

APLICACION DE DOS TECNICAS ESTADISTICAS EN LA INTERPRETACION DE DATOS GEOQUIMICOS EN PROSPECCION MINERA.

GUILLERMO ALFARO

Inst. Invest. Geol., Casilla 10465, Santiago, Chile

RESUMEN

Se presenta, mediante un ejemplo real, la utilización de las técnicas estadísticas conocidas como "análisis de la función discriminante" (discriminant function analysis), en la interpretación de datos geoquímicos, provenientes de un programa de exploración minera. La información utilizada fue obtenida en la prospección geoquímica realizada por la comisión conjunta IIG-Chile y MMAJ-Japón, el año 1979, en la región cordillerana de la provincia de Nuble.

Se hace una presentación matemática simple de ambas técnicas estadísticas y se muestra la simplificación de la interpretación y manejo de los resultados que se logra, mediante el análisis de los componentes principales en relación al análisis de factores (factor analysis). El análisis de la función discriminante se considera como un criterio estadístico de valor para conocer la relación entre el sustrato rocoso y las muestras geoquímicas, cuya proveniencia no es clara debido a falta de información geológica. Ambos desarrollos, aún cuando son aplicados en este trabajo a un caso de exploración geoquímica, pueden ser utilizados en otras ramas de la geología en las cuales se opera con multivariadas.

ABSTRACT

The statistical techniques known as "principal component analysis" and "discriminant function analysis", are presented by means of a practical example, where geochemical data obtained in a mineral exploration survey are used. The information was obtained during a geochemical prospecting carried out by a joint team of the IIG-Chile and MMAJ-Japan, in the Andes of the province of Nuble, Chile, in 1979.

A simple mathematical presentation of both statistical tools is shown. The interpretation and handling of results is easier compared with the factor analysis technique. Discriminant function analysis is considered as a useful statistical criterion in order to decide upon the rock influence in geochemical samples when geological information is not enough. Both statistical techniques can be used in other geological fields handling multivariate statistical analysis.

INTRODUCCION

La prospección geoquímica mediante sedimentos fluviales es una herramienta de exploración minera, que el IIG ha utilizado, sistemáticamente, en los trabajos que ha efectuado en la zona central-sur de Chile. Esta parte del país presenta un buen sistema de drenaje, potente desarrollo de suelo (sector occidental), cubierta de vegetación y una amplia distribución de rocas volcánicas de edad terciaria-

cuaternaria (sector oriental) que, como un gran manto, cubren extensas áreas, ocultando los posibles cuerpos mineralizados más antiguos. Estos factores principales hacen de la geoquímica una herramienta fundamental en la búsqueda de minerales y se debe tratar que, al ser aplicada, se obtenga de ella el mayor provecho. Por esta razón, la determinación analítica de cada muestra se hace

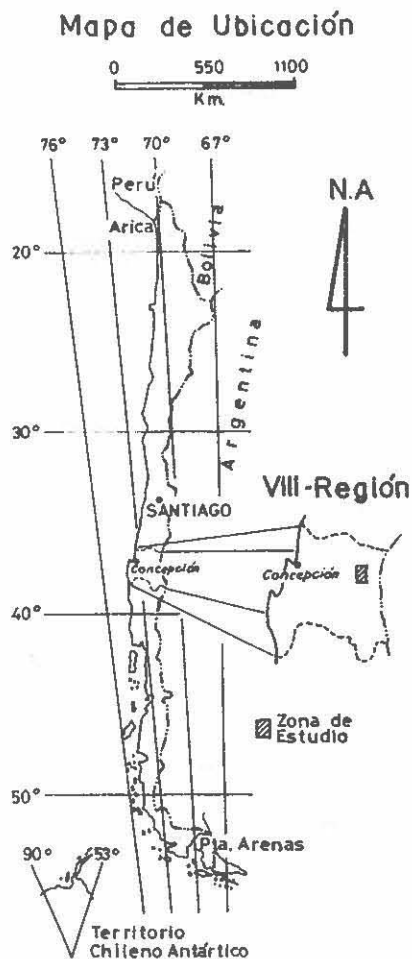


FIG. 1. Mapa de Ubicación.

por varios elementos en trazas (en general más de seis), complicando su posterior interpretación; a esto se suma el hecho que, en general, por la falta de afloramientos, mapeo, etc., la información geológica básica no es la más adecuada para hacer una rápida y efectiva interpretación de los datos geoquímicos. Es entonces cuando la aplicación adecuada de métodos estadísticos facilita la interpretación o respalda aquellas conclusiones logradas en base a la información geológica disponible.

Desde finales de la década de 1960, algunas de dichas técnicas se han aplicado en los trabajos de exploración geoquímica efectuados por el IIG debido, principalmente, al impulso dado por los Drs. B. Page, A. Cruzat y H. Conn. El procesamiento de los datos geoquímicos ha incluido desde los elementales cálculos de estadígrafos básicos

(media, desviación standard, etc.) hasta el complejo análisis de factores (factor analysis) y análisis de regresión, todo esto implementado por los programas de computación disponibles en el IIG.

En este trabajo, se presentan dos nuevas técnicas de interpretación de la información geoquímica, que no han sido utilizadas en Chile y que facilitan el estudio de datos obtenidos en zonas cubiertas y con levantamiento geológico insuficiente. Dichas técnicas son denominadas "análisis de los componentes principales" (principal component analysis) y "análisis de la función discriminante" (discriminant function analysis).

Para mostrar ambos procedimientos, se ha utilizado un conjunto de datos geoquímicos provenientes de la prospección que realizó el año 1979, el Instituto de Investigaciones Geológicas (IIG) conjuntamente con la Metal Mining Agency of Japan (MMAJ) en un área de aproximadamente 600 km², situada al oriente de Chillán (Fig. 1). En total, se obtuvieron 327 muestras de sedimentos fluviales activos, que fueron analizados por Cu, Zn, Mn, Ag, Mo, As y algunas muestras por Au. De los 7 elementos analizados, en este trabajo se usaron sólo Cu, Zn, Mn y Mo; se descartaron Ag y As por presentar gran parte de sus valores bajo el límite de detección y el Au por no haber sido analizado en todas las muestras.

