracción explicada por el plano de

regresión. Este coeficiente se puede calcular según (ver Apendice IV)

$$R^2_{y.12} = \frac{b_1 \Sigma x_1 y + b_2 \Sigma x_2 y}{\Sigma y^2} \qquad (63)$$

La raíz cuadrada R^2 se denomina **coeficiente de correlación múltiple.** Si tanto las Y como las X son variables estocásticas, $R_{y.12}$ aparte de considerarse una medida de la bondad del ajuste del plano de regresión, se puede interpretar también como una medida de la correlación entre Y y su mejor estimador lineal Y', que depende de X_1 y X_2; por lo tanto R es el coeficiente de correlación entre Y y (X_1, X_2). Para el caso en que sólo Y es una variable estocástica (poblaciones tipo I y II) sólo se puede tener $R_{y.12}$; en el caso que tanto las Y como las X sean variables estocásticas se puede calcular $R_{y.12}$, $R_{1.y2}$ y $R_{2.y1}$. Para k variables independientes, la significación del coeficiente de correlación se busca para (k+1) variables y n-(k+1) grados de libertad en la tabla correspondiente.

Cuando se tratan variables que tienen una distribución multivariada (población tipo III) ó normal multivariada (tipo IV) puede interesar conocer la covariación entre ellas. Esto nos lleva a considerar dos casos. Tomemos por ejemplo el caudal medio en función de las precipitaciones

$$Y = b_0 + b_1 X_1 + b_2 X_2$$

Y = caudal total anual para el año i
X_1 = precipitación total para el año i
X_2 = precipitación total para el año i-1

Se puede buscar la covariación entre Y y X_1 ignorando la precipitación del año precedente y se obtiene de esta forma la **correlación total** de Y y X_1

$$r_{y1} = \frac{\Sigma y x_1}{[\Sigma y^2 \Sigma x_1^2]^{1/2}} \qquad r_{ij} = \frac{\Sigma x_i x_j}{[\Sigma x_i^2 \Sigma x_j^2]^{1/2}} \qquad (64)$$

Esto supone que como X_2 no se mantiene constante r_{y1} incluye el efecto de la precipitación precedente y que será diferente según sea la cantidad de mm caída. Del mismo modo se puede hallar la correlación entre Y y X_2. Cuando por el contrario se halla la covariación entre Y y X_1 manteniendo constante las demás variables, se obtiene el **coeficiente de correlación parcial** de Y y X_1 con respecto a X_2 y se denota $r_{y1.2}$. Los primeros dos subíndices indican las variables correlacionadas; el o los restantes las variables cuyos efectos fueron removidos.

$$r_{y1.2} = \frac{r_{y1} - r_{y2} \cdot r_{12}}{[(1 - r^2_{y2})(1 - r^2_{12})]^{1/2}} \qquad (65)$$

Este coeficiente de correlación parcial resulta ser el estimador del coeficiente de correlación parcial de la población $\rho_{y1.2}$.

Una forma general para el cálculo de los coeficientes de correlación parcial parte de la utilización de la matriz de correlación y permite el tratamiento de regresiones múltiples para k variables. Se define una matriz R que corresponde al determinante de la matriz de correlación

$$R = \begin{vmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{vmatrix} = \begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{vmatrix} \qquad (66)$$

con $r_{ij} = r_{ji}$ y $r_{ij} = 1$ si $i=j$

Se define a los menores con su signo o cofactores como R_{ij}

$$R_{11} = \begin{vmatrix} 1 & r_{23} \\ r_{32} & 1 \end{vmatrix} = 1 - r_{23}\, r_{32} = 1 - r^2_{23} \qquad (67)$$

y se puede calcular el coeficiente de correlación parcial como

$$r_{12.3} = \frac{-R_{12}}{[R_{11}\, R_{22}]^{1/2}} \qquad (68)$$

y para 4 variables

$$r_{12.34} = \frac{-R_{12}}{[R_{11}\, R_{22}]^{1/2}} \qquad (69)$$

con la matriz R

$$R = \begin{vmatrix} 1 & r_{12} & r_{13} & r_{14} \\ r_{21} & 1 & r_{23} & r_{24} \\ r_{31} & r_{32} & 1 & r_{34} \\ r_{41} & r_{42} & r_{43} & 1 \end{vmatrix} \qquad (70)$$

$$r_{12} = \begin{vmatrix} r_{21} & r_{23} & r_{24} \\ r_{31} & 1 & r_{34} \\ r_{41} & r_{43} & 1 \end{vmatrix} \qquad (71)$$

y los coeficientes de correlación múltiples resultan ser

$$R_{1.23} = 1 - [R/R_{11}] \qquad (72)$$

$$R_{1.234} = 1 - [R/R_{11}] \qquad (73)$$

En estos casos la variable Y se ha designado con el número 1.

Una vez hallada la recta de regresión interesa saber si los coeficientes de regresión son significativos o no, es decir si son sensiblemente distintos de 0. Nuevamente, si no son significativos no hay regresión entre Y y las X y no se puede predecir Y con mayor precisión que utilizando el plano que pasa por las medias. Se plantea entonces si los coeficientes de regresión parcial de la población son cero

$$H_0 : \beta_i = 0 \; ; \; H_1 : \beta_i \neq 0 \quad o \quad H_1 : \beta_i > 0$$

o también se puede plantear $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$

Para el primer caso se usa el test t, para el segundo el F

a-Prueba t para n-k-1 grados de libertad

$$t = \frac{b_i - \beta_i}{s_{bi}}$$

siendo

$$s_{bi} = CMI \cdot \Sigma x_i^2)^{-\frac{1}{2}} \quad ; \quad CMI = SCI / (n-k-1)$$

s_{bi} = error típico del coeficiente de correlación parcial
CMI = varianza respecto a la regresión

Si el t calculado para un determinado nivel de significación es mayor que b_i, se rechaza la hipótesis nula y el valor de b_i no se debe al azar. Los límites de confianza son

$$\beta_i = b_i \pm t\,\alpha \cdot s_{bi}$$

b- Prueba con la distribución F

Si la regresión es significativa las Y calculadas diferirán de las Y medias sensiblemente y entonces la $SCR = \Sigma (Y' - Y)^2$ será grande mientras que las $SCI = \Sigma (Y - Y')^2$ o residuales serán pequeñas. F entonces resulta la razón de dos estimaciones insesgadas de σ^2, con k y n-k-1 grados de libertad

$$F = \frac{SCR / k}{SCI / n-k-1} = \frac{CMR}{CMI}$$

Si el F calculado es mayor que el tabulado la regresión es significativa y la mejora introducida mediante el plano de regresión no era debida al azar.

OTROS TIPOS DE REGRESION

Si bien la regresión lineal es la más utilizada, existen variables que se ajustan a otros tipos de regresión

1. Regresión polinomial

Su ecuación es

$$Y = b_0 + b_1 X + b_2 X^2 + b_m X^m$$

Si suponemos que cada valor de X con su potencia es una nueva variable

$$X = X_1 \; ; \; X^2 = X_2 \; ; \ldots ; \; X^m = X_m$$

es aplicable el análisis de regresión lineal multiple. Si luego de desarrollar el test surge que β_1 es significativamente distinto de cero mientras que los demás coeficientes β no lo son, no existe evidencia de curvinealidad y se puede trabajar con el modelo más sencillo lineal.

2- Regresión exponencial

$$Y = b_0 \, e^{b_1 X}$$

$$\ln Y = \ln b_0 + b_1 X$$

de modo que la vinculación lineal se obtiene graficando X-ln Y.

3- Regresión logarítmica

$$Y = b_0 + b_1 \ln X$$

Graficando Y-ln X se obtiene una recta y por lo tanto se puede aplicar la regresión lineal.

4- Regresión potencial

$$Y = b_0 X^b$$

$$\ln Y = \ln b_0 + b_1 \ln X$$

Esta función graficada en escala logarítmica para X e Y corresponde a una recta.

APENDICE I

$$b_1 = \frac{\Sigma\ (X - \bar{X})\ (Y - \bar{Y})}{\Sigma\ (X - \bar{X})^2} = \frac{SP_{YX}}{SC_X} \qquad (6a)$$

$$\begin{aligned}
\Sigma(X - \bar{X})\ (Y - \bar{Y}) &= \Sigma\ (XY - X\bar{Y} - \bar{X}Y + \bar{X}\bar{Y}) \\
&= \Sigma\ XY - \bar{Y}\ \Sigma X - \bar{X}\ \Sigma Y + \Sigma\ \bar{X}\bar{Y} \\
&= \Sigma\ XY - n\ \bar{Y}\bar{X} - n\bar{Y}\bar{X} + n\bar{X}\bar{Y} \\
&= \Sigma\ XY - n\ \bar{X}\bar{Y} \\
&= \Sigma XY - \frac{(\Sigma X)\ (\Sigma Y)}{n}
\end{aligned}$$

$$\begin{aligned}
\Sigma\ (X - \bar{X})^2 &= \Sigma\ (X^2 - 2\ X\bar{X} + \bar{X}^2) \\
&= \Sigma X^2 - 2\ \bar{X}\ \Sigma X + \Sigma\bar{X}^2 \\
&= \Sigma X^2 - 2\ n\bar{X}^2 + n\ \bar{X}^2 \\
&= \Sigma X^2 - n\ \bar{X}^2 \\
&= \Sigma X^2 - \frac{(\Sigma X)^2}{n}
\end{aligned}$$

$$b_1 = \frac{\Sigma XY - [\Sigma X\ \Sigma Y]/n}{\Sigma X^2 - [(\Sigma X)^2/n]} = \frac{n\ \Sigma XY - \Sigma X\ \Sigma Y}{n\ \Sigma X^2 - (\Sigma X)^2} \qquad (5.a)$$

APENDICE II

$$Y' = b_0 + b_1 X = \bar{Y} + b_1 (X - \bar{X}) \quad (1)$$

$$Y' - \bar{Y} = b_1 (X - \bar{X}) \quad (2)$$

$$\Sigma (Y - Y')^2 = \Sigma [Y - \bar{Y} - b_1 (X - \bar{X})]^2 =$$

$$= \Sigma (Y - \bar{Y})^2 - 2 b_1 \Sigma (Y - \bar{Y})(X - \bar{X}) + b_1^2 \Sigma (X - \bar{X})^2 \quad (3)$$

De

$$b_1 = \frac{\Sigma (X - \bar{X}) (Y - \bar{Y})}{\Sigma (X - \bar{X})^2} \quad (4)$$

se obtiene

$$b_1 \Sigma (X - \bar{X})^2 = \Sigma (X - \bar{X}) (Y - \bar{Y}) \quad (5)$$

Reemplazando en el segundo término de (3)

$$\Sigma (Y - Y')^2 = \Sigma (Y - \bar{Y})^2 - 2 b_1^2 \Sigma (X - \bar{X})^2 + b_1^2 \Sigma (X - \bar{X})^2 \quad (6)$$

$$= \Sigma (Y - \bar{Y})^2 - b_1^2 \Sigma (X - \bar{X})^2 \quad (7)$$

y según (2)

$$\Sigma (Y - Y')^2 = \Sigma (Y - \bar{Y})^2 - \Sigma (Y' - \bar{Y})^2 \quad (8)$$

Reordenado se obtiene la relación fundamental

$$\Sigma (Y - \bar{Y})^2 = \Sigma (Y - Y')^2 + \Sigma (Y' - \bar{Y})^2 \quad (9)$$

$$r^2 = \frac{\Sigma\ (Y' - \bar{Y})^2}{\Sigma\ (Y - \bar{Y})^2} \qquad (1)$$

$$Y' = b_0 + b_1\ X \qquad (2)$$

$$b_0 = \bar{Y} - b_1\ \bar{X} \qquad (3)$$

$$b_1 = \frac{\Sigma\ (X - \bar{X})\ (Y - \bar{Y})}{\Sigma\ (X - \bar{X})^2} \qquad (4)$$

De (2) y (3)

$$Y' - \bar{Y} = b_1\ (X - \bar{X}) \qquad (5)$$

Reemplazando (5) en (1)

$$r^2 = \frac{b_1^2\ \Sigma\ (X - \bar{X})^2}{\Sigma\ (Y - \bar{Y})^2} \qquad (6)$$

y ahora reemplazando (4) en (6)

$$r^2 = \left[\frac{\Sigma\ (X - \bar{X})\ (Y - \bar{Y})}{\Sigma\ (X - \bar{X})^2}\right]^2 \left[\frac{\Sigma\ (X - \bar{X})^2}{\Sigma\ (Y - \bar{Y})^2}\right]^2 \qquad (7)$$

$$r^2 = \frac{[\ \Sigma\ (X - \bar{X})\ (Y - \bar{Y})\ ]^2}{\Sigma (X - \bar{X})^2\ \Sigma (Y - \bar{Y})^2} = \frac{\Sigma\ (Y' - \bar{Y})^2}{\Sigma\ (Y - \bar{Y})^2} \qquad (8)$$

$$r = \frac{\Sigma\ (X - \bar{X})\ (Y - \bar{Y})}{[\ \Sigma (X - \bar{X})^2\ \Sigma (Y - \bar{Y})^2\ ]^{1/2}} \qquad (9)$$

APENDICE IV

$$R^2_{y.12} = \frac{\Sigma\ (Y' - \bar{Y})^2}{\Sigma\ (Y - \bar{Y})^2} \qquad (1)$$

$$Y' = Y + b_1\ (X_1 - \bar{X}_1) + b_2\ (X_2 - \bar{X}_2) \qquad (2)$$

$$= Y + b_1\ x_1 + b_2\ x_2 \qquad (3)$$

$$Y' - \bar{Y} = b_1\ x_1 + b_2\ x_2 \qquad (4)$$

$$\Sigma\ (Y' - \bar{Y})^2 = \Sigma\ (b_1\ x_1 + b_2\ x_2)^2 \qquad (5)$$

$$= \Sigma\ (b_1^2\ x_1^2 + 2\ b_1\ b_2\ x_1\ x_2 + b_2^2\ x_2^2)$$

$$= b_1\ [b_1\ \Sigma x_1^2 + b_2\ \Sigma x_1 x_2] + b_2\ [b_1\ \Sigma x_1 x_2 + b_2\ \Sigma x_2^2]$$

De acuerdo a las ecuaciones normales

$$b_1\ \Sigma\ x_1^2 + b_2\ \Sigma\ x_1 x_2 = \Sigma\ x_1 y \qquad (6a)$$

$$b_1\ \Sigma\ x_1 x_2 + b_2\ \Sigma\ x_2^2 = \Sigma\ x_2 y \qquad (6b)$$

$$\Sigma\ (Y' - \bar{Y})^2 = b_1\ \Sigma\ x_1 y + b_2\ \Sigma\ x_2 y \qquad (7)$$

Entonces

$$R^2_{y.12} = \frac{b_1\ \Sigma\ x_1 y + b_2\ \Sigma\ x_2 y}{\Sigma\ y} \qquad (8)$$

ANALISIS DE AGRUPAMIENTO Y ORDENACION

Jorge Luis Fasano

ANALISIS DE AGRUPAMIENTO

Cuando se tienen varios objetos o muestras en estudio resulta de interés clasificarlos, es decir identificar grupos con muestras similares. Se puede establecer dos aproximaciones para clasificar: una que corresponde a la de clasificar en sentido estricto, y es la de identificar grupos; la otra ubica muestras u objetos en grupos existentes y se denomina discriminación. A su vez la primera puede dividirse en técnicas de agrupamiento y de ordenación. Las primeras permiten la extracción de grupos discretos que pueden o no ser jerarquizados, los segundos presentan en un espacio de dimensión reducida (2 o 3 dimensiones) las relaciones entre los individuos.

El análisis de agrupamiento (Cluster Analysis) se basa en definir semejanzas o similitudes entre los objetos a clasificar, para los cual se han propuesto una serie de índices o coeficientes de similitud, algunos de los cuales son:

1- **Coeficiente de distancia Euclediana**

$$d_{ij} = \frac{[\Sigma\ (x_{ik} - x_{jk})^2]^{1/2}}{p} \qquad (1)$$

i,j corresponden a los individuos, k designa las variables. Para dos variables (p = 2) d_{ij} corresponde a la línea recta que une al individuo i con el j. Para normalizar los valores d_{ij} se divide por p. El valor d_{ij} va de cero, para el caso de similitud completa a infinito.

