



DIVISION DE EDUCACION CONTINUA  
FACULTAD DE INGENIERIA U.N.A.M.

COMPUTACION APLICADA A LA PLANEACION URBANA

SISTEMAS DE PRONOSTICO

M. en I. Rubén Téllez Sánchez

AGOSTO, 1981

## SISTEMA DE PRONOSTICO

### 1.- INTRODUCCION

Para predecir los valores futuros de una variable se utilizan las series de tiempo, que son un conjunto de observaciones tomadas durante un determinado periodo de tiempo.

La estimación de los valores futuros de las series de tiempo se usan como base para hacer decisiones. Por ejemplo, pronosticar la demanda de productos, inventarios, predecir las características de calidad, instalaciones o equipo utilizado, fuerza de trabajo, circulante, etc.

El pronostico de series de tiempo se aplican ampliamente en sistemas de control de inventarios, en modelos de planeación de producción total, de programación, de control de costos, de información administrativa.

Existen dos formas para estimar valores futuros de series de tiempo: predicción y pronóstico.

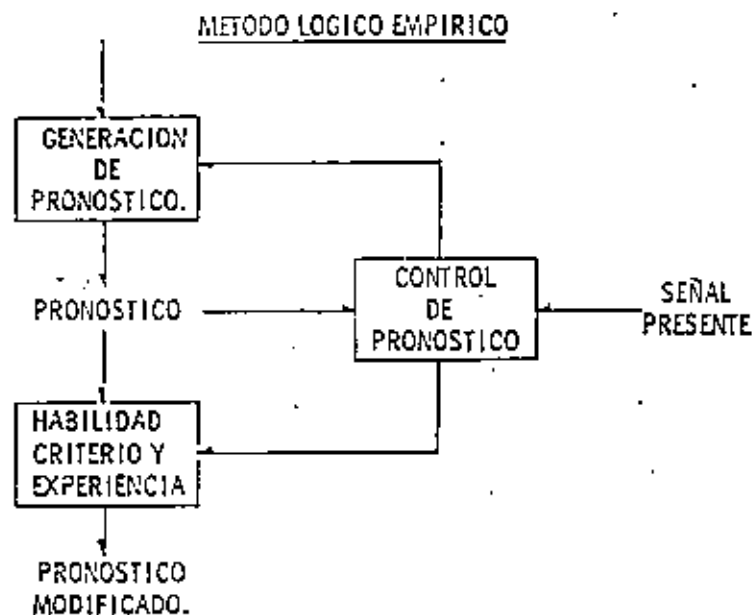
**Predicción.** - Predecir implica prever factores cualitativos, por ejemplo, los efectos al introducir un nuevo producto, competencia, política económica. La predicción requiere habilidad, experiencia y criterio. La predicción es un proceso subjetivo.

**Pronóstico.** - En el pronóstico se analizan datos del pasado para proyectar los al futuro por medio de un modelo matemático. El pronóstico es un proceso objetivo y modificable, se conocen los factores que intervienen, y es la entrada a una predicción. Sin embargo, el pronóstico está limitado

por el tiempo.

El pronóstico de rango corto es más exacto que el de rango largo, su periodo futuro cuando más consta de un año. Su aplicación principal es en la programación de producción, planificación de inventarios y otras actividades. El pronóstico de rango grande tiene un periodo de cinco a diez años y se utiliza en las decisiones de cómo y cuándo se deben construir nuevas plantas.

Para que un pronóstico sea efectivo y eficiente, debe encajar en el sistema de control, como se muestra en la siguiente figura.



Se analizan técnicas de pronóstico para los tres tipos básicos de series de tiempo.

- a). - Proceso constante.
- b). - Tendencia lineal y cuadrática.
- c). - Variación periódica y cíclica y sus combinaciones.

Muchos modelos de series de tiempo se representan por funciones algebraicas o trascendentes de tiempo. Por ejemplo, si las observaciones son de una muestra aleatoria de alguna distribución de probabilidad y si la media de esta no cambia en el tiempo, entonces su modelo constante es:

$$x_t = a + \varepsilon_t \quad (1.1)$$

Donde  $a$  es la media desconocida y  $\varepsilon_t$  es la componente del error aleatorio. Se supone además que :

$$E(\varepsilon_t) = 0 \quad \text{y} \quad V(\varepsilon_t) = \sigma^2$$

Si la media de la distribución cambia linealmente en el tiempo, entonces el modelo de tendencia lineal es:

$$x_t = a + bt + \varepsilon_t \quad (1.2)$$

La variación cíclica se obtiene introduciendo la componente trascendental :

$$x_t = a_0 + a_1 \sin \frac{2\pi t}{12} + a_2 \cos \frac{2\pi t}{12} + \varepsilon_t \quad (1.3)$$

para 12 periodos.

Un problema de pronóstico consiste en estimar los parámetros desconocidos del modelo de series de tiempo para proyectarlos hacia el futuro. Por ejemplo, si  $\hat{a}$  y  $\hat{b}$  son los valores de los parámetros desconocidos  $a$  y  $b$ , en la ecuación 1.2; y si se está al final del periodo  $T$ , entonces el pronóstico para un tiempo futuro  $t+z$  es:

$$\hat{x}_{t+z} = \hat{a} + \hat{b} [t+z]$$

estimando así la tendencia de las componentes  $\hat{a}$ ,  $\hat{b}$ , en el periodo futuro.

En la mayoría de los casos, la exactitud de un pronóstico se determina utilizando la función de error para el periodo  $t$ , y se denota por

$$e_t = x_t - \hat{x}_t$$

para estimar los parámetros del modelo se usa el criterio de mínimos cuadrados para minimizar el error de pronóstico.

## 2.- PREDICCIÓN SUBJETIVA

La metodología de predicción subjetiva es un procedimiento intuitivo, que deberá cumplir con lo siguiente :

- a).- La solución debe de ser en un período corto de tiempo.
- b).- Cuando la relación de beneficios futuros a el costo de técnicas sofisticadas es desfavorable para el tomador de decisiones.
- c).- Cuando el tomador de decisiones piensa que su intuición en una situación particular es más confiable que cualquier función matemática de predicción.

Las características del Criterio subjetivo son :

- a).- No es reproducible.
- b).- Es único para un determinado tomador de decisiones individuales.
- c).- Está basado en un registro de predicciones acumulado en un período de tiempo pasado.

## MODELOS ESTRUCTURALES Y ECONOMETRICOS

Un modelo estructural es un conjunto de funciones matemáticas las cuál... intentan representar relaciones casuales que describen los factores que controlan la variable que desea predecir, así como los medios disponibles para el que realiza la predicción. El modelo debe ser una buena presentación de las observaciones.

Se ilustrará esta metodología con el siguiente ejemplo :

EJEMPLO : (Evaluación de proyectos de inversión en perspectiva).

Para esta evaluación se necesita primero la evaluación del beneficio futuro y por lo tanto del precio del bien que está siendo producido. Seguiremos la nomenclatura que se presenta a continuación para el análisis de nuestro problema.

$Q_t^s$  = Cantidad suministrada al mercado en el tiempo T.

$Q_t^d$  = Cantidad demandada en el mercado en el tiempo T.

