



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE INGENIERIA

**ANALISIS MULTIVARIABLE DE INFORMACION GEOQUIMICA
PARA LA DETECCION DE ANOMALIAS**

TESIS PROFESIONAL

QUE PARA OBTENER EL TITULO DE
INGENIERO GEOFISICO
P R E S E N T A
SANTIAGO MUÑOZ FLAVIO
MEXICO, D. F. 1984



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE INGENIERIA

**ANALISIS MULTIVARIABLE DE INFORMACION GEOQUIMICA
PARA LA DETECCION DE ANOMALIAS**

TESIS PROFESIONAL

QUE PARA OBTENER EL TITULO DE
INGENIERO GEOFISICO

P R E S E N T A

SANTIAGO MUÑOZ FLAVIO

MEXICO, D. F.

1984

I N D I C E

	Pag.
INTRODUCCION -----	1
CAPITULO 1 ELEMENTOS DEL ANALISIS FACTORIAL -----	3
1.1 Introducción al Análisis Factorial -----	3
1.2 Conceptos Fundamentales -----	3
1.2.a Caso Determinístico -----	4
1.2.b Caso Aleatorio -----	4
1.3 Modelos Definidos para el Análisis Factorial	4
1.3.a Modelo para las Variables -----	4
1.3.b Modelo para la Varianza y Covarianza ----	6
1.4 Análisis de Componentes Principales -----	8
1.5 Análisis Factorial Verdadero -----	9
CAPITULO 2 ANALISIS FACTORIAL TIPO - R -----	10
2.1 Introducción -----	10
2.2 Análisis de Componentes -----	11
2.2.a Caso Determinístico -----	11
2.2.b Caso Aleatorio -----	16

	Pag.
2.3 Factores para Componentes Principales -----	17
2.4 Análisis Factorial Verdadero Tipo - R -----	18
CAPITULO 3 ANALISIS FACTORIAL TIPO - Q -----	21
3.1 Introducción -----	21
3.2 Método de Imbrie -----	21
3.3 Desarrollo Matemático -----	23
3.4 Cálculo de los Factores (F) -----	25
CAPITULO 4 ROTACION DE FACTORES -----	27
4.1 Introducción -----	27
4.2 Procedimiento Varimax -----	27
CAPITULO 5 ANALISIS DE RESULTADOS -----	31
5.1 Exploración Geoquímica Superficial -----	31
5.2 Descripción de la Información Geoquímica ---	33
5.3 Definición de Anomalías Geoquímicas -----	34

INDICE DE TABLAS

TABLA 5.1 Conjunto I de Datos -----	38
TABLA 5.2 Conjunto II de Datos -----	41
TABLA 5.3 Matriz de Pesos "A" para el Conjunto I -----	53
TABLA 5.4 Matriz de Pesos "A" para el Conjunto II -----	54
TABLA 5.5 Eigenvalores para el Conjunto I -----	55

	Pag.
TABLA 5.6 Eigenvalores para el Conjunto II -----	56
TABLA 5.7 Matriz de Factores F que se configura para el --- Conjunto I -----	57
TABLA 5.8 Matriz de Factores F que se configura para el -- Conjunto II -----	60

INDICE DE FIGURAS

FIGURA 5.1-A Configuración del Factor Hidrocarburos Ligeros para el Conjunto I de Datos -----	69
FIGURA 5.2-A Configuración del Factor Olefinas para el Con- junto I de Datos -----	70
FIGURA 5.3-A Configuración del Factor Gas Seco para el Con- junto I de Datos -----	71
FIGURA 5.4-A Localización de las Anomalías para el Conjunto I de Datos -----	72
FIGURA 5.1-B Configuración del Factor Hidrocarburos Ligeros para el Conjunto II de Datos -----	73
FIGURA 5.2-B Configuración del Factor Olefinas para el Con- junto II de Datos -----	74
FIGURA 5.3-B Configuración del Factor Gas Seco para el Con- junto II de Datos -----	75

FIGURA 5.4-B Localización de las Anomalías para el Conjunto	
II de Datos-----	76
CONCLUSIONES -----	77
AGRADECIMIENTOS -----	79
BIBLIOGRAFIA-----	80
APENDICE A: Programa de Computación Utilizado para Proce- sar la Información -----	82
APENDICE B: Propiedades de los Eigenvalores y Eigenvectores en Términos de Mínimos Cuadrados -----	96
APENDICE C: Rotación de Sistemas de Coordenadas-----	101

INTRODUCCION

Frecuentemente en distintas áreas científicas es común que existan fenómenos que involucren un número considerable de parámetros, como es el caso de las disciplinas del área de ciencias de la tierra (geofísica, petrolera, geología y minería) donde la información adquiere una presentación multivariable. Por tal motivo se considera que es importante la aplicación de técnicas estadísticas que ayuden a manipular dicha información, y sobre todo que simplifiquen el tiempo de trabajo para analizar e interpretar los resultados.

Una de estas técnicas es el análisis factorial, el cual propone un modelo matemático, que describe la correlación que existe entre un cierto número de variables o atributos en términos de un número relativamente menor de factores fundamentales.

El análisis factorial ha sido y sigue aplicándose en la interpretación de datos geológicos, geofísicos y geoquímicos, entre otros. La información con la cual se demostrará el uso de esta técnica es precisa-

menta de carácter geoquímico y fue obtenida en una de las regiones donde Petróleos Mexicanos (PEMEX) ha efectuado campañas de exploración.

El objetivo de este estudio es señalar las posibles zonas anómalas que indiquen la presencia de hidrocarburos en el subsuelo.

Introduciremos, primeramente, el método del análisis factorial describiendo las distintas formas en que éste puede ser presentado y estudiado para, posteriormente, discutir los resultados obtenidos al aplicarlo en datos geoquímicos cuando el propósito es la detección de anomalías.

I N D I C E

	Pag.
INTRODUCCION -----	1
CAPITULO 1 ELEMENTOS DEL ANALISIS FACTORIAL -----	3
1.1 Introducción al Análisis Factorial -----	3
1.2 Conceptos Fundamentales -----	3
1.2.a Caso Determinístico -----	4
1.2.b Caso Aleatorio -----	4
1.3 Modelos Definidos para el Análisis Factorial	4
1.3.a Modelo para las Variables -----	4
1.3.b Modelo para la Varianza y Covarianza ----	6
1.4 Análisis de Componentes Principales -----	8
1.5 Análisis Factorial Verdadero -----	9
CAPITULO 2 ANALISIS FACTORIAL TIPO - R -----	10
2.1 Introducción -----	10
2.2 Análisis de Componentes -----	11
2.2.a Caso Determinístico -----	11
2.2.b Caso Aleatorio -----	16

	Pag.
2.3 Factores para Componentes Principales -----	17
2.4 Análisis Factorial Verdadero Tipo - R -----	18
CAPITULO 3 ANALISIS FACTORIAL TIPO - Q -----	21
3.1 Introducción -----	21
3.2 Método de Imbrie -----	21
3.3 Desarrollo Matemático -----	23
3.4 Cálculo de los Factores (F) -----	25
CAPITULO 4 ROTACION DE FACTORES -----	27
4.1 Introducción -----	27
4.2 Procedimiento Varimax -----	27
CAPITULO 5 ANALISIS DE RESULTADOS -----	31
5.1 Exploración Geoquímica Superficial -----	31
5.2 Descripción de la Información Geoquímica ---	33
5.3 Definición de Anomalías Geoquímicas -----	34

INDICE DE TABLAS

TABLA 5.1 Conjunto I de Datos -----	38
TABLA 5.2 Conjunto II de Datos -----	41
TABLA 5.3 Matriz de Pesos "A" para el Conjunto I -----	53
TABLA 5.4 Matriz de Pesos "A" para el Conjunto II -----	54
TABLA 5.5 Eigenvalores para el Conjunto I -----	55

	Pag.
TABLA 5.6 Eigenvalores para el Conjunto II -----	56
TABLA 5.7 Matriz de Factores F que se configura para el -- Conjunto I -----	57
TABLA 5.8 Matriz de Factores F que se configura para el -- Conjunto II -----	60

INDICE DE FIGURAS

FIGURA 5.1-A Configuración del Factor Hidrocarburos Ligeros para el Conjunto I de Datos -----	69
FIGURA 5.2-A Configuración del Factor Olefinas para el Con- junto I de Datos -----	70
FIGURA 5.3-A Configuración del Factor Gas Seco para el Con- junto I de Datos -----	71
FIGURA 5.4-A Localización de las Anomalías para el Conjunto I de Datos -----	72
FIGURA 5.1-B Configuración del Factor Hidrocarburos Ligeros para el Conjunto II de Datos -----	73
FIGURA 5.2-B Configuración del Factor Olefinas para el Con- junto II de Datos -----	74
FIGURA 5.3-B Configuración del Factor Gas Seco para el Con- junto II de Datos -----	75

	Pag.
FIGURA 5.4-B Localización de las Anomalías para el Conjunto	
II de Datos-----	76
CONCLUSIONES -----	77
AGRADECIMIENTOS -----	79
BIBLIOGRAFIA-----	80
APENDICE A: Programa de Computación Utilizado para Proce- sar la Información -----	82
APENDICE B: Propiedades de los Eigenvalores y Eigenvectores en Términos de Mínimos Cuadrados -----	96
APENDICE C: Rotación de Sistemas de Coordenadas-----	101

INTRODUCCION

Frecuentemente en distintas áreas científicas es común que existan fenómenos que involucren un número considerable de parámetros, como es el caso de las disciplinas del área de ciencias de la tierra (geofísica, petrolera, geología y minería) donde la información adquiere una presentación multivariable. Por tal motivo se considera que es importante la aplicación de técnicas estadísticas que ayuden a manipular dicha información, y sobre todo que simplifiquen el tiempo de trabajo para analizar e interpretar los resultados.

Una de estas técnicas es el análisis factorial, el cual propone un modelo matemático, que describe la correlación que existe entre un cierto número de variables o atributos en términos de un número relativamente menor de factores fundamentales.

El análisis factorial ha sido y sigue aplicándose en la interpretación de datos geológicos, geofísicos y geoquímicos, entre otros. La información con la cual se demostrará el uso de esta técnica es precisa-

mente de carácter geoquímico y fue obtenida en una de las regiones donde Petróleos Mexicanos (PEMEX) ha efectuado campañas de exploración.

El objetivo de este estudio es señalar las posibles zonas anómalas que indiquen la presencia de hidrocarburos en el subsuelo.

Introduciremos, primeramente, el método del análisis factorial describiendo las distintas formas en que éste puede ser presentado y estudiado para, posteriormente, discutir los resultados obtenidos al aplicarlo en datos geoquímicos cuando el propósito es la detección de anomalías.

CAPITULO 1

ELEMENTOS DEL ANALISIS FACTORIAL

1.1 INTRODUCCION AL ANALISIS FACTORIAL.

El análisis factorial es un método que estudia las intercorrelaciones que existen entre un conjunto de variables u objetos, con la finalidad de encontrar un grupo de variables hipotéticas llamadas factores, los cuales poseen la información esencial de todo el conjunto de variables analizadas, reduciéndose de este modo la complejidad de los datos. Existen varios casos del análisis factorial pero únicamente se explicarán el tipo-R y brevemente el tipo-Q.

1.2 CONCEPTOS FUNDAMENTALES.

El análisis factorial se ha desarrollado por métodos descriptivos e inferenciales de los cuales se han derivado dos casos, uno determinins-

tico y otro aleatorio.

1.2.a Caso Determinístico.

Analiza un conjunto de datos en relación a una población estadística. Los factores obtenidos se interpretan con respecto a los datos muestreados.

1.2.b Caso Aleatorio.

Tenemos este caso cuando el muestreo de objetos se hace en forma aleatoria con relación a alguna población específica.

1.3 MODELOS DEFINIDOS PARA EL ANALISIS FACTORIAL.

Existen dos formas de definir estos modelos:

1.3.a Haciendo comparaciones entre variables.

1.3.b Comparando la varianza y la covarianza de las variables observadas.

1.3.a Modelo para las variables.

El modelo para el análisis factorial se ha desarrollado con relación a la matriz de datos (caso determinístico). La ecuación que lo representa es:

$$X_{(N \times p)} = F_{(N \times k)} A'_{(k \times p)} + E_{(N \times p)} \quad (1.1)$$

X = matriz de datos

F = matriz de factores

A = matriz de pesos asociados a los factores

La ecuación (1.1) representa el modelo matemático para el análisis factorial tipo-R.

En forma general para p variables y k factores, cualquier vector columna de X es:

$$x_i = a_{i1} f_1 + a_{i2} f_2 + \dots + a_{ik} f_k + e_i \quad (1.2)$$

y si aplicamos concepto de combinación lineal para la i -ésima componente, tendremos:

$$c_i = a_{i1} f_1 + a_{i2} f_2 + \dots + a_{ik} f_k \quad (1.3)$$

siendo éste el estimador lineal óptimo de x_i y el cual es llamado parte común o comunalidad de x_i . En este caso, e_i , de la matriz error E , se toma como la suma de las partes no correlacionables, s_i y ϵ_i :

$$c_i = s_i + \epsilon_i \quad (1.4)$$

donde ϵ_i es el error medido y s_i se refiere a la parte específica de x_i .

1.3.b Modelo para la varianza y covarianza.

Si combinamos (1.2), (1.3) y (1.4) obtenemos una nueva ecuación para x_i :

$$x_i = c_i + e_i = c_i + s_i + \epsilon_i \quad (1.5)$$

cuya varianza es:

$$\sigma_{x_i}^2 = \sigma_{c_i}^2 + \sigma_{e_i}^2 = \sigma_{c_i}^2 + \sigma_{s_i}^2 + \sigma_{\epsilon_i}^2 \quad (1.6)$$

y en la cual los términos representan lo que se indica a continuación:

$\sigma_{c_i}^2$ = varianza común o comunalidad

$\sigma_{e_i}^2$ = varianza residual, la cual se refiere como término único

$\sigma_{s_i}^2$ = varianza específica de x_i

$\sigma_{\epsilon_i}^2$ = varianza error de x_i

En una aplicación práctica es difícil conocer estos términos por lo que es necesario estimarlos de alguna forma, siendo una de ellas la que a continuación se desarrollará. De la ecuación (1.3)

$$c_i = a_{11} f_1 + a_{12} f_2 + \dots + a_{1k} f_k$$

por lo tanto:

$$\sigma_{c_i}^2 = \sum_{n=1}^N c_{ni}^2 / N = \sum_{n=1}^N (a_{11} f_{n1} + a_{12} f_{n2} + \dots + a_{1k} f_{nk})^2 / N$$

$$\begin{aligned}
&= a_{i1}^2 \sum f_{n1}^2/N + a_{i2}^2 \sum f_{n2}^2/N + \dots + a_{ik}^2 \sum f_{nk}^2/N + \\
&+ a_{i1}a_{i2} \frac{\sum f_{n1} f_{n2}}{N} + a_{i1}a_{i3} \frac{\sum f_{n1} f_{n3}}{N} + \dots + \\
&+ a_{ik-1} a_{ik} \frac{\sum f_{nk-1} f_{nk}}{N} \tag{1.7}
\end{aligned}$$

Los términos $\sum f_{nj}^2/N$ son la varianza de los factores la cual puede suponerse unitaria, es decir que los factores están en forma estandarizada. Algunas veces $\sum f_{nj} f_{ne}/N$ se toma como la correlación entre factores. Por lo tanto la ecuación (1.7) se simplifica:

$$\sigma_{ci}^2 = a_{i1}^2 + a_{i2}^2 + \dots + a_{ik}^2 + a_{i1}a_{i2}\phi_{12} + \dots + a_{ik-1}a_{ik}\phi_{k-1k} \tag{1.8}$$

donde ϕ_{ij} es la correlación entre factores. Si no existe correlación entre los factores (1.6) es:

$$\sigma_{ci}^2 = a_{i1}^2 + a_{i2}^2 + \dots + a_{ik}^2 \tag{1.9}$$

Por lo tanto la comunalidad para la variable x_i es la suma del i -ésimo renglón de la matriz de pesos asociados a los factores (A).

1.3.b Modelo para la varianza y covarianza.

El modelo que considera a la covarianza supone que la matriz de datos está en forma desviada, es decir:

$$Y = FA' + E \tag{1.10}$$

y

$$Y' = AF' + E' \quad (1.11)$$

En términos de la varianza y covarianza:

$$\frac{1}{N} Y'Y = A \left[\frac{1}{N} F'F \right] A' + A \left[\frac{1}{N} F'E' \right] + \left[\frac{1}{N} E'F \right] A' + \frac{1}{N} E'E \quad (1.12)$$

en la cual Y es una matriz de N renglones que representa el muestreo de una población de p variables. Puesto que el tamaño de las muestras puede ser aumentado, si este se hace indefinido, cualquier término de (1.12) convergerá siempre al valor de la población. Por lo tanto:

$$\frac{1}{N} Y'Y \rightarrow \Sigma, \quad \frac{1}{N} F'F \rightarrow \Phi, \quad \frac{1}{N} F'E=0, \quad \frac{1}{N} E'E \rightarrow \Psi$$

luego entonces la relación (1.12) se simplifica:

$$\Sigma = A\Phi A' + \Psi \quad (1.13)$$

donde:

Σ = matriz de covarianza de $p \times p$ variables observadas

A = matriz de pesos asociados a los factores

Φ = matriz de covarianza para los factores

Ψ = matriz residual de covarianza

1.4 ANALISIS DE COMPONENTES PRINCIPALES.

En el análisis de componentes, los factores se determinan utilizando la máxima varianza de las variables, el término residual es pequeño y no es correlacionable con los factores ni consigo mismo. Los elementos de la matriz residual de covarianza Ψ tienen valores cercanos a cero, por lo tanto, la matriz de covarianza para las variables observadas se obtiene aproximadamente con la ecuación (1.13) sin tomar en cuenta a Ψ , es decir:

$$\Sigma = A\Phi A'$$

1.5 ANALISIS FACTORIAL VERDADERO.

En el análisis factorial verdadero, los factores se definen de acuerdo a la máxima intercorrelación que existe entre las variables, el término residual es pequeño y admite una cantidad considerable de términos únicos, de éstos se utiliza únicamente la parte de la variable que será correlacionada con otras variables. La matriz de covarianza residual Ψ es una matriz diagonal ya que la correlación entre los residuos es cero y la matriz de covarianza se define igual que en la ecuación (1.13).