2-**Coeficiente de correlación**

$$r_{ij} = \frac{\sum^{p} (x_{ik} - \bar{x}_i)\ (x_{jk} - \bar{x}_j)}{[\Sigma\ (x_{ik} - \bar{x}_i)^2]^{1/2}\ [\Sigma\ (x_{jk} - \bar{x}_j)^2]^{1/2}} \qquad (2)$$

Los subíndices i,j corresponden a distintas muestras y k las variables definidas en cada una de ellas. Se observa entonces que $\bar{x}_i$ y $\bar{x}_j$ resultan ser las medias de variables que pueden tener distinta naturaleza y pueden estar medidas en distintas escalas.

3-**Coeficiente de similitud proporcional**

$$\cos \theta_{ij} = \frac{\sum^{p} x_{ik} \cdot x_{jk}}{[\Sigma\ x_{ik}^2]^{1/2}\ [\Sigma\ x_{jk}^2]^{1/2}} \qquad (3)$$

se observa que si x_i y x_j se normalizan (media = 0 y varianza unitaria) el $\cos \theta_{ij} = r_{ij}$. Para dos objetos definidos en un espacio de p-dimensiones, este coeficiente corresponde al coseno del ángulo que forman los dos vectores trazados desde un origen común hasta esos puntos. Este coeficiente varía entre ± 1. Si $\cos \theta = 0$, hay una total disimilitud (vectores ortogonales); si $\cos \theta = 1$ los vectores son colineales y la similitud es completa.

Existen otro tipos de coeficientes de similitud, algunos de los cuales consideran variables cualitativas, codificando su presencia o ausencia, con , por ejemplo, 0 y 1.

Los datos se agrupan en i filas que corresponden a las muestras o individuos estudiados y j columnas que corresponden a las variables analizadas. Así, $X_{1,3}$ corresponde al valor de la variable 3 en la muestra 1.

Si cada elemento x_{ij} se estandariza según

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \qquad (4)$$

se obtiene la matriz de datos estandarizados. A partir de estos valores se construye la matriz de similitud, utilizando algunos de los coeficientes definidos. El modo de obtención de las distintas matrices a partir de la matriza de datos originales se halla detalladamente explicada en Merodio (1985).

Una vez definida la matriz de similitud, y mediante distintas técnicas, se trata de agrupar variables o individuos de acuerdo a su grado de similitud. Existen varias técnicas detalladas por Crisci y López Armengol (1983) según:

1-formen grupos exclusivos o no exclusivos
2-formen grupos jerárquicos o no jerárquicos
3-sean divisos o aglomerativos
4-sean secuenciales o simultáneos

Las técnicas más utilizadas corresponden a las exclusivas, jerárquicas, aglomerativas y secuenciales. La forma de agrupamiento más sencilla es por "grupo par"admitiéndose dos muestras por nivel que de esta forma constituyen un núcleo. Una nueva unidad o un núcleo puede incorporarse posteriormente constituyendo un grupo. Así el núcleo del primer grupo estará formado por las dos unidades observacionales que exhiban el mayor valor de similitud. Seguidamente se incorpora el próximo valor de mayor similitud (puede ser una nueva unidad de observación que se incorpore a un núcleo existente, conformando un grupo; fusión de núcleos existentes o formación de nuevos núcleos). Partiendo de una misma matriz de similitud, los núcleos o grupos que se van conformando sucesivamente pueden diferir según el tipo de procedimiento que se utilice para recalcular la matriz de similitud. Se diferencian tres técnicas

1- Ligamento simple: se utiliza el índice de similitud más parecido entre las unidades a incorporarse y la unidad integrante del grupo o núcleo ya establecido

2- Ligamento completo: el valor de similitud entre la unidad a incorporarse y el grupo o núcleo existente es igual al índice de menor similitud entre el candidato y los componentes del grupo o núcleo.

3- Ligamento promedio: los nuevos índices de similitud se calculan hallando el promedio de los individuos entre el candidato y cada uno de los componentes del grupo o núcleo. La técnica más utilizada para obtener este promedio es la media aritmética no ponderada (UPGMA).

Independientemente del método utilizado para seleccionar los nuevos índices de similitud, se van conformando las matrices derivadas que serán empleadas en el proceso de agrupamiento siguiente. Según la técnica UPGMA, las sucesivas matrices derivadas se calculan siempre a partir de las matrices de similitud original.

La representación gráfica más usada es el dendrograma (Fig.1), y su construcción utilizando los distintos tipos de ligamentos está ejemplificado en Crisci y López Armengol (1983).

Como se mencionó anteriormente, en cada etapa de la construcción de un dendrograma o del agrupamiento, se seleccionan o calculan nuevos índices de similitud, por lo cual, el dendrograma no refleja exactamente la matriz de similitud original. Los agrupamientos obtenidos difieren a su ves según el tipo de

ligamento empleado. A partir de los valores del dendrograma se puede construir una nueva matriz, denominada **matriz cofenética** y mediante el empleo de la fórmula (2) se puede calcular el **coeficiente de correlación cofenética** que permite evaluar cuantitativamente el grado de distorsión de la matriz de similitud cofenética. La técnica de ligamento promedio origina la menor distorsión. Valores de correlación entre las matrices superiores a 0,8 indican una representación significativa de los datos originales de similitud. En general, los coeficientes varían entre 0,6 y 0,95.

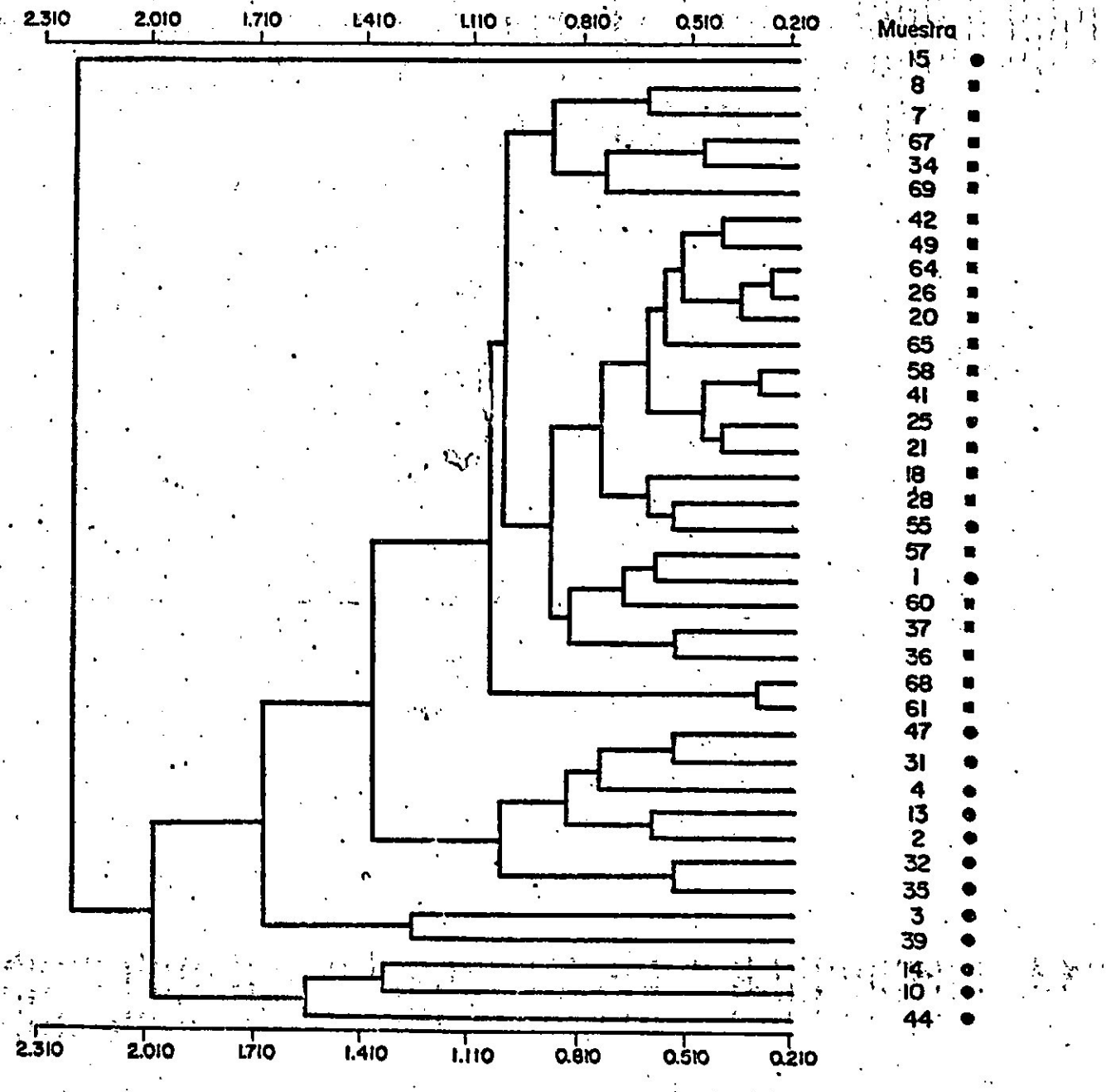


Fig. 1 Análisis de agrupamiento, modo Q

En el ejemplo de la figura 1 se asociaron muestras de agua. En cada una de ellas se determinaron contenidos iónicos y en función de éstos se agruparon las muestras según sus semejanzas. Esta

técnica de agrupamiento se denomina **modo Q.** Pero es también válido hacer la asociación inversa, denominada **técnica R.** En ella se asocian los caracteres o variables y este agrupamiento puede dar información muy valiosa. Por ejemplo en la Fig. 2 se presenta el dendrograma obtenido según la técnica R a partir de la misma matriz de datos originales.

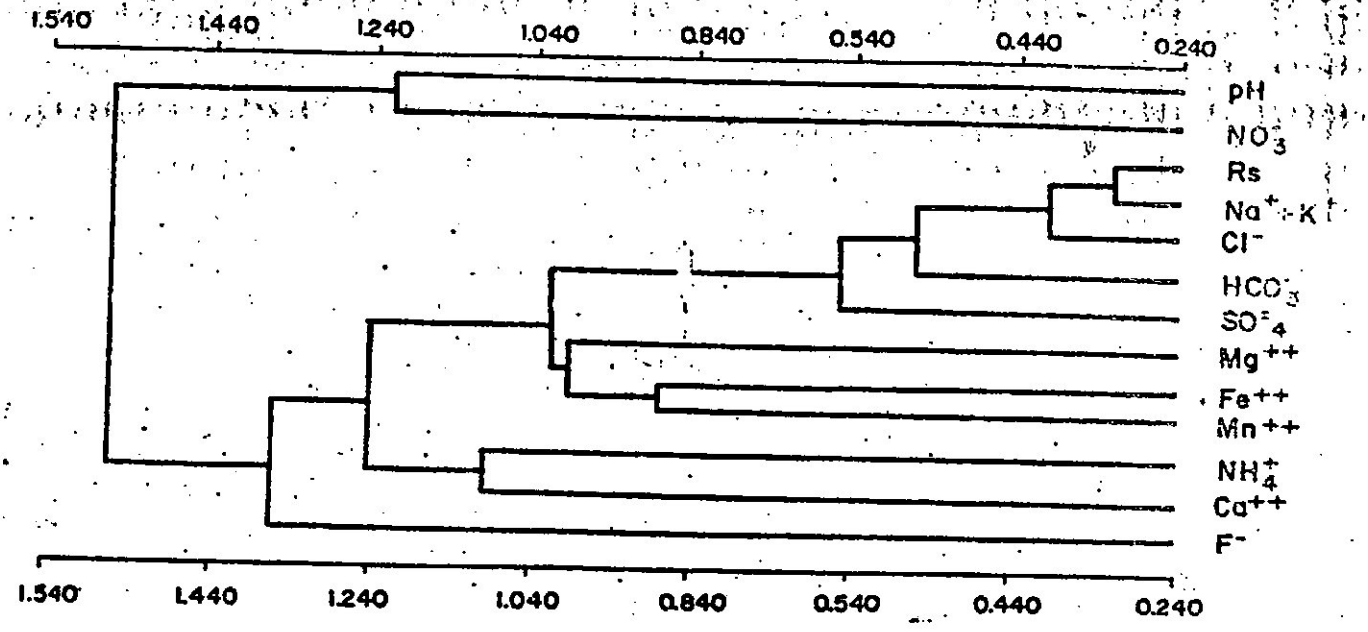


Fig. 2. Análisis de agrupamiento, modo R

Aparte de la asociación entre los elementos mayoritarios con el residuo seco, el núcleo conformado por el Fe y Mn se vincula a procesos de óxido-reducción, en un medio subterráneo, en los que interviene la materia orgánica. La conformación de un grupo con la adición del Mg puede estar indicando una fuente común a partir de minerales básicos. El coeficiente de correlación cofenética es de 0,9498.

ANALISIS DE COMPONENTES PRINCIPALES

Una muestra la podemos imaginar ubicada en un espacio de p dimensiones , donde $p = \Sigma k$ representa el número total de variables. El objetivo del ACP, como otros métodos de ordenación, es reducir el número de dimensiones. Mediante el ACP se construyen nuevas variables -componentes- que resultan de una combinación lineal de las variables observadas sin que ello traiga aparejado una pérdida de información significativa. El ACP permite entonces simplificar el análisis en situaciones que involucren una gran cantidad de observaciones y variables reduciendo su número.

El ACP parte del análisis de variables o caracteres (técnica R) pero su representación gráfica se refiere a las relaciones existentes entre las unidades de observación. Si las unidades de medida de las variables son homogéneas, la matriz de datos originales se transforma en una matriz de varianza-covarianza S; si las unidades difieren , los valores se estandarizan mediante el empleo de la matriz de correlación R.

El primer componente principal se define como aquella combinación lineal de variables que tenga la máxima varianza de todas las funciones derivadas de un dado conjunto de variables. El segundo componente principal, por la combinación lineal de variables que tenga la máxima varianza de todas las funciones lineales de las variables y que sea ortogonal al primer componente. Los coeficientes de cada componente se denominan cargas o pesos (loadings) y las mediciones de cada componente principal sobre cada uno de los individuos, se denominan marcas (scores) del componente. La matriz de cargas L es una matriz (p,p) y la de marcas F de (n,p) siendo n el número de observaciones y p el de variables. Veamos ahora que representan las marcas y cargas.

Si cada columna x_i de la matriz de datos X se representa como un vector en u espacio de p dimensiones, cada fila de X son las coordenadas de las n observaciones en términos de los p vectores que representan a los x_i. Estos vectores x_i pueden ser reemplazados por un nuevo juego de vectores ortogonales, f_i, sin variar por ello la posición relativa de los n puntos. Haciendo referencia a la Fig. 3, dos vectores x_i oblicuos, que representan las variables observadas (ejes WX y YZ) son reemplazados por los ejes ortogonales AB y CD.

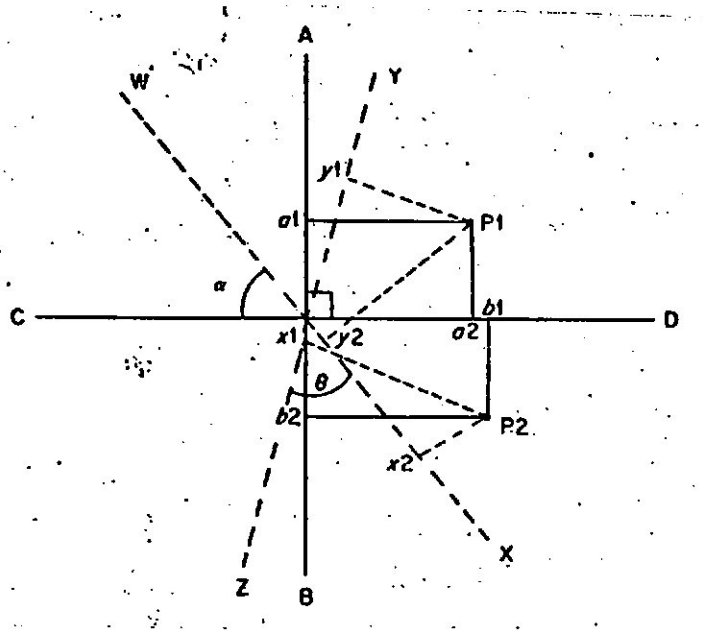


Fig. 3

Un punto P1 tiene coordenadas (y_1,y_2) en los ejes oblicuos y (a_1,a_2) en el nuevo sistema ortogonal. La proyección de los puntos en el eje AB son los elementos de un vector f_i que corresponde a la i-ésima columna de F. Esta proyección corresponde a las marcas de los objetos representados por los puntos sobre las variables observadas (ejes oblicuos) o variables compuestas (ejes ortogonales). Si AB y CD son componentes principales, entonces las proyecciones de P1 y P2 sobre estos ejes corresponden a las marcas de estos puntos en los componentes principales f_{ij} y los ángulos entre los ejes ortogonales y oblucuos corresponden al peso del componente principal. De esta maner, cos α es el peso de la variable representada por el eje WX en el componente principal representado por el eje ortogonal CD. El coseno del ángulo θ entre

pares de vectores es numéricamente equivalente al coeficiente de correlación entre las variables o variables compuestas involucradas si los vectores tiene longitud unitaria. En este caso cos θ es igual a la correlación momento-producto de las variables representadas por los ejes YZ y WX.