$P_t$  = precio en el tiempo  $T$

$W_t$  = Salario pagado, por hora, en la industria en el tiempo  $T$

$Y_t$  = Ingreso del consumidor en el tiempo  $T$

En nuestro ejemplo consideramos que se sigue el modelo más simple en la inter-relación de la economía de la firma por lo que supondremos que el precio está determinado, en un mercado competitivo, por las funciones de oferta-demanda.

En el modelo de oferta-demanda que se postula se considera:

i)  $Q_t^s$  depende del precio  $P_t$  y de  $W_t$

ii)  $Q_t^d$  depende del precio  $P_t$  y del ingreso  $Y_t$  del consumidor

Por lo tanto, matemáticamente el modelo es:

$$Q_t^s = f_s(P_t, W_t)$$

$$Q_t^d = f_d(P_t, Y_t)$$

Para la completa especificación de este modelo deberán realizarse las siguientes etapas.

1a. Etapa: (Especificación de las funciones  $f_s$  y  $f_d$ )

En la práctica se considera que  $f_s$  y  $f_d$  son lineales:

$$Q_t^s = \alpha_0 + \alpha_1 P_t + \alpha_2 W_t \quad (2.1)$$

$$Q_t^d = \beta_0 + \beta_1 P_t + \beta_2 Y_t \quad (2.2)$$

Donde  $\alpha_j$  ( $j=0,1,2$ ) y  $\beta_j$  ( $j=0,1,2$ ) son coeficientes constantes, pero desconocidos.

2a. Etapa: (inferencia de los parámetros del modelo).

Esta etapa consiste en la estimación de los valores de los coeficientes desconocidos  $\alpha_j, \beta_j$  a partir de datos históricos de las variables:

$$Q_t^s, P_t, W_t, Q_t^d, Y_t.$$

En esta inferencia se introducen las variables aleatorias respectivamente, las cuáles representan los errores (perturbancias) asociados en las ecuaciones teóricas propuestas.

Se considera que  $\varepsilon_{s,t}$  y  $\varepsilon_{d,t}$  son variables aleatorias con media cero.

$$Q_t^s = \alpha_0 + \alpha_1 P_t + \alpha_2 W_t + \varepsilon_{s,t} \quad (2.3)$$

$$Q_t^d = \beta_0 + \beta_1 P_t + \beta_2 Y_t + \varepsilon_{d,t} \quad (2.4)$$

$\alpha_j$  y  $\beta_j$  se determinan por medio de la técnica de regresión.

Observaciones:

i). - como el modelo formado por las ecuaciones (2.3), (2.4), se origina en un contexto económico, también es llamado un modelo econométrico.

ii). - la estimación de los coeficientes es un problema en la aplicación de la teoría de la econometría.

3a. Etapa (Predicción de los precios futuros).

Suponiendo que son dados los valores futuros de  $W_t, Y_t, \varepsilon_{s,t}, \varepsilon_{d,t}$  y se desea conocer valores futuros de precio y cantidades suministradas y demandas en el mercado. como en el sistema (2.3), (2.4), se tienen tres incógnitas  $Q_t^s, Q_t^d, P_t$  y dos ecuaciones, se puede resolver este sistema usando la condición de equilibrio en un mercado competitivo perfecto.

$$Q_t^s = Q_t^d$$

$$P_t = \frac{(\beta_0 + \beta_2 Y_t + \varepsilon_{d,t}) - (\alpha_0 + \alpha_2 W_t - \varepsilon_{s,t})}{(\alpha_1 - \beta_1)}$$

Observaciones:

i). - En la práctica no se tiene información sobre los errores (perturbancias), se puede considerar que son iguales a:

$$E(\varepsilon_{s,t}) = 0 \quad y$$

$$E(\varepsilon_{d,t}) = 0$$

ii). -  $W_t$  y  $Y_t$  se consideran conocidas o bien que han sido determinadas fuera - del modelo (variables exógenas).

Su determinación podría ser otro problema de predicción o se usaría pre- dicciones subjetivas (intuitivas) para considerar al salario y al ingreso - como entradas al modelo. Sin embargo, en esta última situación, también podría preferirse confiar directamente en una predicción intuitiva del -- precio, ya que probablemente se tenga más confianza en nuestro "sentim- iento", que tenemos sobre nuestro mercado que en magnitudes tan relativas e intangibles como el ingreso del consumidor.

### Fuentes de error en la predicción cuando se emplea el modelo estructural

- i). - Salarios e ingresos futuros diferirán de los valores supuestos cuando el modelo sea resuelto para obtener predicciones.
- ii). - Valores reales de perturbancias (errores) futuras diferirán de cero.
- iii). - Error de muestreo que se presenta en la estimación de los coeficientes en la etapa segunda.
- iv). - Error de especificación del modelo, es decir, la estructura misma puede ser deficiente en algún aspecto, por ejemplo la relación de oferta-deman- da podría ser no-lineal, en lugar de lineal como se supuso.

### 3.- TÉCNICA DE REGRESIÓN

Técnicas de regresión para estimar los parámetros de algunos modelos de series de tiempo.

Primariamente se analiza un caso particular del modelo general lineal. - Este caso particular es de forma:

$$X_t = a + bt + \varepsilon_t \quad (3.1)$$

cuyas componentes de error son  $E(\varepsilon_t) = 0$ ,  $V(\varepsilon_t) = \sigma^2$  y  $Cov(\varepsilon_i, \varepsilon_j) = 0$ , para  $i \neq j$ .

El modelo lineal general para los parámetros  $a_i$ , se presenta por

$$X_t = \sum_{i=1}^T a_i f_i(t) + \varepsilon_t \quad (3.2)$$

$$t = 1, 2, \dots, T$$

donde  $a_i$  es el coeficiente de los T-ésimos términos del modelo y las variable independientes  $f_i(t)$  son funciones del tiempo.

El modelo general comprende varios casos que se presentarían en la práctica, como los siguientes:

$$X_t = a_0 + a_1 t + a_2 t^2 + \varepsilon_t$$

$$X_t = a_0 + a_1 t + a_2 \sin 2\pi t + a_3 e^{-t} + \varepsilon_t$$

Nótese, que ambos modelos son lineales en los parámetros  $a_i$ , y se puede usar regresión múltiple para estimarlos.

### 3.1. MODELO LINEAL SIMPLE

Definición Sea  $A = \{1, 2, \dots, T\}$  el conjunto indicador histórico de la serie.  $\{X_t : t \in A\}$

El modelo lineal simple para la serie de tiempo  $\{X_t : t \in A\}$  se define por la relación

$$X_t = a + bt + \varepsilon_t, \quad t \in A \quad (3.3)$$

y debe satisfacer las siguientes suposiciones sobre la variable aleatoria  $\varepsilon_t$ , que representa al error del pronóstico cuando se usa la ecuación (3.3):

- i)  $E(\varepsilon_t) = 0$  (i.e., el promedio de error es cero).
- ii)  $V(\varepsilon_t) = \sigma^2$  (i.e., la variancia del error es constante e igual a  $\sigma^2$ ).
- iii)  $Cov(\varepsilon_i, \varepsilon_j) = 0$ , para  $i \neq j$  (i.e., los errores son variables aleatorias no relacionadas).