Para determinar una función discriminante que permitiera separar muestras que tenían como ambiente geológico a granitoides (rocas relacionadas con la mineralización) y aquellas relacionadas con las rocas volcánicas plio-cuaternarias o sedimentarias (estériles), se eligieron 50 y 56 muestras, respectivamente, algunas de las cuales no pudieron ser bien clasificadas por corresponder a zonas con sedimentos derivados de ambos grupos de rocas (granitoides y rocas volcánicas y sedimentarias); esta situación es normal al trabajar en ambientes geológicos y aún más en zonas cubiertas.

Finalmente, el autor desea hacer presente que éste no es un trabajo estadístico, de modo que se ha eliminado al máximo la base matemática de cada una de estas técnicas, entregándose al final una bibliografía en la cual puede obtenerse el desarrollo teórico de cada una de ellas. Los conceptos básicos de matrices, principalmente valor *eigen*, vector *eigen* y su determinación, mencionados en este trabajo, pueden ser consultados en cualquier texto básico sobre la materia. El objetivo de este trabajo

es presentar ambas herramientas estadísticas a aquellos colegas cuyas actividades se relacionan,

principalmente, con exploración geoquímica.

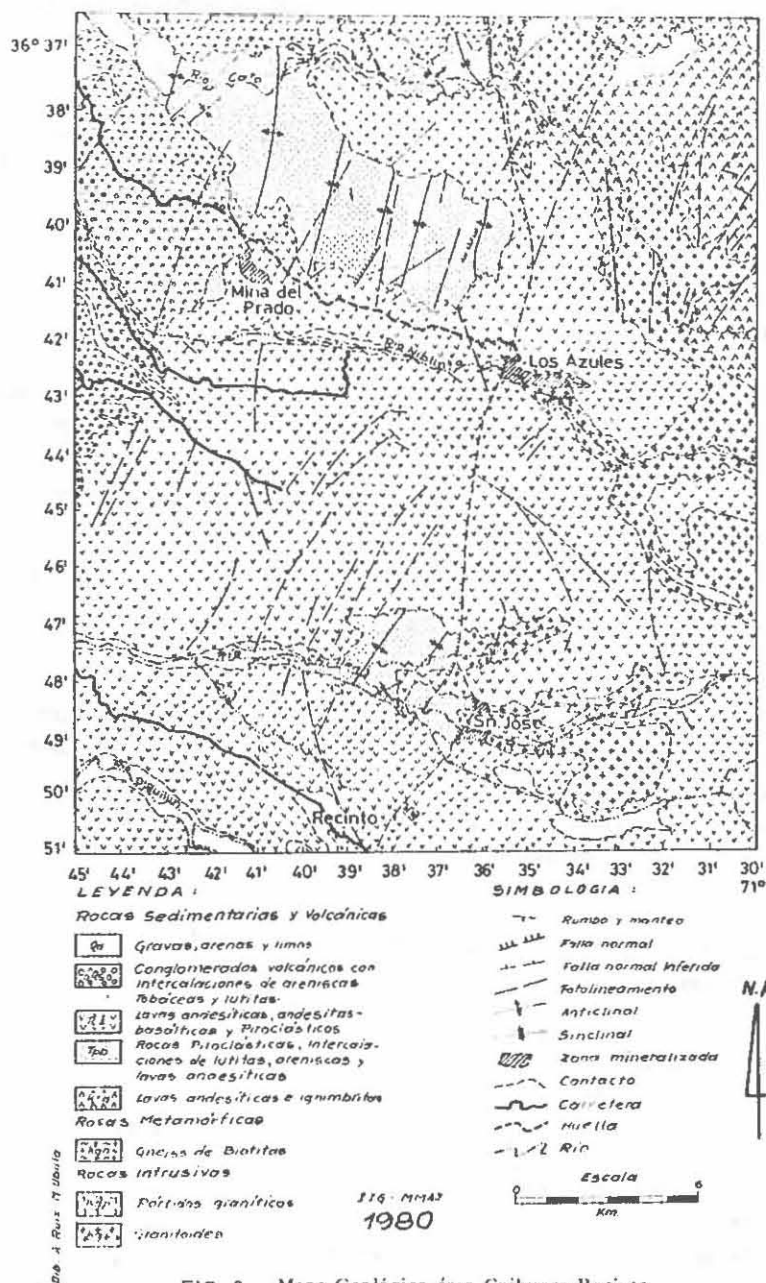


FIG. 2. Mapa Geológico área Coihueco-Recinto.

BOSQUEJO GEOLOGICO DEL AREA EXPLORADA

En la región estudiada se han reconocido rocas sedimentarias, volcánicas, metamórficas e intrusivas (IIG-MMAJ, 1980, Fig. 2).

Alrededor del 75% del área está cubierta por lavas andesítico-basálticas y rocas pirolásticas, con estructura monoclinal y buzamiento al oeste;

estas rocas han sido asignadas al Terciario superior-Cuaternario y se les conoce como Formación Cola de Zorro (González y Vergara, 1960).

En la parte central-norte y central-sur, afloran rocas piroclásticas con intercalaciones de lutitas, areniscas y lavas andesíticas, bien estratificadas y afectadas por un conjunto de sinclinales y anticlinales, cuyos ejes tienen rumbo norte-sur. Esta secuencia ha sido denominada Unidad Volcánico-Sedimentaria Piernas Blancas y se asigna, en base a su contenido faunístico, al Eoceno.

En el extremo noreste del área estudiada, aflora una secuencia de lavas andesíticas e ignimbritas, cuya edad radiométrica corresponde al Cretácico Superior y se correlacionaría con la Formación Abanico (González y Vergara, 1960).

Las únicas rocas metamórficas observadas afloran en el curso medio del río Niblinto, en la localidad de Los Azules; litológicamente, es un gneis

de biotita cuya edad es indeterminada.

Las rocas intrusivas mapeadas corresponden a granitoides, pórfidos graníticos y diques granodioríticos. La edad del granitoide, en base a dataciones K-Ar es 69 m.a. (Cretácico Superior), a él se asocia la mineralización de cobre y molibdeno que se conoce en la región.

Existen dos zonas principales con mineralización; la denominada Minas del Prado (Lat. 36°38'S; Long. 71°41'W), que consiste en una brecha volcánica silicificada con oro y pirita, cuya erosión ha permitido la formación de placeres auríferos, y el prospecto cuprífero San José (Lat. 36°48'S; Long. 71°37'W), con mineralización y alteración hidrotermal tipo pórfido cuprífero (IIG-MMJ, 1978). Sólo Minas del Prado ha sido explotado artesanalmente, en tanto que el prospecto San José está siendo explorado. En otras partes del área se conocen sólo pequeñas manifestaciones de cobre y molibdeno.