Un problema que se plantea es la posición de los ejes que deben ser seleccionados de tal manera que la variación a lo largo de cada eje sea la máxima posible, siempre con la restricción de la condición de ortogonalidad. Así, el eje I representa la mayor y más importante dimensión de variabilidad, seguido a continuación por el eje II. En la Fig.4 se graficaron un conjunto de observaciones sobre las que se han realizado mediciones en el plano que contiene a los dos ejes principales. La dispersión a lo largo del eje I es mayor que en el eje II. La utilización del criterio de máxima varianza permite la selección de un único juego de ejes ortogonales que son los ejes principales de un cuerpo geométrico que encierra la nube de puntos. Es útil recalcar que la forma de este cuerpo dependerá del peso dado a cada variable que variará según se parta de una matriz de dispersión (varianza-covarianza) o de una de correlación. Independientemente de que se parta de una matriz S o R , los autovectores (eigenvectors) de esas matrices corresponden a los ejes principales de la matriz de datos X. La raiz cuadrada de los autovalores (eigenvalues) constituyen el factor de escala de los autovectores. La dispersión o varianza a lo largo de un eje principal la mide el autovalor. Si la matriz está estandarizada (R) su traza es igual a p (número de columnas) y es igual a la suma de los autovalores de R. Por lo tanto, cada autovalor brinda la proporción de varianza asociada a una determinada dimensión. La traza de la matriz S, en cambio, da la varianza total en las unidades originales.

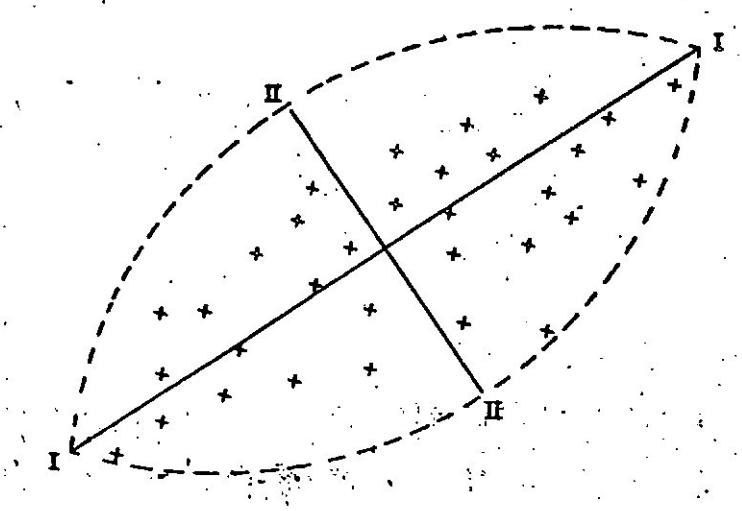


Fig. 4

Supongamos que X es la matriz de datos (n.p). Cada columna $x_{.j}$ de X debe transformarse en un nuevo conjunto de variables $f_{.j}$ que posea la propiedad de ortogonalidad y máxima varianza y que resulta de la combinación lineal de los $x_{.j}$. Es decir

$$f_{ij} = x_{i1}\, a_{11} + x_{i2}\, a_{21} + \ldots\ldots + x_{ip}\, a_{p1} \quad (i = 1,2,3,\ldots,n) \quad (5)$$

o en forma matricial

$$F = X A \quad (n.p)\ (n.p)\ (p.p) \qquad (6)$$

F representa la matriz de las marcas de los componentes principales o sea sus coordenadas y A la matriz de los pesos de los componentes principales donde se agrupan los autovectores. Como se observa en la ecuación anterior, para una muestra i donde se han realizado p observaciones, cada variable tendrá un peso distinto en el nuevo eje vinculado al valor de los a_{ij}. Cuanto más alto sea, la variable a la que afecta tendrá una mayor contribución. Pero se había mencionado que existen dos restricciones, de máxima varianza y ortogonalidad, que deben ser respetadas. Por lo tanto deben encontrarse los valores de a_{ij} que satisfagan ambos requicitos.

La condición de máxima varianza se logra eligiendo los elementos a_{ij} tales que la suma de cuadrados de los elementos de la matriz F sean máximos.

$$F' F = A' X' X A = \text{máximo} \qquad (7)$$

sujeta a la condición de ortonormalidad

$$A' A = I \quad o \quad a_{ij} a_{ik} = \delta_{jk} \qquad (8)$$

El símbolo ' indica transposición, I es la matriz identidad y δ es el delta de Kronecker.

Estas dos condiciones se cumplen cuando los vectores columnas $a_{.j}$ son los autovectores de longitud unitaria correspondientes al autovalor más grande de la matriz X'X. El autovector 1 corresponde al autovalor más alto, el autovector 2 al valor siguiente y así sucesivamente en orden decreciente de magnitud.

Volviendo a la ecuación (7) y siendo A la matriz de los correspondientes autovectores unitarios, se puede demostrar que la varianza de los componentes principales $f_{.j}$ es numéricamente igual a los elementos d_i de la matriz diagonal D de autovalores de X'X. La suma de la varianza es entonces la traza (D).

Dada una matriz simétrica cuadrada P se puede expresar en función de sus autovalores D y sus autovectores A como

$$P = A D A' \qquad (9)$$

X'X es una matriz cuadrada simétrica (ya sea R o S) y por lo tanto substituyendo (9) en (7)

$$F'F = A' (A D A') A \qquad (10)$$

y como A es una matriz ortonormal A A'= A' A = I por lo que (10) se puede expresar como

$$F'F = I D I = D \qquad (11)$$

El porcentaje de la varianza total debida a un componente principal i es

$$\frac{d_i \cdot 100}{\text{traza } D} \qquad (12)$$

A partir de los elementos de la matriz de autovectores normalizadosse puede calcular la correlación entre un componente principal f_j y la variable original x_i

$$l_{ij} = \frac{(d_j)^{1/2} \cdot a_{ij}}{(\Sigma x_{ik}^2)^{1/2}} \qquad (i,j = 1,2,\ldots,p) \qquad (13)$$

Los elementos l_{ij} se ordenan en una matriz L que generalmente se denomina matriz de pesos de componentes principales. El cuadrado de l_{ij} denota la proporción de la varianza de x_i contribuída por el componente principal j. Para un i dado la suma de los cuadrados de $l_{ij} = 1$.

Como en general se parte no de la matriz de datos originales sino de sus valores estandarizados Z, la matriz X'X se convierte en la de correlación R y por lo tanto como el denominador de (13) es igual a la unidad, la matriz L se calcula

$$L = A\, D^{1/2} \qquad (14)$$

con A, la matriz de autovectores de longitud unitaria y $D^{1/2}$ la raíz cuadrada de los autovalores. Al multiplicar se llevan los vectores a sus longitudes originales.

A veces las columnas de F se estandarizan para que su varianza sea unitaria, obteniéndose una matriz F*

$$F^* = Z\, L\, D^{-1} \qquad (15)$$

y según (14) la expresión (15) queda

$$F^* = Z\, A\, D^{-1/2} = F\, D^{-1/2} \qquad (16)$$

es decir, F* resulta de la postmultiplicación de la matriz de datos estandarizados Z por la matriz A de autovectores normalizados y postmultiplicado este producto por la matriz diagonal de la recíproca de la raíz cuadrada de los autovalores d_i (= se divide cada componente por su desvío estandard).

Con los mismos datos del ejemplo para el análisis de agrupamiento se presentan en la tabla siguiente las cinco variables que contribuyen en mayor proporción a los tres primeros factores. El primer factor, vinculado a los iones mayoritarios, explica el 62,5 % de la varianza total. El componente 2 (13,2 % de la varianza total) destaca una segunda variabilidad de las muestras vinculada a procesos químicos espacialmente circunscriptos (ambiente reductor correspondiente al antiguo fondo de la laguna Corrientes). En la figura 5 se ubicaron las muestras en el plano de los dos primeros componentes.

Variable	Unidad	Rango	Contribución a los Componentes I	II	III
COMPONENTE I					
Residuo seco	mg/l	525,0 -3430,0	-.951	-.260	.111
$Na^+ + K^+$	mg/l	80,0 -1293,0	-.895	-.203	.311
HCO_3^-	mg/l	405,0 -2196,0	-.884	-.028	.123
Cl^-	mg/l	64,0 -1120,0	-.869	-.319	.140
SO_4^{-2}	mg/l	12,0 - 630,0	-.840	-.306	.006
COMPONENTE II					
Mn^{+2}	mg/l	0,02- 3,0	-.649	.551	.028
Fe^{+2}	mg/l	0,05- 2,0	-.667	.533	.001
pH		7,3 - 8,2	.242	-.497	.695
NH_4^+	mg/l	0,02- 2,2	-.195	-.457	-.494
NO_3^-	mg/l	0,0 - 65,0	.620	-.385	-.144
COMPONENTE III					
pH		7,3 - 8,2	.242	-.497	.695
Ca^{+2}	mg/l	17,0 - 224,0	-.438	-.369	-.633
NH_4^+	mg/l	0,02- 2,2	-.195	-.457	-.494
$Na^+ + K^+$	mg/l	80,0 -1293,0	-.895	-.203	.311
F^-	mg/l	0,05- 2,8	-.052	.104	.310
Mg^{+2}	mg/l	17,0 - 249,0	-.623	.258	-.307
EIGEN-VALOR			5,8789	1,7114	1,4827
PORCENTAJE DE TRAZA			45,22	13,16	11,41

Tabla 1. Contribución a los componentes de las cinco variables de mayor peso

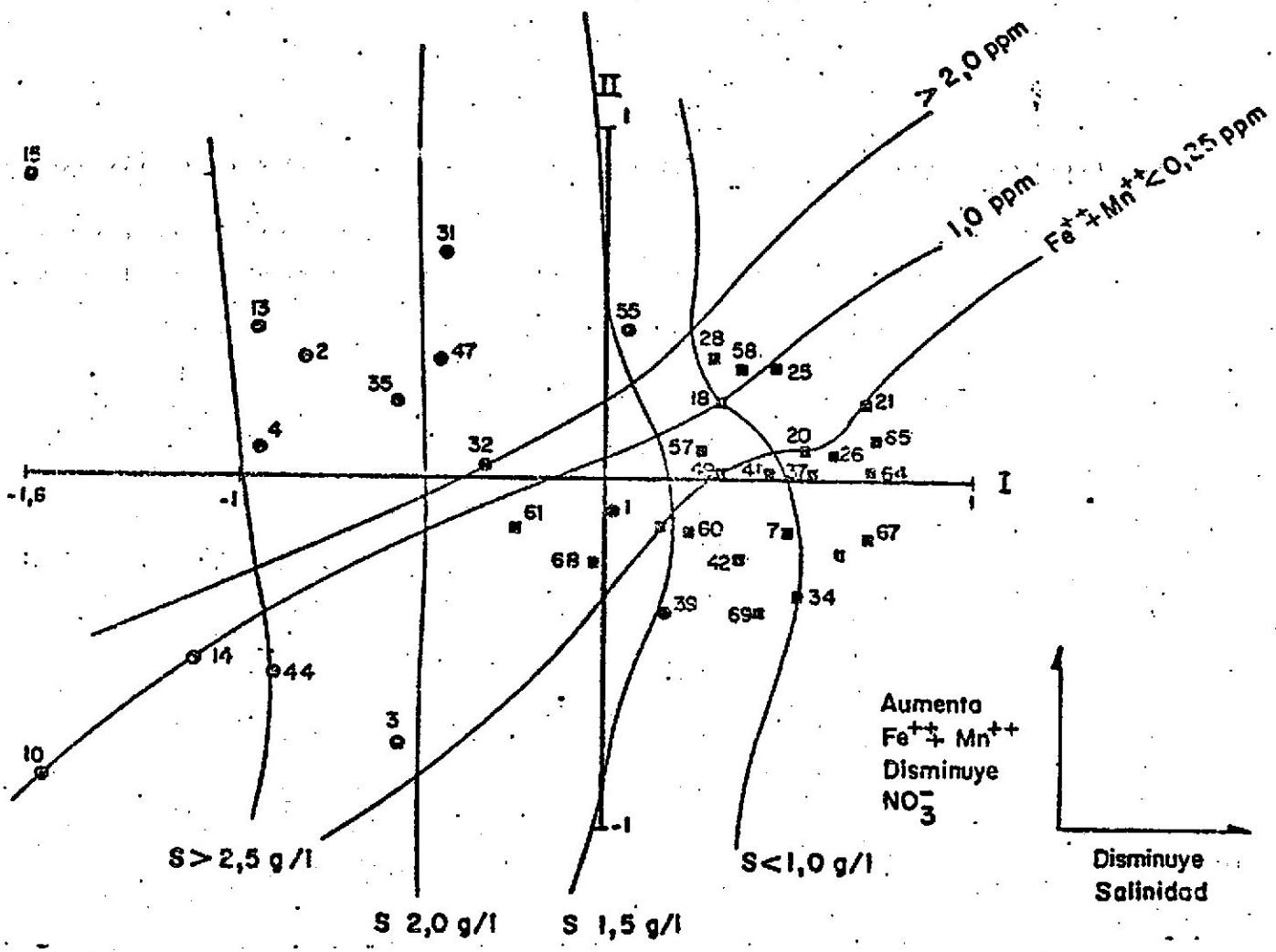


Fig. 5 Proyección de las muestras en el plano de los componentes I y II.

GEOESTADISTICA

Emilia Maria Bocanegra

APLICACION DE LA GEOESTADISTICA A LA HIDROLOGIA SUBTERRANEA

INTRODUCCION

La teoría de transporte de flujo y de transporte de masa en un medio poroso es la más utilizada en Hidrología Subterránea para el estudio del comportamiento de diversos parámetros. Se basa en considerar un volumen infinitesimal de referencia y representar sus variaciones espaciales y temporales por medio de funciones continuas que cumplen las leyes de la mecánica del continuo.

Se puede aplicar la teoría de campos para las variables y parámetros hidrogeológicos ya que representan las propiedades continuas del medio poroso tales como la porosidad θ, la conductividad hidráulica K, la tramisividad T, el coeficiente de almacenamiento S y la dispersividad D.

Los principios de dinámica de fluidos en medios porosos se expresan por medio de ecuaciones en derivadas parciales considerando que sus parámetros son funciones deterministicas en el espacio. La solución de estas ecuaciones se obtiene mediante modelos matemáticos, siendo los más usados el de diferencias finitas y de elementos finitos, que asignan valores medios de los parámetros a las celdas y elementos respectivamente. Pero los fenómenos hidrogeológicos presentan una variabilidad espacial errática a la que hay que añadir la variabilidad que introducen las determinaciones experimentales. Si además estos valores son empleados para estimar los parámetros en otras zonas del acuífero donde no fueron determinados, se introduce una incertidumbre adicional. La única forma de tener en cuenta la aleatoriedad de la variabilidad de los parámetros hidrogeológicos y la incertidumbre asociada con la insuficiente información sobre su distribución espacial, es mediante su interpretación probabilística en contraste con su descripción deterministica. De esta forma cada parámetro es interpretado como una variable aleatoria que puede adoptar un conjunto infinito de valores de acuerdo con una distribución de probabilidad.