El siguiente teorema ilustra la forma de estimar los parámetros  $\hat{a}$  y  $\hat{b}$  del modelo (3.3), usando el criterio de mínimos cuadrados, y también proporciona el método de obtener el pronóstico para un tiempo futuro  $T+\tau$ .

**TEOREMA.** Sea  $\{X_t: t \in A\}$  una serie de tiempo con  $A = \{1, 2, \dots, T\}$  como conjunto indicador histórico. Si el modelo elegido para esta serie es de la forma (3.3), entonces:

$$\hat{a} = \frac{2(2T+1)}{T(T-1)} \sum_{t=1}^T X_t - \frac{6}{T(T-1)} \sum_{t=1}^T tX_t \equiv \hat{a}(T) \quad (3.4)$$

$$\hat{b} = \frac{12}{T(T^2-1)} \sum_{t=1}^T tX_t - \frac{6}{T(T-1)} \sum_{t=1}^T X_t \equiv \hat{b}(T) \quad (3.5)$$

ii) el pronóstico para un tiempo futuro  $T+\tau$ , es:

$$\hat{X}_{T+\tau} = \hat{a}(T) + \hat{b}(T)[T+\tau] \quad (3.6)$$

**Demostración:**

El criterio de mínimos cuadrados, consiste en elegir aquellos valores de  $\hat{a}$  y  $\hat{b}$  que minimicen la suma de los cuadrados de los residuos, la cual se indica por  $SS_E$ , i.e.,

$$SS_E = \sum_{t=1}^T [X_t - \hat{a} - \hat{b}t]^2$$

Para minimizar  $SS_E$ ,  $\hat{a}$ ,  $\hat{b}$  deberán satisfacer que

$$\frac{\partial SS_E}{\partial \hat{a}} = -2 \sum_{t=1}^T [X_t - \hat{a} - \hat{b}t] = 0 \quad \text{y} \quad \frac{\partial SS_E}{\partial \hat{b}} = -2 \sum_{t=1}^T [X_t - \hat{a} - \hat{b}t](t) = 0$$

Estas dos condiciones dan origen a dos ecuaciones en las incógnitas  $\hat{a}$  y  $\hat{b}$  que al resolverlas, se obtienen los resultados mostrados en la parte i) del teorema. Para la demostración de la parte ii) basta con sustituir los valores de los estimadores  $\hat{a}$ ,  $\hat{b}$  en el modelo lineal simple, y evaluar  $X_t$  en el tiempo  $t = T+\tau$ .

**Observación:**

Los valores de los estimadores  $\hat{a}$  y  $\hat{b}$ , se calcularán en base a un número  $T$  de datos históricos, por lo que su valor

depende de este número  $T$  de datos disponibles. Por esta razón los valores de  $\hat{a}$  y  $\hat{b}$  se han indicado por  $\hat{a}(T)$  y  $\hat{b}(T)$  para hacer incipie que están calculados en base a un número  $T$  de datos.

**Ejemplo:**

El registro de la demanda semanal de un producto nuevo es mostrado en la siguiente tabla. Estos datos se usan para estimar los parámetros en el modelo de tendencia lineal.

Semana ( $t$ )	Demanda ( $X_t$ )
1	10
2	12
3	15
4	18
5	20

Aplicando la primera parte del teorema anterior se tiene:

$$\sum_{t=1}^5 X_t = 75 \quad \sum_{t=1}^5 tX_t = 251$$

$$\hat{a} = \frac{2(11)}{5(4)} (75) - \frac{6}{5(4)} (251) = 7.2$$

$$\hat{b} = \frac{12}{5(24)} (251) - \frac{6}{5(4)} (75) = 2.6$$

Aplicando la segunda parte del teorema, la ecuación del pronóstico es:

$$\hat{X}_{5+\tau} = 7.2 + 2.6(5+\tau)$$

El pronóstico de la demanda para la siguiente semana, es decir  $\tau=1$ , es:

$$\hat{X}_6 = 7.2 + 2.6(6) = 22.8 \approx 23.$$

### 3.2 - EL MODELO GENERAL LINEAL.

Definición.- El modelo general lineal se define por la relación:

$$x_t = \sum_{i=1}^n \beta_i f_i(t) + \varepsilon_t, \quad t \in A, \quad t=1, 2, \dots, T \quad (3.7)$$

donde  $A = \{1, \dots, T\}$  es el conjunto indicador histórico, las  $\beta_i$  son los parámetros de regresión,  $f_i(t)$  son funciones independientes del tiempo, y se cumplen las siguientes condiciones:  $E(\varepsilon_t) = 0$ ,  $V(\varepsilon_t) = \sigma_t^2$ , y  $Cov(\varepsilon_i, \varepsilon_j) = 0$ , para  $i \neq j$ .

NOTAS: Representaciones matriciales:

i) Representación matricial del modelo general lineal (3.7):

Sea  $\beta = (\beta_i)$  un vector columna  $n \times 1$ ,  
 $f(t) = (f_i(t))$  un vector columna  $n \times 1$ ,  
 $x(t) = (x_i)$  un vector renglón  $1 \times T$ ,  
 $E(\varepsilon_t)$  un vector renglón  $1 \times T$ , y  
 $F$  una matriz definida por:

$$F = \begin{bmatrix} f_1(1) & f_1(2) & \dots & f_1(T) \\ f_2(1) & f_2(2) & \dots & f_2(T) \\ \vdots & \vdots & \ddots & \vdots \\ f_n(1) & f_n(2) & \dots & f_n(T) \end{bmatrix}$$

La notación  $\beta'$  indica la transpuesta de  $\beta$ .

La relación (3.7) se puede expresar matricialmente por cualquiera de las siguientes formas:

$$x_t = \beta' f(t) + \varepsilon_t, \quad t \in A \quad (3.8)$$

$$X = \beta' F + E \quad (3.9)$$

ii) Representación matricial del residuo:

Sea  $\hat{\beta} = (\hat{\beta}_i)$  un vector  $n \times 1$ , un estimador de

$\beta = (\beta_i)$

$$\text{sea } e_t = x_t - \hat{x}_t = x_t - \hat{\beta}' f(t), \quad t \in A \quad (3.10)$$

el  $t$ -ésimo residuo.

Si  $e = (e_t)$  es un vector  $1 \times T$ , la representación matricial del residuo, i.e. de la ecuación (3.10) es

$$e = X - \hat{\beta}' F \quad (3.11)$$

iii) Representación matricial de la suma con prioridades de los cuadrados de los residuos:

$$\text{Sea: } SS_E = \sum_{t=1}^T w_{tt} e_t^2 \quad (3.12)$$

la suma con prioridades de los cuadrados de los residuos, donde  $w_{tt}$  es el factor de prioridad asociado al  $t$ -ésimo residuo.