MUESTREO GEOQUIMICO, ANALISIS DE LAS MUESTRAS Y ESTADIGRAFOS BASICOS

Simultáneamente con el levantamiento geológico regional, se recolectaron 327 muestras de sedimentos fluviales activos. Con el objeto de determinar el error de muestreo, por cada 10 muestras se tomó un duplicado. El análisis por Cu, Zn y Mn se hizo por espectrometría de absorción atómica y por Mo, mediante colorimetría. La precisión y sensibilidad en la determinación de cada elemento se señala en la Tabla 1.

Los estadígrafos básicos para cada elemento de toda la población se indican en la Tabla 2.

TABLA 1 Precisión y sensibilidad en la determinación de Cu, Zn, Mn y Mo.

	Precisión analítica (o/o)	Precisión de muestreo (o/o)	Sensibilidad (ppm)
Cu	± 4	± 6	10
Zn	± 3	± 7	20
Mn	± 3	± 5	5
Mo	± 30	± 40	1

TABLA 2 Estadígrafos básicos.

	$\bar{X}$	S	$\bar{X} + S$	$\bar{X} + 2S$	$\bar{X} + 3S$
Cu	38,25	17,00	55,25	72,25	89,25
Zn	78,25	35,61	113,86	149,47	185,08
Mn	586,98	259,33	846,31	1.105,64	1.364,97
Mo	2,95	1,83	4,78	6,61	8,44

La matriz de correlación para los elementos considerados aparece en la Tabla 3.

TABLA 3 Matriz de Varianza-Covarianza.

	Cu	Zn	Mn	Mo
Cu	1,00	0,04	-0,002	0,07
Zn	0,04	1,00	0,64	0,04
Mn	-0,002	0,64	1,00	-0,02
Mo	0,07	0,04	-0,02	1,00

Como criterio de selección de las anomalías probables se usó la $\bar{X}+3S$ de la población tamizada. Al llevar estos valores al plano se observa un sector anómalo de cobre y molibdeno en los ríos Piernas Blancas, San Juanito y San José, relacionado con el prospecto cuprífero San José y distribución errática de los otros elementos en la mitad occidental del área estudiada, con pequeña tendencia a agruparse los valores altos de Mn y Zn, en los esteros que drenan el cuerpo mineralizado con oro de Minas del Prado (Fig. 3).

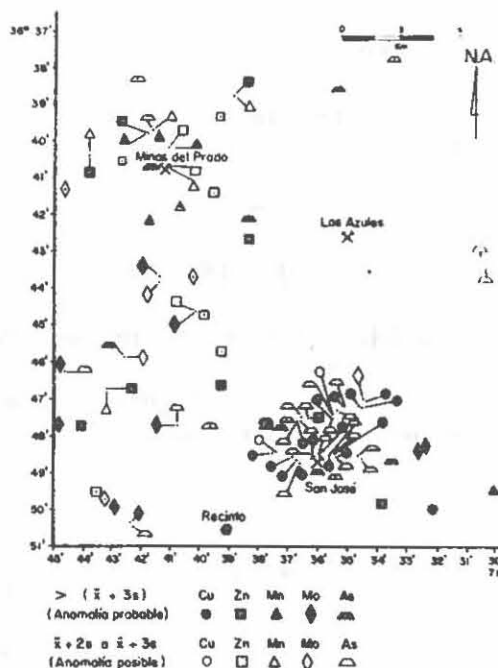


Fig. 3: Muestras anómalas en Cu, Zn, Mn, Mo y As.

ANÁLISIS DE LOS COMPONENTES PRINCIPALES

INTRODUCCION

Una de las técnicas utilizadas en la interpretación de procesos en que intervienen multivariantes es el conocido Análisis de Factores (Facto Analysis). Esta herramienta estadística se caracteriza por su complicado tratamiento matemático y peculiar jerga, lo cual la hace difícil de manejar y aún más, de "sentir" los resultados, para quienes la usan sólo ocasionalmente. Al contrario, el análisis de los componentes principales, sin recurrir a las técnicas matemáticas de rotación de ejes en espacios n-dimensionales, usa los mismos principios matemáticos del análisis de factores presentando, en forma simple, las relaciones entre multi-variables (por ejemplo los elementos en trazas considerados en cualquier trabajo geoquímico) lo que permite llevarlos a un plano mediante el artificio de obtención de puntajes (scores).

El análisis de los componentes principales consiste en representar las variables utilizando

aquellos vectores *eigen*, obtenidos de la matriz de varianza-covarianza, que mejor explican la varianza de la población (de aquí proviene el nombre de "componentes principales"). La estructura de la matriz permite que la varianza de todos los elementos, tomados en forma individual, corresponda a los elementos diagonales de la matriz; es, por lo tanto, fundamental comprender el significado físico del vector *eigen* y del valor *eigen*, conceptos por sí complejos para un no-matemático. Para hacer esta presentación se ha tomado, por su sencillez el ejemplo dado por Davis (1973) descrito a continuación.

Consideremos un sistema de ecuaciones simultáneas expresadas en forma de matriz:

$$[A] \cdot [X] = \lambda [X] \quad (1)$$

Este problema es idéntico a resolver un sistema de ecuaciones sumultáneas tal como:

$$[A] \cdot [X] = [B] \quad (2)$$

$$\text{en que } [B] = \lambda [X] \quad (3)$$

debiéndose encontrar ahora los valores de λ que satisfagan las ecuaciones.

La ecuación (1) puede ser escrita en la forma:

$$([A] - \lambda [I]) \cdot [X] = 0 \quad (4)$$

En que $\lambda [I]$ es λ veces una matriz de identidad del mismo tamaño que $[A]$

Expresada la ecuación (4) en forma general para un conjunto de ecuaciones simultáneas se tendrá:

$$(a_{11} - \lambda) X_1 + a_{12} X_2 + \dots + a_{1n} X_n = 0$$

$$a_{21} X_1 + (a_{22} - \lambda) X_2 + \dots + a_{2n} X_n = 0 \quad (5)$$

$$\dots \dots \dots a_{n1} X_1 + a_{n2} X_2 + \dots \dots + (a_{nn} - \lambda) X_n = 0$$

Si se asume que los valores de las "x" del sistema de ecuaciones (5) es distinto a cero, entonces el determinante de los coeficientes de la matriz es cero:

$$|A - \lambda I| = \begin{vmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} - \lambda \end{vmatrix} = 0 \quad (6)$$

Los valores de λ que satisfagan la ecuación (6) se denominan "valores *eigen*" y las correspondientes soluciones para $[x]$ son llamados "vectores *eigen*".