CAPITULO 2

ANALISIS FACTORIAL TIPO - R

2.1 INTRODUCCION.

Este tipo de análisis estudia las intercorrelaciones entre variables. En la matriz de datos X los renglones constituyen una muestra aleatoria de las observaciones de la población, además, de X se calcula la matriz de covarianza de las muestras, S . Para el caso aleatorio se tratará de ajustar el modelo de la población al de los datos muestreados; en el caso determinístico el modelo de la población es aproximadamente igual al de los datos y para ambos, S o X , se aproximan por matrices de rango más bajo.

Para el caso determinístico es importante la matriz de pesos asociados a los factores (A) y la matriz de factores (F). En el caso aleatorio únicamente interesa (A) y la matriz de covarianza (Φ) de los factores.

Enseguida se hará el estudio considerando el análisis de componentes y el análisis de factores por separado, una marca (\wedge) sobre un símbolo de la población indicará que la muestra es estimada.

2.2 ANALISIS DE COMPONENTES

2.2.a Caso Determinístico.

Empezaremos considerando nuevamente que la matriz de datos X está en forma desviada, es decir:

$$Y = FA' + E \quad (2.1)$$

Por conveniencia asumiremos que todos los elementos de Y se dividen por \sqrt{N} por lo que la matriz de covarianza S se simplifica en:

$$S = Y'Y$$

En este caso nos aproximaremos al modelo de la población aplicando el método de mínimos cuadrados pesados a la matriz de datos Y (consultar Apéndice B). Calcularemos $F_{(N \times k)}$ y $A_{(p \times k)}$ tomando en cuenta que $k < p$, de tal forma que el cuadrado de todos los elementos de la matriz:

$$E = Y - FA' \quad (2.2)$$

sea lo más pequeña posible.

La solución para este problema de acuerdo a lo que se enuncia en el Apéndice B queda establecida a partir de la siguiente relación:

$$\widehat{FA} = \gamma_1 v_1 u_1' + \gamma_2 v_2 u_2' + \dots + \gamma_k v_k u_k'$$

y si consideramos que:

$$V_k = \begin{bmatrix} v_1 & v_2 & \dots & v_k \end{bmatrix}, \quad U_k = \begin{bmatrix} u_1 & u_2 \dots & u_k \end{bmatrix}$$

$$\text{y } \Gamma_k = \text{diag} (\gamma_1, \gamma_2, \dots, \gamma_k)$$

finalmente tendremos que la solución propuesta toma una forma más condensada:

$$\widehat{FA} = V_k \Gamma_k U_k'$$

De esta ecuación se notará que el producto \widehat{FA} , no siempre da una solución única para F y A, ya que es claro que hay infinidad de soluciones para ambos, por ejemplo, simplemente con reemplazar a F por $F^* = FT$ y A por $A^* = AT^{-1}$, donde T es una matriz de transformación no singular. Para resolver este problema propenderemos dos casos diferentes de solución haciendo la diferencia únicamente si los factores están escalados o no.

Solución 1

Consideramos que:

$$\hat{F} = V_k \text{ y } \hat{A} = U_k \Gamma_k \quad (2.9)$$

Esto es consecuencia de saber que la suma de las columnas de Y son igual a cero y que por lo tanto las V_k también son cero. De esta manera, F está en forma desviada y la matriz de covarianza de factores se obtiene con la ecuación siguiente:

$$\hat{F}'\hat{F} = V_k' V_k = I$$

en la cual los factores son incorrelacionables y estandarizados.

Para la matriz de pesos asociados a cada factor tendremos:

$$\hat{A}'\hat{A} = \Gamma_k' U_k' U_k \Gamma_k = \Gamma_k'^2 = \Lambda_k$$

donde Λ_k es una matriz diagonal de $k \times k$ y sus elementos son $\lambda_1 = \gamma_1^2$, $\lambda_2 = \gamma_2^2 \dots \lambda_k = \gamma_k^2$, los eigenvalores de S . Si los elementos de E son cercanos a cero, entonces la matriz de datos puede definirse aproximadamente con la siguiente expresión:

$$Y \approx \hat{F}\hat{A}'$$

y la matriz de covarianza S es igual a:

$$S = Y'Y \approx \hat{A}'\hat{F}'\hat{F}\hat{A} = \hat{A}'\hat{A}$$

Resumiendo, para llegar a esta solución se puede proceder tomán

do en cuenta el siguiente orden:

- Calculamos la matriz de covarianza S
- Calculamos los k eigenvalores Λ_k y los correspondientes eigenvectores U_k de S.
- Calculamos $A = U_k \Lambda_k^{1/2}$
- Calculamos $F = Y \hat{A} \Lambda_k^{-1}$

Solución 2

Consideremos que:

$$\hat{A} = U_k, F = V_k \Gamma_k \quad (2.4)$$

La matriz de covarianza para los factores queda como:

$$\hat{F}'\hat{F} = \Gamma_k' V_k' V_k \Gamma_k = \Gamma_k^2 = \Lambda_k$$

es decir es una matriz diagonal, con elementos iguales a los eigenvalores de S en orden descendente de magnitud. Para la matriz de pesos asociados a cada factor tendremos que:

$$\hat{A}'\hat{A} = U_k' U_k = I$$

y para la matriz de covarianza

$$S = Y'Y \approx \hat{A}'\hat{F}'\hat{F}\hat{A} = \hat{A}'\Lambda_k\hat{A}$$

De la misma forma que en el caso anterior, el orden que se propone para llegar a esta solución es el siguiente:

- Calculamos la matriz de covarianza S .
- Calculamos los eigenvalores λ_k y los correspondientes eigenvectores U_k de S .
- Hacemos $\hat{A} = U_k$
- Calculamos $\hat{F} = Y\hat{A}$

Para esta solución no es necesario escalar las columnas, por lo que resulta más simple hacer el cálculo.

En la solución 1 lo que se obtiene es la correlación entre variables y factores, faltando únicamente dividir cualquier renglón de A por la desviación estándar de las variables. La matriz de correlación entre variables y factores es:

$$\hat{C} = D_S^{-1} \hat{A} = D_S^{-1} U_k \Lambda_k^{1/2}$$

en la que D_S es una matriz diagonal, cuyos elementos están dados por la desviación estándar de las variables.

Para la solución 2 los factores no son correlacionables, tienen diferente varianza, y las columnas de \hat{A} no son directamente comparables. La matriz de covarianza entre los factores está dada por Λ_k , pero

la covarianza entre variables y factores es A_k y su correlación correspondiente se obtiene como lo indica la siguiente ecuación.

$$\hat{C} = D_S^{-1} \hat{A} \Lambda_k \Lambda_k^{-1/2} = D_S^{-1} U_k \Lambda_k^{1/2}$$

2.2.b Caso aleatorio.

Para este caso el modelo queda definido con la siguiente ecuación:

$$\Sigma = AA' + \Psi \quad (2.5)$$

En el análisis de componentes los elementos de E y Ψ se suponen pequeños, por lo tanto un criterio razonable que se adapta al modelo de los datos es hacer $S-AA'$ lo más pequeña posible por la misma conveniencia del método. En el análisis de componentes escogeremos A de tal forma que el cuadrado de todos los elementos de la expresión anterior sea mínima. Para este caso también nos auxiliaremos del método de mínimos cuadrados pesados para aproximarnos al modelo de la población (consultar Apéndice B). La solución para este problema es:

$$\hat{AA}' = \lambda_1 u_1 u_1' + \lambda_2 u_2 u_2' + \dots + \lambda_k u_k u_k' \quad (2.6)$$

o en forma matricial

$$\hat{AA}' = U_k \Lambda_k U_k'$$

donde Λ_k es una matriz diagonal cuyos elementos son los k eigenvalores

extraídos de S y las columnas U_k son los correspondientes eigenvectores.

La solución para A puede tomarse como:

$$\hat{A} = U_k \Lambda_k^{1/2} \quad (2.7)$$

siendo este resultado el mismo que se obtuvo en el caso determinístico.

Notaremos además que:

$$\hat{A}'\hat{A} = \Lambda_k \quad (2.8)$$

La solución (2.7), es la que corresponde a los componentes principales, pero no es la única que resuelve el problema planteado, por lo que se propone una más general con la siguiente ecuación:

$$\hat{A} = U_k \Lambda_k^{1/2} T \quad (2.9)$$

donde T es matriz ortogonal de orden $k \times k$. La solución (2.9) corresponde a una transformación ortogonal de los componentes principales.

2.3 FACTORES PARA COMPONENTES PRINCIPALES

Consideremos un vector de observación, x . Entonces si los k componentes principales se ajustan razonablemente, aproximadamente tendremos que:

$$x \approx \hat{A}f$$

donde f es el vector de componentes principales. Premultiplicando esta ecuación por \hat{A}' :

$$\hat{A}' x \approx \hat{A}' \hat{A} f$$

o de (2.8)

$$\hat{A}' x \approx \Lambda_k f$$

por lo tanto

$$f \approx \Lambda_k^{-1} \hat{A}' x$$

o de (2.7)

$$f \approx \Lambda_k^{-1/2} U_k x \quad (2.10)$$

Tomando la traspuesta de esta expresión y escribiendo para todos los vectores observados de las muestras, obtenemos finalmente:

$$\hat{F} \approx X U_k \Lambda^{-1/2} \quad (2.11)$$

Es decir los factores son el producto de los k primeros eigenvec-
tores con la matriz de datos, escalados por la raíz cuadrada de la matriz
inversa de los eigenvalores.

2.4 ANALISIS FACTORIAL VERDADERO TIPO-R

La ecuación de este modelo es igual a la descrita por (2.5), solamente que ahora Ψ es una matriz diagonal con elementos no iguales y representados por las varianzas únicas, las que se estiman de los datos al mismo tiempo que la matriz de pesos asociadas a cada factor, A . Un criterio a seguir para aproximarnos al modelo de los datos es minimizar la suma de los cuadrados de todos los elementos de $S - AA' - \Psi$, i.e. minimizando:

$$\text{tr} (S - AA' - \Psi)^2 \quad (2.12)$$

Si Ψ es conocida, AA' se ajustará como $S - \Psi$, con lo cual escogemos las columnas de A como los eigenvectores de $S - \Psi$ correspondientes a los k eigenvalores y escalados como la suma de los cuadrados de cada columna correspondiente a los eigenvalores. En la práctica Ψ es desconocida, pero puede estimarse de los datos, esto último de estimar Ψ equivale a tomar las comunalidades. Los elementos de $S - \Psi$ son estimados como las comunalidades. Tradicionalmente, el método usado consiste en escoger la comunalidad de cualquier variable con relación al cuadrado de los múltiplos de los coeficientes de correlación de una variable con las otras variables. Esto se demuestra tomando:

$$\hat{\Psi} = \text{diag} (S^{-1})^{-1} \quad (2.13)$$

es decir el conjunto de Ψ_S , es igual al recíproco de los elementos de la diagonal de S^{-1} . Una vez seleccionada Ψ , podremos empezar a hacer

las estimaciones sobre A . Este método es llamado Factores Principales. Cuando se haya estimado A a partir de A podremos escoger el mejor estimador de Ψ con la siguiente ecuación

$$\hat{\Psi} = \text{diag} (S - \hat{A}\hat{A}')$$

Esta relación puede ser usada como base para obtener un nuevo estimador de A , continuando con el proceso iterativo hasta que el estimador A converja a un valor estable.

CAPITULO 3:

ANALISIS FACTORIAL TIPO-Q

3.1 INTRODUCCION

El análisis factorial tipo-Q está diseñado para buscar interrelaciones entre objetos. Su desarrollo analítico establece coeficientes de similitud a través de una matriz de $N \times N$ llamada de similaridad, la cual da el grado de semejanza entre todos los posibles pares de N objetos.

La parte analítica de esta técnica puede desarrollarse siguiendo varios métodos, pero en este trabajo únicamente se hará énfasis en uno de ellos, el método de Imbrie.

3.2 METODO DE IMBRIE

La técnica seguida por este método es definir similaridad de acuerdo a las proporciones de sus componentes. Es decir este autor define

un índice de similitud proporcional, el cual da el grado de semejanza entre dos objetos, evaluando de esta forma la relación proporcional que existe entre sus constituyentes.

Para dos objetos cualesquiera (definidos como vectores renglón de la matriz de datos), el coeficiente de similitud proporcional, Cos° , se determina con la siguiente expresión:

$$\text{Cos}^\circ_{nm} = \frac{\sum_{j=1}^p x_{nj} \cdot x_{mj}}{\sqrt{\sum_{j=1}^p x_{nj}^2 \cdot \sum_{j=1}^p x_{mj}^2}} \quad (3.1)$$

Como observamos únicamente se calcula el coseno del ángulo entre dos vectores situados en un espacio p - dimensional.

Cuando tenemos un conjunto de N objetos, el Cos°_{nm} puede calcularse para cualquier par posible de los mismos, no sin antes haberlos arreglado en una matriz de asociación de $N \times N$ que será representada por la letra H .

El Cos°_{nm} se calculará siguiendo el camino que a continuación se desarrollará. Primero definiremos:

$$w = \frac{x_{nj}}{\sqrt{\sum_{j=1}^p x_{nj}^2}} \quad (j=1, \dots, p; n=1, \dots, N) \quad (3.2)$$

Esta ecuación indica que se divide el elemento de un renglón por la raíz cuadrada de la suma de todos los renglones normalizados, con lo cual la matriz de datos queda elevada el cuadrado:

$$\sum_{j=1}^p w_{nj}^2 = 1 \quad \text{para } n=1, \dots, N$$

por lo tanto

$$\text{Cos}^{\oplus}_{nm} = \sum_{j=1}^N w_{nj} w_{mj} \quad (3.3)$$

Sin embargo es más conveniente utilizar la notación matricial con lo que la expresión anterior queda:

$$W_{(N \times p)} = D^{-1/2} X \quad (3.4)$$

esta ecuación considera que cualquier vector renglón de W es de longitud unitaria. La matriz de asociación es:

$$H = WW' = D^{-1/2} XX'D^{-1/2} \quad (3.5)$$

3.3 DESARROLLO MATEMATICO

Para recalcar y ampliar un poco más el análisis descrito anteriormente se hará el siguiente desarrollo:

La matriz de datos, X , es una matriz con renglones normalizados:

$$W = D^{-1/2} X$$

en la cual $D = \text{diag}(XX')$.

La matriz de asociación H está definida por la siguiente ecuación

$$H = WW'$$

La matriz de datos normalizados W también se puede expresar aproximadamente como el producto de la matriz de pesos asociada a cada factor por la matriz de factores:

$$W \approx AF' \quad (3.6)$$

A - es una matriz de $N \times k$

F - es una matriz de $p \times k$

k - es aproximadamente el rango de W

La relación entre W , H , A y F está dada por:

$$H = WW' = AF'FA' \quad (3.7)$$

De la ecuación (3.7) podemos decir que F es ortonormal ya que si realizamos la operación que se indica a continuación obtenemos la matriz identidad:

$$FF' = I$$

lo que conduce a definir una nueva ecuación para H:

$$H = AA' \quad (3.8)$$

La matriz H es una matriz simétrica que también puede definirse con la siguiente relación:

$$H = U\Lambda U' \quad (3.9)$$

donde U es la matriz de eigenvectores y Ψ es la matriz diagonal asociada a los eigenvalores. Por lo tanto:

$$H = U\Lambda U'$$

y

$$\hat{A} = U_k \Lambda_k^{1/2} \quad (3.10)$$

Es decir la matriz de pesos asociada a los factores es la matriz de eigenvectores, escalada por la raíz cuadrada de eigenvalores.

3.4 CALCULO DE LOS FACTORES. (F)

Análogamente al tipo - R, la matriz de Factores se define de la siguiente forma:

$$W \approx \hat{A}F'$$

Premultiplicando por \hat{A}' tendremos:

$$\hat{A}'W \approx \hat{A}'\hat{A}F'$$

pero como: $\hat{A}'\hat{A} = \Lambda$

entonces: $\hat{F}' = \Lambda^{-1} \hat{A}'W$

o bien: $\hat{F} = W'\hat{A}\Lambda^{-1}$

CAPITULO 4

ROTACION DE FACTORES

4.1 INTRODUCCION

Para el estudio de la rotación de factores se han tomado en cuenta dos criterios: uno cualitativo y otro cuantitativo. De éstos, únicamente consideraremos el segundo, el cual comprende el método Varimax.