Ahora bien, el estudio de las variables aleatorias es el objetivo de la estadística clásica. Son sus propiedades esenciales la repetición indefinida de un test que asigna un valor numérico a la variable y la independencia de cada test respecto al previo. (Ejemplo, el juego de cara o cruz al tirar una moneda). Surge claramente que las variables hidrogeológicas no cumplen con estas 2 propiedades, dado que cada muestra es única y no es independiente de las muestras vecinas. Analicemos estos conceptos para el caso de un muestreo hidroquímico: la muestra se extrae una sola vez y no es posible repetir el test. Es posible extraer una segunda muestra y una tercera, pero no tenemos la certeza que tenga exactamente la misma composición química que la primera. No obstante podríamos asumir que existe la posibilidad aparente de repetir el test, aunque el test no sea el mismo; es levemente diferente porque las muestras se obtienen de coordenadas ligeramente diferentes.

No obstante asumir la posibilidad de repetición, la 2a. propiedad no puede ser respetada; dos muestras vecinas no son independientes, por ejemplo, tienden a altas conductividades eléctricas si provienen de una zona salinizada. Esta tendencia expresa el grado de continuidad de la variación de la conductividad en el agua subterránea de la región.

Igual consideración podría hacerse para cualquier variable hidrogeológica. Es decir que en general estas variables presentan un cierto grado de continuidad debido a un control geológico, de manera que no son tan caóticas como para eliminar la posibilidad de

una estimación, ni tan regulares como para permitir el empleo exclusivo del método determinístico.

En la solución de la ecuación del flujo subterráneo mediante métodos numéricos, los parámetros del modelo se tienen que estimar a partir de valores medidos en un número finito de puntos del acuífero. Dada la incertidumbre asociada a este proceso de estimación, los parámetros estimados se convierten en variables aleatorias, el modelo numérico es estocástico y sus predicciones son inciertas.

Para minimizar la incertidumbre de las predicciones del modelo, los parámetros deben ser estimados de manera tal que la varianza de los errores de estimación sea mínima.

Cuando se emplean modelos de flujo determinísticos o estocásticos, es siempre necesario analizar los estadísticos de la variación espacial de los datos.

A tal efecto la Geoestadística ofrece la posibilidad de analizar la variabilidad espacial de los parámetros y variables hidrogeológicas.

La Geoestadística, de acuerdo con Matheron (1963) es la aplicación de las variables aleatorias al reconocimiento y estimación de los fenómenos naturales y se basa en considerar tanto la dispersión de las muestras como el carácter espacial de su distribución.

En hidrología subterránea, los objetivos más importantes de esta ciencia son:

- La estimación de variables y parámetros hidrogeológicos
- El diseño óptimo de redes y campañas de muestreo

VARIABLES REGIONALIZADAS

Una variable distribuida en el espacio y con una estructura espacial, se dice que está regionalizada.

Es frecuente observar en una variable regionalizada, dos aspectos complementarios y aparentemente contradictorios: 1) un aspecto aleatorio asociado con las variaciones erráticas e impredecibles de la variable regionalizada de un punto a otro, y 2) un aspecto general que refleja en cierta forma las características estructurales del fenómeno regionalizado. La Fig. 1 que muestra el contenido de dureza en el agua subterránea de una cuenca, ilustra estos conceptos.

Las **variables regionalizadas** se caracterizan por tener una distribución geográfica que depende de su continuidad espacial y que puede ser evaluada determinísticamente, así como una componente estocástica que determina su valor en cada punto. Sus características cualitativas fundamentales (Matheron, 1963) son: la **localización**, dado que ocupan una posición en una **región** dentro de la cual cumplen con ciertas hipotesis y se manifiestan físicamente a través de un **soporte geométrico**, (i.e. volumen de la muestra), la **continuidad** en su variación espacial expresada a través de diferencias más o menos importantes entre los valores de dos muestras vecinas y la **anisotropía** dado que existe una dirección preferencial a lo largo de la cual la variable no varía significativamente, mientras que lo hace rapidamente según una dirección transversal.

VARIOGRAMA Y ANALISIS ESTRUCTURAL

Los diferentes aspectos de la distribución espacial de la variable regionalizada pueden estudiarse por medio de una herramienta matemática, el **variograma**, y viene expresado por (Journel & Huijbregts, 1978)

$$r(h) = 1/2 * E\{[Z(x+h) - Z(x)]^2\}$$

siendo: h = vector distancia x = posición de una observación puntual Z = valor de la variable en una posición. Las variables regionalizadas son funciones aleatorias que responden localmente a una determinada distribución de probabilidad y regionalmente a una cierta estructura espacial.

Si consideramos por ejemplo, que en un punto determinado x_1 de un acuífero fue medido el potencial hidráulico de 10 m, es decir $z(x_1) = 10$ m, esta cantidad es la realización particular de cierta variable aleatoria en el punto x_1. El grupo de todos los niveles piezométricos $z(x_1), z(x_2),..., z(x_n)$, para todos los puntos x_1, $x_2,...,$ x_n dentro del acuífero, puede ser considerado como una realización particular del conjunto de variables aleatorias. Este conjunto de variables aleatorias es la **función aleatoria Z(x).**

La geoestadística interpreta cada valor $z(x_i)$ como una realización particular de una variable aleatoria, que a su vez forma parte de cierta función aleatoria Z(x). Para que esta interpretación probabilística tenga sentido, según señalan Journel & Huijbregts (1978), es necesario inferir estadísticamente la función de distribución de la variable aleatoria; ésto no es posible dado que se cuenta con una sola realización y es imposible obtener más de una debido a la falta de repetición de la variable regionalizada. Para resolver el problema es necesario introducir ciertas hipotesis relacionadas con la función aleatoria Z(x).

La inferencia de una ley de probabilidad de una función aleatoria requiere asumir que existe un cierto grado de homogeneidad espacial,por ejemplo, suponer que la función aleatoria es estacionaria puede pensarse como equivalente a que la función aleatoria se "repite" en el espacio, y que esta "repetición" proporciona la información equivalente a muchas realizaciones de la misma función aleatoria Z(x), permitiendo la posibilidad de inferencia estadística. Esta es la hipótesis de estacionariedad estricta según la cual la ley de distribución de la función aleatoria es invariante respecto a la traslación del vector h.

En Geoestadística lineal se emplea un tipo de estacionariedad de 2º orden llamada hipótesis intrínseca. Una función aleatoria intrínseca, es aquella cuyos incrementos $|Z(x+h) - Z(x)|$ tienen esperanza matemática y varianza definidas e independientes de x (Girardi,1987)

$$E[Z(x)] = m$$

$$Var[Z(x+h) - Z(x)] = E\{[Z(x+h) - Z(x)]^2\} = 2\,r(h)$$

La Geoestadística, además de la esperanza matemática y el variograma, aplica a las variables regionalizadas los operadores varianza o varianza a priori y covarianza; que bajo hipotesis de estacionariedad de 2º orden, toman la forma:

$$Var[Z(x)] = E\{[Z(x) - m]^2\} = C(0)$$

$$Cov[Z(x), Z(x+h)] = E[Z(x).Z(x+h) - m^2] = C(h)$$

pudiéndose demostrar que:

$$r(h) = C(0) - C(h)$$

Esta ecuación revela que tanto el variograma como la covarianza definen la estructura de autocorrelación entre las variables aleatorias Z(x+h) y Z(x).

El **análisis estructural** es el estudio del comportamiento de la variable regionalizada. Su variabilidad espacial está reflejada por el semivariograma experimental calculado a partir de las observaciones puntuales, en el que se grafica una distancia h en abcisas y en ordenadas el valor medio del cuadrado de la diferencia entre dos muestras obtenidas a una distancia h una de la otra (Matheron, 1963).

El variograma representa las siguientes características de regionalización de una variable geológica:

a) **Zona de influencia** a partir de la cual la autocorrelación es nula. Se puede definir el rango cuando el variograma se aproxima asintóticamente a una meseta, como la separación mínima entre pares de observaciones estadísticamente independientes (Fig.2).

b) **Continuidad:** la pendiente del variograma cerca del origen representa el grado de regionalización de la variable. Se pueden observar cuatro comportamientos diferentes (Fig.3).

1 - El variograma es **parabólico** en el origen y representa una variable regionalizada con alta continuidad, tal como el potencial hidráulico en acuíferos poco explotados en zonas de llanuras.

2 - Es el tipo **lineal**, caracterizado por una tangente oblicua en el origen, y representa una variable que tiene una continuidad media, es el tipo más común para tenores en depósitos metalíferos.

3 - En algunos casos aparece una discontinuidad en el origen, llamada **efecto pepita**, debido a que existen fuentes de variabilidad con rango menor a la distancia entre observaciones.

4 - Es el caso límite que corresponde a una variable aleatoria clásica.

La Fig 4. muestra ejemplos de variogramas con diferentes tipos de comportamiento en el origen. (Delhomme, 1978)

c) **Anisotropía:** se presenta cuando la función variograma depende de la dirección y el módulo del vector h.

Modelos de variogramas: Las dos características principales del variograma son su comportamiento al origen y la presencia o ausencia de una meseta para valores de h>a. Atendiendo a estas características , las principales funciones empleadas en la representación de variogramas son las siguientes (Fig.4):

1 - Modelos con meseta y tipo lineal al origen
- Esférico $\gamma(h) = C_0/2[3(h/a)-(h/a)^3]$ $0<h<a$
- Exponencial $\gamma(h) = C_0[1 - e^{-(h/a)}]$

2 - Modelos con meseta y tipo parabólico al origen
- Gaussiano $\gamma(h) = C_0[1 - e^{-(h/a)^2}]$

3 - Modelos sin meseta o monómicos
- del tipo $\gamma(h) = k.h^{\theta}$ $\theta \in [0,2]$
- logaritmico $\gamma(h) = C_1[\log h + C_2]$

4 - Modelos con crecimiento no monótono
- Efecto agujero $\gamma(h) = C_0[1 - (\text{sen } h)/h]$

El ajuste de un modelo teórico al variograma muestreal se hace por mínimos cuadrados o a sentimiento.

VARIOGRAMA DE TRANSMISIVIDADES

El estudio de la variabilidad espacial de la transmisividad está recibiendo una atención prioritaria debido a la aparición de modelos estocásticos y de las implicaciones de la variabilidad de la transmisividad en las ecuaciones de transporte de flujo.

Dado que las transmisividades parecen seguir una distribución log-normal, y a que el variograma de log T parece tener un mejor comportamiento que el de T en el sentido que permite reconocer mejor la existencia de una estructura espacial, que generalmente se trabaja con log T.

Delhomme (1978) cita el caso del acuífero de arenas eocenas en la Aquitania Francesa. Los variogramas obtenidos empleando T o log T se indican en la Fig. 6 , como puede apreciarse el variograma de T podría indicar una escasa correlación espacial, mientras que el de log T muestra una clara estructura.

Dado que las transmisividades suelen variar espacialmente con una gran amplitud, es que la mayoría de los variogramas de log T presentan efecto pepita. Esto se agudiza por la magnitud de los errores que se presentan en los ensayos de bombeo y por el hecho de que la base de datos de T suele ampliarse mediante una regresión lineal de log T con las capacidades específicas, normalmente disponibles en muchos más pozos. Ambos factores dan lugar a errores en las mediciones puntuales que se traducen en un efecto pepita.

La Fig. 7 muestra los variogramas experimentales y teóricos correspondientes a log T de cuatro acuíferos.

VARIOGRAMAS DE NIVELES PIEZOMETRICOS

Los niveles piezométricos constituyen una variable continua y derivable, salvo en los puntos en que haya un salto brusco de transmisividad. Al ser una variable muy regular de la que se suele disponer de una cantidad razonable de datos, los variogramas de niveles son continuos en el origen.

Sin embargo el flujo de agua subterránea impone una marcada tendencia por lo que los niveles no suelen ser funciones estacionarias. En consecuencia el variograma de niveles piezométricos suele estar no acotado.

Existe la posibilidad de restar la tendencia a los valores medidos y en este caso, los niveles suelen ser estacionarios y su variograma alcanza una meseta.

Otra propiedad de los variogramas de niveles es que la variación suele ser mucho mayor en la dirección paralela al flujo que en la perpendicular. Por tanto, los variogramas son anisótropos. Bocanegra y Fasano (1990) presentan esta propiedad del potencial hidráulico en el sector oeste de la cuenca de la laguna Mar Chiquita, prov. de Buenos Aires (Fig. 8).

KRIGEAGE ORDINARIO Y VARIANZA DE ESTIMACION

El krigeage ordinario es un procedimiento de estimación de una variable regionalizada en un punto $Z^*(x_0)$ en función de las observaciones próximas $Z(x_i)$ y de ponderadores geoestadísticos δ_i.

Las condiciones que se imponen a la estimación son:

- Linealidad: $Z^*(x_0) = \sum_{i=1}^{n} \delta_i \, Z(x_i)$

- No sesgo: $\Sigma \, \delta_i = 1$ o bien $E(Z^*) = E(Z)$

- Mínima varianza: $E|(Z^* - Z)^2|$ es mínima

Para optimizar la estimación de los ponderadores δ_i se minimiza el error empleando el método de los multiplicadores de Lagrange, lo que conduce a un sistema lineal de n+1 ecuaciones con n+1 incógnitas, n ponderadores y el parámetro de Lagrange μ. (David,1977; Journel & Huijbregts,1978),

$$\sum_{j=1}^{n} \delta_j \, \Gamma(x_i - x_j) + \mu = \Gamma(x_i - x_0)$$

$$\sum_{i=1}^{n} \delta_i = 1$$

donde:

$\Gamma(x_i - x_j)$ = semivarianza correspondiente a la distancia h entre los puntos x_i y x_j.

$\Gamma(x_i - x_0)$ = semivarianza correspondiente a la distancia entre los puntos x_i y x_0.

En la forma matricial, el sistema de ecuaciones de krigeage se expresa:

$$\begin{bmatrix} \Gamma(x_1,x_1) & \Gamma(x_1,x_2) & \Gamma(x_1,x_3) & \dots & \Gamma(x_1,x_n) & 1 \\ \Gamma(x_2,x_1) & \Gamma(x_2,x_2) & \Gamma(x_2,x_3) & \dots & \Gamma(x_2,x_n) & 1 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \Gamma(x_n,x_1) & \Gamma(x_n,x_2) & \Gamma(x_n,x_3) & \dots & \Gamma(x_n,x_n) & 1 \\ 1 & 1 & 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_n \\ \mu \end{bmatrix} = \begin{bmatrix} \Gamma(x_1,x_0) \\ \Gamma(x_2,x_0) \\ \vdots \\ \Gamma(x_n,x_0) \\ 1 \end{bmatrix}$$

La varianza de la estimación se calcula reemplazando los valores de los ponderadores obtenidos por krigeage en (Davis,1973):

$$\sigma_k^2(Z) = \mu + \sum_{i=1}^{n} \delta_i \, \Gamma(x_i - x_0)$$

El krigeage es un estimador exacto, esto es, si se trata de estimar Z en un punto de observación, el resultado será el valor medido, con incertidumbre nula.

En el sector oeste de la laguna Mar Chiquita se han calculado los niveles piezométricos por krigeage a partir de un modelo teórico lineal de variograma. La estimación de la superficie freática a partir de una malla regular es la misma que la que se deriva de los niveles observados con distribución irregular.

La varianza de la estimación muestra valores bajos especialmente en las zonas más densamente muestreadas. Los efectos de borde producen alta varianza de la estimación. (Fig.8).

DISEÑO DE REDES DE MUESTREO

En la mayoría de la ciencias aplicadas la obtención de datos suele ser un proceso lento y costoso. Por ello es difícil determinar la importancia de establecer los puntos de medida de forma que proporcionen la mayor información sobre la variable a estudiar.

Las ecuaciones de krigeage no dependen de los valores medidos de las variables, sino solamente de sus posiciones y del variograma, si dichas ecuaciones dependieran de los valores medidos, el interpolador no sería lineal. Por lo tanto la varianza de la estimación sólo depende del variograma y de los coeficientes de ponderación, solución de las ecuaciones de krigeage. El hecho de que la varianza pueda calcularse antes de hacer mediciones es una propiedad extraordinariamente útil para el diseño de redes de observación.

En conclusión, el krigeage es un excelente método para la localización de puntos de medición minimizando la incertidumbre.

En el estudio geoestadístico de la piezometría en Mar Chiquita, Bocanegra y Fasano (1990) concluyen que para el diseño de muestreo empleado el error medio es 1.6 m para distancias entre pozos de 2 a 2.5 km en promedio. Considerando que el variograma es una función lineal, la varianza de estimación es también una función lineal de la distancia entre pozos. Duplicando ésta, la varianza de estimación también se duplicará obteniéndose valores de 5 m o σ =2.7 m para pozos espaciados entre 4 y 5 km. Si se distancian entre 6 y 7 km, σ = 2.75 m. Es decir que duplicando o triplicando la distancia entre pozos lo que equivale a reducir el número de observaciones entre 4 y 9 veces, se obtiene un error en la estimación prácticamente coincidente con la equidistancia de 2.5 m, usualmente empleada en la cartografía IGM para zonas de llanura.