Como normalmente se conocen los coeficientes A_{ij} de la matriz, es posible calcular los valores λ que cumplen con las condiciones de la ecuación (6). Para visualizar mejor la definición anterior se calcularán los valores *eigen* de una matriz simétrica de orden 2 cuyo determinante tiene la forma:

$$\begin{vmatrix} A_{11} - \lambda & A_{12} \\ A_{21} & A_{22} - \lambda \end{vmatrix} = 0$$

$$(A_{11} - \lambda)(A_{22} - \lambda) - A_{21} A_{12} = 0$$

agrupando los elementos conocidos A_{ij} se tendrá la ecuación:

$$\lambda^2 + \alpha_1 \lambda + \alpha_2 = 0$$

Así, los valores *eigen* de la matriz

$$\lambda^2 + \alpha_1 \lambda + \alpha_2 = 0$$

Así, los valores *eigen* de la matriz

$$\begin{vmatrix} -5 & -2 \\ -6 & -3 \end{vmatrix}$$

deben satisfacer la ecuación $\lambda^2 - 2\lambda - 3 = 0$

entonces $\lambda_1 = 3$ y $\lambda_2 = -1$

Los vectores *eigen* se obtienen al reemplazar los valores " λ " en la ecuación (1):

$$[A - \lambda_1 I] X_1 = \begin{vmatrix} 5-3 & 2 \\ -6 & -3-3 \end{vmatrix} \begin{vmatrix} X_{11} \\ X_{12} \end{vmatrix} = \begin{vmatrix} 0 \\ 0 \end{vmatrix}$$

Los valores X_{11} y X_{12} (elementos del vector *eigen*) se obtienen al resolver las ecuaciones $2X_{11} + 2X_{12} = 0$ y $-6X_{11} - 6X_{12} = 0$. Para λ_2 se procede en forma similar.

Sin perder de vista su utilización en la interpretación de datos geoquímicos a partir de la matriz de varianza-covarianza cuyo valor es normalmente mayor que dos (por ejemplo, al utilizar los elementos Cu, Mo, Pb, Zn, Ag en muestras de sedimentos fluviales), debido a la complejidad y extensión que el cálculo implica, sólo es posible obtener resultados prácticos mediante la utilización de computadoras; esto explica por qué sólo en los últimos años se han aplicado estos conceptos al campo de la geología.

Supongamos el caso real de una matriz de varianza-covarianza $[A]$, bidimensional, cuyos elementos podrían corresponder a Cu y Mo:

$$[A] = \begin{vmatrix} 6 & 8 \\ 8 & 6 \end{vmatrix}$$

Esta matriz elemental tiene valores *eigen* = 14 y -2, los cuales pueden ser representados como los ejes mayor y menor de un elipsoide que inclu-

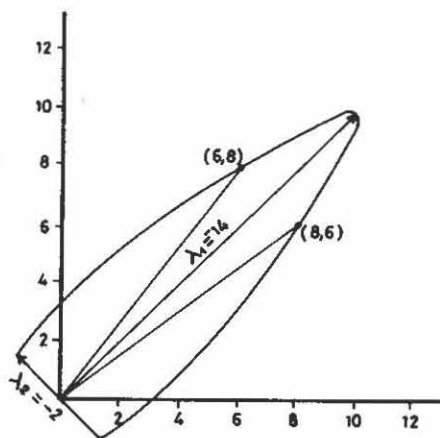


FIG. 4. Representación gráfica de los valores "eigen" (Davis, 1973).

ye los puntos de coordenadas (6,8) y (8,6) (Fig. 4). La relación gráfica muestra que el eje mayor es claramente más grande que el eje menor. Si los puntos coinciden, como por ejemplo en la matriz:

$$[A] = \begin{vmatrix} 4 & 8 \\ 4 & 8 \end{vmatrix}$$

el eje menor de la elipse (segundo valor *eigen*) se vuelve igual a cero y el elipsoide se transforma en una línea recta (Fig. 5a). Al contrario, si la matriz corresponde a dos vectores perpendiculares tales como:

$$[A] = \begin{vmatrix} -4 & 8 \\ 8 & 4 \end{vmatrix}$$

ambos valores *eigen* son iguales en longitud y se obtiene una circunferencia (Fig. 5b). Asociado a cada valor *eigen* existe un vector *eigen*, que puede ser interpretado como la pendiente del eje mayor y menor del elipsoide (cuando la matriz simétrica tiene rango mayor que 3, los vectores son perpendiculares en un espacio de n-dimensiones).

Explicando en forma simple el concepto físico de valor *eigen* y vector *eigen*, consideremos una matriz de varianza y covarianza de un conjunto de muestras de sedimentos analizados por Cu y Mo (se consideran dos elementos para trabajar en un espacio bidimensional):

	Cu	Mo
Cu	20,3	15,6
Mo	15,6	24,1

Al calcular los vectores *eigen* de la matriz, se obtendrán los ejes principales del elipsoide cuyas longitudes estarán representadas por los valores *eigen*. El método del análisis de los componentes principales consiste en encontrar dichos ejes y sus longitudes. Como se utiliza una matriz simétrica con "m" variables (matriz de varianza y covarianza de los elementos considerados), los "m" vectores *eigen* obtenidos serán ortogonales entre sí.

En el ejemplo anterior se obtienen dos vectores *eigen*:

$$I = \begin{vmatrix} 0,66 \\ 0,75 \end{vmatrix} \quad II = \begin{vmatrix} 0,75 \\ 0,66 \end{vmatrix}$$

y los valores *eigen* 37,9 y 6,5, respectivamente (Fig. 6). La varianza total de la población es igual a la suma de la varianza de Cu y Mo (igual a la suma de los elementos de la diagonal de la matriz):

$$\text{Varianza total} = 20,3 + 24,1 = 44,4 \text{ (Fig. 6)}$$

El Cu contribuye con 20,3/44,4 (46%) de la varianza total y el Mo, con 24,1/44,4 (58%). Como una de las propiedades de los valores *eigen* es que su suma es igual a la suma de los elementos de la diagonal de la matriz o traza (en el ejemplo, la traza corresponde a las varianzas de los elementos), la suma de ambos valores *eigen* es igual a 37,9 + 6,5 = 44,4 y representa la longitud total de los dos ejes principales, por lo tanto, representan también la varianza total de la población (varianza del Cu y Mo). Sin embargo, el primer eje representa aproximadamente el 86% de la varianza (37,9/44,4) y el segundo eje representa sólo el 14% (6,5/44,4); si se compara estos valores con la contribución a la varianza total debida a las variables originales, es posible comprender la utilidad del análisis de los componentes principales. Se logra, entonces, representar con pocos ejes (número de componentes principales menor que el número de variables originales) y en forma más eficiente, la varianza total de la población, facilitan-

do así la interpretación, cuando se trabaja con un número grande de elementos.