Si  $W$  es una matriz definida por:

$$W = \begin{bmatrix} w_{11} & 0 & \dots & 0 & 0 \\ 0 & w_{22} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & w_{TT} \end{bmatrix}$$

entonces la representación matricial de  $SS_E$  es:

$$\begin{aligned} SS_E &= (eW)(eW)' = eWw'e = eW^2e = (X - \hat{\beta}'F)W^2(X - \hat{\beta}'F)' \\ &= XW^2X' - \hat{\beta}'FW^2X' - XW^2F'\hat{\beta}' + \hat{\beta}'FW^2F'\hat{\beta}' \quad (3.13) \end{aligned}$$

El estimador de mínimos cuadrados debe satisfacer que

$$\frac{\partial SS_E}{\partial \hat{\beta}} = -2FW^2X' + 2FW^2F'\hat{\beta}' = 0$$

$\therefore$  la ecuación normal de mínimos cuadrados es:

$$FW^2F'\hat{\beta}' = FW^2X' \quad \text{o} \quad G\hat{\beta}' = g$$

TEOREMA. Para el modelo general lineal (3.7) se tiene:

a) el estimador  $\hat{\beta}' = (\hat{\beta}_i)$  de  $\beta = (\beta_i)$ , usando el



el criterio de mínimos cuadrados es:

$$\hat{\beta} = G^{-1} g \equiv \hat{\beta}(T) \quad (3.16)$$

donde  $G = (FW)(FW)'$  es la matriz de los cuadrados de los pesos.

y  $g = FW^2 x'$  es el vector de la suma de los pesos.

En notación escalar:

$$G_{ij} = \sum_{t=1}^T W_{it}^2 f_i(t) f_j(t), \text{ para } i, j = 1, 2, \dots, n$$

$$g_i = \sum_{t=1}^T W_{it}^2 x_t f_i(t), \text{ para } i = 1, 2, \dots, n$$

b) el pronóstico para un tiempo futuro  $T+t$ , es:

$$x_{T+t} = \hat{\beta}'(T) f(T+t) \quad (3.17)$$

c) el estimador  $\hat{\beta}$  es insesgado, i.e.

$$E(\hat{\beta}) = E(G^{-1} g) = (FW^2 F')^{-1} FW^2 E(x') = F'^{-1} F' F F^{-1} a = a$$

$$\therefore E(\hat{\beta}) = a$$

Variancia-covariancia

d) la matriz de variancia-covariancia del estimador  $\hat{\beta}$  es:

$$V = E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = \sigma_\epsilon^2 G^{-1} F W^2 (F W^2)' G^{-1} \quad (3.18)$$

Si  $W$  es la matriz unitaria (i.e. todos los datos reciben los mismos pesos), entonces:

$$V = \sigma_\epsilon^2 G^{-1} = \sigma_\epsilon^2 (F F')^{-1} \quad (3.19)$$

e) Si  $\sigma_\epsilon^2$  es desconocido, un estimador  $\hat{\sigma}_\epsilon^2$  es

$$\hat{\sigma}_\epsilon^2 = \frac{ee'}{T-n} = \frac{\sum_{t=1}^T (x_t - \hat{x}_t)^2}{T-n} \quad (3.20)$$

4. TECNICAS DE PROMEDIOS MOVILES.

Existen dos técnicas de pronósticos basadas en promedios móviles. La técnica de promedios móviles consiste en aplicar el criterio de mínimos cuadrados a un conjunto de datos históricos de tamaño fijo, donde a cada dato se le asigna el mismo factor de prioridad (o peso). Se analizarán las técnicas de promedios móviles para procesos constantes y de tendencia lineal.

4.1 PROCESOS CONSTANTES.

Definición. un proceso constante para la serie en el tiempo  $\{x_t : t \in A\}$  es un modelo de la forma

$$x_t = a + \epsilon_t \quad (4.1)$$

donde  $a$  es un parámetro desconocido, constante en cualquier segmento local de tiempo, pero es posible que en diferentes intervalos largos de tiempo varíe,  $\epsilon_t$  es una variable aleatoria tal que:

$$E(\epsilon_t) = 0 \text{ y } V(\epsilon_t) = \sigma_\epsilon^2.$$

TECNICA de promedios móviles simples.

Teorema.- Para el proceso constante (4.1) se tiene que:

i) el estimador  $\hat{a}$  de  $a$ , tomando únicamente  $N$  de los  $T$  datos del conjunto indicador histórico  $A = \{t : t = 1, 2, \dots, T\}$ , y usando el criterio de mínimos cuadrados está dado por:

$$\hat{a} = \frac{1}{N} \sum_{t \in T+N} x_t = M_T \quad (4.2)$$

o por su forma alternativa:

$$M_T = M_{T-1} + \frac{x_T - x_{T-N}}{N}$$

ii) el pronóstico para un tiempo futuro  $T+t$ , es:

$$x_{T+t} = \hat{a} = M_T \quad (4.4)$$

Demost. usando el criterio de mínimos cuadrados y considerando que sólo se usarán los  $N$  más recientes datos del conjunto histórico se tiene que la suma de los cuadrados de los residuos,  $SS_E$ , está dada por:

$$SS_E = \sum_{t=T-N+1}^T (x_t - \bar{a})^2$$

El valor de  $\bar{a}$  que minimiza  $SS_E$ , y satisface que  $\frac{\partial SS_E}{\partial \bar{a}} = 0$ , lo cual implica (4.2).

### Observación.

El estimador  $M_T$  para  $a$ , es el promedio de las  $N$  más recientes observaciones del conjunto histórico  $\{x_t: t \in A\}$ . Por esta razón,  $M_T$  se le llama promedio móvil de periodo  $N$ . En cada periodo la observación más vieja es descartada y la más nueva es añadida al conjunto.

Ejemplo: la demanda de un gato hidráulico en los últimos 6 días es:

$$x_1 = 19, x_2 = 24, x_3 = 22, x_4 = 19, x_5 = 20, x_6 = 16.$$

El estimador  $M_6$  para  $a$  es:

$$M_6 = \frac{1}{6} [19 + 24 + 22 + 19 + 20 + 16] = 20$$

por lo tanto  $\bar{a} = M_6 = 20$ .

El pronóstico para el siguiente día es  $\hat{x}_7 = \bar{a} = 20$ . Suponiendo que al otro día se observa una demanda  $x_7 = 20$ , entonces eliminando  $x_1$  y agregando  $x_7$  para obtener un nuevo estimador:

$$M_7 = \frac{1}{6} [24 + 22 + 19 + 20 + 16 + 22] = 20.5$$

o usando la ecuación (3.22)

$$M_T = M_{T-1} + \frac{x_T - x_{T-N}}{N} \quad \therefore M_7 = 20 + \frac{22 - 19}{6} = 20.6$$

NOTA: El comportamiento del método de promedio móvil es una función de  $N$ , el número de observaciones que se desean promediar. Se analiza cuando es conveniente elegir un valor grande de  $N$  y cuando un valor pequeño.

i) Si el parámetro  $a$  del proceso cambia súbitamente de un valor  $a = a_1$  a otro  $a = a_2$ , se necesitará que  $N$  sea pequeña para que al calcular  $M_T$  no se tomen datos más viejos ya que éstos no proporcionarían una buena evaluación del nuevo parámetro  $a = a_2$ . Obsérvese que, si se elige un valor grande de  $N$  entonces se estarían tomando los primeros datos los cuales ya no representan el comportamiento del proceso. Por esta razón se dice que cuando  $N$  es grande, el método de promedios móviles reacciona lentamente a cambios súbitos del parámetro  $a$  del proceso.

ii) Si el parámetro  $a$  del proceso no cambia, entonces se desea que  $N$  sea grande, ya que para  $N$  grande se obtendrá un estimador  $\bar{a}$  con menor variancia, puesto que:

$$V(\bar{a}) = \frac{\sigma^2}{N} \rightarrow 0, \text{ cuando } N \rightarrow \infty.$$

lo cual es una característica deseable para un estimador.