El procedimiento varimax realiza una rotación ortogonal o rígida de los ejes factoriales no rotados. Como sucede con los ejes no rotados, los ejes Varimax son ortogonales una respecto del otro.

4.2 PROCEDIMIENTO VARIMAX

Kaiser (1958) desarrolló lo que es quizá el más popular de todos los procedimientos analíticos. Su método recurre a la simplificación de las columnas de la matriz factorial no rotada. Un factor simplificado es

aquel que tiene pocos pesos altos y muchos ceros, o valores cercanos a cero. Esta noción puede ser interpretada en relación a la varianza de los pesos factoriales, por lo tanto podemos decir que cuando los elementos de un factor dado se aproxima a unos y ceros, la varianza del factor será máxima.

Para evitar complicaciones debido a los signos de los factores pesados, utilizaremos el cuadrado de la varianza de los mismos. Por lo tanto el cuadrado de la varianza de los factores pesados para el j -ésimo factor está dada por:

$$S_j^2 = \frac{1}{p} \sum_{i=1}^p (b_{ij}^2)^2 - \frac{1}{p^2} \left(\sum_{i=1}^p b_{ij}^2 \right)^2 \quad (4.1)$$

Cuando la varianza es máxima, el factor obtenido es más simple. Para la matriz de factores pesados, debemos procurar que la varianza de cada factor individual, S_T^2 , sea máxima.

$$S_T^2 = \sum_{j=1}^k S_j^2 = \frac{1}{p} \sum_{j=1}^k \sum_{i=1}^p b_{ij}^4 - \frac{1}{p^2} \sum_{j=1}^k \left(\sum_{i=1}^p b_{ij}^2 \right)^2 \quad (4.2)$$

Por otro lado Kaiser (1958) sugirió que cualquier renglón de la matriz sea normalizada a una longitud unitaria antes de calcular la varianza. Simplemente si tomamos en cuenta que la suma de los elementos elevados al cuadrado de un renglón de la matriz de factores es igual a la comu

nalidad de la variable, la normalización puede llevarse a cabo dividiendo cualquier elemento de un renglón por la raíz cuadrada de la comunalidad asociada. Después de la rotación, los renglones son llevados a sus longitudes originales.

La cantidad final maximizada para producir una estructura simple a través del procedimiento varimax es:

$$S_V^2 = p^2 S_T^2 = p \sum_{j=1}^k \sum_{i=1}^p (b_{ij}/h_i)^4 - \sum_{j=1}^k \left(\sum_{i=1}^p b_{ij}^2/h_i^2 \right)^2 \quad (4.3)$$

de donde escribiremos, por razones de simplicidad, h_i^2 para la comunalidad.

Para cualquier par de factores, j y 1 , la cantidad maximizada es:

$$S^2(v_{j1}) = p \left[\sum_{i=1}^p (b_{ij}/h_i)^4 + \sum_{j=1}^p (b_{i1}/h_i)^4 \right] - \left[\sum_{i=1}^p (b_{ij}^2/h_i^2)^2 - \sum_{j=1}^p (b_{i1}^2/h_i^2)^2 \right] \quad (4.4)$$

Para maximizar (4.4) rotaremos los ejes factoriales j y 1 a través de algún ángulo θ_{j1} . Para encontrar este ángulo θ_{j1} que maximice (4.4) seguiremos los siguientes pasos:

Recordaremos primero que todas las rotaciones rectas de los ejes se hacen

siguiendo el proceso descrito en el Apéndice C.

$$\begin{aligned} b_{lj} &= a_{lj} \cos \theta_{jl} + a_{ll} \sin \theta_{jl} \\ b_{ll} &= a_{lj} \sin \theta_{jl} + a_{ll} \cos \theta_{jl} \end{aligned} \quad (4.5)$$

donde a_{lj} y a_{ll} es el peso de la variable l sobre los factores no rotados j y l ; y b_{lj} y b_{ll} son los pesos de los factores rotados θ_{jl} grados.

Para maximizar (4.4) se sustituyen las b 's y diferenciamos con respecto a θ_{jl} .

La determinación de θ_{jl} para cualquier par posible de factores j y l definen un ciclo. La matriz de pesos rotados después de un ciclo puede obtenerse de:

$$B = AT_{12} T_{13} \dots T_{jl} \dots$$

donde T_{jl} representa la matriz de transformación de la forma (4.5), derivada de la rotación de los factores j y l . Este proceso continúa hasta que S_v converge a un valor estable. La matriz de transformación final puede verse como un operador que transforma la matriz de factores no rotados, A , a una matriz de factores varimax, B .

CAPITULO 5

ANALISIS DE RESULTADOS

5.1 EXPLORACION GEOQUIMICA SUPERFICIAL

El objetivo fundamental de la exploración geoquímica es la detección de la presencia de tendencias químicas anómalas, o anomalías geoquímicas, relacionadas con trampas de hidrocarburos o con cuerpos mineralizados. En el caso particular de la exploración petrolera, la prospección geoquímica superficial abarca, entre otros aspectos, el análisis cromatográfico de núcleos extraídos de pozos someros perforados en un área de interés, siguiendo cierto patrón preestablecido. Los núcleos extraídos pueden provenir de diversas profundidades (3m, 15m, hasta 30m), o bien pueden estar localizados en zonas litológicamente distintas (suelos, lutitas, etc.). El área de interés pudo haber sido definida en base a campañas geológicas y/o geofísicas realizadas con anterioridad, las cuales per-

mitieron suponer la posible presencia de una trampa de hidrocarburos en el subsuelo. El patrón de perforación puede fijarse de acuerdo a las características morfológicas de la trampa (o trampas) interpretadas según experiencia en otras áreas geológicamente similares. Cuando el objetivo sea configurar la información recabada, o efectuar estimaciones estadísticas insesgadas, el patrón de perforación más eficiente será aquél en el cual las localizaciones de los pozos estén uniformemente distribuidas en el área de interés.

El análisis cromatográfico proporciona partes por millón (ppm) de distintos hidrocarburos presentes en los núcleos—metano, etano, etileno, propano, . . . , y gasolinas. Las partes por millón de un hidrocarburo, metano por ejemplo, se consideran integradas por dos componentes, una asociada a las ppm del metano generado localmente, o "background", y otra formada por las ppm del metano que ha migrado a través de cientos o miles de metros de roca, desde la trampa hasta el estrato superficial muestreado. Aún cuando la existencia de la difusión vertical de hidrocarburos no ha sido probada satisfactoriamente (Hitchon, 1974), la premisa fundamental en geoquímica sigue siendo que, dado que no existe roca totalmente impermeable, tal migración puede ocurrir en un cierto intervalo de tiempo geológico. Esta componente migratoria constituye precisamente la anomalía a detectar.

En general:

$$z_o = z_b + z_a$$

donde:

z_o , es la fracción observada de un hidrocarburo,

z_b , es la fracción del mismo hidrocarburo generado localmente y

z_a , es la fracción migrada o anómala del hidrocarburo.

Por otra parte, cuando se ha muestreado cromatográficamente cierto número de hidrocarburos, es deseable que en la detección objetiva de anomalías se hagan intervenir el máximo de hidrocarburos y no únicamente aquellos que muestren mayor contraste de valores. El problema consiste en interpretar simultáneamente n variables espaciales las cuales definen un espacio n -dimensional. El análisis factorial permite reducir tal espacio n -dimensional en uno de dimensiones mucho menor en el cual nuevas variables o factores adquieren otro sentido interpretativo pero directamente asociado al de las variables originales. En el nuevo espacio dimensional generado por los factores es posible inferir la presencia de anomalías; tal inferencia puede lograrse configurando los factores o bien a través de técnicas geoestadísticas como la descrita por Berlanga (1984).

5.2 DESCRIPCION DE LA INFORMACION GEOQUIMICA

El análisis factorial puede aplicarse directamente cuando la infor-

mación disponible está clasificada en atributos o variables y en localizaciones u objetos formando ésta una matriz. La información geoquímica empleada en el presente estudio consistió de 44 grupos de mediciones efectuadas en igual número de localizaciones. Dentro de cada grupo de mediciones se detectaron partes por millones de las ocho siguientes variables. (1) metano, (2) etano, (3) etileno, (4) propano, (5) propileno, (6) butano, (7) isobutano y (8) gasolina. Estas mediciones provenientes de núcleos a 15 metros de profundidad con mayoría en el porcentaje de lutitas constituy6 lo que en adelante denominaremos "Conjunto I de Datos". Simultáneamente, 184 grupos de mediciones obtenidas en las mismas ocho variables geoquímicas y extraídas a tres metros de profundidad constituyendo el "Conjunto II de Datos".

Las tablas 5.1 y 5.2 muestran un listado de tales conjuntos. La división en dos conjuntos atendió razones de carácter litológico y difusional.

5.3 DEFINICION DE ANOMALIAS GEOQUIMICAS

La técnica de análisis factorial fue aplicada en los dos conjuntos de datos descritos anteriormente. Con mayor claridad, en el conjunto II que en el conjunto I, puede observarse, según indican los "pesos" de mayor valor absoluto (Tabla 5.3 y 5.4) que el factor 1 se identifica con el etano, propano, butano, isobutano y gasolina; el segundo factor se asocia con el etileno y propileno, y el factor 3 con el metano. Tales asociaciones

sugieren los siguientes calificativos para los factores: (1) Factor de hidrocarburos ligeros, (2) olifeinas y (3) gas seco. Dichos factores de ninguna manera pudieron haber sido detectados o medidos durante la campaña geoquímica ya que ellos representan ahora un conjunto claramente interpretable de atributos. Los tres factores alcanzan a reproducir cerca del 89% de la variabilidad total mostrada por los ocho variables originales (Tabla 5.5), mientras que los factores deducidos del conjunto II simulan más del 97 por ciento de la variabilidad total (de las variables originales) (Tabla 5.6). En este sentido los factores del conjunto II captan mejor los rasgos estructurales que los del conjunto I. Valores de los factores en cada localización de pozo en los conjuntos de datos I y II son mostradas en las tablas 5.7 y 5.8, respectivamente. En la detección de anomalías geoquímicas mediante la configuración de factores, aquellas asociadas al conjunto I fueron interpretadas independientemente de los relacionados con el conjunto II. Para tal efecto las Figuras 5.1-A a 5.3-A muestran planos configurados de los factores 1, 2 y 3 pertenecientes al conjunto I. Las zonas sombreadas denotan "porciones" anómalas interpretables subjetivamente a partir de la observación directa de los valores de las variables originales que intervienen en la definición de un factor, esto es, mediante la ubicación de los máximos y mínimos de las variables originales y la posterior asociación de estos a los valores respectivos en los factores. En el caso del factor hidrocarburo ligero, la porción anómala se identificó con

los máximos valores (mayor que 1.0). La relación etileno/propileno ($C_2^= / C_3^=$) permitió relacionar zonas anómalas con los mínimos del factor olifeinas. Análogamente a la identificación de anomalías magnéticas en geofísica, el factor gas no mostró anomalías en zonas de máximos y mínimos, esto es, en porción fuera del rango de valores (-2.0, 2.0). Analizando conjuntamente los tres planos de factores pueden inferirse "zonas de anomalía de primer-orden, segundo-orden y tercer-orden", es decir zonas interpretadas como anómalas por uno, dos o tres factores, simultáneamente. Según esta clasificación, cuanto más grande sea el orden de la anomalía, mayor será su importancia. La anomalía de tercer orden que se ilustra en la Figura 5.4-A estaría respaldada por las ocho variables geoquímicas originalmente involucradas en este estudio, como se interpretó, ya que dichas variables fueron descritas por los tres factores.

Interpretaciones similares efectuadas en los factores deducidos del conjunto de datos II, aquel perteneciente a la información de los pozos a tres metros de profundidad, permiten deducir zonas de anomalías (Figuras 5.1-B a 5.3B) los cuales al integrarse fueron clasificados en anomalías de primer-orden, segundo-orden y tercer-orden, según se ilustra en la Figura 5.4B. Con el objeto de evitar los problemas de configuración inherentes a la mala distribución de los pozos en este último conjunto, los datos de cada factor fueron agrupados en ternas y promediados. Cada valor promedio se ubicó en la localización del dato central de la terna corres-

pendiente.

Comparando las anomalías de tercer-orden inferidas a partir de la información proveniente de pozos a 15 metros de profundidad con las anomalías deducidas de los datos de pozos a tres metros no se observa correlación local alguna, hecho que no se puede explicar, al menos con la información geológica disponible. No se ha estudiado, por ejemplo, la posible presencia de corrientes subterráneas de agua los cuales pudieran afectar las concentraciones de hidrocarburos en los estratos superficiales muestreados. Si tales corrientes no existieran en el área de estudio, la difusión vertical no varía, obviamente, el mecanismo generador de anomalías superficiales.

Sin embargo, lo que se observa al comparar las anomalías de las figuras 5.4-A y 5.4-B, es una clara tendencia lineal entre las anomalías de segundo y tercer-orden pertenecientes a ambos conjuntos. Tal tendencia coincide, según se ha confirmado, con la ubicación de una falla geológica, de aquí que la asociación de estas anomalías con la presencia de hidrocarburos en el subsuelo sea cuestionable. Este no fue sin embargo, el problema que se intentó resolver con este trabajo, sino como ya se mencionó, el objetivo fue la identificación de anomalías considerando la inclusión de todas las variables geoquímicas. Sin obtenerse ninguna mejora aparente en la ubicación de las anomalías, los factores fueron rotados según el criterio varimax y configurados.

TABLA 5.1
CONJUNTO I DE DATOS

Observación	Metano	Etano	Etileno	Propano	Propileno	Etileno/ Propileno	Butano	Isobutano	Gasolina
	1	2	3	4	5	6	7	8	9
1	943.0	98.7	50.3	37.1	25.9	1.94	5.6	8.1	20.2
2	1338.0	352.8	187.4	169.2	131.2	1.42	16.6	53.5	102.5
3	919.9	303.5	141.2	121.2	67.8	2.10	14.8	39.2	104.4
4	464.3	102.2	73.1	38.3	37.8	1.9	3.8	7.2	31.1
5	590.6	98.8	51.4	36.2	26.1	1.96	4.2	7.1	28.1
6	590.6	98.8	51.4	36.2	26.1	1.96	4.2	7.1	28.1
7	5742.0	1231.0	683.7	480.4	463.2	1.4	47.3	139.5	223.4
8	836.9	172.0	78.2	69.0	43.7	1.8	11.5	24.6	218.0
9	726.1	122.0	57.7	40.2	21.6	2.7	5.5	12.7	93.7
10	1012.0	219.5	95.2	95.0	54.4	1.7	9.5	22.2	101.2
11	863.2	118.4	54.0	31.4	21.8	2.5	3.2	4.1	48.9
12	1145.0	204.7	87.7	95.9	47.9	1.8	10.2	24.3	89.3
13	849.3	213.6	111.2	97.3	58.5	1.9	9.6	25.5	65.9

(continuación)

14	70.0	10.6	6.2	3.8	2.2	2.7	.3	.4	1.5
15	702.2	148.8	78.1	43.1	34.8	2.3	7.1	10.3	35.1
16	1203.0	164.3	90.0	72.4	49.7	1.8	8.1	22.4	110.9
17	1315.0	373.7	158.6	161.4	119.9	1.32	18.5	163.5	247.5
18	1704.0	223.4	86.0	108.9	60.8	1.43	15.6	42.6	189.8
19	1833.0	391.4	152.5	168.7	91.9	1.6	16.1	53.0	150.7
20	2215.0	318.9	94.6	150.1	61.8	1.5	26.3	56.0	209.4
21	366.8	108.1	62.3	53.0	46.0	1.3	4.7	18.8	60.3
22	625.6	82.7	43.1	38.7	33.8	1.3	5.0	15.8	72.4
23	453.1	114.4	50.5	45.0	35.8	1.4	4.5	11.8	31.7
24	709.1	63.4	34.3	24.8	23.7	1.4	4.7	9.3	51.3
25	736.0	63.0	27.4	26.1	15.4	1.8	4.2	9.0	37.6
26	1056.0	66.4	11.0	19.8	2.7	4.07	5.7	2.6	8.9
27	375.7	85.0	59.5	32.0	35.5	1.68	3.4	8.2	9.5
28	273.4	55.5	37.0	15.0	15.0	2.4	1.4	2.5	15.6
29	318.7	51.1	36.8	13.8	13.2	2.7	1.9	2.0	4.1
30	22220.0	3843.0	121.3	1007.0	10.9	11.1	540.3	100.0	15.6

(continuación)

31	69269.0	1701.0	123.6	161.5	8.9	13.8	53.1	315.2	387.1
32	13305.0	357.0	22.4	30.8	3.9	5.7	10.2	3.0	72.2
33	11683.0	223.2	41.1	35.7	3.7	11.1	6.3	1.3	11.5
34	18612.0	362.7	33.5	38.6	1.5	22.3	12.7	2.8	8.6
35	1761.0	558.1	239.0	233.1	150.7	1.6	23.7	76.0	215.3
36	517.5	223.6	124.2	115.4	104.3	1.2	11.4	50.0	204.1
37	1217.0	349.0	157.9	168.4	117.0	1.34	20.7	70.1	321.8
38	58765.0	341.2	49.0	14.6	6.4	7.6	2.8	30.0	20.0
39	8422.0	590.5	65.5	103.0	16.6	3.9	28.7	8.5	76.3
40	9779.0	270.3	97.8	112.0	59.8	1.6	15.3	43.2	222.0
41	371.6	96.8	56.6	43.9	34.8	1.6	5.0	12.4	89.3
42	1375.6	247.3	117.3	126.2	78.6	1.5	16.1	52.2	207.5
43	1107.0	183.3	80.2	83.5	51.4	1.55	9.8	28.4	133.5
44	20528.0	1896.0	72.3	879.2	25.2	2.8	258.8	315.2	387.1