Es por esto que los costos de la densificación de la red de observación no compensarían la mejoría en la información, por el contrario una red con pozos distanciados en promedio 4.5 km daría resultados muy semejantes con un incremento del error estandard de 0.6 m. Este error es posible disminuirlo, manteniendo constante la separación promedio entre pozos si éstos se ubican en forma más regular.

BIBLIOGRAFIA

BOCANEGRA,E.M. y J.L.FASANO, 1990. Continuidad espacial de variables hidrogeológicas en el sector sudeste bonaerense: Aplicación de la Geoestadística. XIV Congreso Nacional del Agua. Córdoba. (en prensa).

BRO,P.B., 1985. Aplicación de la Geoestadística a la modificación de una red de medición de niveles piezométricos. X Congreso Nac. del Agua. Mendoza.

DAVID,M. 1977. Geostatistical ore reserve estimation. Elsevier Scientific Publishing Co.

DAVIS,J. 1973. Statistics and Data Analysis in Geology. John Wiley & Sons.

DELHOMME,J.P., 1978. Kriging in the hydrosciences. Ads. Water Resour., 1(5). 251-266.

DUNLAP,L.E and J.M.SPINAZOLA. 1984. Interpolating water table altitudes in West Central Kansas using kriging techniques. U. S.Geol Survey Water Supply. Paper 2238.

GIRONDI,J.P.1987. Geoestadística. Universidad Nacional de San Juan. San Juan. 38 p.

JOURNEL,A.G. and C.J.HUIJBREGTS. 1978. Mining Geostatistics. Academic Press. London.

MATHERON,G. 1963. Principles of Geostatistics. Economic Geology. 58: 1246-1266.

MYERS,D.E., BEROVICH,C.L., BUTZ,T.R. y V.E. KANE, 1982. Variogram models for regional groundwater geochemical data. J. Int. Assoc. Math. Geol. 14 (6) :629-644.

OLEA,R.A., 1977. Measuring spatial dependance with semivariograms. Kansas Geol. Sur., Series on Spatial Analysis 3 Univ. Kansas: 29 p.

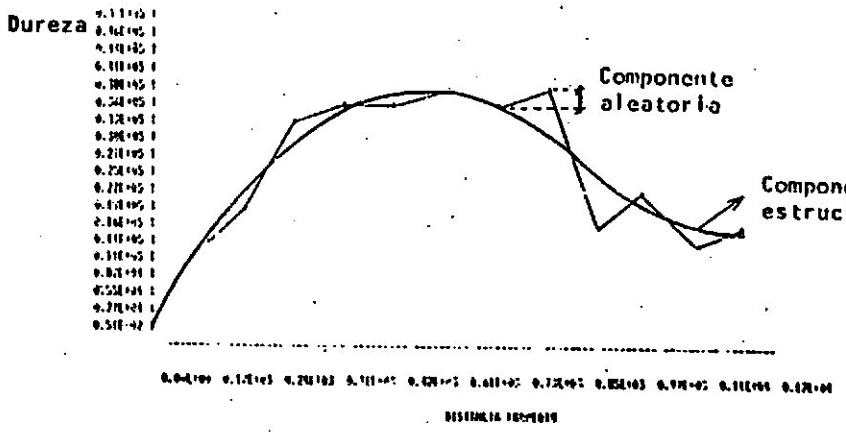


Figura 1

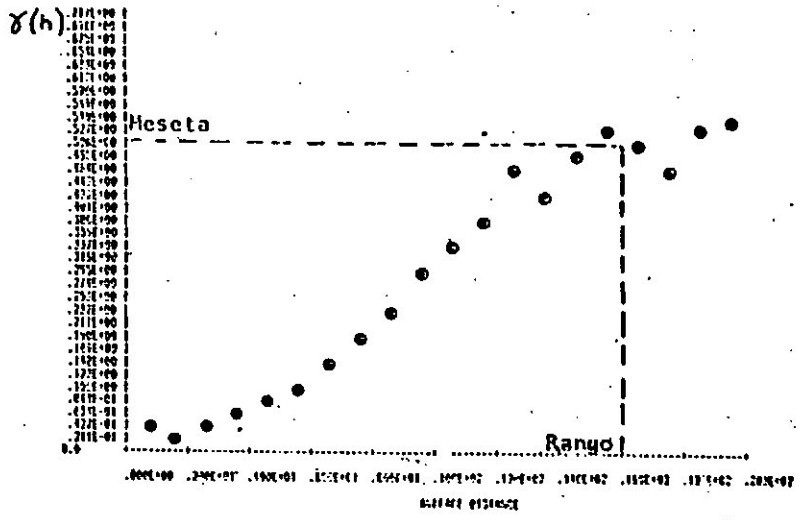


Figura 2

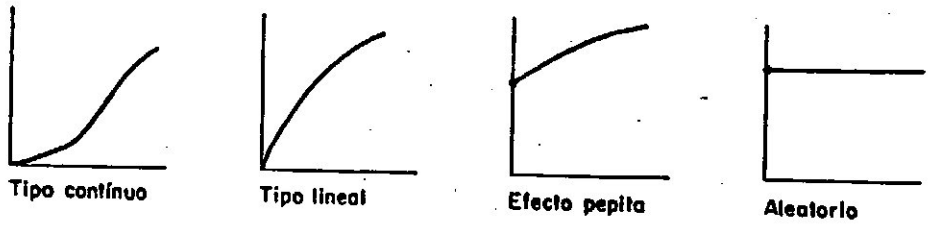
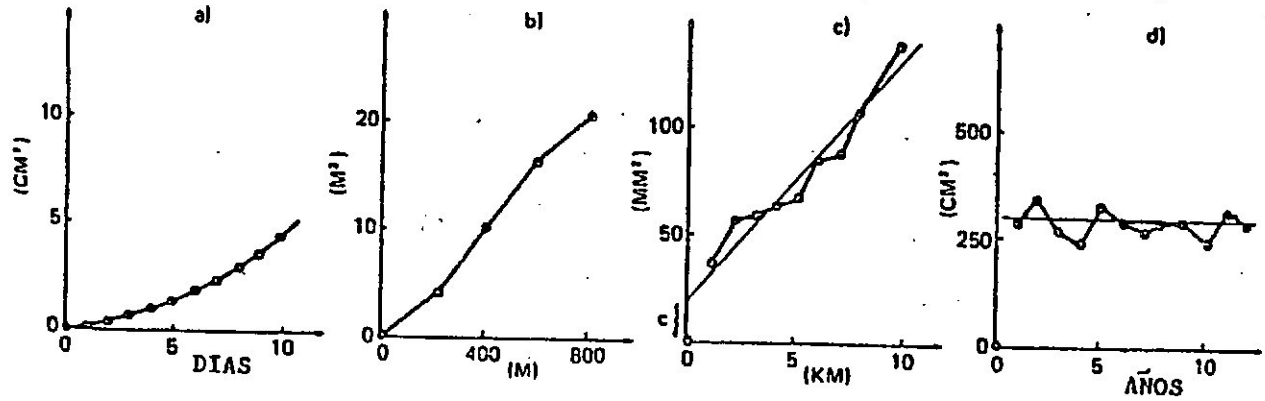


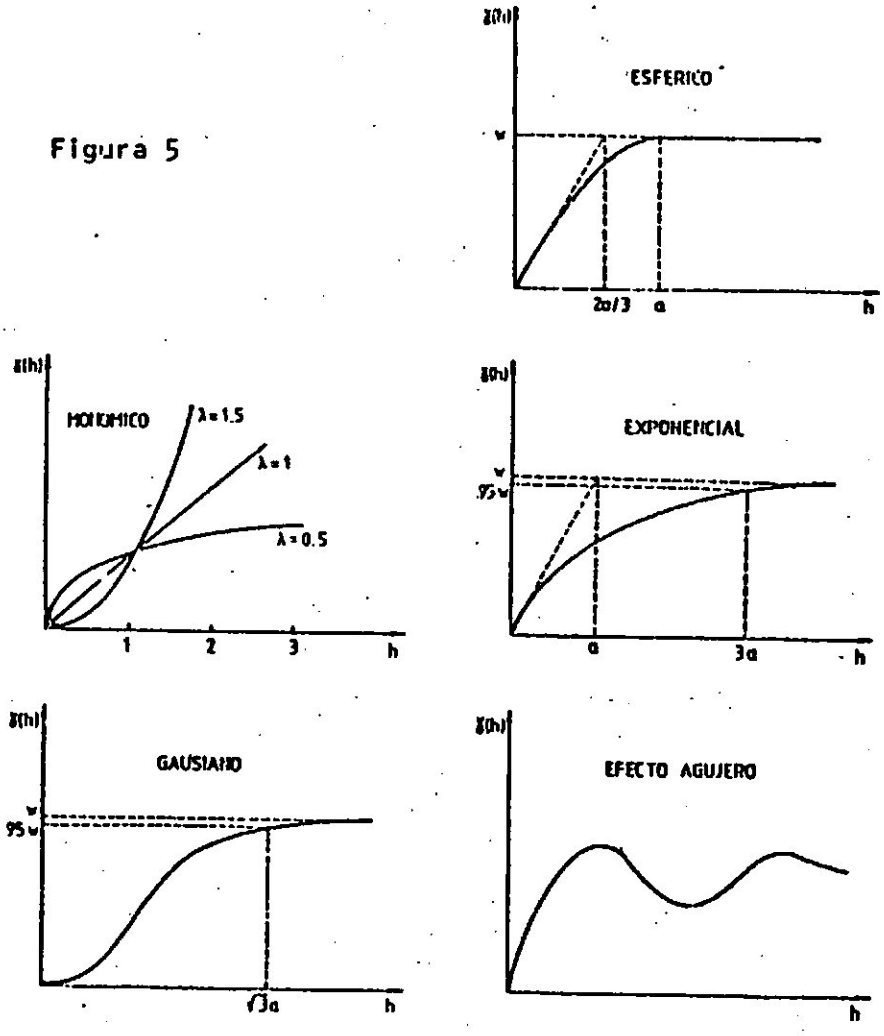
Figura 3

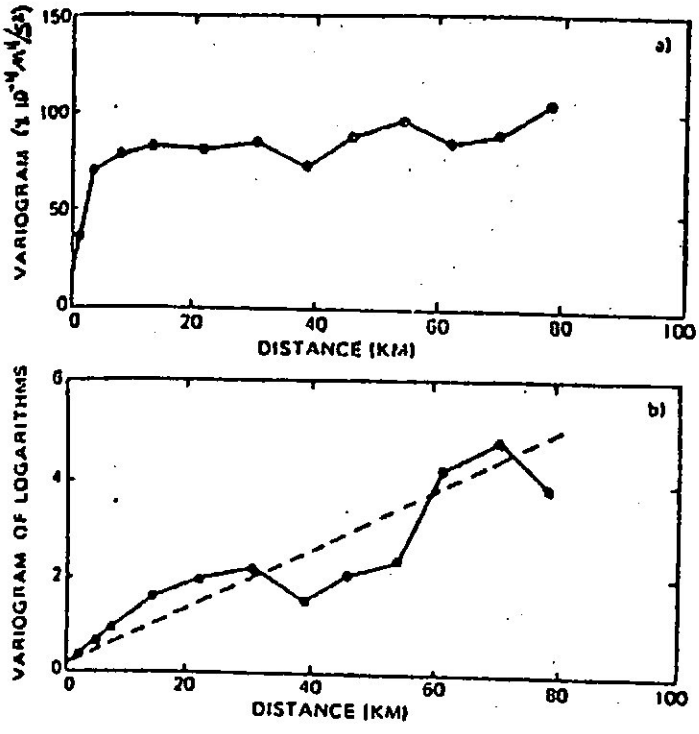


Ejemplos de variogramas con diferentes tipos de comportamiento en el origen: a) Niveles en un pozo profundo en función del tiempo; b) espesor de una formación geológica; c) precipitación en función de la distancia; y d) precipitaciones medias anuales. (Según Delhomme, 1978).

Figura 4

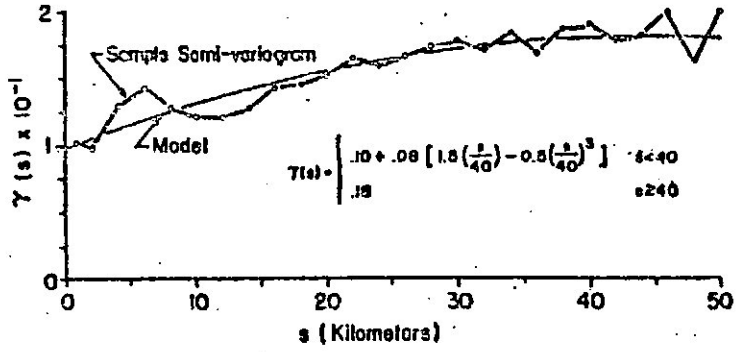
Figura 5

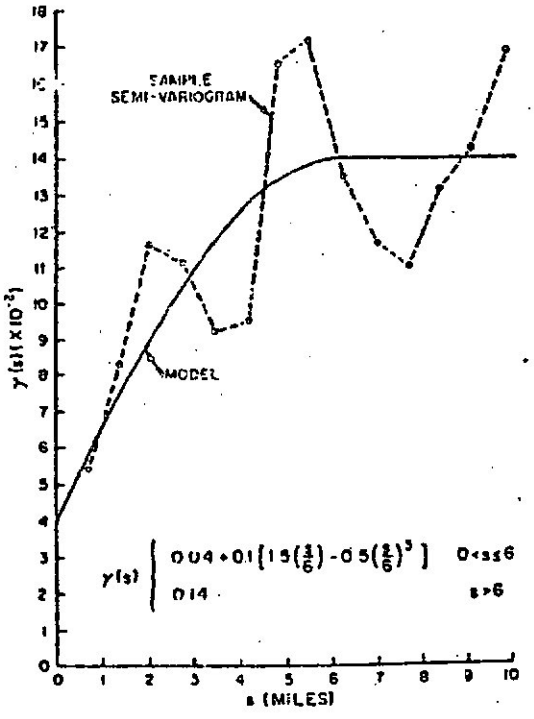


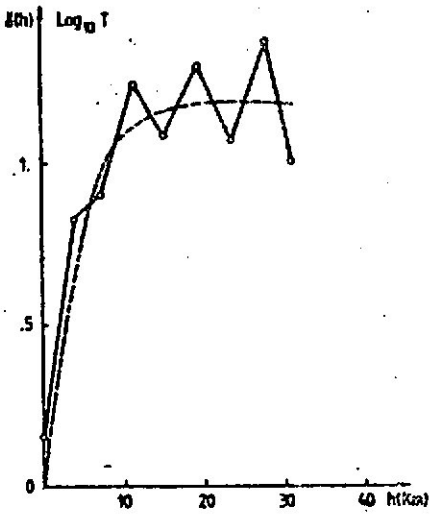


Influencia de la transformación logarítmica (Aquitania, Francia).
a) Variograma de transmisividades.
b) Variograma de log-transmisividades.
(Tomada de Delhomme, 1978).