## 4.2 - PROCESOS CON TENDENCIA LINEAL.

Definición: Un proceso con tendencia lineal para la serie de tiempo  $\{x_t: t \in A\}$ , es un modelo de la forma:

$$x_t = a + bt + \varepsilon_t \quad (4.5)$$

donde  $a$  y  $b$  son parámetros constantes desconocidos en cualquier intervalo de tiempo, y  $\varepsilon_t$  es una variable aleatoria con  $E(\varepsilon_t) = 0$  y  $V(\varepsilon_t) = \sigma^2$ .

TECNICA DE PROMEDIOS MOVILES DOBLES

TEOREMA. Para el proceso constante lineal (4.5), se tiene:

1- Si  $\hat{a}$  y  $\hat{b}$  son estimadores de  $a$  y  $b$ ; y  $\hat{a}' = \hat{a} + \hat{b}'\bar{t}$ ,

donde

$$\bar{t} = \frac{1}{N} \sum_{t=T-N+1}^T t, \text{ entonces: } \hat{a}'(T) = \hat{a}' = \frac{1}{N} \sum_{t=T-N+1}^T X_t = M_T \quad (4.6)$$

$$\hat{b}(T) = \hat{b} = \frac{12}{N(N^2-1)} \left[ \frac{N-1}{2} X_t + \frac{N-3}{2} + \dots + \dots - \frac{N-3}{2} X_{T-N+2} - \frac{N-1}{2} X_{T-N+1} \right] W_T \quad (4.7)$$

2- Las formulas recursivas para los estimadores  $\hat{a}'(T)$  y  $\hat{b}(T)$

son:

$$\hat{a}'(T) = M_T = M_{T-1} + \frac{1}{N} (X_T - X_{T-N}) \quad (4.8)$$

$$\hat{b}(T) = W_T = W_{T-1} + \frac{12}{N(N^2-1)} \left[ \frac{N-1}{2} X_T + \frac{N-1}{2} X_{T-N} - N M_{T-1} \right] \quad (4.9)$$

3- Si se desea encontrar  $\hat{a}$  en lugar de  $\hat{a}'$ , entonces:

$$\hat{a}(T) = M_T - W_T \left( T - \frac{N-1}{2} \right) \quad (4.10)$$

4- El pronostico para un tiempo futuro  $T+t$ , se puede calcular por cualquiera de las siguientes formulas:

i) 
$$\hat{X}_{T+t} = \hat{a}(T) + \hat{b}(T)(T+t) = \hat{a}'(T) + \hat{b}(T) \left[ \frac{N-1}{2} + t \right] = M_T + W_T \left[ \frac{N-1}{2} + t \right] \quad (4.11)$$

ii) 
$$\hat{X}_{T+t} = \hat{X}_T + \hat{b}(T)t = 2M_T - M_T^{[2]} + t \left( \frac{2}{N-1} \right) (M_T - M_T^{[2]}) \quad (4.12)$$

donde:

$$M_T = \frac{1}{N} \sum_{t=T-N+1}^T X_t \quad (4.13)$$

$$M_T^{[2]} = \frac{1}{N} \sum_{t=T-N+1}^T M_t \quad (4.14)$$

$$= M_{t-1}^{[2]} + \frac{1}{N} [M_t - M_{t-N}]$$

ala cantidad  $M_t^{[2]}$  se le llama promedio móvil doble. A las cantidades  $M_T$  se les llama promedio móvil simple.

Ejemplo: Las ventas semanales de un acondicionador de aire de 5000 BTU. se encuentran en la segunda columna de la siguiente tabla:

semana	demanda	$M_t$	$M_t^{[2]}$	$\hat{X}_t$
1	10			
2	12			
3	15			
4	17			
5	16	13.4		
6	17	15.2		
7	18	16.4		
8	21	17.6		
9	23	17.4	16.40	
10	20	20.2	17.76	23.70
11	22	20.8	18.85	23.86
12	24	22.0	20.00	23.00
13	23	22.4	20.96	24.56
14	21	22.0	21.48	22.78
15	25	23.0	22.04	24.44

El proceso que genera estos datos puede ser adecuadamente aproximado por un modelo de tendencia lineal, y para pronosticar las ventas en la siguiente semana, se utiliza un promedio móvil doble de cinco semanas. Los promedios móviles simples  $M_t$  se calculan para cada semana desde  $t=5, 6, \dots, 15$  y se muestran en la Tercera columna. Los promedios móviles  $M_t^{[2]}$  se calculan usando (4.14) ó (4.15) para  $t=9, 10, \dots, 15$ . Los pronosticos hechos al final de la semana  $t$ , para la siguiente semana se calculan con la ecuación (4.12) para  $t=1$ , y se muestran en la última columna para el período  $t+1$ .

Por ejemplo:

$$\hat{X}_{10} = 2M_9 - M_9^{[2]} + (1)(\frac{2}{4})(M_9 - M_9^{[2]}) =$$

$$= 2(17.4) - (16.4) + \frac{1}{2}(17.4 - 16.4) = 23.90$$

$$\hat{X}_{11} = 2M_{10} - M_{10}^{[2]} + (1)(\frac{2}{4})(M_{10} - M_{10}^{[2]}) = 2(20.2) - (17.76) + \frac{1}{2}(20.2 - 17.76) = 23.26$$

### 5- TÉCNICAS DE SUAVIZACIÓN EXPONENCIAL.

Las técnicas de suavización exponencial, que en seguida se analizan, son las más ampliamente usadas debido a su exactitud y eficiencia computacional. En el desarrollo de estas técnicas se usa el criterio de mínimos cuadrados con pesos (prioridades).

#### 5.1- SUAVIZACIÓN EXPONENCIAL SIMPLE PARA UN PROCESO CONSTANTE.

Técnica de suavización exponencial simple para un proceso constante.

##### TEOREMA

$$X_t = a + E_t, \quad t \in A = \{1, 2, \dots, T\} \quad (5.1)$$

$E_t$  = componente del error aleatorio.