TABLA 5.2
 CONJUNTO II DE DATOS

Observación	Metano	Etano	Etileno	Propano	Propileno	Etileno/ Propileno	Butano	Isobutano	Gasolina
	1	2	3	4	5	6	7	8	9
1	638.0	143.0	75.3	72.0	53.1	1.42	7.2	27.4	72.4
2	1183.0	238.3	123.3	118.9	88.9	1.39	11.7	46.8	130.0
3	506.9	122.4	78.6	59.7	55.3	1.42	5.6	22.2	60.7
4	871.1	149.3	75.7	72.8	51.6	1.47	6.4	26.1	61.4
5	551.1	112.3	69.1	58.1	52.2	1.32	6.0	23.7	116.2
6	521.2	140.0	82.8	69.8	59.1	1.4	6.4	27.2	78.9
7	643.4	104.8	60.6	47.3	49.4	1.23	4.9	18.3	62.9
8	268.5	57.8	49.0	28.8	38.4	1.28	3.0	11.0	36.3
9	679.4	119.6	62.2	54.9	42.0	1.48	5.2	18.8	47.8
10	451.5	83.0	50.1	39.7	37.6	1.33	4.0	15.2	53.0
11	1284.0	312.5	194.9	183.6	136.3	1.43	16.8	64.6	266.1
12	135.7	21.6	14.1	10.5	11.5	1.23	1.4	4.0	12.3

(continuación)

13	911.3	226.1	100.5	108.4	67.2	1.5	10.2	33.3	53.6
14	755.2	154.8	76.4	70.1	49.1	1.56	7.4	26.4	101.6
15	575.0	116.0	62.5	49.6	40.6	1.54	6.0	16.5	40.9
16	588.7	118.5	65.3	50.8	44.2	1.48	5.2	15.6	21.1
17	911.3	226.1	100.5	108.4	67.2	1.50	10.2	33.3	53.6
18	595.3	68.1	28.9	27.2	19.9	1.45	3.0	9.1	42.7
19	665.5	121.4	57.4	55.3	44.1	1.30	7.4	20.8	96.4
20	873.0	203.0	91.8	93.7	59.4	1.55	9.0	31.2	54.9
21	7445.0	1852.0	974.2	791.8	664.3	1.47	69.7	232.4	559.5
22	978.2	272.3	142.4	124.8	96.8	1.47	12.6	44.3	81.8
23	864.8	209.3	100.3	93.4	68.8	1.50	8.3	31.4	48.5
24	779.0	211.0	101.3	94.5	69.7	1.47	9.2	33.2	66.2
25	21950.0	1129.0	171.2	451.7	101.7	1.68	117.9	159.4	524.0
26	1352.0	296.5	144.1	121.5	85.5	1.19	11.0	29.6	80.6
27	685.0	145.3	71.4	64.9	48.4	1.48	6.6	21.5	46.2
28	4850.0	659.5	177.5	288.8	104.1	1.71	49.3	113.3	251.6
29	435.5	67.5	36.7	26.6	24.6	1.49	2.6	8.9	27.2

(continuación)

30	2540.0	195.6	70.6	80.2	47.8	0.88	14.8	27.0	130.9
31	547.4	129.9	81.2	61.2	62.4	1.30	6.5	24.3	73.4
32	5715.0	530.0	139.3	215.1	81.9	1.70	38.2	90.0	268.9
33	41063.0	6154.0	204.8	2558.0	19.3	10.61	758.4	697.6	2323.0
34	116054.0	5767.0	202.0	2443.0	33.0	6.12	724.6	848.8	2365.0
35	12338.0	5303.0	365.0	2288.0	161.7	2.26	670.3	959.3	2782.0
36	34235.0	2015.0	82.1	850.2	23.7	3.46	251.8	346.7	1116.0
37	182331.0	6400.0	254.8	1727.0	36.1	7.06	635.2	477.0	1137.0
38	75268.0	3754.0	248.6	1083.0	93.3	2.66	383.0	329.0	1511.0
39	163663.0	3658.0	235.8	1030.0	73.4	3.21	347.8	296.3	134.6
40	38285.0	3823.0	200.9	1282.0	41.7	4.82	434.7	278.8	671.7
41	107136.0	7169.0	135.5	3517.0	49.8	2.72	930.8	1280.0	4084.0
42	53765.0	3579.0	134.0	1728.0	57.9	2.31	426.5	577.0	2323.0
43	19599.0	1136.0	149.6	509.3	84.0	1.78	111.5	162.6	674.8
44	52653.0	2940.0	65.8	1243.0	24.6	2.67	345.8	396.9	1464.0
45	32638.0	1889.0	157.8	945.4	90.9	1.74	230.9	364.4	1069.0
46	3892.0	220.6	31.2	104.1	18.6	1.68	23.3	36.1	207.7

(continuación)

47	13341.0	592.8	70.1	248.6	39.9	1.76	67.5	80.9	339.6
48	560.9	122.8	65.7	61.3	39.9	1.65	5.5	21.8	78.8
49	300.3	23.1	8.6	10.0	6.0	1.43	1.9	3.2	14.1
50	1192.0	292.4	147.1	148.9	84.2	1.75	13.5	52.6	137.6
51	739.2	204.8	91.7	99.2	52.4	1.75	8.7	34.1	82.3
52	1017.0	111.8	57.4	49.6	33.2	1.73	4.9	17.8	65.7
53	870.3	102.5	49.6	42.9	34.1	1.45	5.9	17.4	84.6
54	1945.0	316.9	140.1	139.3	90.0	1.56	18.9	54.4	162.5
55	501.0	84.0	48.5	37.2	35.4	1.37	4.5	16.4	75.5
56	790.2	195.6	97.5	84.7	64.6	1.51	8.6	31.4	61.9
57	1148.0	227.3	102.8	101.6	68.7	1.50	14.2	42.9	133.8
58	1329.0	332.0	159.4	149.6	111.6	1.43	15.8	59.4	189.0
59	1225.0	359.9	175.9	166.2	126.3	1.39	16.8	67.1	234.7
60	865.3	158.3	62.1	67.1	45.6	1.36	9.3	26.7	89.2
61	998.8	286.2	145.6	126.0	104.5	1.39	14.4	49.9	167.5
62	918.2	207.4	93.3	93.4	53.8	1.59	9.2	26.9	122.0
63	66917.0	3130.0	106.8	1138.0	38.6	2.77	351.1	313.4	1157.0

(continuación)

64	855.0	160.7	73.9	74.8	47.0	1.57	800.9	25.4	153.6
65	134185.0	6898.0	141.3	3053.0	45.8	3.09	805.8	926.9	2932.0
66	4226.0	284.5	40.9	116.0	21.8	1.88	28.0	36.0	191.8
67	10857.0	948.8	150.7	418.8	78.7	1.91	103.7	133.2	473.2
68	116845.0	5452.0	97.2	2479.0	28.9	3.36	741.9	880.6	2129.0
69	37199.0	2045.0	65.0	955.2	31.7	1.75	249.8	301.0	1186.0
70	10115.0	546.4	48.5	242.9	57.6	0.84	3.2	77.0	240.0
71	20742.0	1019.0	79.7	456.0	45.4	1.76	112.8	141.3	510.2
72	21842.0	1180.0	35.2	495.1	17.4	2.02	164.3	199.2	589.2
73	25931.0	1504.0	123.6	602.3	70.2	1.76	147.8	175.0	673.3
74	19078.0	1155.0	78.3	515.6	38.0	2.06	128.4	161.4	586.0
75	14377.0	742.5	68.0	317.2	37.1	1.83	78.6	104.7	367.5
76	49241.0	2729.0	123.0	1260.0	66.8	1.84	334.0	429.5	1401.0
77	160643.0	8405.0	158.3	3853.0	55.7	2.84	1097.0	1341.0	4509.0
78	47526.0	2192.0	107.3	814.3	50.3	2.13	240.1	247.6	1041.0
79	14776.0	763.3	2.7	343.2	2.5	1.08	103.6	117.5	410.9
80	135555.0	3905.0	117.5	1570.0	45.2	2.6	392.9	470.0	1505.0

(continuación)

81	26898,0	2169,0	84,2	1056,0	37,7	8,83	253,4	329,4	881,6
82	98789,0	6285,0	114,0	2910,0	41,2	8,77	744,0	884,4	2739,0
83	72897,0	4838,0	126,2	2319,0	61,1	8,07	644,4	845,4	2901,0
84	13547,0	780,0	103,3	351,2	52,8	1,96	86,5	135,0	386,0
85	792,8	206,4	81,0	95,4	46,1	1,76	10,1	29,4	118,4
86	12927,0	611,8	112,7	269,1	69,9	1,61	56,1	84,8	364,5
87	13869,0	541,0	91,0	209,8	51,5	1,77	39,6	61,4	254,6
88	1257,0	250,5	104,5	123,4	71,1	1,47	16,3	46,1	185,6
89	10860,0	3277,0	1517,0	1711,0	956,2	1,59	163,2	687,0	2236,0
90	8374,0	2744,0	1598,0	1407,0	1148,0	1,39	130,8	599,6	1744,0
91	714,2	126,8	65,2	59,1	65,7	0,99	7,7	24,2	71,2
92	871,6	271,9	154,3	145,3	104,6	1,47	13,2	58,4	135,6
93	721,7	209,4	128,0	111,8	88,8	1,44	10,9	46,4	106,8
94	681,5	187,0	97,0	92,0	64,6	1,50	8,4	36,8	123,5
95	1263,0	81,0	40,8	38,6	28,2	1,45	4,8	15,7	71,8
96	571,0	109,2	66,2	51,7	47,4	1,40	4,7	18,7	40,0
97	705,7	60,3	38,7	29,0	29,3	1,32	3,1	13,7	57,0

(continuación)

98	357.0	69.1	44.9	33.3	34.7	1.29	4.0	13.3	36.6
99	2030.0	103.9	51.6	47.9	38.4	1.34	7.3	17.6	69.6
100	277.8	66.7	37.0	30.1	31.5	1.17	2.8	11.3	39.2
101	1152.0	425.6	223.4	211.0	163.1	1.37	20.1	81.2	134.5
102	596.9	123.1	63.9	64.0	48.7	1.31	6.8	24.0	47.6
103	977.6	122.9	58.0	59.1	44.1	1.32	8.8	23.8	91.1
104	1224.0	305.2	140.4	140.2	98.8	1.42	13.2	45.2	89.5
105	660.2	111.7	58.0	50.8	47.2	1.23	6.4	19.9	52.7
106	592.6	104.2	52.5	47.3	36.1	1.45	5.4	17.8	61.9
107	723.4	159.5	72.9	69.0	47.3	1.54	7.4	22.5	63.7
108	690.3	150.9	85.0	74.1	62.3	1.86	8.0	25.5	67.2
109	583.8	120.1	59.1	58.5	48.1	1.23	8.1	24.0	105.1
110	244.6	77.9	48.7	42.7	37.1	1.31	4.4	17.8	72.5
111	680.3	121.9	67.5	58.9	50.6	1.33	6.3	22.6	67.4
112	213.8	52.7	39.1	27.7	31.4	1.25	3.6	12.4	39.0
113	1035.0	207.2	121.2	93.5	96.8	1.25	11.7	37.0	95.2
114	2024.0	389.9	162.1	172.0	140.5	1.15	20.7	58.4	120.2

(continuación)

115	1193.0	346.4	141.8	171.0	99.0	1.43	16.4	54.3	73.5
116	1825.0	240.1	111.5	100.2	76.5	1.46	11.4	30.0	52.4
117	2152.0	292.2	106.6	100.7	77.0	1.38	15.2	33.1	60.5
118	3242.0	325.4	118.5	131.5	76.8	1.54	25.4	37.0	84.0
119	8242.0	618.9	116.0	240.6	71.8	1.62	55.7	80.3	173.9
120	6905.0	363.6	41.3	78.2	12.2	3.39	28.8	9.2	9.0
121	18015.0	2676.0	182.6	1402.0	104.0	1.76	595.4	485.2	934.0
122	103319.0	3957.0	191.7	1303.0	108.0	1.78	438.7	419.5	1462.0
123	16466.0	1365.0	79.9	536.0	44.4	1.80	162.1	165.0	610.0
124	93603.0	4742.0	173.8	1349.0	95.0	1.83	466.8	366.0	588.5
125	11529.0	399.0	66.5	106.6	47.6	1.40	22.7	32.8	243.7
126	22427.0	1939.0	112.3	839.7	68.9	1.63	227.0	285.3	887.8
127	46674.0	2317.0	94.0	728.6	46.8	2.01	249.0	214.8	447.5
128	65724.0	1530.0	128.7	353.3	81.6	1.58	105.3	97.0	544.4
129	8062.0	267.7	52.5	61.2	29.9	1.76	15.8	16.4	87.3
130	65638.0	4112.0	152.0	1343.0	33.6	4.52	522.8	356.4	1407.0
131	57787.0	2882.0	164.5	1031.0	59.0	2.79	362.2	306.0	845.6

(continuación)

132	6882.0	110.0	54.1	39.7	45.7	1.13	4.6	16.6	97.7
133	20751.0	1327.0	108.0	592.9	58.0	1.86	268.4	195.4	305.1
134	84250.0	10964.0	30.0	5778.0	39.4	0.76	1782.0	2241.0	2409.0
135	9755.0	600.0	137.0	327.0	115.3	1.19	86.3	141.0	417.7
136	891.8	132.6	79.4	79.0	64.5	1.23	12.3	31.8	110.3
137	277.4	70.0	61.0	44.7	55.6	1.10	6.2	20.5	70.0
138	791.0	203.9	102.0	103.7	72.0	1.42	10.5	33.1	25.7
139	525.1	106.1	55.5	52.5	41.0	1.35	5.9	18.5	44.0
140	1243.0	352.4	140.1	149.7	99.2	1.41	16.0	55.2	162.0
141	839.4	144.3	60.0	72.2	44.6	1.35	9.6	26.6	84.6
142	904.6	187.7	111.9	93.7	92.5	1.30	11.6	39.8	148.6
143	517.9	154.1	100.7	77.2	95.5	1.05	7.6	30.2	107.1
144	312.7	72.2	49.1	38.5	42.2	1.16	4.5	17.2	70.7
145	740.3	84.4	86.8	43.0	39.0	2.23	6.7	18.4	54.3
146	744.8	141.0	73.0	72.7	57.8	1.26	8.4	30.6	90.1
147	1049.0	126.0	74.1	58.8	54.4	1.36	6.7	23.7	76.5
148	4027.0	250.9	56.8	145.5	36.8	1.54	49.5	76.8	438.3

(continuación)

149	1336.0	416.7	140.5	166.2	126.0	1.12	16.3	67.8	338.0
150	268.5	52.9	31.7	26.0	25.4	1.25	2.8	11.7	39.2
151	617.4	100.7	63.2	47.6	46.6	1.36	4.4	19.9	53.3
152	571.8	157.3	104.9	77.0	76.2	1.38	7.8	31.3	86.0
153	914.7	233.4	113.1	106.7	77.5	1.46	11.4	41.4	78.5
154	445.6	108.5	59.6	52.3	46.0	1.30	5.6	22.0	69.0
155	1121.0	220.2	106.5	101.1	78.4	1.38	13.8	42.8	122.0
156	797.0	286.8	128.4	127.3	45.3	2.83	13.2	51.1	254.7
157	61468.0	1303.0	244.5	306.2	116.2	2.10	64.5	61.3	232.6
158	29743.0	1044.0	90.0	233.7	59.4	1.52	81.4	53.7	273.4
159	64788.0	4090.0	162.3	1657.0	94.3	1.72	469.3	536.1	1963.0
160	86850.0	1756.0	97.6	656.9	47.6	2.05	172.7	221.6	655.2
161	42196.0	1921.0	168.7	515.4	71.7	2.35	168.5	130.8	520.2
162	24185.0	1501.0	97.7	395.5	33.1	2.95	132.1	96.9	537.1
163	46551.0	3215.0	113.6	1469.0	29.6	3.84	433.6	542.5	564.4
164	2103.0	300.1	124.7	136.0	88.8	1.40	18.2	49.3	204.0
165	13524.0	738.0	170.4	258.0	100.5	1.70	38.1	82.8	306.7

(continuación)