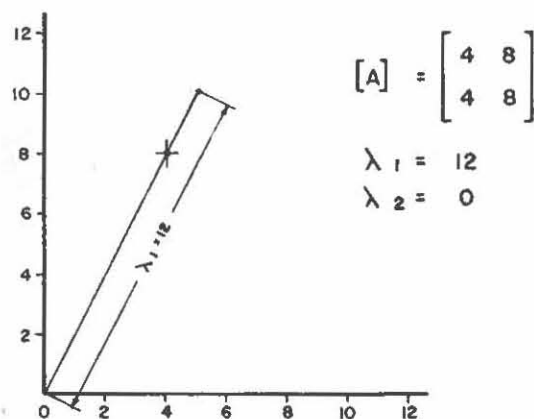


FIG. 5a. Elipse para $\lambda_1 = 12$ y $\lambda_2 = 0$ (Davis, 1973).

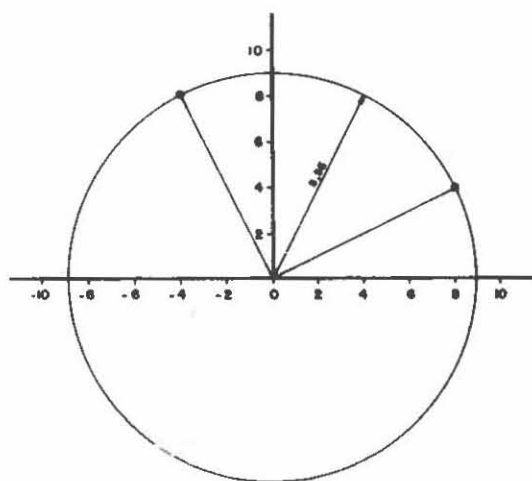


Fig. 5b: Elipse para $|\lambda_1| = |\lambda_2|$ (Davis, 1973)

FIG. 5. Casos extremos en la representación de valores "eigen".

Un criterio que se sugiere para disminuir el número de componentes principales es eliminar aquellos que representen menos de un 20% de la varianza original. Para operar en unidades de desviación standard, la dimensión de los componentes principales se obtiene multiplicando los elementos

Vector eigen	Longitud	Pendiente
I = OP ₁	37,9	$\begin{bmatrix} 0,66 \\ 0,75 \end{bmatrix} = 48^\circ,6$
II = OP ₂	6,5	$\begin{bmatrix} -0,75 \\ -0,66 \end{bmatrix} = 41^\circ,4$

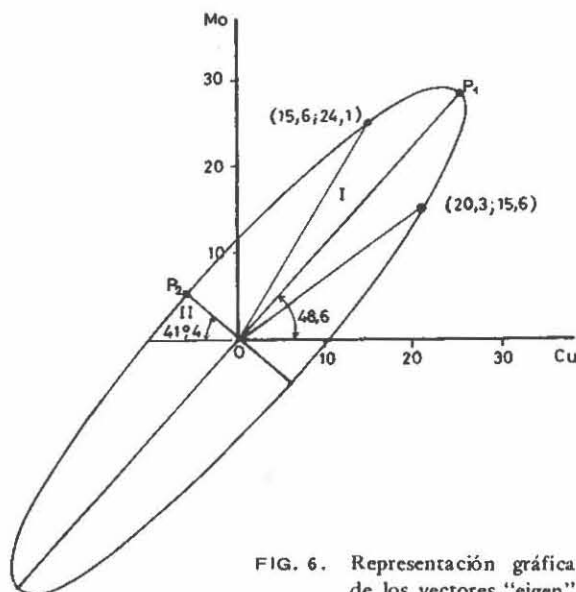


FIG. 6. Representación gráfica de los vectores "eigen" (modificado de Davis, 1973).

de cada vector *eigen* por la raíz cuadrada de cada valor *eigen*. Si se compara esta técnica, que no necesita rotación de ejes, criterios especiales para elegir el número de ejes, etc., con el análisis de factores, se comprenderá mejor su simplicidad.

Cada observación original (por ejemplo, contenido de elemento en trazas en una muestra geoquímica) puede ser convertida en un puntaje (score), proyectándola en cada uno de los ejes principales, esto se logra multiplicando los valores observados de las variables (por ejemplo, valores de Cu y Mo en el caso anterior), por los elementos correspondientes del primer vector *eigen* (primer puntaje), luego por los elementos del segundo vector *eigen* (segundo puntaje) y así, sucesivamente, en el caso de considerarse más de dos componentes principales. En el ejemplo precedente, los puntajes serían:

$$\text{Puntaje I} = (X_{1i} \times 20,3) + (X_{2i} \times 15,6)$$

$$\text{Puntaje II} = (X_{1i} \times 20,3) - (X_{2i} \times 15,6)$$

donde: X_{1i} = valor de la variable 1, muestra i

X_{2i} = valor de la variable 2, muestra i

TABLA 4 Listado de Valores *Eigen*, Vectores *Eigen* y Componentes Principales.

Valor <i>Eigen</i> I	=	1,6411			
Vector <i>Eigen</i> I	=	0,5324	0,7075	0,7042	0,02428
Comp. Princ. I	=	0,0682	0,9064	0,9022	0,0311
Valor <i>Eigen</i> II	=	1,0725			
Vector <i>Eigen</i> II	=	0,6079	0,0168	-0,0941	0,7097
Comp. Princ. II	=	0,7228	0,0174	-0,0975	0,7450
Valor <i>Eigen</i> III	=	0,9300			
Vector <i>Eigen</i> III	=	0,7124	-0,0204	-0,0092	-0,7014
Comp. Princ. III	=	0,0870	-0,0196	-0,0088	-0,6764
Sumas de los valores <i>eigen</i>	=	3,644			
Residuales	=	0,356			
Total de la varianza explicada	=	91,10%			

Una de las ventajas que se logra con los puntajes es que si, por alguna consideración, se decide eliminar una de las variables (por ejemplo, para simplificar la representación de los datos geoquímicos), la pérdida en la varianza total es menor que si se consideran los datos en su forma original. Esto se observa claramente en el ejemplo usado, donde cada variable, en su forma original, aporta el 46% y 54% de la varianza total, respectivamente, por lo tanto, al eliminar una de dichas variables, en el mejor de los casos, sólo se pierde el 46%; por el contrario, al considerar las variables como puntajes, sólo se pierde el 14%.