Se supone  $E(E_t) = 0$ ,  $V(E_t) = \sigma_t^2$ , es decir, como el definido por (4.1), se tiene:

El estimador  $\hat{a} = \hat{a}(T) \equiv S_T$  del parámetro  $a$ , se puede calcular por cualquiera de las siguientes 3 ecuaciones:

$$i) \hat{a} = \frac{(1-\beta)}{(1-\beta^{T+1})} \sum_{t=1}^T \beta^{T-t} X_t \equiv \hat{a}(T) \equiv S_T \quad (5.2)$$

donde  $\beta$  es una constante dada entre 0 y 1.

ii) Si  $\alpha = 1 - \beta$ , entonces  $\hat{a} \equiv S_T$ , está dada por

$$S_T = \alpha X_T + (1-\alpha) S_{T-1} \quad (5.3)$$

La operación definida por (5.3) es llamada suavización exponencial simple, al valor  $S_T$  se le llama valor alisado o estimador de suavización, y a la constante  $\alpha$  se le llama constante de suavización.

iii) Otra expresión para encontrar  $\hat{a} \equiv S_T$ , es

$$S_T = \alpha \sum_{k=0}^{T-1} \beta^k X_{T-k} + \beta^T S_0 \quad (5.4)$$

donde  $\alpha = 1 - \beta$ ,  $i = 2, 3, \dots$

El proceso da un estimador insesgado de  $a$ , puesto que

$$E(S_T) = E\left[\alpha \sum_{k=0}^{T-1} \beta^k X_{T-k}\right] = \alpha \sum_{k=0}^{T-1} \beta^k E[X_{T-k}] = \alpha \sum_{k=0}^{T-1} \beta^k a = a.$$

2.- El pronóstico para un tiempo futuro  $T+t$ , en base a un conjunto histórico indicado  $A^h = \{t = 1, 2, \dots, T\}$ , es

$$\hat{X}_{T+t} = S_T \quad (5.5)$$

Demostración. El criterio de mínimos cuadrados con prioridades para encontrar el estimador  $\hat{a}$ , consiste en elegir aquel valor  $\hat{a}$  que minimice la suma con prioridades (pesos) de los cuadrados de los residuos:

$$SS_E = \sum_{t=1}^T \beta^{T-t} e_t^2 = \sum_{t=1}^T \beta^{T-t} (X_t - \hat{X}_t)^2$$

$$SS_E = \sum_{t=1}^T \beta^{T-t} (X_t - \hat{a})^2, \quad 0 < \beta < 1$$

donde  $\beta^{T-t}$  es el peso (o prioridad) dado al residuo  $e_t$ .

Derivando  $SS_E$  con respecto a al valor  $\hat{a}$ , e igualando a cero se obtiene el resultado (5.2). El resultado (5.3) se obtiene realizando operaciones algebraicas en la ecuación (5.2). Existe un procedimiento heurístico para obtener el resultado (5.3), que se usa frecuentemente. (ver referencia 9).

Observaciones:

- 1- Otro nombre para la técnica de suavización exponencial simple, es suavización exponencial de primer orden.
- 2- La ecuación que en la práctica se utiliza para estimar  $\hat{S}_t = S_t$ , es la ecuación (5.3). Para aplicar esta ecuación se requiere conocer el valor de la constante  $\alpha$ . El valor que se elige de  $\alpha$  es, generalmente, entre 0.1 y 0.30. El criterio para seleccionar este valor se presentará más adelante.
- 3- La técnica de suavización exponencial requiere del conocimiento de un valor inicial  $S_0$ , para emplear la fórmula (5.3), en forma iterativa (ver la ecuación 5.4). Si se tienen datos históricos disponibles,  $S_0$  se puede tomar como el promedio de los más recientes datos. Si no existen datos históricos disponibles, entonces se hará un pronóstico subjetivo de  $S_0$ .

Ejemplo: Un contratista desea pronosticar el número de instalaciones de calentadores de agua por semana. Él tiene los siguientes datos disponibles:

Semana	1	2	3	4	5	6	7	8	9	10
No de instalaciones	15	18	10	12	20	17	22	16	14	20

Examinando gráficamente los datos, él decide suponer un modelo constante y utilizar suavización exponencial simple. Suponiendo arbitrariamente un  $\alpha = 0.1$ , entonces

$$S_{11} = (0.1)X_{11} + (0.9)S_{10}$$

Sin embargo no hay un valor inicial  $S_{10}$  (notese que  $S_{10}$  representaría a  $S_0$  si redefinimos el origen del origen del conjunto indicador histórico como  $t=10$ ). Para estimar el valor inicial  $S_0$ , se tomará el promedio de las demandas de las diez primeras semanas, que es igual a 16.6. Por lo tanto, es razonable considerar

$$S_{10} = 16.6 \approx 17$$

Entonces el promedio para cualquier tiempo futuro  $10+t$ , será:

$$\hat{X}_{10+t} = S_{10} = 17$$

Suponga que el número real de instalaciones en la semana 11 fue de 15. Entonces:

$$S_{11} = \alpha X_{11} + (1-\alpha)S_{10}$$

$$S_{11} = (0.1)(15) + (0.9)(16.60) = 16.44 \approx 16$$

y el pronóstico para un tiempo futuro  $11+t$ , basándose en un conjunto indicador histórico  $A = \{t: t=1, 2, \dots, 11\}$ , es:

$$\hat{X}_{11+t} = S_{11} = 16.44 \approx 16.$$

En la siguiente tabla aparecen los pronósticos para periodos de tamaño  $t=1$ , contra los valores reales de las instalaciones de calentadores:

Demanda y pronóstico de instalaciones de calentadores de agua.

semana (t)	$X_t$	$S_t$	$\hat{X}_t$
1	15		
2	18		
3	10		
4	12		
5	20		
6	17		
7	22		
8	16		
9	14		
10	20	16.50	
11	15	16.44	17
12	12	16.00	16
13	16	15.60	16
14	20	16.40	16
15	22	16.76	16
16	17	16.96	17
17	15	16.76	17
18	10	16.08	17
19	16	16.07	16
20	20	16.46	16
21	19	16.71	16
22	24	17.44	17
23	18	17.50	17
24	15	17.25	17
25	20	17.53	17

52. SUAVIZACION EXPONENCIAL DOBLE PARA PROCESOS CON TENDENCIA LINEAL.

Técnica de suavización exponencial doble.  
Teorema. - Para un proceso con tendencia lineal.

$x_t = a + bt + E_t$  ,  $t \in A = \{1, 2, \dots, T\}$   
como el definido por (4.5),  $E(E_t) = 0$   
 $V(E_t) = \sigma^2$

∴ se tiene:

- 1- Los estimadores  $\hat{a}$  y  $\hat{b}$  de  $a$  y  $b$  son respectivamente  
 $\hat{b} = \frac{\alpha}{\rho} (S_T - S_T^{[2]}) = \hat{b}(T)$  (5.6)  
 $\hat{a} = 2S_T - S_T^{[2]} - T \frac{\alpha}{\rho} [S_T - S_T^{[2]}] = \hat{a}(T)$  (5.7)

donde  $S_T = \alpha x_T + (1-\alpha) S_{T-1}$  (5.8)

y  $S_T^{[2]} = \alpha S_T + (1-\alpha) S_{T-1}^{[2]}$  (5.9)

A  $S_T^{[2]}$  se le llama estimador doblemente suavizado exponencialmente. La notación  $S_T^{[2]}$  no significa que  $S_T$  se eleve al cuadrado, sino que es un símbolo para indicar el resultado de la ecuación (5.9)

- 2- El pronóstico para un tiempo futuro  $T+\tau$ , usando suavización exponencial doble se calcula por:

$\hat{x}_{T+\tau} = \hat{x}_T + \tau \hat{b}(T)$  (5.10)

$\hat{x}_{T+\tau} = (2 + \tau) S_T - (1 + \tau) S_T^{[2]}$  (5.11)

donde  $\delta = \tau(\alpha/\rho)$  (5.12)

- 3- Los valores iniciales  $S_0$  y  $S_0^{[2]}$ , para iniciar la suavización exponencial doble se calculan por:

$S_0 = \hat{a}(0) - \frac{\rho}{\alpha} \hat{b}(0)$  (5.13)

$S_0^{[2]} = \hat{a}(0) - 2 \frac{\rho}{\alpha} \hat{b}(0)$  (5.14)

donde  $\hat{a}(0)$ ,  $\hat{b}(0)$  son estimados de datos históricos usando regresión lineal. Si no se dispone de datos históricos entonces se hacen estimaciones subjetivas de  $\hat{a}(0)$  y  $\hat{b}(0)$ .