166	112098.0	5013.0	238.1	2314.0	40.6	5.86	677.8	76.9	1344.0
167	2567.0	208.6	56.0	34.6	20.4	2.75	34.6	20.4	275.4
168	28755.0	1381.0	221.5	628.1	144.8	1.53	160.7	232.9	822.2
169	100145.0	7592.0	269.8	4081.0	91.0	2.96	1182.0	1607.0	2092.0
170	39796.0	1793.0	116.8	845.8	46.4	2.52	243.8	327.8	1062.0
171	28297.0	1459.0	3.7	617.8	4.6	0.80	183.6	211.7	527.5
172	90943.0	5313.0	16.5	2305.0	19.0	0.87	683.2	770.0	1359.0
173	53388.0	2927.0	161.2	1335.0	40.7	3.96	379.0	450.2	857.1
174	100917.0	6021.0	213.9	2690.0	29.7	7.20	779.2	898.6	1668.0
175	76197.0	3555.0	185.6	1601.0	35.6	5.21	467.2	545.3	1021.0
176	57014.0	3204.0	145.7	1462.0	35.9	4.06	418.9	478.3	1087.0
177	42002.0	1521.0	149.8	695.3	59.2	2.53	194.1	275.1	445.3
178	29437.0	2121.0	103.7	1000.0	28.7	3.61	284.6	347.6	802.0
179	65586.0	4216.0	167.1	2040.0	36.6	4.57	578.0	708.2	1353.0
180	67552.0	3291.0	171.8	1403.0	43.7	3.93	410.5	470.5	884.9
181	91783.0	4853.0	196.2	2147.0	35.0	5.59	662.5	779.6	1164.0
182	48374.0	3179.0	160.0	1469.0	40.9	3.91	416.6	498.3	995.6

(continuación)

183	174862.0	6126.0	374.7	2097.0	89.9	4.17	721.0	624.2	2464.0
184	82769.0	6227.0	190.8	2939.0	31.4	6.08	861.5	1048.0	2213.0

TABLA 5.3
MATRIZ DE PESOS "A" PARA EL CONJUNTO I

	FACTORES		
	Hidrocarburos Ligeros	Olefinas	Gas Seco
	1	2	3
Metano	.2676	-.5303	-.5456
Etano	.8410	-.4833	-.1599
Etileno	.6170	.6329	-.4491
Propano	.9489	-.2423	.1129
Propileno	.5163	.7760	-.3348
Butano	.7676	-.5573	.1036
Isobutano	.8550	.1422	.3357
Gasolina	.5921	.4953	.4336

TABLA 5.4
MATRIZ DE PESOS "A" PARA EL CONJUNTO II

	FACTORES		
	Hidrocarburos Ligeros	Olefinas	Gas Seco
	1	2	3
Metano	.8514	-.1826	-.4876
Etano	.9900	-.0716	-.0385
Etileno	.3228	.9375	-.0832
Propano	.9862	-.0548	.1163
Propileno	.1329	.9866	.0143
Butano	.9550	-.1778	.1025
Isobutano	.9604	-.0224	.2324
Gasolina	.9351	.0407	.0454

TABLA 5.5
EIGENVALORES PARA EL CONJUNTO I

	Factor	Eigenvalor	% de Variabilidad	% de Variabilidad Acumulada
Número de	1	3.9975	49.9683	49.9683
Factores	2	2.1525	26.9057	76.8740
Considerados	3	.9613	12.0160	88.8900
	4	.7089	8.8613	97.7512
	5	.1520	1.8997	99.6509
	6	.0143	.1790	99.8300
	7	.0101	.1267	99.9566
	8	.0035	.0434	100.0000

TABLA 5.6
EIGENVALORES PARA EL CONJUNTO II

	Factor	Eigenvalor	% de Variabilidad	% de Variabilidad Acumulada
Número de	1	5.5084	68.8548	68.8548
Factores	2	1.9276	24.0951	92.9498
Considerados	3	.3265	4.0814	97.0312
	4	.1507	1.8834	98.9147
	5	.0447	.5583	99.4730
	6	.0223	.2790	99.7520
	7	.0133	.1656	99.9177
	8	.0066	.0823	100.0000

TABLA 5.7
MATRIZ F QUE SE CONFIGURA PARA EL CONJUNTO I

	Coordenadas		Hidrocarburos Ligeros	Olefinas	Gas Seco
1	653.	6474.	-.6213	.1830	-.0315
2	960.	6070.	.3391	.7827	-.4628
3	1286.	5672.	.0270	.3825	-.0330
4	1591.	5280.	-.5560	-.0217	-.1278
5	1892.	4884.	-.6164	-.1499	.0063
6	4294.	4058.	-.6164	-.1499	.0063
7	2827.	3666.	2.9301	3.5919	-3.3525
8	3463.	2868.	-.0984	.4308	.8240
9	3787.	3472.	-.4742	-.0028	.3424
10	4412.	1709.	-.2051	.2146	.1129
11	294.	4500.	-.5992	-.1220	.0762
12	771.	4370.	-.2400	.1345	.1335
13	1248.	4229.	-.2167	.2010	-.1187
14	1747.	4102.	-.8725	-.4006	.1743
15	3204.	3691.	-.5067	-.0412	-.1120
16	3734.	3569.	-.2518	.2329	.1979
17	4707.	3288.	.9288	1.1232	1.1022

(continuación)

18	5681.	3013.	.0452	.4221	.6648
19	6161.	2866.	.3046	.5869	.0807
20	1779.	7956.	.2477	.4192	.7920
21	1843.	7477.	-.4461	.0621	.1076
22	1901.	6978.	-.5121	-.0155	.2800
23	1962.	6481.	-.5614	-.0989	.0181
24	2012.	5965.	-.6244	-.1360	.2186
25	2969.	5468.	-.6718	-.2302	.2208
26	2118.	4968.	-.7903	-.4253	.1618
27	2177.	4481.	-.6250	-.1115	-.1491
28	2255.	3950.	-.7418	-.2415	.0433
29	2291.	3440.	-.7669	-.2796	-.0065
30	2429.	2449.	3.4267	-3.7303	-.3062
31	2475.	1968.	.5631	-1.9897	-3.2781
32	2524.	1481.	-.4954	-.5651	-.1468
33	89.	3162.	-.6232	-.5809	-.4356
34	443.	3556.	-.5422	-.8087	-.7019
35	1401.	4680.	.8747	1.1952	-.1285
36	2743.	6219.	.2251	.8197	.4495
37	3077.	6599.	.6862	1.1781	.9080
38	2217.	1405.	-.2529	-1.3370	-2.2199
39	3119.	1860.	-.2030	-.4518	-.1439

(continuación)

40	3705.	2193.	.1709	.3716	.4368
41	4791.	2781.	-.4670	.0622	.2838
42	5240.	3009.	.2161	.6404	.6016
43	5375.	8084.	-.1822	.2599	.3883
44	5783.	3308.	3.3949	-1.0365	3.1229

TABLA 5.8
MATRIZ DE FACTORES F PARA EL CONJUNTO II

	Hidrocarburos Ligeros	Olefina	Gas Seco
	1	2	3
1	-.6432	-.0890	.1908
2	-.5743	.1971	.1834
3	-.6526	-.0686	.1730
4	-.6452	-.0951	.1749
5	-.6446	-.1074	.1974
6	-.6398	-.0408	.1856
7	-.6654	-.1440	.1791
8	-.6916	-.2222	.1791
9	-.6662	-.1732	.1735
10	-.6799	-.2239	.1861
11	-.4789	.6030	.1821
12	-.7262	-.4361	.1987
13	-.6152	.0383	.1710
14	-.6366	-.1032	.1849
15	-.6704	-.1784	.1693
16	-.6734	-.1551	.1600

(continuación)

17	-.6152	.0389	.1710
18	-.7000	-.3603	.1847
19	-.6544	-.1769	.1972
20	-.6264	-.0188	.1722
21	.3560	5.0210	-.3704
22	-.5740	.2834	.1537
23	-.6232	.0374	.1609
24	-.6170	.0532	.1722
25	-.1182	.2812	-.1662
26	-.5819	.2383	.0957
27	-.6556	-.1207	.1718
28	-.3827	.3818	.2200
29	-.7009	-.3177	.1774
30	-.6150	-.1322	.1407
31	-.6448	-.0305	.1783
32	-.4383	.1851	.1527
33	1.9887	-.3816	2.2015
34	2.1937	-.4873	-.8294
35	1.9064	.8245	3.1849
36	.3515	-.4041	.2316
37	1.8736	-.4668	-4.8969
38	.9413	.1952	-1.3939

(continuación)

39	.9312	-.1637	-5.1456
40	.6318	-.1197	-.0488
41	3.2011	-.6744	1.4315
42	1.2544	-.2464	.5819
43	-.0966	.1530	-.0028
44	.7380	-.5434	-.0757
45	.9784	.1108	.2587
46	-.5954	-.3761	.1862
47	-.3993	-.2209	-.0199
48	-.6560	-.1720	.1867
49	-.7282	-.4762	.1972
50	-.5507	.2404	.1711
51	-.6200	-.0490	.1877
52	-.6670	-.2251	.1632
53	-.6676	-.2420	.1821
54	-.5389	.2432	.1638
55	-.6753	-.2377	.1919
56	-.6248	.0209	.1671
57	-.5896	.0514	.1960
58	-.5206	.3944	.1826
59	-.4922	.5049	.1989
60	-.6433	-.1597	.1995

(continuación)

61	-.5473	.9273	.1802
62	-.6144	-.0375	.1683
63	.7005	-.4093	-.9610
64	-.1308	-.3817	1.0881
65	2.6811	-.6990	-.8988
66	-.5827	-.3393	.1636
67	-.2301	.1585	.1965
68	2.1172	-.8084	-.6279
69	.3716	-.4236	.0947
70	-.4845	-.1758	.0441
71	-.1902	-.2149	-.0470
72	-.1046	-.4860	.1701
73	-.0033	-.0179	-.1306
74	-.1384	-.2554	.1064
75	-.3442	-.2503	.0298
76	.7337	-.1855	.0535
77	3.8251	-.7942	-.2003
78	.3504	-.2460	-.5735
79	-.3331	-.5944	.1698
80	1.3158	-.5554	-3.0363
81	.3213	-.3342	.5093
82	2.3454	-.6767	.2823

(continuación)

83	1.9590	-.4085	.9516
84	-.2973	-.0857	.1056
85	-.6192	-.1059	.1950
86	-.3744	.0341	-.0481
87	-.4460	-.1031	-.1676
88	-.5677	.0659	.2159
89	1.5814	7.7537	.4841
90	1.3595	8.8484	.1319
91	-.6493	-.0614	.1966
92	-.5454	.3513	.1923
93	-.5842	.2114	.1873
94	-.6088	.0212	.1978
95	-.6774	-.2929	.1700
96	-.6678	-.1378	.1710
97	-.6891	-.2916	.1835
98	-.6896	-.2509	.1876
99	-.6626	-.2214	.1375
100	-.6954	-.2861	.1946
101	-.4661	.7925	.1768
102	-.6588	-.1398	.1932
103	-.6507	-.1768	.1930
104	-.5648	.2849	.1560

(continuación)

105	-.6658	-.1620	.1858
106	-.6713	-.2251	.1872
107	-.6483	-.1217	.1746
108	-.6383	-.0222	.1742
109	-.6488	-.1545	.2100
110	-.6753	-.2292	.2070
111	-.6548	-.1207	.1833
112	-.6956	-.2805	.1971
113	-.5944	.2271	.1595
114	-.5137	.5220	.1595
115	-.5513	.2871	.1866
116	-.6076	.1073	.1132
117	-.6027	.0940	.1212
118	-.5666	.1182	.0926
119	-.4400	.0591	.0725
120	-.6273	-.3913	-.0526
121	.7533	.1158	1.5730
122	1.2090	.0038	-2.0247
123	-.0961	-.2316	.2528
124	1.0033	-.1309	-1.9283
125	-.5253	-.1708	-.1588
126	.2180	-.0864	.4557

(continuación)

127	.1931	-.3112	-.6907
128	.0167	-.0298	-2.0018
129	-.6145	-.2767	-.0930
130	1.0202	-.3781	-.6254
131	.5813	-.1401	-.7725
132	-.6382	-.1924	-.0536
133	-.0419	-.1494	.2095
134	4.5376	-1.4599	5.8912
135	-.2988	.2970	.2450
136	-.6243	-.0289	.2047
137	-.6655	-.1144	.2034
138	-.6230	.0636	.1697
139	-.6715	-.1959	.1881
140	-.5361	.2852	.1946
141	-.6454	-.1700	.2053
142	-.5877	.1843	.1931
143	-.6151	.1688	.1882
144	-.6759	-.2057	.2029
145	-.6617	-.1171	.1334
146	-.6363	-.0751	.2036
147	-.6472	-.0865	.1651
148	-.4842	-.2308	.3240

(continuación)

149	-.4765	.4063	.2621
150	-.7003	-.3273	.2005
151	-.6670	-.1491	.1768
152	-.6211	.0949	.1714
153	-.5981	.1186	.1787
154	-.6622	-.1617	.1986
155	-.5900	.1044	.1928
156	-.5473	.0229	.2078
157	-.0948	.4681	-2.1788
158	-.3065	-.1329	-.7592
159	1.2743	-.0638	-.0309
160	.2950	-.3549	-2.3072
161	.0532	.0608	-.9790
162	-.1481	-.2381	-.3325
163	.7396	-.4434	.4227
164	-.5401	.1964	.1749
165	-.3620	.3308	-.1904
166	1.4376	-.3097	-2.6573
167	-.5935	-.2947	.1655
168	.1214	.5754	-.1694
169	3.2802	-.2746	2.3642
170	.3428	-.2149	-.0729

(continuación)

171	-.0380	-.6634	.0332
172	1.6779	-1.0033	-.0219
173	.7125	-.2397	-.1532
174	2.1191	-.4824	-.1370
175	1.0590	-.2975	-.6633
176	.8592	-.3300	-.0914
177	.1205	-.0604	-.5235
178	.3331	-.3360	.4221
179	1.3826	-.3784	.4440
180	.8550	-.2506	-.6214
181	1.6240	-.4181	-.3821
182	.8196	-.2492	.2508
183	2.3600	.1177	-3.9534
184	2.3545	-.5249	1.1951

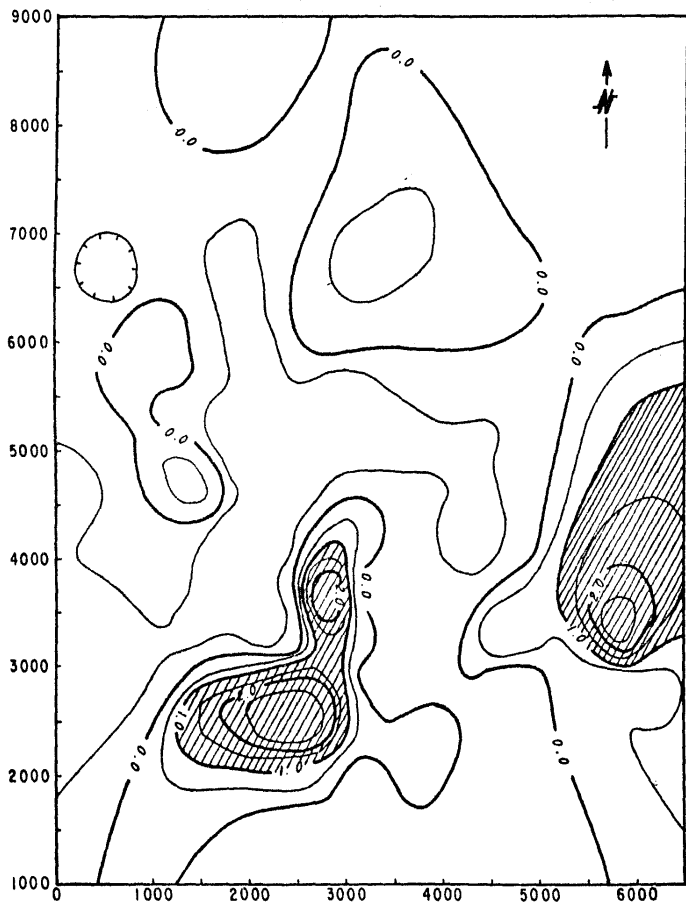


FIGURA 5.1-A Configuración del Factor Hidrocarburos Ligeros para el Conjunto 1 de Datos.

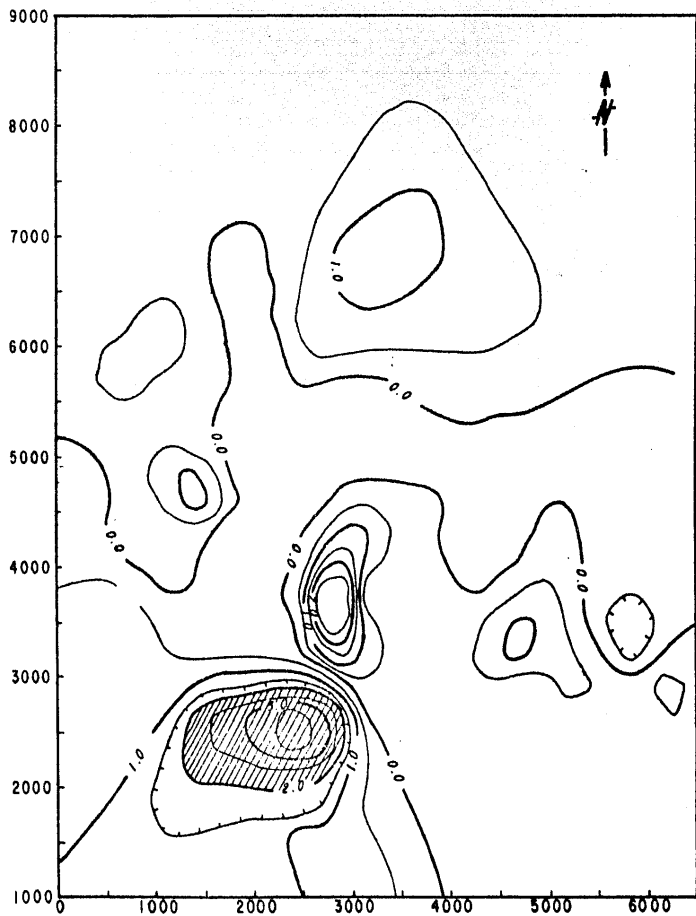


FIGURA 5.2-A Configuración del Factor Otefeñas para el Conjunto I de Datos.