ANÁLISIS DE LOS COMPONENTES PRINCIPALES DE LAS MUESTRAS GEOQUÍMICAS DE LA REGIÓN DE CHILLAN.

El procesamiento de los resultados geoquímicos de las muestras de sedimentos fluviales usadas en este trabajo, se hizo mediante un computador FACOM 230 (equivalente a IBM 360), de fabricación japonesa, perteneciente a la Facultad de Ciencias, Universidad de Tohoku, Japón.

El programa usado, que utiliza lenguaje FORTRAN 4, calcula la media y desviación standard de las observaciones, entrega la matriz de correlación y un listado de los valores *eigen*, vectores *eigen* y componentes principales que en forma individual, representa más del 20% de la varianza

original. Finalmente el programa entrega un listado de los puntajes.

Los resultados obtenidos para la población geoquímica estudiada, considerando las variables Cu, Zn, Mn, y Mo se muestra en la Tabla 4.

El componente principal I explica la asociación Zn-Mn; el componente principal II, el conjunto Cu-Mo y el componente principal III, el Cu.

- Puntaje I (Zn-Mn):** Se distribuye en dos zonas: una franja central con puntajes altos, relacionados geológicamente con el granitoide y los sedimentos generados parcialmente a partir de él; los valores bajos se distribuyen en las extensiones norte y sur del cuadrángulo donde se conocen rocas graníticas. En zonas con mineralización aparece el puntaje I con mayor dispersión (valores comprendidos entre -0,3 y -0,9), este fenómeno se explica por la influencia conjunta de rocas estériles (sedimentos terciarios y rocas volcánicas modernas) y cuerpos mineralizados. En el área del prospecto San José se observa claramente esta situación.
- Puntaje II (Cu-Mo):** Los valores más extremos (menores que -0,8) se sitúan en una franja central en la cual hay rocas intrusivas con mineralización diseminada o en venillas y sedimentos generados a partir de estas rocas. En la parte norte del área mapeada, donde predominan rocas sedimentarias terciarias y sedimentos cuaternarios la distribución del puntaje es heterogénea (valores

altos y bajos del puntaje II mezclados).

- c. Puntaje III (Cu): Presenta una clara distribución areal: los valores altos se sitúan en la franja central, relacionada con los granitoides y zonas mineralizadas (Los Azules, San José y Minas del Prado) y los sedimentos derivados de ellos (aunque no exclusivamente). Esta franja coincide con el sistema de distribución de los valores más bajos del puntaje II.

El área achurada en la Fig. 7 representa la zona de superposición de alto puntaje III (Cu) y bajo puntaje II (Cu-Mo) y de valores altos y bajos de puntaje I (Zn-Mn).

CONCLUSIONES

- a. El resultado del análisis de los componentes principales de los datos geoquímicos estudia-

dos, hecho a la luz de la información geológica disponible, permite ubicar una franja central, correspondiente aproximadamente a un tercio del área total estudiada, en la cual es posible esperar zonas mineralizadas.

- b. El análisis de los componentes principales facilita la interpretación de la información geoquímica, cuando se trabaja con un gran número de muestras y elementos.
- c. La limitación más notoria del análisis de los componentes principales es función de los coeficientes de correlación entre las variables; cuando ellos son bajos (próximos a cero), la posibilidad de lograr menos componentes principales que variables disminuye y, por lo tanto, la dimensionalidad del problema se mantiene. Esta limitación es similar para el análisis de factores.

ANALISIS DE LA FUNCION DISCRIMINANTE

INTRODUCCION

El análisis de la función discriminante es el proceso estadístico que permite asignar una muestra, de origen desconocido, a uno de dos grupos previamente definidos. Esta situación es muy común en exploración geoquímica, en zonas con gran desarrollo de suelo, como sucede en la región sur de Chile. La definición del análisis de la función discriminante puede aclararse con el siguiente ejemplo: al realizar la exploración geoquímica por cromo en cuerpos ultramáficos, que intruyen los esquistos paleozoicos del sur de Chile, se presentó la dificultad de trazar el contacto entre ambas unidades, debido al potente suelo que lo homogeniza; sin embargo, se dispuso de información geoquímica confiable (contenidos de Cr, Ni, Co, Al, Fe, Mg), obtenida en suelos desarrollados tanto sobre los esquistos como sobre los cuerpos ultramáficos. A partir de esta información, se obtuvo una función discriminante que permitió asignar las muestras de origen incierto a una u otra unidad, entregando así otro antecedente para facilitar el trazado del contacto y, por lo tanto, restringir el área de exploración.

Una función discriminante transforma los valores originales de una muestra en un puntaje discriminante, el cual indica la posición de la muestra

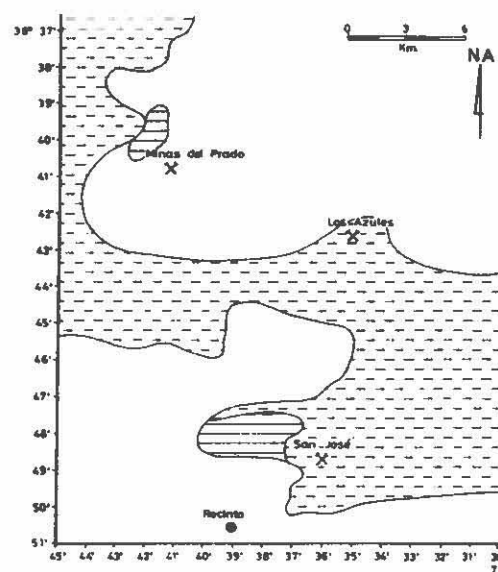
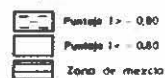


Fig. 7: Representación del Puntaje I (Zn-Mn)



a lo largo de una recta. Dicha recta está definida por la función discriminante.