Ejemplo: Un analista de investigación de operaciones de un centro de cómputo de tiempo compartido, desea pronosticar los ingresos para su compañía ha estado en operación dos años, sin embargo, él considera que los ingresos de estos dos años no indican las operaciones comerciales actuales, ya que la compañía no ha llegado a estar operativamente establecida. Él piensa que los ingresos se incrementarán linealmente con el tiempo, y además sus mejores estimadores subjetivos para los parámetros de esta relación lineal son (en miles de dólares):  $\hat{a}(0) = 95$  y  $\hat{b}(0) = 1.0$ .

El analista decide usar suavización exponencial doble con  $\alpha = 0.1$ . Usando estos estimadores, los valores iniciales requeridos para la suavización exponencial son:

$S_0 = 95 - \frac{(0.9)}{(0.1)}(1) = 86$

$S_0^{[2]} = 95 - 2 \frac{(0.9)}{(0.1)}(1) = 77$

Por lo tanto, el pronóstico (en miles de dólares) para el mes 1 es:

$\hat{x}_1 = [2 + (1) \frac{(0.1)}{(0.9)}] S_0 - [1 + (1) \frac{(0.1)}{(0.9)}] S_0^{[2]}$   
 $= (2.11)(86) - (1.111)(77) = 95.999 \approx 96$

Suponga que el ingreso real en el mes 1 fue 98, entonces los estimadores suavizados serían:

$S_1 = \alpha x_1 + (1-\alpha) S_0 = (0.1)(98) + (0.9)(86) = 87.20$

$S_1^{[2]} = \alpha S_1 + (1-\alpha) S_0^{[2]} = (0.1)(87.20) + (0.9)(77) = 78.02$

y el pronóstico para el 2do año sería:

$x_2 = (2.11) S_1 - (1.111) S_1^{[2]} = (2.11)(87.20) - (1.111)(78.02) = 96.399 \approx 96$

Los ingr. mensuales para el siguiente año y sus pronósticos se presentan en la siguiente tabla, y graficados en la fig. 5.1

Pronósticos de ingresos mensuales usando series de tiempo expresional doble.

Mes (t)	$X_t$	$S_t$	$S_t^{LAT}$	$\hat{X}_t$
0		86.00	77.00	
1	98	87.20	78.02	46
2	94	87.88	79.01	97
3	97	88.99	80.03	98
4	104	90.49	81.05	99
5	108	92.24	82.17	101
6	100	93.02	83.26	103
7	106	94.32	84.36	104
8	104	95.29	85.45	105
9	118	97.54	86.66	106
10	104	98.70	87.87	110
11	102	99.53	88.48	111
12	116	100.73	90.16	110

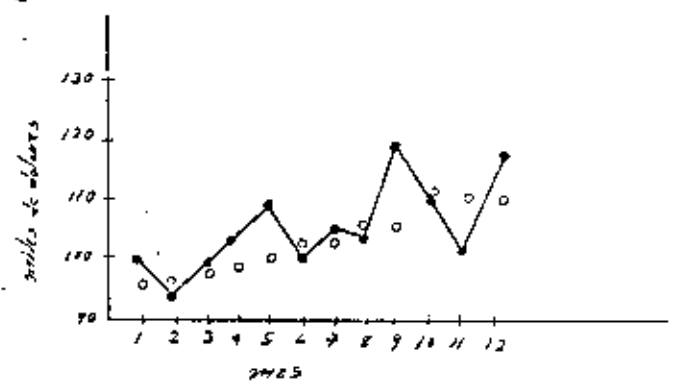


fig. 5.1

o = ingresos actuales.  
— = pronóstico.

### 5.3 - METODO DE WINTER PARA VARIACIONES ESTACIONALES.

Para el pronóstico de una serie de tiempo con variación estacional se usa el método de winter.

**Teorema.** - Considerando un modelo estacional cuya estación o ciclo, está formado por L periodos, y que sigue la relación:

$$X_t = (a + bt) C_t + E_t \quad (5.15)$$

donde

a es la señal base y se le llama componente permanente.

b es la componente de tendencia lineal.

$E_t$  es la componente del error aleatorio.

$C_t$  es la componente estacional (o factor estacional) para el período t, y satisface:

$$\sum_{t=1}^L C_t = L \quad (5.16)$$

i) El procedimiento para revisar periódicamente los estimadores de los parámetros del modelo y hacer el pronóstico, se presenta a continuación. Al final de cualquier período T, después de observar  $X_T$ , se realizan los siguientes pasos:

i) Se revisa (o se estima) el estimador de la componente permanente:

$$\hat{a}(T) = \alpha \left[ \frac{X_T}{C_T(T-L)} \right] + (1-\alpha) [\hat{a}(T-1) + \hat{b}(T-1)] \quad (5.17)$$

donde  $0 \leq \alpha \leq 1$ , es una constante de suavización.

ii) Se revisa el estimador de la componente de tendencia:

$$\hat{b}(T) = \beta [\hat{a}(T) - \hat{a}(T-1)] + (1-\beta) \hat{b}(T-1) \quad (5.18)$$

donde  $0 < \beta < 1$ , es una segunda constante de suavización.

iii) Se revisa el estimador del factor estacional para el período T:

$$C_t(T) = \gamma \left[ \frac{X_T}{\hat{a}(T)} \right] + (1-\gamma) \hat{C}_T(T-L) \quad (5.19)$$

donde  $0 < \gamma < 1$ , es la tercera constante de suavización.

iv) El  $\hat{x}_{T+t}$  nóstico para cualquier período  $T+t$ , es:

$$\hat{x}_{T+t} = [\hat{a}(T) + t\hat{b}(T)]\hat{c}_{T,t} (T+t-L) \quad (5.20)$$

2- Los valores iniciales  $\hat{a}(0)$ ,  $\hat{b}(0)$  y  $\hat{c}_t(0)$  para  $t=1,2,\dots,L$ , necesarios para iniciar esta técnica, se estiman de la siguiente manera:

i) Si se disponen de datos de dos estaciones anteriores, entonces

$$\hat{b}(0) = \frac{\bar{x}_2 - \bar{x}_1}{L} \quad (5.21)$$

donde  $\bar{x}_1$  es la demanda promedio de la primera estación y  $\bar{x}_2$  es la demanda promedio durante la más reciente estación.

ii) Los factores estacionales iniciales pueden primeramente calcularse por

$$\hat{c}_t(0) = \frac{x_{t+L} - x_t}{L\hat{b}(0)}, \quad t=1,2,\dots,L \quad (5.22)$$