Figura 6







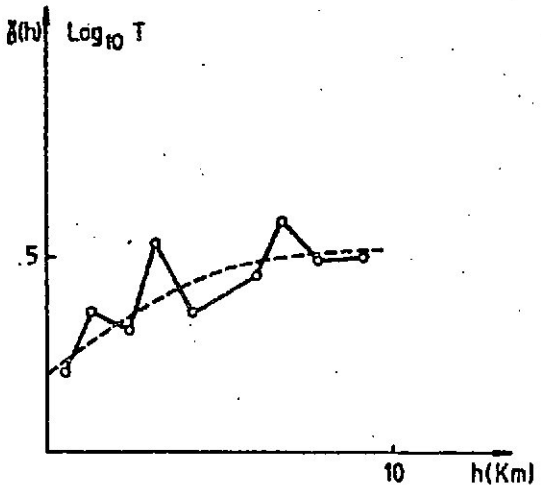


Figura 7

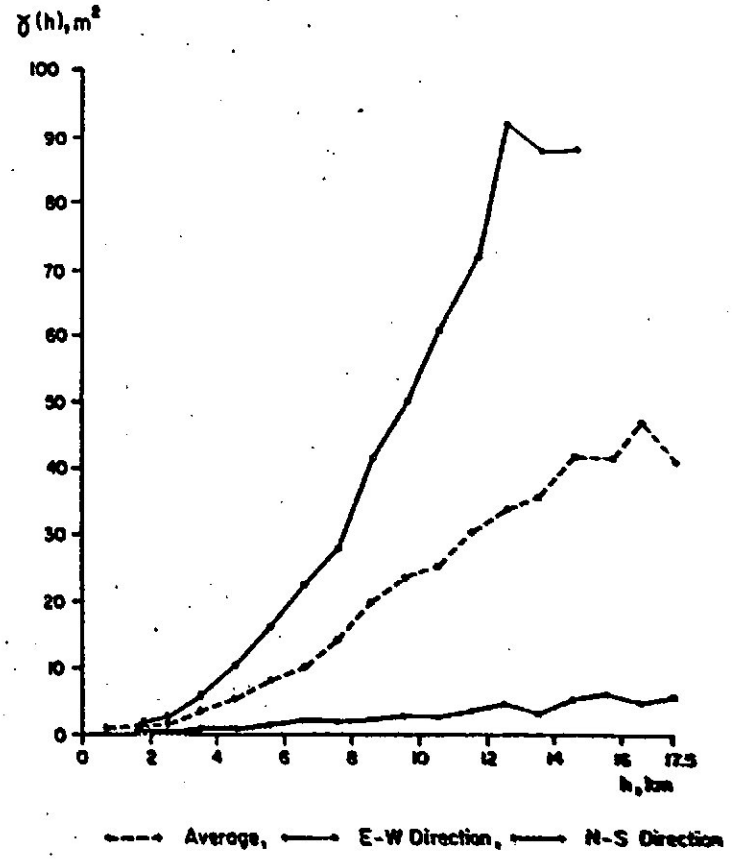
Variogramas experimentales y teóricos de log-transmisividades.
a) Cuenca del Tajo (Fennessy, 1982).
b) Avra Valley, Arizona (USA) (Clifton, 1981).
c) Acuífero Baton (De Marsily, 1980).
d) Acuífero calizo (De Marsily, 1980).

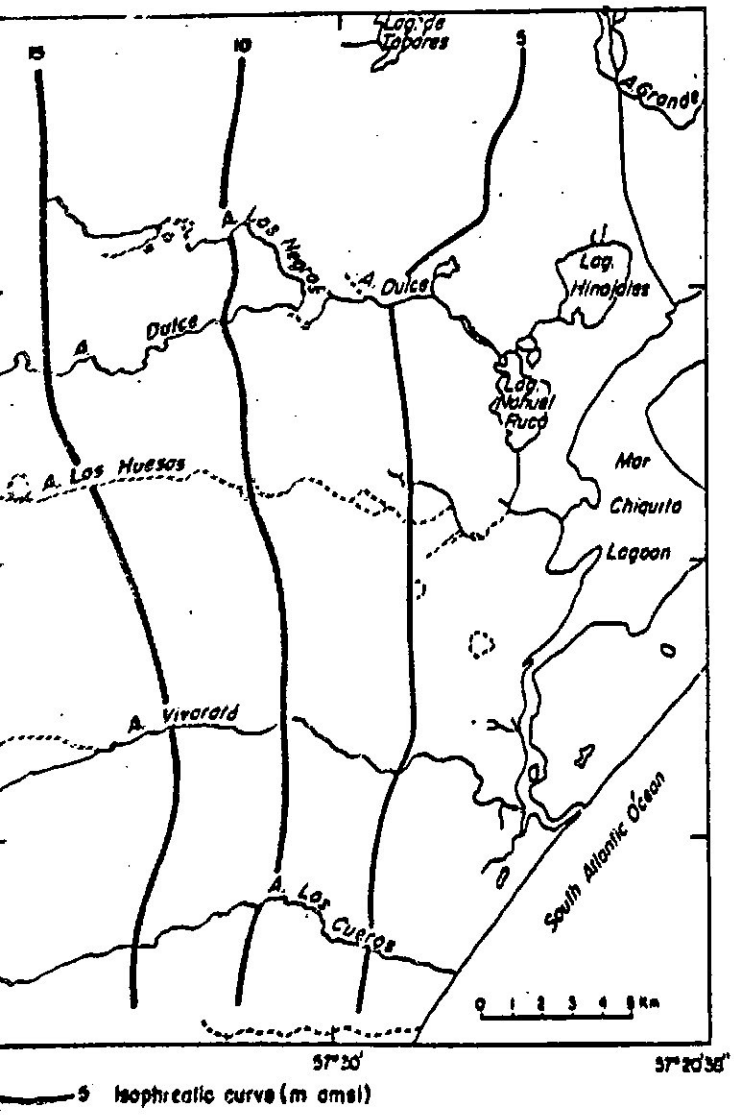
SECTOR OESTE DE LA CUENCA DE LA LAGUNA MAR CHIQUITA

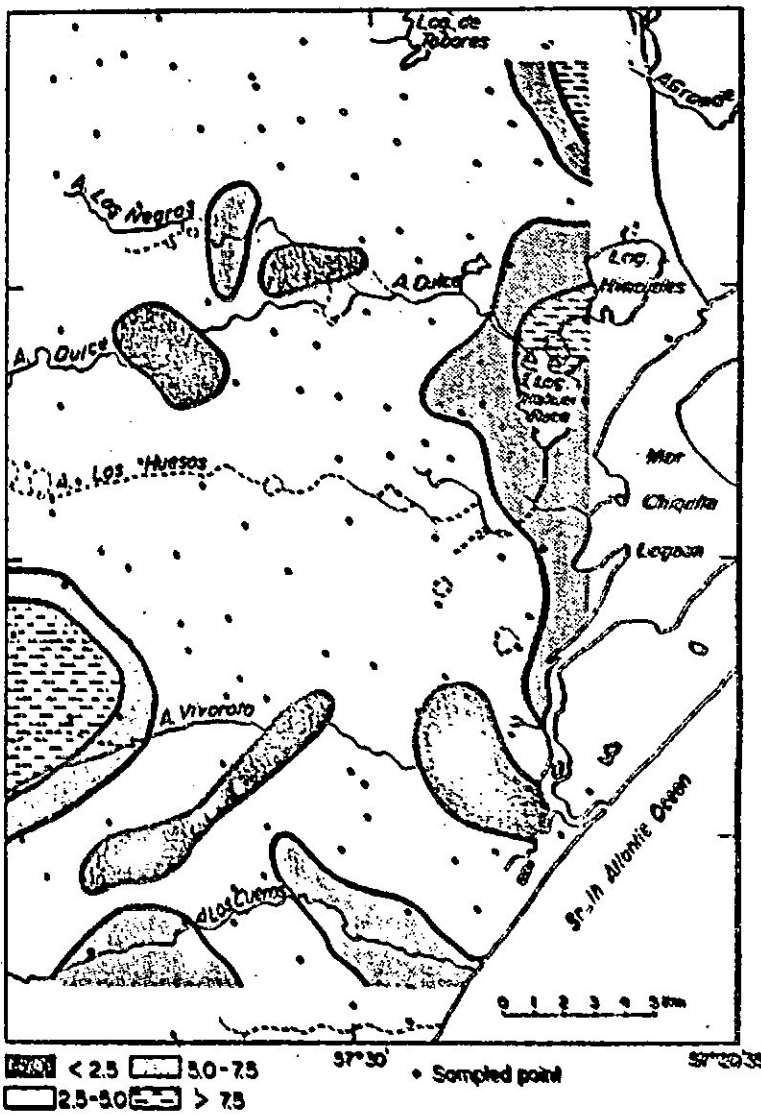
- Variograma Experimental
- Curvas isofreáticas obtenidas por krigeage
- Varianza de la estimación

(Bocanegra y Fasano,1989)

Figura 8







APENDICE I

EMPLEO DEL PROGRAMA DE CALCULO DEL VARIOGRAMA EXPERIMENTAL

El programa MAREC12 para el cálculo del variograma experimental de 12 variables regionalizadas ha sido adaptado por E.Bocanegra a partir del programa MAREC2.(David,1977).

Está escrito en lenguaje FORTRAN y ha sido implementado en una IBM PC. Tiene dos archivos para entrada de datos y salida de resultados y es interactivo para la selección del nombre de la variable (IDEF) y de su límite superior (BORN) a fin de desestimar valores altos que puedan suponerse erráticos.

Para la lectura de datos, las variables numéricas tienen formato libre. La descripción de variables es la siguiente:

Parámetros	Definición
ICOM	Titulos
ILOG	>0 ,variograma de log Z
IDIR	número de direccipnes (máx 10)
STEP	paso de la distancia h
IFIN	total de variables leidas (máx 12)
YCHEL	parámetros para caracterizar
XCHEL	la representación del variograma
PSI	ángulo de tolerancia (máx 10)
PHI	dirección
INOM	caracterización de la muestra
Y	coordenadas de las muestras
X	(máx 500)
Z	variable regionalizada
NVAR	número de variable leida

Se adjunta copia parcial de archivo de salida de resultados.

APENDICE II

CARACTERISTICAS DEL PROGAMA DE KRIGEAGE Y VARIANZA DE ESTIMACION

El programa KRIGE para el krigeage de datos sobre una red irregular bidimensional ha sido adaptado por E.Bocanegra para su empleo en una computadora HP1000, a partir del programa original de Per Bro (1985).

Dado la elevada capacidad de memoria requerida y a la incompatibilidad de los sistemas, no ha sido posible instalarlo en una computadora IBM PC. No obstante queda a disposición de los participantes un listado del mismo con la descripción de variables.

El programa estima la variable y la varianza sobre los puntos de una grilla rectangular partiendo de una red de datos irregulares, para diferentes modelos de variogramas teóricos: lineal, esférico, exponencial, gaussiano y doble esférico. Es necesario definir todos los parámetros del modelo teórico.

Se adjunta copia parcial de un archivo de salida.

V A R I O G R A M A

MAR CHIQUITA
NIVELES PIEZOMETRICOS
(CON UN CAMPO DE 180. GRADOS EN CADA DIRECCION)

hole
0.

STEP IN KM = 0.2000E+01

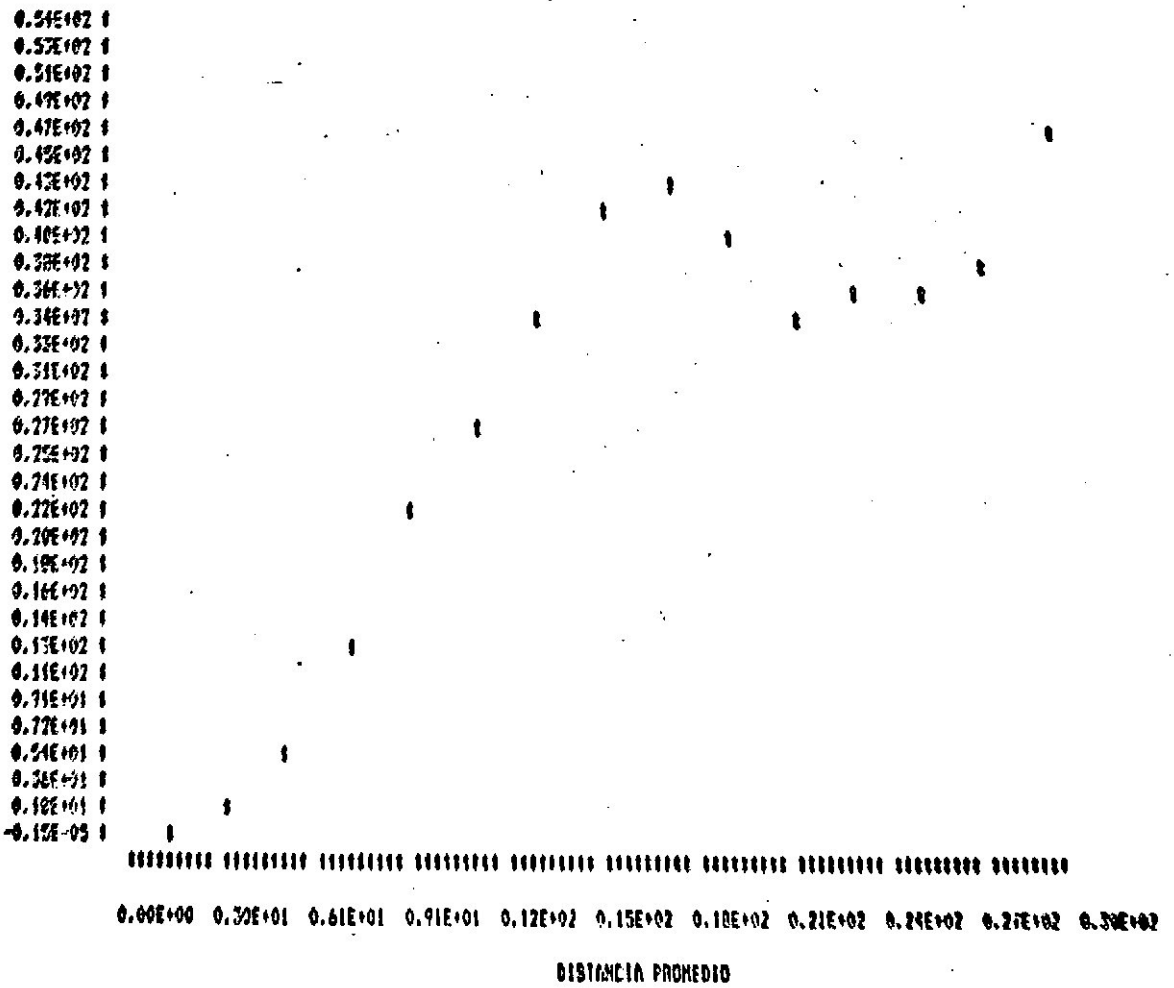
LIMITE SUPERIOR DE I = 0.4000E+02

MEDIA GENERAL DE I = 0.9184E+01

VARIANZA GENERAL DE I = 0.2960E+02

DISTANCIA EN KM	NO. OF PARES	DRIFT	VARIOGRAMA	DISTANCIA PROMEDIO
0 ---- 2	70	-0.348E+00	0.1113E+01	1.5
2 ---- 4	230	-0.169E+00	0.2963E+01	3.1
4 ---- 6	341	-0.324E+00	0.6833E+01	5.0
6 ---- 8	441	-0.243E+00	0.1280E+02	7.0
8 ---- 10	505	-0.748E-01	0.2218E+02	9.0
10 ---- 12	517	-0.591E+00	0.2893E+02	11.0
12 ---- 14	476	-0.973E-01	0.3580E+02	13.0
14 ---- 16	448	0.708E-01	0.4249E+02	15.0
16 ---- 18	379	0.171E+00	0.4417E+02	17.0
18 ---- 20	329	-0.687E+00	0.4145E+02	19.0
20 ---- 22	282	-0.152E+01	0.3500E+02	21.0
22 ---- 24	232	-0.125E+01	0.3775E+02	22.9

MAR CHIQUITA - NIVELES PIEZOMETRICOS



0065	117	4.10	32.30	17.10
0066	107	7.20	22.76	14.50
0067	64	7.82	22.20	13.20
0068	111	8.80	22.84	11.75
0069	109	9.20	24.00	11.55
0070	75	13.70	23.16	6.10
0071	19	15.40	23.40	4.10
0072	18	15.00	24.00	5.00
0073	17	17.00	24.00	1.60
0074	113	4.44	21.50	17.00
0075	110	7.38	24.80	14.55
0076	15	12.20	25.60	6.10
0077	16	15.46	25.30	2.24
0078	11	16.60	25.80	1.30
0079	15	17.34	25.60	1.50
0080	25	17.60	25.80	.80
0081	100	.60	26.40	21.60
0082	65	3.50	27.00	17.60
0083	98	5.54	27.80	16.00
0084	97	6.80	26.70	14.00
0085	12	12.24	26.76	6.00
0086	10	16.30	26.40	1.00
0087	8	16.96	26.20	.40
0088	9	10.74	27.40	1.90
0089	102	1.38	28.60	19.90
0090	93	8.72	28.70	10.60
0091	5	15.00	28.20	1.90
0092	4	15.84	29.00	2.20
0093	7	17.76	28.40	.25
0094	103	3.80	31.20	16.70
0095	95	7.90	30.00	9.10
0096	96	10.04	31.20	9.30
0097	101	1.80	33.20	19.20
0098	63	12.00	32.80	4.60
0099	2	13.76	33.40	.47