La Fig. 10, tomada de Davis (1973), señala claramente la operatoria de la función discriminante. La figura muestra el caso simple de dos grupos A y

B constituidos por m_A y m_B muestras, respectivamente, en las cuales se han determinado las variables x_1 y x_2 (por ejemplo, Cu y Mo).

La función discriminante buscada debe transformar las variables originales, de modo que se obtenga una recta, entre los dos grupos, que corresponda a la máxima separación entre ellos a la

[D]: diferencia $[A_j] - [B_j]$, es decir la diferencia entre las "m" variables,

Resulta la matriz y, encontrados los coeficientes, la ecuación de la función discriminante "R" será:

$$R = \lambda_1 \phi_1 + \lambda_2 \phi + \dots + \lambda_m \phi_m$$

donde: ϕ_m = valor medio de la variable "m".

La pendiente de la recta, en un ejemplo bidimensional (2 variables) es igual a λ_2/λ_1 . El medio entre los dos grupos de muestras (R_0) se obtiene al reemplazar los coeficientes de la ecuación por el término cuya forma general es:

$$\phi_j = \frac{\bar{A}_j + \bar{B}_j}{2}$$

El valor R_0 señala exactamente la mitad entre el centro del grupo A y el centro del grupo B. El centro del grupo A, para el caso de dos variables (por ejemplo, Cu y Mo) se obtiene de:

$$R_A = \lambda_1 \bar{A}_1 + \lambda_2 \bar{A}_2$$

y el centro del grupo B:

$$R_B = \lambda_1 \bar{B}_1 + \lambda_2 \bar{B}_2$$

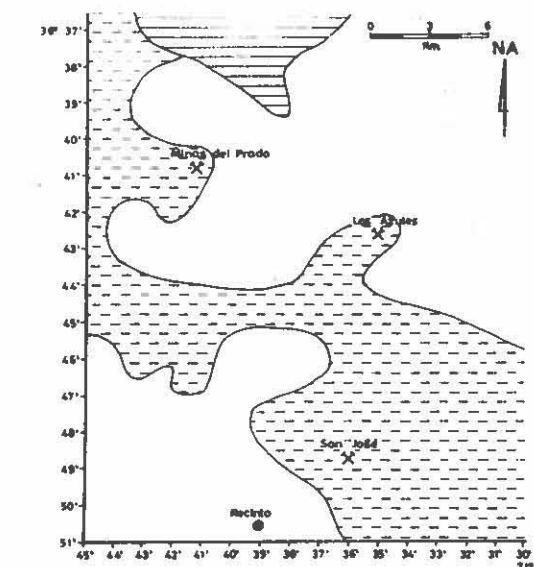


Fig. 8: Representación del Puntaje II (Cu-Mo)



vez que sea mínima la dispersión de los elementos de cada grupo. La función discriminante entrega las coordenadas de dicha recta.

El método más simple para encontrar la función es haciendo un "análisis de regresión", tomándose como variable dependiente, la diferencia de las medias entre las variables de los dos grupos. Expresada en forma de matriz, la ecuación a resolver es de la forma

$$[S_p] \cdot [\lambda] = [D]$$

donde: $[S_p]$: es la matriz de varianza-covarianza de las "m" variables.

$[\lambda]$: por convención, los coeficientes de la función discriminante (puede compararse con los coeficientes " β " de las ecuaciones de regresión).

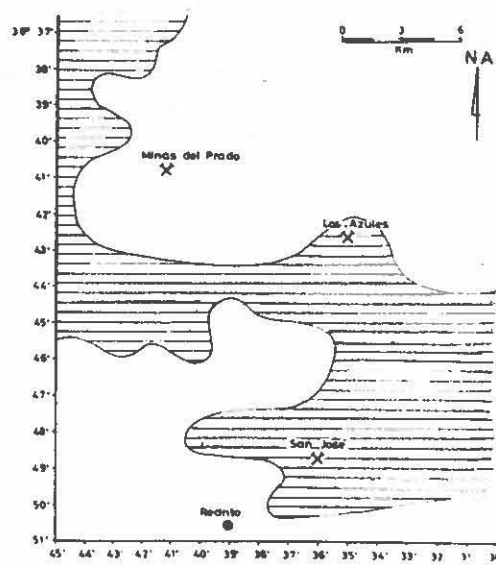


Fig. 9: Representación del Puntaje III (Cu)



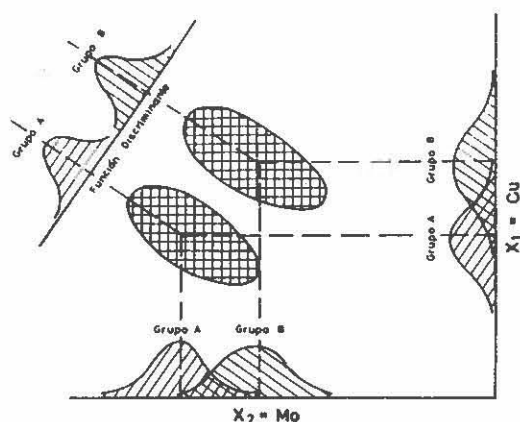


Fig. 10: Representación gráfica de la función del discriminante (Davis, 1973)

Al efectuarse la proyección de los puntajes en la recta, se observa que algunas muestras del grupo A caen entre R_O y R_B y, a su vez algunas muestras del grupo B caen entre R_O y R_A ; dichas muestras deben considerarse mal clasificadas por la función discriminante (Davis, 1973).

La distancia entre R_A y R_B lleva el nombre de "distancia de Mahalanobis" (D^2) en honor al matemático que la definió.

Para ubicar las muestras de origen desconocido, debemos reemplazar los valores ϕ por aquéllos que corresponden a la muestra en la ecuación de la función discriminante, y dibujar sobre la recta el puntaje obtenido. Esta herramienta estadística funcionará en mejor forma en la medida que la dispersión de las variables sea mínima en cada grupo (baja varianza) y que las medias de cada variable, en cada grupo, estén claramente separadas.

Un estudio más detallado y restricciones en el análisis de la función del discriminante, puede ser encontrado en los textos señalados en la bibliografía.

ANÁLISIS DEL DISCRIMINANTE CON MUESTRAS DE SEDIMENTOS FLUVIALES DE LA REGIÓN DE CHILLAN.