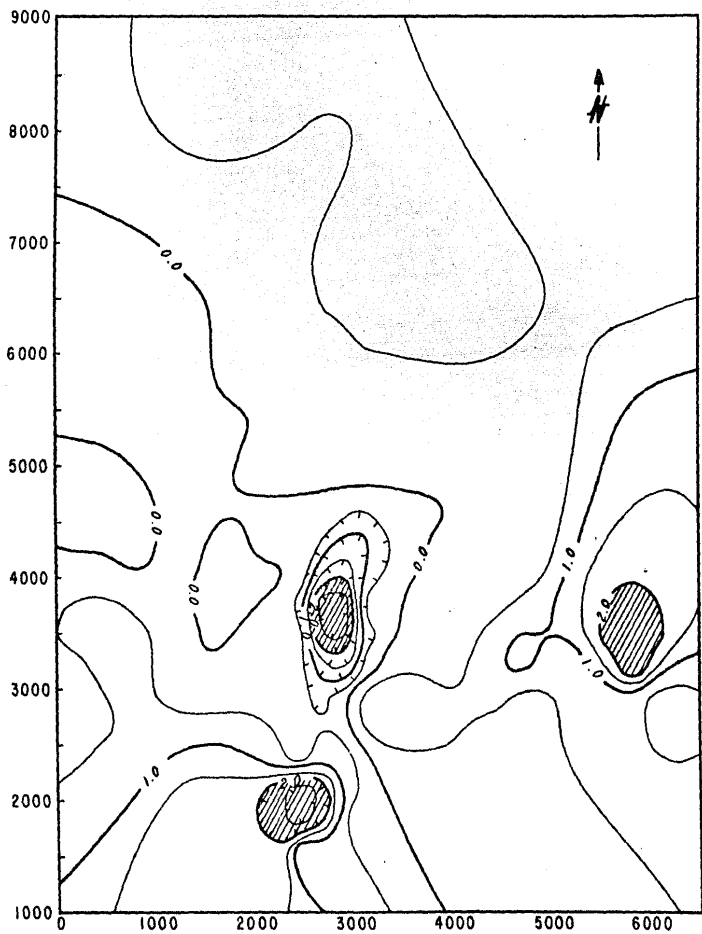


FIGURA 5.3-A Configuración del Factor Gas Seco para el Conjunto I de Datos.

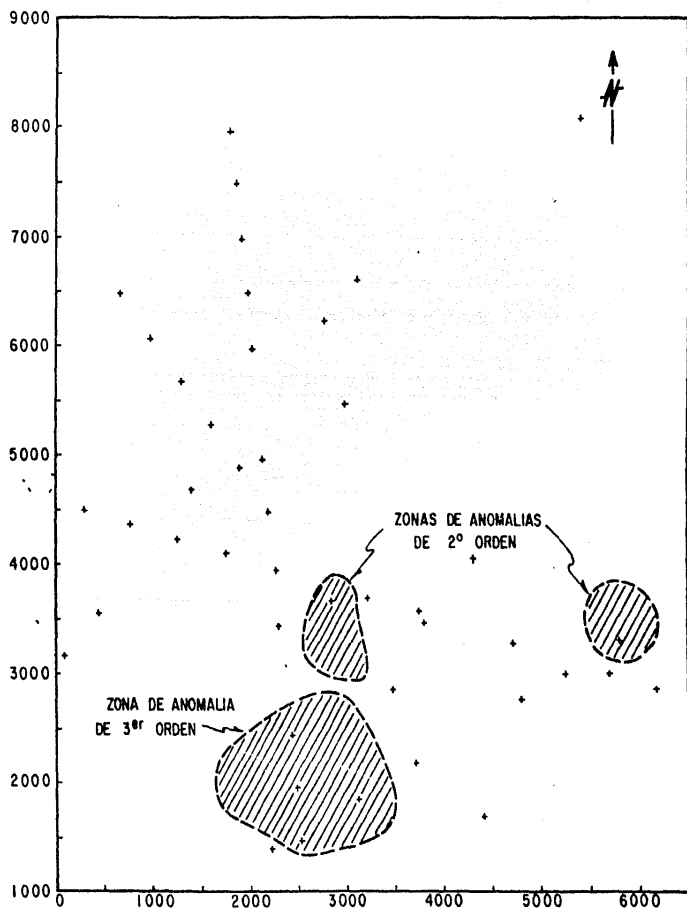


FIGURA 5. 4-A Localización de las Anomalías para el Conjunto I de Datos.

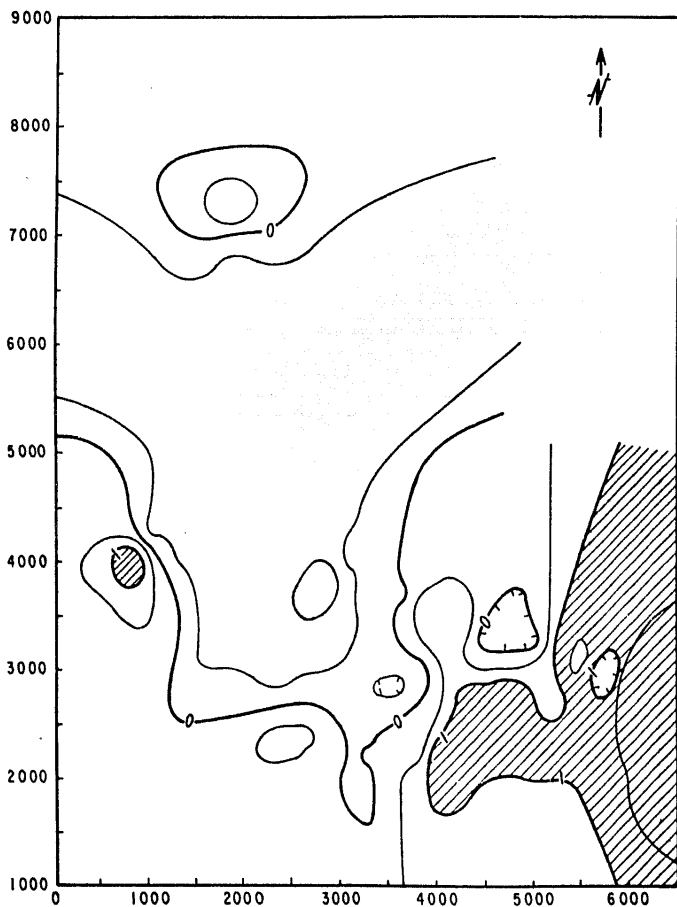


FIGURA 5.1.-B Configuración del Factor Hidrocarburos Ligeros para el Conjunto II de Datos,

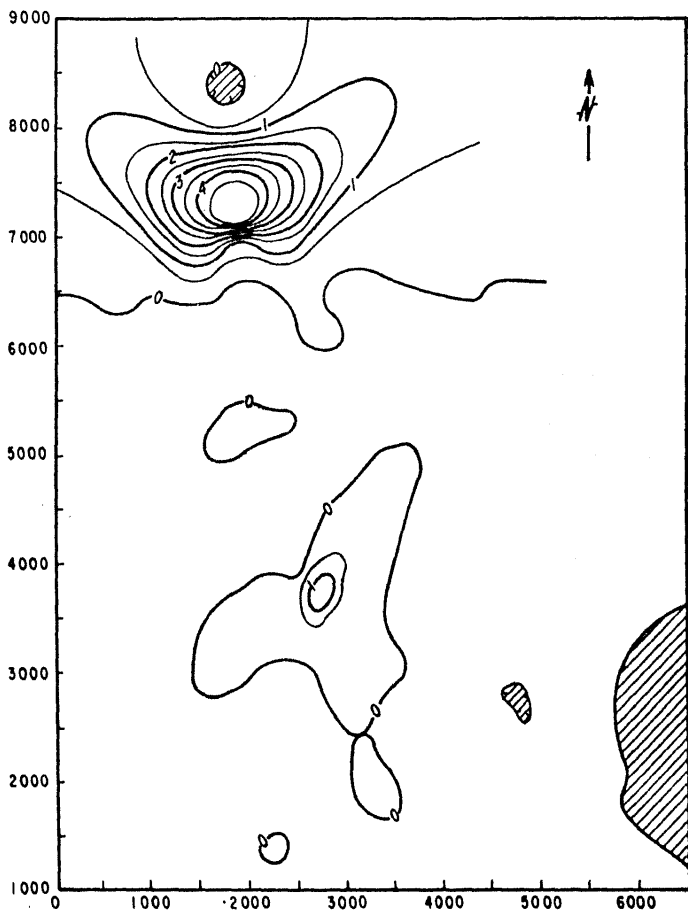


FIGURA 5.2-B Configuración del Factor Olefñas para el Conjunto II de Datos.

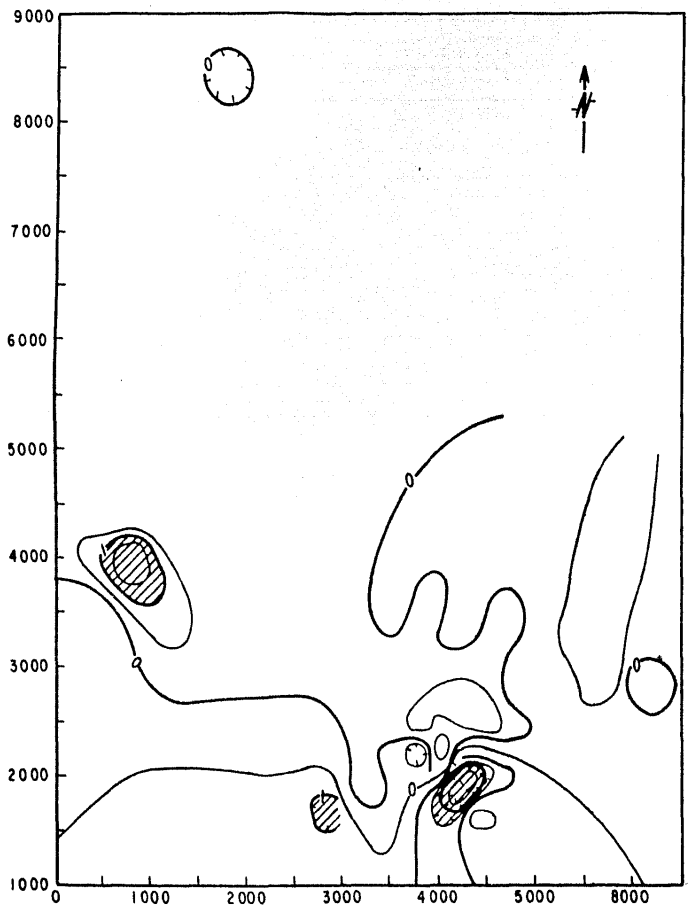


FIGURA 5.3-B Configuración del Factor Gas Seco para el Conjunto II de Datos.

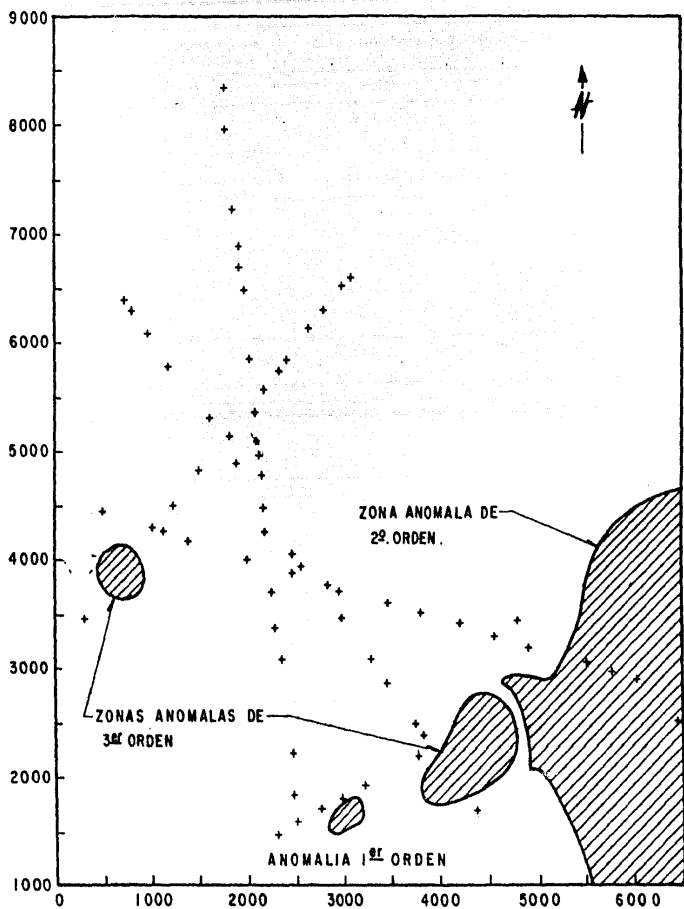


FIGURA 5.4-B Localización de las Anomalías para el Conjunto II de Datos.

CONCLUSIONES

El análisis factorial es una rama de la estadística multivariable que proporciona una opción más en el análisis de datos.

Con este método se integra de una forma más sencilla toda la información disponible en términos de un número menor de variables o factores fundamentales. Tal afirmación queda confirmada al analizar el conjunto I y el conjunto II de datos presentados en el capítulo anterior, en los que después de tener ocho variables y 184 observaciones respectivamente, estos se redujeron al aplicar el método del análisis factorial, únicamente a tres variables o factores fundamentales con el mismo número de observaciones.

Por otra parte la eficiencia de esta técnica se basa en que sus formulaciones matemáticas pueden programarse en computadoras, además de que existen paquetes estadísticos como el SPSS (STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES) instalados en los diferentes sistemas de cómputo que facilitan el acceso a muchos profesionales en ramas

no afines con la ingeniería.

Finalmente uno de los inconvenientes que posiblemente limitarían el uso de esta técnica, es una cantidad excesiva de información, que aumentaría el tiempo de cómputo e indiscutiblemente los costos.

AGRADECIMIENTO

Al Dr. Juan M. Berlanga G. ya que sin su interés ni su acertada guía, hubiera resultado imposible la realización de este trabajo.

Al Ing. Raimundo Martínez A. por su contribución en la configuración de mapas que son complemento fundamental para esta tesis.

Al Instituto Mexicano del Petróleo por ser el centro donde se procesó toda la información.

A la Sra. Elia Yolanda Méndez Galindo por la forma tan correcta en que mecanografió este trabajo.

BIBLIOGRAFIA

- Davis, J.C. , 1970, *Statistics and Data Analysis in Geology*: Jhon Wiley & Sons, Inc. , New York, 550 p.
- Harman, H.H. , 1967, *Modern Factor Analysis*, 2nd: Univ. Chicago Press, Chicago, 474 p.
- Hitchon, B. , 1974, *Application of Geochemistry to the Search for Crude Oil and Natural Gas*. A contribution on: *Introduction to Exploration Geochemistry* by A.A. Levinson, Aplied Publishing Ltd. , 509 - 545.
- Horvitz, L. , 1974, *Geochemical Techniques for Petroleum Exploration*: UN ESCAP, CCOP/SOPAC Tech. Bull. 3: 261 - 271.
- Horvitz, L. , 1972, *Vegetation and Geochemical Prospecting for Petroleum*: *The American of Petroleum Geologist Bulletin*, Vol. 56, No. 5, May, 1972, pags. 924 - 940.

Klovan, J.E., Geological Factor Analysis. Elsevier Scientific Publishing Company, New York, 178 p.

McCammon, R., 1974, The statistical treatment of Geochemical Data. A Contribution on: Introduction to Exploration Geochemistry by A.A. Levinson, Applied Publishing Ltd., 469 - 508.

APENDICE A

PROGRAMA DE COMPUTACION UTILIZADO PARA PROCESAR LA INFORMACION

FACTOR ANALYSIS

C THE PROGRAM ACCEPTS AN N BY M DATA MATRIX WHERE N IS THE
 C NUMBER OF OBSERVATIONS AND M IS THE NUMBER OF VARIABLE. IF
 C THE FIRST OPTION IS 1, AN M BY M SIMILARITY MATRIX BETWEEN
 C COLUMNS WILL BE COMPUTED. IF THIS OPTION IS 2, AN N BY N
 C SIMILARITY MATRIX BETWEEN ROWS WILL BE COMPUTED.
 C IF THE OPTION IS 0, THE PROGRAM CALLS EXIT, AS THE PROGRAM
 C LOOPS BACK AND RESTARTS AFTER COMPLETION OF AN ANALYSIS.
 C IF THE SECOND OPTION IS 1, A STANDARDIZED COVARIANCE
 C (CORRELATION) MATRIX IS CREATED. IF THIS OPTION IS 2, A
 C COSINE THETA SIMILARITY MATRIX IS CREATED. THE THIR OPTION
 C SPECIFIES THE NUMBER OF FACTORS TO BE RETAINED.