0100
0101
0102 CENTRO DE GEOLOGIA DE COSTAS
0103 MAR DEL PLATA, ARGENTINA
0104
0105 KRIGEAJE DE DATOS BIDIMENSIONALES
0106
0107
0108
0109 MODELO LINEAL
0110 GAMMA(H) = 0.000000 + 3.000000 H
0111 DISTANCIAS (H) EN KILOMETROS
0112
0113
0114
0115 RECTANGULO DE DATOS
0116 X MINIMO 1.6000 X MAXIMO 22.8400
0117 Y MINIMO .2600 Y MAXIMO 33.4000
0118
0119 RECTANGULO INTERPOLADO
0120 X MINIMO 1.6000 X MAXIMO 19.0000
0121 Y MINIMO 1.0000 Y MAXIMO 31.2000
0122
0123 RADIO DE BUSQUEDA 8.6722
0124
0125 NUMERO MINIMO DE PUNTOS REQUERIDO PARA UNA INTERPOLACION 3
0126 NUMERO MAXIMO DE PUNTOS REQUERIDO PARA UNA INTERPOLACION 10
0127
0128
0129
0130

NUMBER OF DATA POINTS RETAINED 98

MATRIX OF SEARCH RECTANGLES

1	5	8	11	16	16	20
23	27	30	35	38	43	47
49	53	55	58	61	64	66
69	72	75	77	80	84	93
94	94	96	97	97	97	99

KRIGLAJE DE PUNTOS

DISTANCIA ENTRE INTERPOLACIONES 1.9333
NUMERO DE FILAS EN LA GRILLA INTERPOLADA 16
NUMERO DE COLUMNAS EN LA GRILLA INTERPOLADA 18

XGLN	YGLN	TKRIGE	ESTVAR
1.60	31.20	18.79292	5.14561
3.53	31.20	16.66963	2.36306
5.47	31.20	14.34769	5.94775
7.40	31.20	10.77808	5.47590
9.33	31.20	9.44394	3.15534
11.27	31.20	7.01975	4.51205
13.20	31.20	3.79871	6.24472
15.13	31.20	1.84514	5.80260
17.07	31.20	1.14933	8.15150
19.00	31.20	.99784	13.27675
1.60	29.27	19.32936	3.24293
3.53	29.27	17.11056	5.33238
5.47	29.27	14.75179	5.26474
7.40	29.27	11.11593	3.52253
9.33	29.27	9.70684	3.00926
11.27	29.27	7.40007	6.37163
13.20	29.27	4.48536	6.49389
15.13	29.27	2.29078	3.88153
17.07	29.27	1.11302	3.73250
19.00	29.27	1.09085	7.19433
1.60	27.33	20.06395	3.71550
3.53	27.33	17.60660	1.73679
5.47	27.33	16.00184	2.16708
7.40	27.33	12.97426	3.50226
9.33	27.33	10.41468	5.14923
11.27	27.33	7.61744	4.62436
13.20	27.33	4.47402	4.23481
15.13	27.33	1.97473	3.24480
17.07	27.33	.70748	3.21569
19.00	27.33	1.74412	1.44191
1.60	25.40	20.02069	5.58296
3.53	25.40	17.82069	4.27811
5.47	25.40	16.03081	4.21499
7.40	25.40	14.12476	2.55715
9.33	25.40	11.00087	4.55260
11.27	25.40	7.74124	3.98745
13.20	25.40	5.04547	3.85411
15.13	25.40	2.42187	1.77718
17.07	25.40	1.49071	1.39232
19.00	25.40	1.05442	3.83362
1.60	23.47	19.13189	5.88344
3.53	23.47	17.30656	6.25581
5.47	23.47	15.94532	4.50325
7.40	23.47	14.25187	2.67781

ANALISIS DISCRIMINANTE SERIES TEMPORALES

Federico Ignacio Isla

ANALISIS DE DISCRIMINANTES

Y SERIES TEMPORALES

INTRODUCCION

La estadística convencional limita su análisis a dos o tres variables debido principalmente a inconvenientes de graficación. El análisis multivariado representaba originalmente un cálculo tedioso y de difícil comprensión.

Hoy día, con la ayuda de computadoras y "paquetes estadísticos" (BMDP, SPSS, SYSTAT, NTSYS, SAS), al menos el cálculo es más simple. Con sólo cerciorarnos de un correcto cálculo, nuestra tarea como usuarios se limita a comprender y aprovechar los resultados. Es propósito de este capítulo, explicar las pruebas clásicas, no en su procedimiento matemático, sino qué hacen, para qué sirven, cómo se usan o qué se les puede pedir. Estas son todas las preguntas que debe conocer un usuario novato, ya que existen numerosos condicionamientos y opciones de uso..., pero eso supera los objetivos de nuestro curso introductorio.

INTERVALO DE MEDICION

La utilidad de las variables depende de los distintos intervalos de medición:

Cualitativas

Nominales: Es el menor nivel de medición. Son simples nombres, sin nivel de comparación. Las dicotomías son un caso especial, donde el nivel de medición queda a consideración del investigador (Hull y Nie, 1981).

Ordinales: o Ranking. Hay un orden según categorías (grande, mediano, chico).

Cuantitativas

Intervalos: Las distancias entre las categorías son fijas. Por ejemplo, entre 30 y 31°C existe la misma distancia que entre 80 y 81°C, pero es imposible definir un 0°C porque es una convención física (que no necesariamente implica nada de calor).

Relaciones: Similar al nivel de intervalos, pero posee 0 físicamente definido.

ESTANDARIZACION

Cuando se aplica cualquier análisis multivariado utilizando variables cuantitativas, es conveniente estandarizar la matriz; es decir, reconvertir cada dato en una variación de su nivel medio según la fórmula

$Xijs = (Xij - Xi) / Si$, donde Xij: datos originales,
Xijs: datos estandarizados,
Xi: media de la variable i, y
Si: desvío standard de la variable i
(Crisci y López Armengol, 1983).

ANALISIS DE DISCRIMINANTES

Puede realizarse según dos o más grupos o variables nominales dependientes. A su vez, las variables independientes pueden ser ingresadas en el análisis DIRECTAMENTE o A PASOS ("stepwise"). Es conveniente efectuar previamente una prueba de independencia de variables (correlación), de modo de no ingresar al cálculo variables que puedan contener información semejante.

Análisis de discriminantes directo

El objetivo matemático de esta prueba es pesar y combinar linealmente ecuaciones discriminantes (factores) de tal modo que los grupos sean forzados a ser lo más estadísticamente diferentes (Fig. 1). De otro modo, es encontrar la línea que mejor separe los grupos en términos de las proyecciones de los centroides grupales (Mather, 1976).

Las funciones de discriminación siguen la fórmula:

$$D1 = d1 \cdot Zx + d2 \cdot Zy + \ldots d3 \cdot Zz$$

, donde D1 es el score de la función discriminante 1,

d = coeficiente de "peso", y

Z = valores estandarizados de las variables x, y, z.

El # de funciones discriminantes = # de grupos - 1,

= # de variables (si hay más grupos que variables) ó,

según criterio de tolerancia (o significación) prefijado.

Para probar la significación de la prueba (D de Mahalanobis, lambda de Wilks, o V de Rao) pueden utilizarse tests estadísticos. Los coeficientes de peso pueden ser interpretados como una regresión múltiple, o como en el análisis factorial; es decir, que identifican las variables que mejor contribuyen a la diferenciación de la función (discriminante). El comportamiento del grupo según las ecuaciones discriminantes puede apreciarse a través de las coordenadas de sus centroides.

Como la orientación espacial de las ecuaciones discriminantes puede ser arbitraria, una opción VARIMAX permite rotar los ejes manteniendo fijas las ubicaciones relativas de casos y centroides grupales.

Una vez que son establecidas las ecuaciones discriminantes, puede realizarse un ANALISIS CLASIFICATORIO que permite clasificar nuevos casos en los grupos ya establecidos o testear la significación de la prueba en ubicar correctamente los grupos ya establecidos. Estiman así la probabilidad de un caso de ser clasificado en un grupo determinado.

Análisis de discriminación "a pasos"

El método "a pasos" selecciona la variable de mejor discriminación de acuerdo a criterios ya establecidos de antemano

(valor mínimo de Lambda de Wilks, mínima distancia de Mahalanobis entre grupos, mayor relación F entre grupos, mayor incremento de la correlación múltiple promedio, mayor incremento del valor V de Rao; Klecka, 1981). Luego, una segunda variable es incorporada al análisis, y así sucesivamente las variables son ingresadas a la prueba según su contribución en la discriminación. Puede suceder que variables que han ingresado al análisis pierdan "peso" en la discriminación y, por lo tanto, son retiradas del análisis. Cuando, según criterio seleccionado, las variables no utilizadas no producen discriminación alguna, la prueba termina su selección "a pasos" y procede a dar los resultados finales.

PROCEDIMIENTO SYSTAT

El procedimiento del paquete SYSTAT permite realizar el análisis de discriminantes según ciertas etapas:

1. Se modelan linealmente las variables dependientes como respuesta a una constante y la variable independiente (o variable de agrupamiento). Testeando el efecto de la variable de categorización (o de agrupamiento) se deben guardar los "scores" canónicos (factors), los coeficientes o distancias de Mahalanobis (distance) y las probabilidades de clasificación (PROB1).

2. Ese archivo puede listarse en el módulo de datos (DATA) categorizando los distintos grupos según números (1,2,3 y 4) y archivando exclusivamente la filiación grupal de cada caso y aquella pronosticada (PREDICT) de la ecuación discriminante calculada.

3. En el módulo de tablas (TABLES), se plotea este último archivo de modo de evaluar la efectividad en la correcta predicción de los casos de cada grupo.

4. Finalmente, en el módulo de graficación, se plotean también cada uno de los casos según los factores principales calculados o de mayor significación.

EJEMPLO

Se tomaron distintas muestras de agua en pozos ubicados en la cuenca de la laguna Mar Chiquita, que constituían 4 ambientes o áreas distintivas de la cuenca: a. Zona Serrana del Sistema de Tandilia (n=9), b. Planicie Pampeana, cuenca del Arroyo Grande (n=7), c. Planicie Pampeana, cuenca del Arroyo Vivoratá (n=8) y d. médanos de Mar Chiquita (n=10). Se realizó un análisis de discriminantes con el objeto de:

1. ¿ Es posible reconocer o discriminar los distintos ambientes o grupos mediante el análisis de sus datos hidroquímicos?

2. ¿ Cuales son las variables más importantes (alcalinidad, cloruros, sulfatos, Ca, Na, Mg, K) para reconocer esos grupos?

3. ¿ Que porcentaje de las muestras están efectivamente caracterizando al grupo al que pertenecen?

4. ¿ Cuales son las relaciones hidroquímicas entre los ambientes o grupos?

En el método directo todas las variables son combinadas linealmente (como componentes principales) pero buscando la mayor discriminación. Este procedimiento fue ejecutado utilizando el método A PASOS del paquete BMDP. Sólo 3 variables fueron necesarias (Na, SO4 y Ca en ese orden) y de acuerdo a las razones F (4.0) que justificaban la inclusión en la prueba. El análisis de coeficientes de correlación entre variables indicó que algunas variables poseían esencialmente la misma información; por ejemplo, Na y Cl (0.93) ó Na y SO4 (0.93). Un 81% de las muestrasfueron correctamente clasificadas, aunque existe aún cierta superposición entre los campos grupales (Fig. 2).

Se resalta la importancia de esta ecuación discriminante para analizar la afinidad de una muestra no bien definida arealmente o anómala, **utilizando sólo 3 análisis químicos (Na, SO4 y Ca).**

SERIES TEMPORALES

Una variable medida en diferentes niveles conforma una SERIE, SECUENCIA o CADENA. Es posible estimar la ecuación de toda variable temporal si se posee un registro suficientemente extenso, completo y medido en intervalos iguales. Si esos niveles son espaciados irregularmente, su análisis estadístico puede realizarse mediante **el análisis de la regresión**, es decir el análisis de la variable en función de su espaciamiento (variable x). Si el espaciamiento o nivel de medición es uniforme, se puede recurrir a análisis de tendencias por mínimos cuadrados, autocorrelaciones, correalciones cruzadas, series de Fourier, o promedios móviles ("auto-regressive integrated moving average"; ARIMA). Cada una de estas pruebas permiten un análisis diferente de nuestra variable aunque a veces los resultados suelen parecer similares. Por ejemplo, algunas pruebas hacen hincapie en la frecuencia de la variable (Fourier), mientras que otras analizan sus valores en función de los valores próximos (autocorrelación, promedios móviles), o en tendencias sistemáticas de sus valores (mínimos cuadrados).

AUTOCORRELACION Y CORRELACION CRUZADA

Muchas veces, parte de una secuencia parece repetirse, es decir que la variable tiene cierta ciclicidad (Davis, 1970). Si comparamos una variable (y) consigo misma variando sucesivamente las unidades de posicionamiento en distancia (x) o en tiempo (t), la AUTOCORRELACION es entonces la correlación entre una serie temporal y esa misma serie desplazada un cierto intervalo ("lag") de tiempo o distancia.

$$r1 = \frac{COV\ (Yi;\ Yi+1)}{Sy^2}$$

El correlograma, en cambio, grafica la relación entre el coeficiente de autocorrelación según cada uno de los "lags".

experimentados (Davis, 1973; Fig. 3).

Otras veces interesa buscar la correlación entre dos secuencias, por ejemplo entre perfiles geoeléctricos de 2 perforaciones de secuencias estratigráficas supuestamente semejantes. La CORRELACION CRUZADA entonces consiste en el desplazamiento de ambas series temporales de modo de encontrar su máxima correlación.

$$rm = \frac{COV\ 1,2}{S1\ .\ S2}$$

El correlograma analiza la relación entre el coeficiente de correlación en función de la distancia que se desplazan ambas series (Davis, 1973)

PROCEDIMIENTO DE BOX-JENKINS

Cualquier serie de tiempo puede ser analizada según el procedimiento ideado por Box y Jenkins (1976) que permite la **IDENTIFICACION** del modelo más apto, la **ESTIMACION** de los parámetros estadísticos y la **PREDICCION** (FORECAST) de futuros valores. Originalmente, la serie G que Box y Jenkins utilizaban para su modelo -demanda de pasajeros de una aerolínea-, poseía un incremento exponencial y una componente estacional con picos hacia el final del verano (Box y Jenkins, 1976). Es decir que la serie G no es estacionaria y es heteroscedástica (la variancia es diferente a lo largo de la serie), lo cual debe considerarse para su modelado.

Dadas las características de la variabilidad de y, se propuso 1. una transformación logarítmica, 2. diferenciación de los datos según la componente estacional y 3. diferenciación de los datos según componentes no estacionales.

El paquete SYSTAT permite transformaciones estandard. La **diferenciación** (DIFFERENCE) consiste en la sustracción de cada valor del siguiente en la serie. De este modo, el primer valor de la serie se pierde porque no tiene valor anterior alguno. Este procedimiento es muy útil para transformar las series en ESTACIONARIAS (Wilkinson, 1986).

Otras veces interesa de modo semejante sustraer un valor anterior, pero según un lag determinado. Por ejemplo, los valores de enero de 1989 deben sustraerse a los de enero de 1990 de acuerdo a un lag de 12 (meses).

Otras veces, las series tienen mucho "ruido" o una componente aleatoria importante. En ese caso, el procedimiento SMOOTH "suaviza" la serie; el más conocido es el de PROMEDIOS MOVILES donde debe especificarse la cantidad de datos a promediarse.

La **SERIE DE FOURIER** es un caso especial de transformación donde los datos son descompuestos en sus componentes armónicas (seno y coseno).

El procedimiento ARIMA se aconseja cuando existe evidencia que los datos son función de otros anteriores y sus errores, y no de los efectos periódicos y su "ruido" (Wilkinson, 1986). Merece destacarse que existen ciertas pruebas estadísticas que han tenido

mayor aceptación que otras en diferentes países. La aproximación por mínimos cuadrados (TREND SURFACE ANALYSIS) es preferida en Estados Unidos e Inglaterra, cuando los franceses y sudafricanos prefieren las aproximaciones por promedios móviles, especialmente el krigeage. No obstante, las fórmulas son diferentes no sólo en el procedimiento matemático de los datos (Davis, 1973).