Para presentar en forma práctica el análisis del discriminante, se tomaron dos grupos de muestras de sedimentos que fueron analizados por Cu, Zn, Mn y Mo. Ambos conjuntos fueron clasificados como provenientes de diferente ambiente geológico, en base a la información geológica disponible. Así, el grupo A está formado por 50 muestras que

proviene de una zona con granitoides y el grupo B (56 muestras) fue obtenido de una zona en que predominan rocas volcánicas terciarias y volcanitas asignadas al Terciario superior-Cuaternario. En base al levantamiento geológico y geoquímico se había concluido que aquellas muestras provenientes de áreas predominantemente graníticas, presentaban mayor interés en exploración, por lo tanto, se deseaba disponer de un criterio que permitiera discernir el origen de las muestras, en atención a la falta de exposición de la geología.

El procesamiento de la información se hizo en un Computador FACOM 230 de la Facultad de Ciencias, Universidad de Tohoku, Japón.

El programa utilizado calcula los valores de R_A , R_B y R_O , la distancia Mahalanobis (D^2) y los valores de las constantes λ_1 a λ_n de la ecuación del discriminante; finalmente, entrega una lista de los puntajes para las muestras de cada grupo (grupos A y B).

Para el área estudiada se obtuvieron los valores señalados en las Tablas 5 y 6.

TABLA 5 Índices discriminantes R_A , R_O , R_B y distancia de Mahalanobis (D^2)

R_A	R_O	R_B	D^2
-33,24	-21,51	-9,79	-11,72

TABLA 6

Variable	Coficiente (λ_i)	Media del grupo A* ($\bar{X}_A$)	Media del grupo B* ($\bar{X}_B$)
Cu	-0,0101	86,80	32,96
Zn	-0,0218	90,20	78,40
Mn	0,0241	417,00	660,00
Mo	-10,5784	3,90	2,24

* Cálculos hechos fuera del programa.

Los puntajes se presentan gráficamente en la Fig. 11; puede observarse que algunas muestras están mal clasificadas, esto es normal, al trabajar con fenómenos geológicos, sin embargo, los grupos A y B se separan nítidamente.

Por ejemplo, se deseaba confirmar la posición respecto a la geología de las muestras 36 y 72 (ambas con Mo alto) cuyos contenidos en elementos en traza eran:

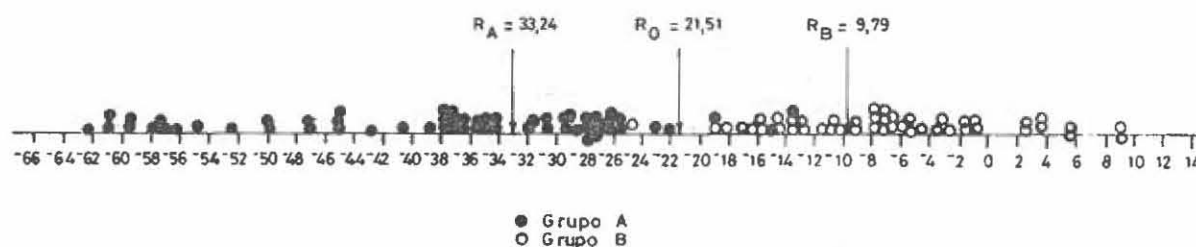


FIG. 11. Puntaje de la función del discriminante.

Muestra	Cu (ppm)	Zn (ppm)	Mn (ppm)	Mo (ppm)
36	19	42	445	14
72	31	51	476	11

Al reemplazar los valores analíticos en la ecuación de la función discriminante R , se obtiene:

$$R_{36} = -(0,01 \times 19) - (0,02 \times 42) + (0,02 \times 445) - (10,58 \times 14) = -138,50$$

$$R_{72} = -(0,01 \times 31) - (0,02 \times 51) + (0,02 \times 476) - (10,58 \times 11) = -106,34$$

Llevados dichos valores a la recta que representa la función discriminante (Fig. 11) se observa que las muestras pertenecen al grupo A (sedimentos provenientes de zonas con predominio de granitoides).

CONCLUSIONES

- El análisis de la función discriminante es una técnica estadística que permite discernir la ubicación de una muestra geoquímica en uno de dos grupos previamente definidos. Aunque se ha presentado como una ayuda en la interpretación de datos geoquímicos, puede ser utilizada en paleontología, petrografía, etc.
- El análisis de la función discriminante, aunque opera con 2 poblaciones de "m" variables, reduce la dimensionalidad del problema a una variable.

AGRADECIMIENTOS

El autor agradece al Dr. H. Kanno, del Instituto de Mineralogía, Petrología y Geología Económica, Universidad de Tohoku, Japón, el haberlo introdu-

cido en el campo de la estadística aplicada a la exploración geoquímica.

REFERENCIAS

- AGTEMBERG, F.D. 1974. *Geomathematics*. Elsevier Sci. Publ. Co., Amsterdam.
- CONN, H.; PAGE, B. 1973. Prospección geoquímica de cromo-níquel en rocas ultrabásicas, Provincias de Cautín y Valdivia, Chile. *Inst. Invest. Geol.* (inédito), 55 p. Santiago.
- CRUZAT, A. 1974. Prospección geoquímica del yacimiento Gallehué, Provincia de Malleco. *Inst. Invest. Geol.* (inédito), 22 p. Santiago.
- DAVIS, J.C. 1973. *Statistics and data analysis in Geology*. John Wiley & Sons Inc., New York.
- FUENZALIDA, H. 1965. Orografía. In *Geografía Económica de Chile*. (Texto refundido). Corp. Fomento de la Producción (CORFO), p. 7-33. Santiago.
- HARBAUGH, J.; MERRIAM, D. 1968. *Computer applications in stratigraphic analysis*. John Wiley & Sons Inc., New York.

**INSTITUTO DE INVESTIGACIONES GEOLOGICAS;
METAL MINING AGENCY OF JAPAN. 1978.**
Informe de reconocimiento geológico del área
Coihueco(Nuble) - Lonquimay (Malleco), zona
Central-Sur de Chile. (inédito) 352 p. Santiago-
Tokyo.

MILLER, R.L.; KAHN, J.S. 1965. Statistical analysis

in the Geological Sciences. John Wiley & Sons
Inc., New York.

PAGE, B.; CONN, H. 1973 Investigación sobre métodos
de prospección geoquímica en el yacimiento tipo
cobre porfírico El Abra, Provincia de Antofagasta,
Chile. Inst. Invest. Geol. (inédito), 74 p. Santiago.