C TO PERFORM R-MODE ANALYSIS, SELECT 1 ON OPTION ONE AND 1 ON
 C OPTION TWO. TO PERFORM Q-MODE FACTOR ANALYSIS, SELECT 2
 C ON OPTION ONE AND 2 OPTION TWO.

FORMAT OF CONTROL CARD

C COL 1-3 0 = END OF JOB
 C 1 = DO NOT TRANSPOSE DATA MATRIX
 C 2 = TRANSPOSE DATA MATRIX
 C COL 4-6 1 = CALCULATE CORRELATION MATRIX
 C 2 = CALCULATE COSINE THETA MSTRIX

```

C
C      COL 7-9 NUMBER OF FACTORS DESIRED
C
C
C      =====
C
C      DIMENSION X(184,184), FSCORE(184,184)
C      DIMENSION A1(184,184), A2(184,184), A3(184,184)
C      MD=184
C      ND=184
C      MM=184
C
C...  READ CONTROL CARD
C
C      READ (5,1000) ITRAN,ISIM,L
C      IF (ITRAN.LE.0) CALL EXIT
C
C...  READ AND PRINT INPUT DATA MATRIX
C
C      CALL READM (X,N,M,ND,MD)
C      CALL PRINTM (X,N,M,ND,MD)
C      WRITE (10,2001)
C
C...  IF CORRELATION MATRIX IS TO BE CALCULATED
C      STANDARDIZE INPUT DATA MATRIX AND PRINT STANDARDIZED DATA
C      IF (ISIM .NE. 1) GO TO 2
C      CALL STAND (X,N,M,ND,MD)
C      CALL PRINTM(X,N,M,ND,MD)
C      WRITE(10,2008)
C
C...  TRANSPOSE DATA MATRIX (IF REQUIRED)
C
C      2 IF (ITRAN .NE. 2) GO TO 3
C      MT=M
C      IF (N .GT. M ) MT=N
C      DO 110 I=1,MT
C      DO 110 J=I,MT
C      XS=X(I,J)
C      X(I,J)=X(J,I)
C      X(J,I)=XS
C 110 CONTINUE
C      KT=M
C      M=N
C      N=MT
C
C...  CALCULATE AND PRINT SIMILIRITY MATRIX
C
C      3 IF (ISIM .EQ. 1) CALL RCOEF(X,N,M,ND,MD,A1,MM)
C      IF (ISIM .EQ. 2) CALL CTHETA(X,N,M,ND,MD,A1,MM)
C      CALL PRINTM(A1,M,M,MM,MM)

```



```

      WRITE(10,2002)
C
C... SAVE CORRELATION MATRIX
C
      IF (ISIM .NE. 1) GO TO 4
      DO 111 I=1,M
      DO 111 J=1,M
      A3(I,J)=A1(I,J)
      111 CONTINUE
C
C... CALCULATE EIGENVALUES AND EIGENVECTORS
C
      4 CALL EIGENJ(A1,A2,M,MM)
C... MOVE EIGENVALUES TO FIRST COLUMN
C... CALCULATE SUM OF EIGENVALUES
C
      SUME=0.0
      DO 100 I=1,M
      A1(I,1)=A1(I,I)
      SUME=SUME+A1(I,1)
      100 CONTINUE
C
C... CLACULATE PORCENT CONTRIBUTION OF EACH EIGENVALUE
C
      SUMEE=0.0
      DO 101 I=1,M
      2010 FORMAT (1H0,4X,'REPRODUCED CORRELATION MATRIX')
      2011 FORMAT (1H0,4X,'RESIDUAL CORRELATION MATRIX')
      END
C
*FTN,F50 ,TPF$.02
C SUBROUTINE TO READ A MATRIX
C HAVING N ROWS AND M COLUMNS
C
C SUBROUTINE READM(A,N,M,N1,M1)
C DIMENSION A(184,184)
C READ SIZE OF MATRIX
C READ (5,1000) N,M
C READ MATRIX ONE ROW AT A TIME
C DO 100 I=1,N
C READ(5,1001) (A(I,J),J=1,M)
C 100 CONTINUE
C RETURN
C 1000 FORMAT (2I3)
C 1001 FORMAT (22X,7F8.0/F8.0)
C END
*FTN,IS ,TPF$.03
C SUBROUTINE PRINTM(A,N,M,N1,M1)
C DIMENSION A(184,184)
C PRINT MATRIX OUT IN STRIPS OF 10 COLUMNS
C DO 100 IB=1,M,10

```

```

:      IE=IB+9
:      IF (IE-M) 2,2,1
:      1 IE=M
:C     PRINT HEADINGS
:      2 WRITE (10,2000) (I,I=IB,IE)
:      DO 101 J=1,N
C     PRINT ROW OF MATRIX
      WRITE (10,2001) J,(A(J,K),K=IE,IE)
101  CONTINUE
100  CONTINUE
      RETURN
2000 FORMAT (1H1,1X,10I12)
2001 FORMAT (1H0,15,10F12.4)
      END
@FTN,FSO ,TPF$.04
C
C     SUBROUTINE TO SUBSTRACT TWO MATRICES
C     B FROM A TO FORM C. ALL HAVE N ROWS AND M COLUMNS.
C
      SUBROUTINE SUBM(A,B,C,N,M,N1,M1)
      DIMENSION A(184,184),B(184,184),C(184,184)
      DO 100 I=1,N
      DO 101 J=1,M
      C(I,J)=A(I,J)-B(I,J)
101  CONTINUE
100  CONTINUE
      A1(I,2)=A1(I,1)*100.0/SUMEE
      SUMEE=SUMEE+A1(I,1)
      A1(I,3)=SUMEE*100.0/SUMEE
101  CONTINUE
C
C... PRINT EIGENVALUES AND PORCENT CONTRIBUTION
C
      CALL PRINTM(A1,M,3,MM,MM)
      WRITE(10,2003)
C... PRINT EIGENVECTORS
C     NOTE... EIGENVECTORS ARE STORED COLUMNWISE
C
      CALL PRINTM(A2,M,M,MM,MM)
      WRITE(10,2004)
C
C... CALCULATE AND PRINT FACTOR LOADINGS
C
      DO 102 I=1,M
      DO 102 J=1,L
      A2(I,J)=A2(I,J)*SQRT(A1(J,1))
102  CONTINUE
      CALL PRINTM(A2,M,M,MM,MM)
      WRITE(10,2005)
C

```

C... CALCULATE AND PRINT REPRODUCED CORRELATION MATRIX AND
 C RESIDUAL CORRELATION MATRIX

C
 IF (ISIM .NE. 1) GO TO 5
 DO 112 I=1,M
 DO 112 J=1,M
 DET=0.0
 DO 113 K=1,L
 DET=DET+A2(I,K)*A2(J,K)
 113 CONTINUE
 A1(I,J)=DET
 A1(J,I)=DET
 112 CONTINUE
 CALL PRINTM(A1,M,M,MM,MM)
 WRITE(10,2010)
 CALL SUBM(A3,A1,A3,M,M,MM,MM)
 CALL PRINTM(A3,M,M,MM,MM)
 WRITE(10,2011)

C
 C... CALCULATE AND PRINT FACTOR SCORES

C
 5 DO 103 I=1,L
 DO 103 J=1,L
 DET=0.0
 DO 104 K=1,M
 DET=DET+A2(K,I)*A2(K,J)
 104 CONTINUE
 A3(I,J)=DET
 A3(J,I)=DET
 103 CONTINUE
 CALL MINV(A3,A1,L,MM,DET)
 CALL MMULT(A2,A1,A3,M,L,L,MM,MM,MM,MM,MM,MM,MM,MM)
 CALL MMULT(X,A3,FScore,N,M,L,ND,MD,MM,MM,ND,MD)
 CALL PRINTM(FSCORE,N,L,ND,MD)
 WRITE(10,2007)

C
 C... ROTATE FACTOR MATRIX

C
 CALL VARMAX(A2,M,L,MM)

C
 C... PRINT ROTATE FACOR MATRIX

C
 CALL PRINTM(A2,M,L,MM,MM)
 WRITE(10,2006)

C... CALCULATE AND PRINT VARIMAX FACTOR SCORE

C
 DO 105 I=1,L
 DO 105 J=1,L
 DET=0.0
 DO 106 K=1,M
 DET=DET+A2(K,I)*A2(K,J)

```

106 CONTINUE
A3(I,J)=DET
A3(J,I)=DET
105 CONTINUE
CALL MINV(A3,A1,L,MM,DET)
CALL MMULT(A2,A1,A3,M,L,L,MM,MM,MM,MM,MM,MM)
CALL MMULT(X,A3,FSCORE,N,N,L,ND,MD,MM,MM,ND,MD)
CALL PRINTM(FSCORE,N,L,ND,MD)
WRITE (10,2009)
1000 FORMAT (3I3)
2001 FORMAT (1H0,4X,'INPUT DATA MATRIX -',1X,
1'COLUMNS = VARIABLES,ROWS = OBSERVATIONS')
2002 FORMAT (1H0,4X,'SIMILARITY MATRIX')
2003 FORMAT (1H0,4X,'COLUMN 1 = EIGENVALUES,',2X,
1'COLUMN 2 = PERCENT OF TRACE',/,
2 5X,'COLUMN 3 = CUMULATIVE PERCENT OF TRACE')
2004 FORMAT (1H0,4X,'PRINCIPAL AXIS MATRIX -',1X,
1'COLUMNS = EIGENVECTORS, ROWS = VARIABLES')
2005 FORMAT (1H0,4X,'FACTOR LOADINGS-',1X,
1'COLUMNS = FACTORS, ROWS = VARIABLES')
2006 FORMAT (1H0,4X,'ROTATED FACTOR MATRIX -',1X,
1'COLUMNS = FACTORS, ROWS = VARIABLES')
2007 FORMAT (1H0,4X,'FACTOR SCORES -',1X,
1'COLUMNS = FACTORS, ROWS = OBSERVATIONS')
2008 FORMAT (1H0,4X,'STANDARDIZED INPUT DATA MATRIX -',1X,
1'COLUMNS = VARIABLES, ROWS = OBSERVATIONS')
2009 FORMAT (1H0,4X,'VARIMAX FACTOR SCORES -',1X,
1'COLUMNS = FACTORS, ROWS = OBSERVATIONS')
RETURN
END
)FTN,F50 ,TPF$.05
C
C SUBROUTINE FOR MULTIPLICATION OF MATRIX A BY MATRIX B
C TO GIVE MATRIX C. A IS L ROWS BY N COLUMNS.
C B IS N ROWS BY M COLUMNS, AND C WILL BE L ROWS BY M COLUMNS.
C
C SUBROUTINE MMULT(A,B,C,L,N,M,NA,MA,NB,MB,NC,MC)
DIMENSION A(184,184),B(184,184),C(184,184)
DO 100 I=1,L
DO 101 J=1,M
C(I,J)=0.0
DO 102 K=1,N
C(I,J)=C(I,J)+A(I,K)*B(K,J)
102 CONTINUE
101 CONTINUE
100 CONTINUE
RETURN
END
)FTN,F50 ,TPF$.06
C
C SUBROUTINE TO FIND INVERSE OF MATRIX A. B IS THE INVERSE

```

```

C OF A. A IS REDUCED TO THE IDENTITY MATRIX.
C A AND B ARE N X L. DET IS THE DETERMINANT OF A.
C
SUBROUTINE MINV(A,B,N,N1,DET)
DIMENSION A(184,184),B(184,184)
C SET B TO IDENTITY MATRIX AND SAVE THE ORIGINAL A MATRIX
C
DO 100 I=1,N
DO 101 J=1,N
B(I,J)=0.0
101 CONTINUE
B(I,I)=1.0
100 CONTINUE
DET=1.0
C
C CALCULATE INVERSE
C
DO 102 I=1,N
DIVIDE ITH ROW OF A AND B BY A(I,I)
DIV=A(I,I)
DET=DET*DIV
DO 103 J=1,N
A(I,J)=A(I,J)/DIV
B(I,J)=B(I,J)/DIV
103 CONTINUE
C REDUCED THE ITH COLUMN OF A TO ZERO
DO 104 J=1,N
IF (1-J) 1,104,1
1 RATIO=A(J,I)
DO 105 K=1,N
A(J,K)=A(J,K)-RATIO*A(I,K)
B(J,K)=B(J,K)-RATIO*B(I,K)
105 CONTINUE
104 CONTINUE
102 CONTINUE
RETURN
END
:QFTN,FSO ,TPF$.07
:
C SUBROUTINE TO STANDARDIZE THE COLUMNS OF A DATA MATRIX
:
SUBROUTINE STAND(X,N,M,N1,M1)
DIMENSION X(184,184)
:
C... STANDARDIZE IEACH COLUMN OF THE MATRIX
:
DO 100 I=1,M
:
C... CALCULATE MEAN AND STANDART DEVIATION OF COLUMN
:

```

```

:      SX=0.0
:      SXX=0.0
:      DO 101 J=1,N
:      SX=SX+X(J,I)
:      SXX=SXX+X(J,I)**2
: 101 CONTINUE
:      XM=SX/FLOAT(N)
:      SD=SQRT((SXX-SX*SX/FLOAT(N))/FLOAT(N-1))
: C
: C... SUBTRAC MEAND FROM EACH ELEMENT IN COMLUMN, THEN
: C DIVICE RESULT BY THE STANDART DEVIATION.
: C
:      DO 102 J=1,I,
:      X(J,I)=(X(J,I)-XM)/SD
: 102 CONTINUE
: 100 CONTINUE
:      RETURN
:      END
: @FTN,F50 ,TPF$.08
: C
: C SUBROUTINE TO CALCULATE THE MATRIX OF CORRELATIONS
: C BETWEEN COLUMNS OF DATA MATRIX X
: C
: C SUBROUTINE RCOEF(X,N,M,I1,M1,A,M2)
: C DIMENSION X(184,184), A(184,184)
: C AN=N
: C
: C... CALCULATE CORRELATION COEFFICIENT BETWEEN COLUMNS I AND J
: C
:      DO 100 I=1,M
:      DO 100 J=I,M
: C
: C... ZERO SUMS
: C
:      SX1=0.0
:      SX2=0.0
:      SX1X1=0.0
:      SX2X2=0.0
:      SX1X2=0.0
: C
: C... CALCULATE SUMS, SUMS OF SQUARES AND SUM OF CROSS-PRODUCT
: C OF COLUMNS I AND J
: C
:      DO 101 K=1,M
:      SX1=SX1+X(K,I)
:      SX2=SX2+X(K,J)
:      SX1X1=SX1X1+X(K,I)**2
:      SX2X2=SX2X2+X(K,J)**2
:      SX1X2=SX1X2+X(K,I)*X(K,J)
: 101 CONTINUE

```

```

C
C... CALCULATE CORRELATION COEFFICIENT AND STORE IN MATRIX A
C
R=(SX1X2-SX1*SX2/AN)/
1 SQRT((SX1X1-SX1*SX1/AN)*(SX2X2-SX2*SX2/AN))
A(I,J)=R
A(J,I)=R
100 CONTINUE
RETURN
END
$FTN,FSO ,TPF$.09
C
C SUBROUTINE TO CALCULATE THE MATRIX OF COSINE THETA SIMILARITY
C COEFFICIENTS BETWEEN COLUMNS OF DATA MATRIX X
C
SUBROUTINE CTHETA(X,N,M,I1,M1,A,M2)
DIMENSION X(184,184), A(184,184)
C
C... CALCULATE COSINE THETA BETWEEN COLUMNS I AND J
C
DO 100 I=1,M
DO 100 J=1,M
C
C... ZERO SUMS
SX1X1=0.0
SX2X2=0.0
SX1X2=0.0
C
C... CALCULATE SUMS OF SQUARES AND SUM OF CROSS PRODUCT
C
DO 101 K=1,N
SX1X1=SX1X1+X(K,I)**2
SX2X2=SX2X2+X(K,J)**2
SX1X2=SX1X2+X(K,I)*X(K,J)
101 CONTINUE
C
C... CALCULATE COSINE THETA AND STORE IN MATRIX A
C
A(I,J)=SX1X2/SQRT(SX1X1*SX2X2)
A(J,I)=A(I,J)
100 CONTINUE
RETURN
END
$FTN,FSO ,TPF$.10
C
C SUBROUTINE TO CALCULATE THE EIGENVALUES AND EIGENVECTORS
C OF AN N*N, SYMMETRIC MATRIX.
C
C UPON COMPLETION THE EIGENVALUES ARE STORED IN THE DIAGONAL
C ELEMENTS OF MATRIX A (IN DESCENDING ORDER). THE EIGENVECTORS
C ARE STORED BY COLUMNS IN MATRIX B.