En el prpcedimiento Box-Jenkins, una vez efectuada la transformación logarítmica y establecidos los rangos de diferenciación (DIFFERENCING) tanto estacional como no estacional, se logra reducir el desvío estandard de la serie. Al graficar así la nueva serie diferenciada se demuestra su transformación como estacionaria y ya estaríamos en condiciones de modelarla (ESTIMATE) reconociendo las componentes estacionales y efectos de valores vecinos (promedios móviles).

Una vez conocidos todos los parámetros que definen la serie, fundamentalmente los que definen las componentes estacional y no estacional de los promedios móviles, se pueden predecir (FORECAST) futuros valores de la serie (Hull y Nie, 1981).

CORRELACION CANONICA

El análisis factorial, a diferencia del análisis de la regresión, manipula las variables sin distinguir aquellas dependientes de las independientes. Las que "explican" y las "explicadas" son "mezcladas" de modo que las que están más relacionadas contribuirán en los mismos factores (Warwick, 1981).

El procedimiento más obvio entonces sería separar las variables dependientes de las independientes y ejecutar un análisis factorial separadamente. Sin embargo...

1) La regresión múltiple permitiría reconocer el efecto de varias variables independientes en una dependiente, pero cómo proceder si fueran varias variables dependientes.

2) El análisis factorial elige factores de modo de conectar variables pero maximizando la variancia que ellos explican. No puede distinguir variables independientes de dependientes (Warwick, 1981).

La solución del problema puede aproximarse mediante la CORRELACION CANONICA. El usuario, basándose en su experiencia, debe ingresar dos grupos de variables: Uno de independientes y otro de dependientes. La prueba combina linealmente ambos grupos de variables de manera de maximizar la correlación entre éstos. Estas combinaciones o variantes canónicas (CANONICAL VARIATES) manejan componentes principales, excepto que su criterio de selección ha sido arbitrario (inducido por el usuario). Mientras que el PCA suma la variancia posible, la CC busca la máxima correlación entre los grupos de variables. De modo que las dos primeras variantes canónicas poseen la mayor correlación posible y el segundo por la mayor correlación remanente.

Como los pesos o cargas en el PCA, los coeficientes ponderan la importancia de las variables en las variantes canónicas. El coeficiente de correlación canónica mide la correlación entre las variantes canónicas, y su cuadrado (equivalente al eigen valor) representa el porcentaje de la variancia que una variante canónica es explicada por la otra (Warwick, 1981).

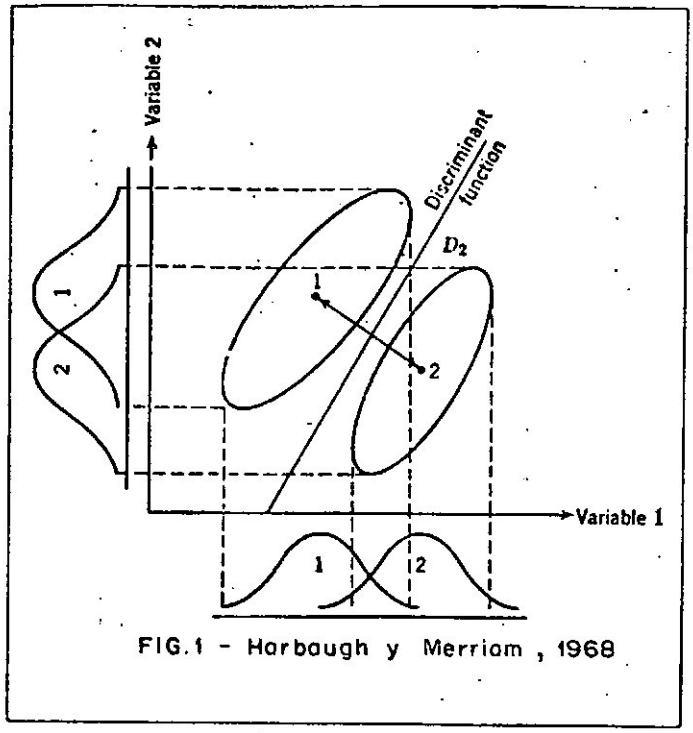


FIG.1 - Harbaugh y Merriam , 1968

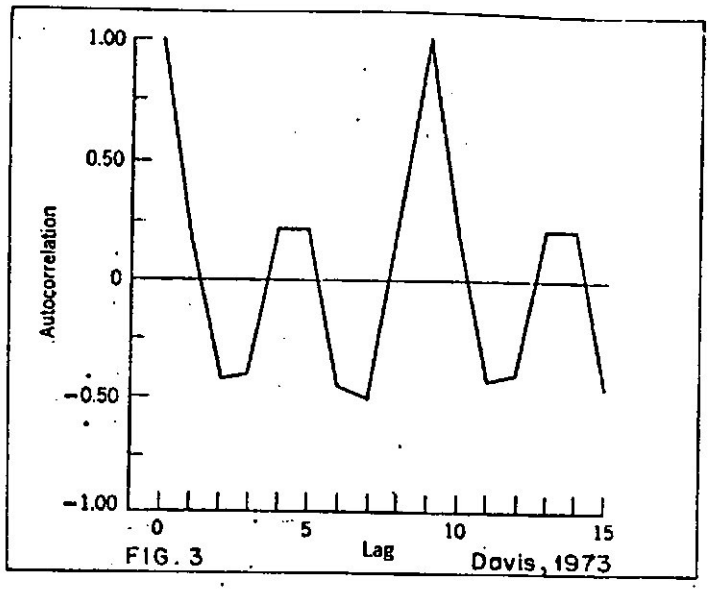
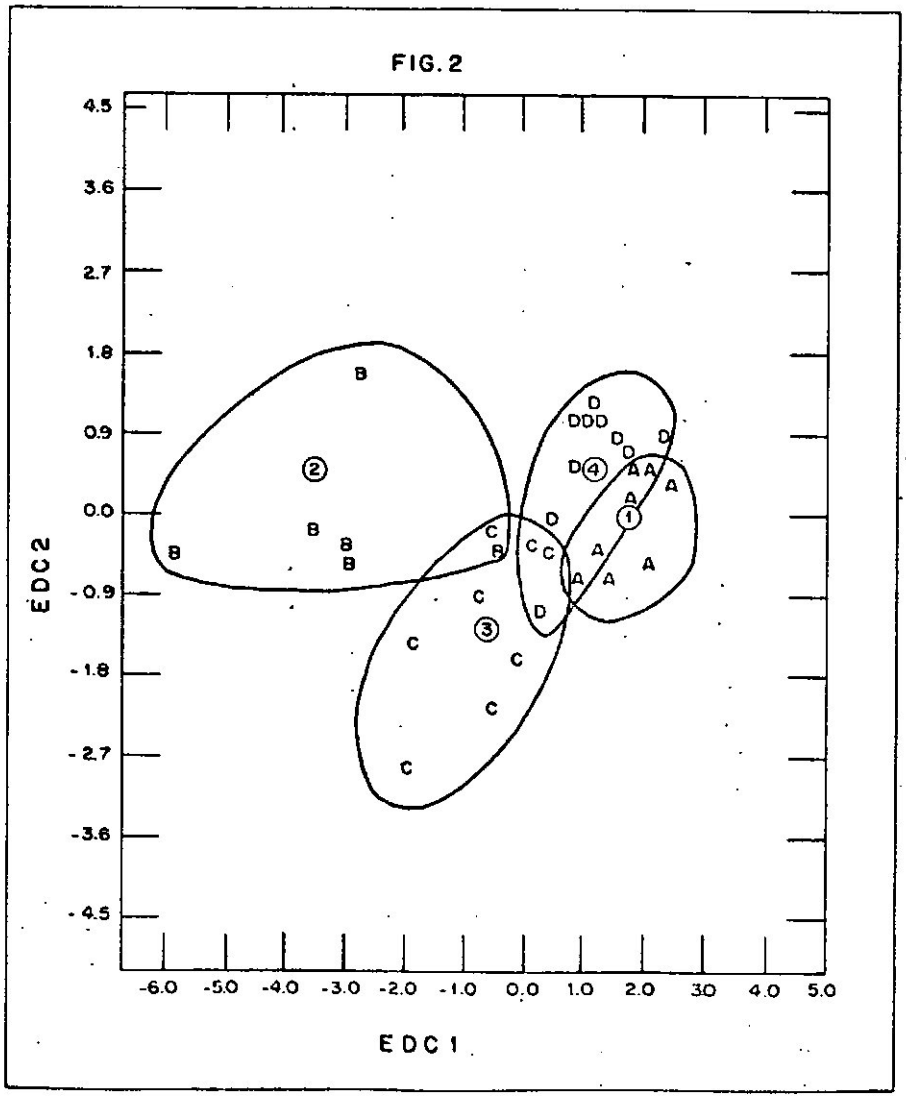


FIG. 3 Davis, 1973

FIG. 2



BIBLIOGRAFIA

TEXTOS

BOX, G. E. P. y JENKINS, G. M., 1976. Time series analysis: Forecasting and control. San Francisco, Holden-Day.

CRISCI, J. V. y M. F. LOPEZ ARMENGOL, 1983. Introducción a la teoría y práctica de la taxonomía numérica. OEA, Serie de Biología. Monografía nº 26.

DAVIS, J. C., 1973. Statistics and data analysis in geology. J. Wiley & Sons Inc., New York.

DIM, J.O. y KUHOUN, F.J., 1981. En Hull, C.H. y Nie, N.H. Special Package for the Social Sciences. Update 7-9. Mc Graw-Hill Co, 398-432.

DIXON, W. J. y BROWN, M. B., 1979. BMDP-79. Biomedical computer programs P-Series. University of California Press, Berkeley, 979 pp.

HARBAUGH, J. W. y MERRIAM, D. F., 1968. Computer applications in stratigraphic analysis. J. Wiley and sons, 282 pp.

HULL, C.H. y NIE, N.H., 1981. Special Package for the social Sciences. Update 7-9. Mc Graw-Hill Co.

KLECKA, W.R., 1981. Discriminant analysis. En Hull, C.H. y Nie, N.H. Special Package for the Social Sciences. Update 7-9. Mc Graw-Hill Co., 434-467.

KRUMBEIN, W. C. y F. A. GRAYBILL, 1965. An introduction to statistical models in Geology. Mc Graw-Hill Co., New York.

MATHER, P. M., 1976. Computational methods of Multivariate Analysis in Physical Geography. J. Willey & Sons Inc., Londres.

MENDIO, J. C., 1985. Métodos estadísticos en Geología. A.G.A. Serie B, Didáctica y Complementaria Nº 13. Buenos Aires.

RICKMERS, A. D. y H. N. TODD, 1974. Introducción a la estadística. CECSA, Barcelona.

MARWICK, P. V., 1981. Special Package for the Social Sciences. Update 7-9, Mc Graw Hill Co., 515-527.

WINER, B. J., 1962. Statistical Principles in Experimental Design. Mc Graw Hill Co., New York.

WILKINSON, L., 1986. SYSTAT: The system for statistics. Evanston, IL., Systat Inc.

YAMANE, T., 1979. Estadística. Harla, México

ARTICULOS VARIOS

AHRENS, L. H., 1954 a. The lognormal distribution of the elements. Geochim. Cosmochim., Acta 5: 49-73

----------, 1954 b. The lognormal distribution of the elements II. Geochim. Cosmochim., Acta 6: 121-131.

ALHAJJAR, B. J., J. M. HARKIN y G. CHESTERS, 1989. Detergent formula effect on transport of ntrient to ground water from septic systems. Ground Water 27(2): 209-219

ALTHER, G. A., 1979. A simplified statistical sequence applied to routine water quality analysis: a case history. Ground water, 17: 556-571.

AUGE, M. P. y C. E. ZURITA, 1988. Caracterizacion hidrogeológica de 9 de Julio y alrededores, Pcia. de Buenos Aires. Actas: 619-629. II JOrn. Geol. Bonaerenses, Bahia Blanca.

BATBEDAT, A. y J. L. FASANO, 1986. Utilization de l'approche pyramidale haute pour préciser l'etude chimique des eaux souterraines. Cahiers de D.E.A.: B/F1 - B/F14. Université des Sciences et Techniques du Languedoc, Montpellier, Francia.

BOLYARD, T. H., G. M. HORNBERGER, R. DOLAN y B. P. HAYDEN, 1979. Freshwater reserves of Mid-Atlantic Coast Barrier Islands. Environ. Geology 3(1): 1-12.

CAZES, P., SOLETY e Y. VUILLAUME, 1970. Exemple de traitement statistique des données hydrochimiques. Bull. B.R.G.M. (II série), Sec. III (4): 75-90.

CECH, I. y C. KREITLER. Radon distribution in domestic water of Texas. Replay to discussion by P. T. King y J. A. Connor. Gorund Water 27 (3):404-407.

DEPETRIS, P. J., 1980. Hydrochemical aspects of the Negro River; Patagonia, Argentina. Earth Surf. Processes 5: 131-186.

DUROVIC, S., 1959.Contibution to the lognormal distribution of elements. Geochim. Cosmochim. Acta 15: 330-336.

DUTTON, A. R., B. C. RICHTER y W. KREITLER, 1989. Brine discharge and Salinization, Concho River Watershed, West Texas. Ground Water 27 (3): 357-383.

FASANO, J. L., 1986. Aplicación de los métodos de agrupamiento y ordenación al estudio químico de las aguas subterráneas. Cahiers D.E.A.: JF1-JF15. Université des Sciences et Techniques du Languedoc, Montpellier, Francia.

GIBBONS, R., 1987. Statistical Models for the analysis of volatile organic compounds in waste disposal sites. Ground Water 25(5): 572-590.

GONZALES, N. y M. A. FERNANDEZ, 1988. Empleo del análisis numérico (multivariante) en el tratamiento de problemas geohidrológicos regionales. Actas: 549-557. II Jor. geol. Bonaerenses, Bahia Blanca.

HAGERTY, D. J. y K. LIPPERT, 1982. Rising Ground Water. Problem or resource. Ground Water 20(2): 217-223.

HOUSTON, J. F. T. y R. T. LEWIS, 1988. The Victoria Province Drought Relief Project, II. Borehole Yield Relationships. Ground Water 26(4): 419-426.

KASHYAP, D., P. DACHADESH y L. S. J. SINHA, 1988. An optimization Model for Analysis of Tests Pumping Data. Ground Water 26(3): 289-297.

LARRIESTRA, C., 1979. Análisis sedimentológico de la Formación Puelches en el área Gran Rosario (Prov. de Santa Fe). Rev. Asoc. Arg. Min. Petrol. y Sed., 10, 1/2, 57/64.

LAWRENCE, F. W. y S. B. UPCHURCH, 1976. Identification of geochemical patterns in ground water by numerical analysis. En J. A. Saleem (ed), Advances in ground water hydrology, Minneapolis, Amer. Water Resour. Assn.: 199-214.

----------., 1982. Identificatiuon of recharge areas using geochemical factor analysis. Ground water 20 (6): 680-687.

LE MARECHAL, A. y H. TEIL, 1973. Application de quelques traitements statistiques aux données hydrochimiques des sources thermominérales du Cameroun. Cah. ORSTROM, sér. Géol. 2: 217-234.

MATALAS, N. C. y B. J. REINER, 1967. Some coments in the use of factor analysis. Water Res. Res. 3(1) 213-223.

NELSON, J. y R. C. WARD. 1981. Statistical consideration and sampling techniques for the ground water quality monitoring . Ground water 19 (6): 617-625.

PUCCI, A. A. y J. A. E. MURASHIGE, 1987. Applications of Universal Kriging to an Aquifer Study in New Jersey. Ground Water 25 (6): 672-679.

RAJAGOPAL, K., 1987. Large data bases and regional ground water quality assessments. An Iowa case study. Ground water 25 (4): 415-426.

----------., 1988.Influence of outlying observations on selected estimates or parameters of distributions. Ground water 26 (3): 325-332.

USUNOFF, E. y A. G. GUZMAN, 1989. Multivariate analysis in hydrochemistry. An example of the use of factor and correspondence analysis. Ground water 27(1) : 27-34.

VISWANA THAn, M. N., 1983. The rainfall/water table level relationship of an unconfines aquifer. Ground water 21 (1): 49-56.

WILLIAMS, T. A. y A. K. WILLIAMSON, 1989. Estimating water table altitudes for regional ground water flow modeling, U.S. gulf Coast. Ground water 27 (3): 333-340.