```

```

:C
:C   EIGENVALUE A(I,I) CORRESPONDS TO EIGENVECTOR (B(J,I),J=1,N)
:C
:C   SUBROUTINE EIGENJ (A,B,N,N1)
:C   DIMENSION A(184,184),B(184,184)
:C
:C... CALCULATE INITIAL AND FINAL NORMS
:C   SET B TO IDENTITY MATRIX
:C
:C   ANORM=0.0
:C   DO 100 I=1,N
:C   DO 101 J=1,N
:C   IF (I-J) 2,1,2
:C   1 B(I,J)=1.0
:C   GO TO 101
:C   2 B(I,J)=0.0
:C   ANORM=ANORM+A(I,J)*A(I,J)
:C 101 CONTINUE
:C 100 CONTINUE
:C   ANORM=SQRT(ANORM)
:C   FNORM=ANORM*1.0E-09/FLOAT(N)
:C
:C... INITIALIZE INDICATORS AND COMPUTE THRESHOLD
:C
:C   THR=ANORM
:C 23 THR=THR/FLOAT(N)
:C   3 IND=0
:C... SCAN DOWN COLUMNS FOR OFF-DIAGONAL ELEMENTS
:C   GREATER THAN OR EQUAL TO THRESHOLD
:C
:C   DO 102 I=2,N
:C   I1=I-1
:
:
:   DO 103 J=1,I1
:   IF(ABS(A(J,I))-THR) 103,4,4
:C
:C... COMPUTE SIN AND COS
:C
:C   4 IND=1
:C   AL=-A(J,I)
:C   AM=(A(J,J)-A(I,I))/2.0
:C   AO=AL/SQRT(AL*AL+AM*AM)
:C   IF (AM) 5,6,6
:C   5 AO=-AO
:C   6 SINX=AO/SQRT(2.0*(1.0+SQRT(1.0-AO*AO)))
:C   SINX2=SINX*SINX
:C   COSX=SQRT(1.0-SINX2)
:C   COSX2=COSX*COSX

```



```

:C
:C... ROTATE COLUMNS I AND J
:C
      DO 104 K=1,N
      IF (K-J) 7,10,7
      7 IF (K-I) 8,10,8
      8 AT=A(K,J)
      A(K,J)=AT*COSX-A(K,I)*SINX
      A(K,I)=AT*SINX+A(K,I)*COSX
      10 BT=B(K,J)
      B(K,J)=BT*COSX-B(K,I)*SINX
      B(K,I)=BT*SINX+B(K,I)*COSX
      104 CONTINUE
      XT=2.0*A(J,I)*SINX*COSX
      AT=A(J,J)
      BT=A(I,I)
      A(J,J)=AT*COSX2+BT*SINX2-XT
      A(I,I)=AT*SINX2+BT*COSX2+XT
      A(J,I)=(AT-BT)*SINX+COSX+A(J,I)*(COSX2-SINX2)
      A(I,J)=A(J,I)
      DO 105 K=1,N
      A(J,K)=A(K,J)
      A(I,K)=A(K,I)
      105 CONTINUE
      103 CONTINUE
      102 CONTINUE
      IF(IND) 20,20,3
      20 IF(THR-FNORM) 25,25,23
C
C... SORT EIGENVALUES AND EIGENVECTORS
C
      DO 110 I=2,N
      J=I
      29 IF(A(J-1,J-1)-A(J,J)) 30,110,110
      30 AT=A(J-1,J-1)
      A(J-1,J-1)=A(J,J)
      A(J,J)=AT
      DO 111 K=1,N
      AT=B(K,J-1)
      B(K,J-1)=B(K,J)
      B(K,J)=AT
      111 CONTINUE
      J=J-1
      IF (J-1) 110,110,29
      110 CONTINUE
      RETURN
      END
@FTN,FSO ,TPF$.11
:C
:C... SUBROUTINE TO PERFORM A VARIMAX ROTATION ON A
:C FACTOR LOADING MATRIX

```

```

:C
: SUBROUTINE VARMAX(F,M,L,M1)
: DIMENSION F(184,184), H(184)
:
:C
:C... INITIALIZE
:C
: WRITE(10,2001)
: SQR2=1.0/SQRT(2.0)
: XM=M
: L1=L-1
: NIT=-1
: NCM=0
:C
:C... NORMALIZE ROWS OF FACTOR MATRIX
:C
:C CALCULATE INITIAL COMUNALITIES
:C
: DO 100 I=1,M
: SUMH=0.0
: DO 101 J=1,L
: SUMH=SUMH+F(I,J)**2
: 101 CONTINUE
: H(I)=SUMH
: DO 102 J=1,L
: F(I,J)=F(I,J)/SQRT(SUMH)
: 102 CONTINUE
: 100 CONTINUE
:C
:C... CALCULATE AND PRINT VARIANCE BEFORE EACH ITERATION
:C
: 1 TVF=0.0
: DO 103 I=1,L
: SF1=0.0
: SF2=0.0
: DO 104 J=1,M
: SF1=SF1+F(J,I)**2
: SF2=SF2+F(J,I)**4
: 104 CONTINUE
: TVF=TVF+(XM*SF2-SF1*SF1)/(XM*XM)
: 103 CONTINUE
: IF (NIT .LT. 0) GO TO 2
:C
:C... IF VARIANCE IS THE SAME AFTER 5 SUCCESSIVE ITERATIONS,
:C... VARIMAX ROTATION IS COMPLETE
:C
: IF(ABS(TVF-TVI) .GT. 0.000001) GO TO 2
: NCM=NCM+1
: IF (NCM .GE. 5) GO TO 50
: 2 NIT=NIT+1
: TVI=TVF
: WRITE (10,2002) NIT,TVF

```

```

: C... ROTATE COLUMNS I AND J
:   DC 105 I=F,L1
:   L2=I+1
:   DO 106 J=L2,L
: C
: C... CALCULATE PARAMETERS REQUIRED TO FIND ROTATION ANGE
: C
:   A=0.0
:   B=0.0
:   C=0.0
:   D=0.0
:   DO 107 K=1,M
:   X=F(K,I)
:   Y=F(K,J)
:   U=(X+Y)*(X-Y)
:   V=2.0*X*Y
:   A=A+U
:   B=B+V
:   C=C+(U+V)*(U-V)
:   D=D+2.0*U*V
: 107 CONTINUE
:   XM=D-(2.0*A*B)/XM
:   XC=C-(A+A*B)/XM
:   XR=SQRT(XM*XM+XC*XC)
:   IF (XR .LE. 0.001) GO TO 106
: C
: C... CALCULATE SIN AND COS OF ROTATION ANGLE
: C
:   COS4T=XO/XR
:   COS2T=SQRT((1.0+COS4T)/2.0)
:   COS1T=SQRT((1.0+COS2T)/2.0)
:   SIN1T=SQRT(1.0-COS1T*COS1T)
: C
: C... IF ROTATION ANGLE VERY SMALL, NO NEED TO ROTATE
: C
:   IF (SIN1T .LE. 0.001) GO TO 106
:   IF (XM .LT. 0.0) SIN1T=-SIN1T
: C
: C... PERFORM ROTATION ON COLUMNS I AND J
: C
:   DO 108 K=1,M
:   X=F(K,I)
:   Y=F(K,J)
:   F(K,I)=X*COS1T+Y*SIN1T
:   F(K,J)=Y*COS1T-X*SIN1T
: 108 CONTINUE
: 106 CONTINUE
: 105 CONTINUE
:   GO TO 1
: 50 WRITE (10,2004)

```

```

:      WRITE (10,2001)
:C
:C... UN-NORMALIZE ROWS OF ROTATED FACTOR MATRIX
:C      CALCULATE FINAL COMMUNALITIES
:C      PRINT INITIAL AND FINAL COMMUNALITIES
:C
:      DO 110 I=1,M
:      SUMH=0.0
:      DO 111 J=1,L
:      F(I,J)=F(I,J)*SQRT(H(I))
:      SUMH=SUMH+F(I,J)**2
: 111 CONTINUE
:      D=H(I)-SUMH
:      WRITE (10,2003) I,H(I),SUMH,D
: 110 CONTINUE
:      WRITE (10,2005)
:      RETURN
: 2001 FORMAT (1F1)
: 2002 FORMAT ( )
: 2003 FORMAT ( )
: 2004 FORMAT (1F0,4X,'NUMBER OF VARIMAX ITERATIONS',1X,
: 1 'AND VARIANCE AT EACH STEP')
: 2005 FORMAT (1F0,4X,'COLUMN 1 = INITIAL COMMUNALITY',2X,
: 1 'COLUMN 2 = COMMUNALITY AFTER ROTATION',/,
: 2 5X,'COLUMN 3 = DIFFERENCE')
:      END

```

APENDICE B

PROPIEDADES DE LOS EIGENVALORES Y EIGENVECTORES EN TERMINOS DE MINIMOS CUADRADOS

Cualquier matriz simétrica S de orden $p \times p$ puede ser representada en términos de sus eigenvalores $\lambda_1 > \lambda_2 > \dots > \lambda_r$ y eigenvectores u_1, u_2, \dots, u_r como:

$$S = \lambda_1 u_1 u_1' + \lambda_2 u_2 u_2' + \dots + \lambda_r u_r u_r' \quad (1)$$

donde r es el rango de S . De la misma manera, se puede demostrar que cualquier matriz rectangular X de orden $N \times p$ y rango r puede ser representada en función de sus valores singulares $\gamma_1 > \gamma_2 > \dots > \gamma_r$ y vectores v_1, v_2, \dots , y u_1, u_2, \dots, u_r como:

$$X = \gamma_1 v_1 u_1' + \gamma_2 v_2 u_2' + \dots + \gamma_r v_r u_r' \quad (2)$$

Aún más fundamental que estas dos representaciones son las pro-

propiedades de mínimos cuadrados en términos de estas, lo cual se relaciona con la solución a los dos siguientes problemas.

Problema 1

Dada una matriz Gramiana simétrica (una matriz Gramiana tiene eigenvalores no negativos) $S_{(p \times p)}$ de rango r , encontrar una matriz simétrica gramiana T con un rango menor, $k < r$, que se aproxima a S en función de los mínimos cuadrados.

Problema 2

Dada una matriz rectangular $X_{(N \times p)}$ de rango r , encontrar una matriz W de rango k que se aproxime a X en términos de mínimos cuadrados.

La solución a estos dos problemas es:

$$T = \lambda_1 u_1 u_1' + \lambda_2 u_2 u_2' + \dots + \lambda_k u_k u_k'$$

y

$$W = \gamma_1 v_1 u_1' + \gamma_2 v_2 u_2' + \dots + \gamma_k v_k u_k'$$

(3)

respectivamente. Las matrices T y W son llamadas "matrices de aproximación con mínimos cuadrados de rango mínimo" de S y X , respectivamente.

Según las propiedades de mínimos cuadrados podemos decir que

T se puede representar como una aproximación de S ya que la suma de los cuadrados de todos sus elementos cuando hacemos $S - T$ es mínima. Algo similar sucede con W que también puede aproximarse a X en el sentido de que la suma de los cuadrados de sus elementos cuando consideramos $X - W$ es mínima.

Un cálculo final o una buena aproximación es en ambos casos:

$$\lambda_{k+1} + \lambda_{k+2} + \dots + \lambda_r;$$

es decir, este cálculo es la suma de $r - k$ eigenvalores mínimos de $X'X$.

Un cálculo relativo es:

$$\frac{\lambda_{k+1} + \lambda_{k+2} + \dots + \lambda_r}{\lambda_1 + \lambda_2 + \dots + \lambda_r} \quad (4)$$

o alguna relación funcional que se use para obtener una buena aproximación.

La importancia de estos dos problemas es que son el punto de partida para analizar el caso más común, el cual es cuando $r = p$. En el problema 1, S es la matriz de covarianza de p variables linealmente dependientes. La matriz T también representa una matriz de covarianza de p variables, y puesto que el rango de T es $k < p$, estas p variables son linealmente dependientes sobre las k variables. Aquí, las p variables originales, cuya matriz de covarianza es S, puede generarse apro-

ximadamente por k variables. Si consideramos el problema 2, estas ideas se aclararán. La matriz original, X , de orden $N \times p$ puede escribirse como:

$$X = V \Gamma U'$$

donde V es de orden de $N \times p$, con columnas ortonormales, Γ es una matriz diagonal de orden $p \times p$, y U es una matriz cuadrada ortonormal de orden $p \times p$. Con una aproximación de un rango menor, W , puede expresarse como sigue:

$$W = V_k \Gamma_k U_k'$$

donde V_k está formada por las primeras columnas de V , Γ_k consta de las k primeros renglones y columnas de Γ y U_k está formada por las primeras columnas de U . Sencillamente $W \approx X$, por lo tanto

$$X = V_k \Gamma_k U_k \quad (5)$$

post-multiplicando a esta matriz por $U_k \Gamma_k^{-1}$ obtendremos:

$$V_k = X U_k \Gamma_k^{-1} \quad (6)$$

La matriz $U_k \Gamma_k^{-1}$, de orden p por k , representa una transformación de los renglones de X en p -espacio Euclídiano a un k -espacio Euclídiano y la ecuación (5) indica que existe una transformación de la

matriz X de N por p a una matriz V_k de N por k . La matriz X representa N puntos en un p -espacio Euclídiano el cual puede aproximarse a un k -espacio Euclídiano. La matriz V_k representa las coordenadas de N puntos en k -espacio Euclídiano.

APENDICE C

ROTACION DE SISTEMAS DE COORDENADAS

En la figura 2.a, los elementos del vector x dan las coordenadas de un punto $P(x_1, x_2)$ con respecto a los ejes coordenados para dos variables.

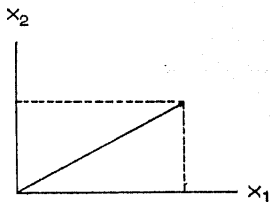


Figura 2.a

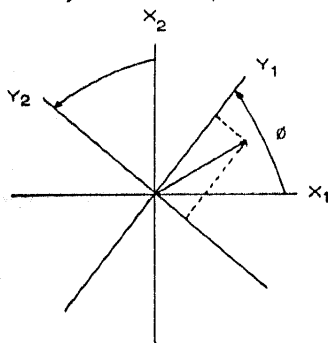


Figura 2.b

Suponiendo que por alguna razón necesitamos rotar los ejes θ° en

dirección contraria al movimiento de las manecillas del reloj, como lo indica la figura 2.b. Por lo tanto nuestro problema es encontrar las coordenadas del punto $P_1(x_1, x_2)$ con respecto a los ejes Y_1 y Y_2 .

De los conceptos fundamentales de trigonometría, las coordenadas del punto P_1 están dadas por las siguientes ecuaciones:

$$y_1 = \cos \theta x_1 + \sin \theta x_2 \quad (1)$$

$$y_2 = \sin \theta x_1 + \cos \theta x_2$$

Estas dos ecuaciones pueden ser escritas en forma matricial:

$$y' = x' T \quad (2)$$

donde $T =$

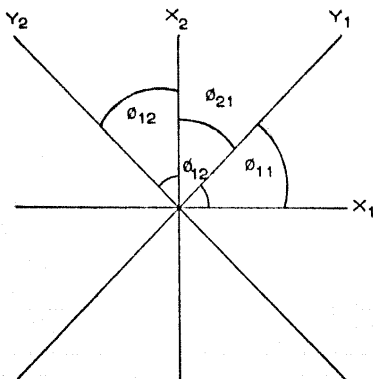
$$\begin{matrix} \cos \theta & - \sin \theta \\ \sin \theta & \cos \theta \end{matrix}$$


Figura 2.c

Consideremos un caso más general según lo indica la figura 2.c. Aquí, los ángulos θ_{ij} son los ángulos entre el i -ésimo eje de referencia y el j -ésimo nuevo eje. Para una rotación recta, y un espacio bidimensional, deduciremos las siguientes relaciones:

$$\theta_{12} = \theta_{11} + 90^\circ$$

$$\theta_{21} = \theta_{11} - 90^\circ$$

$$\theta_{22} = \theta_{11}$$

Las funciones trigonométricas que relacionan estos ángulos son:

$$\begin{aligned} \text{sen } \theta_{11} &= \text{sen } (\theta_{21} + 90^\circ) = \text{cos } \theta_{21} \\ - \text{sen } \theta_{11} &= \text{sen } (\theta_{12} - 90^\circ) = \text{sen } (90^\circ - \theta_{12}) = \text{cos } \theta_{12} \end{aligned} \quad (3)$$

sustituyendo estas ecuaciones en (2) tendremos:

$$\begin{aligned} y_1 &= \text{cos } \theta_{11} x_1 + \text{cos } \theta_{21} x_2 \\ y_2 &= \text{cos } \theta_{12} x_1 + \text{cos } \theta_{22} x_2 \end{aligned} \quad (4)$$

Todo este procedimiento puede ser generalizado para cualquier número de ejes ortogonales y sus equivalentes rotados. El sistema de ecuaciones para p ejes es:

$$\begin{aligned} y_1 &= \text{cos } \theta_{11} x_1 + \text{cos } \theta_{21} x_2 + \dots + \text{cos } \theta_{p1} x_p \\ y_p &= \text{cos } \theta_{1p} x_1 + \text{cos } \theta_{2p} x_2 + \dots + \text{cos } \theta_{pp} x_p \end{aligned} \quad (5)$$

Sustituyendo $t_{ij} = \cos \theta_{ij}$, donde el subíndice i indica los ejes del sistema coordinado original y j los ejes del sistema rotado, las ecuaciones anteriores se resumen en la siguiente expresión:

$$y' = x' T \quad (6)$$

Para un conjunto de N vectores renglón en la matriz $X_{(n \times p)}$ la ecuación:

$$Y = XT \quad (7)$$

dará en Y las coordenadas de todos los N vectores renglones en términos de los p ejes rotados.

La matriz T es una matriz de transformación que convierte X en Y .