Sin embargo, este procedimiento, generalmente, da resultados pobres debido a la alta variabilidad del patrón de la demanda. Por esto se recomienda estimar inicialmente  $\hat{a}(0)$  por (5.22), que no puede ser usado para estimar la componente permanente  $\hat{a}$  (ver ecuación (5.20)), y después se vuelven a estimar en términos de  $\hat{a}(0)$  y  $\hat{b}(0)$ , ver (5.2+).

iii) Para estimar  $\hat{a}(0)$ , se usa:

$$\hat{a}(0) = \frac{\sum_{t=1}^{2L} x_t + 3L^2 \hat{b}(0) - 2\hat{b}(0) \sum_{t=1}^L t \hat{c}_t(0)}{2L} \quad (5.23)$$

para  $t=1,2,\dots,L$

iv) Para revalorizar los factores estacionales iniciales, basándose en los valores de  $\hat{a}(0)$  y  $\hat{b}(0)$ , se usa:

$$\hat{c}_t(0) = \frac{1}{2} \left[ \frac{x_t}{\hat{a}(0) - (2L-2)\hat{b}(0)} + \frac{x_{t+L}}{\hat{a}(0) - (L-t)\hat{b}(0)} \right] \quad (5.24)$$

para  $t=1,2,\dots,L$ .

NOTA: Otro procedimiento para estimar los factores estacionales consiste en dividir la demanda en cada período, entre la demanda promedio en la estación. Este procedimiento es muy adecuado cuando no aparece la componente de tendencia (i.e., cuando  $b=0$ ).

Ejemplo: La demanda para sistemas de aire acondicionado con características de 5000 BTU y 110V, es estacional, con una mayor demanda en los meses de primavera y verano. Datos históricos para 1970 se encuentran disponibles y aparecen en la tabla I. Suponemos que la componente de tendencia lineal es cero, por lo tanto los factores estacionales se calculan dividiendo la demanda mensual entre la demanda mensual promedio durante el año. Estos factores estacionales aparecen en la última columna de la tabla I.

TABLA I- Datos históricos (1970)

MES	Demanda	Factores estacionales estimados.
Enero	1	0.18
Febrero	2	0.24
Marzo	3	0.60
Abril	8	0.96
Mayo	11	1.32
Junio	13	1.56
Julio	18	2.16
Agosto	15	1.80
Septiembre	9	1.08
Octubre	6	0.72
Noviembre	5	0.60
Diciembre	4	0.48
<b>Total</b>	<b>100</b>	<b>12.00</b>

Los parámetros iniciales son:

$\hat{b}(0) = 0$ , por supuesto que no existe esta componente.

$$\hat{a}(0) = \frac{100}{12} = 8.3$$

Las constantes de suavización que se eligen son  $\alpha = 0.2$ ,



$$\beta = 0.1 \quad \alpha = 0.5$$

Los valores para  $\hat{a}(t)$ ,  $\hat{b}(t)$ ,  $\hat{c}(t)$  se calculan con las ecuaciones (5.11), (5.16), (5.19) respectivamente. El pronóstico con la ecuación (5.20). Los resultados aparecen en la tabla II. Por ejemplo, en enero:

$$\hat{I}_{\text{ener}} = [\hat{a}(0) + \hat{b}(0)] \hat{C}_{\text{ener}} (\text{en } 1970) = (8.3 + 0)(0.48) = 3.98 \approx 4.$$

Ya que la demanda real en enero de 1971, fue de 5, se tiene de (5.17), que:

$$\begin{aligned} \hat{a}(\text{ener}) &= \alpha \left[ \frac{I_{\text{ener}}}{\hat{C}_{\text{ener}} (\text{en } 1970)} \right] + (1-\alpha) [\hat{a}(0) + \hat{b}(0)] = \\ &= 0.2(5) / 0.48 + (0.8)(8.3) = 8.72 \end{aligned}$$

De (5.18), se tiene:

$$\begin{aligned} \hat{b}(\text{ener}) &= \rho [\hat{a}(\text{ener}) - \hat{a}(0)] + (1-\rho) \hat{b}(0) = \\ &= 0.1(8.72 - 8.3) + 0.9(0) = 0.043. \end{aligned}$$

y de la (5.19), se tiene:

$$\begin{aligned} \hat{C}_{\text{ener}} (\text{en } 1971) &= \gamma \left[ \frac{I_{\text{ener}}}{\hat{a}(\text{ener})} \right] + (1-\gamma) \hat{C}_{\text{ener}} (\text{enero } 1970) \\ &= 0.5 \left( \frac{5}{8.72} \right) + 0.5(0.48) = 0.53 \end{aligned}$$

El pronóstico para Febrero sería

$$\begin{aligned} \hat{I}_{\text{Fe}} &= [\hat{a}(\text{ener}) + \hat{b}(\text{ener})] \hat{C}_{\text{Fe}} (\text{Fe. } 1970) \\ &= (8.72 + 0.043)(0.24) = 2.1 \end{aligned}$$

Los elementos restantes de la tabla II, se calculan de manera similar.

TABLA II - Resultados calculados para el año actual = 1971

MES	Demanda real	$\hat{a}(t)$	$\hat{b}(t)$	$\hat{C}_t(t)$	Pronóstico realizado un periodo anterior
Enero	5	8.72	0.043	0.53	4.0
Febrero	4	10.34	0.200	0.31	2.1
Marzo	7	10.77	0.723	0.63	6.3
Abril	7	10.25	0.149	0.82	10.6
Mayo	15	10.57	0.158	1.37	13.7
Junio	17	10.78	0.121	1.56	16.8
Julio	24	10.77	0.164	2.17	23.6
Agosto	18	10.71	0.142	1.73	20.0
Septiembre	12	11.06	0.143	1.08	11.7
Octubre	7	10.91	0.114	0.58	8.1
Noviembre	8	11.49	0.121	0.65	6.6
Diciembre	6	11.82	0.178	0.49	5.8
Total	130				127.3

6-5. SUAVIZACIÓN ADAPTATIVA.

La suavización adaptativa es un procedimiento equivalente al de la suavización exponencial, y fue desarrollado por Brown. En esta técnica se usan funciones polinomiales y exponenciales. El empleo de polinomios produce un pronóstico aproximadamente igual al de la suavización múltiple, excepto que en aquella, los parámetros del modelo se obtienen directamente.

El uso más importante de la suavización adaptable es con funciones trigonométricas, como seno y coseno. Por ejemplo, el modelo de una onda senoidal es:

$$x = a \sin \omega t$$

donde  $a$  = amplitud y origen en  $t=0$   
Para  $\lambda$  periodos, y en radianes, se tiene:

$$x = a \sin \omega(t + \lambda)$$

Usando un par equivalente seno-coseno, entonces tenemos:

$$x = a \sin \omega(t + \lambda) = a \cos \omega \lambda \sin \omega t + a \sin \omega \lambda \cos \omega t \\ = a_1 \sin \omega t + a_2 \cos \omega t$$

Para un modelo polinomial de 4 parámetros, se tiene:

$$x_t = a_1 + a_2 t + a_3 \sin \frac{2\pi t}{12} + a_4 \cos \frac{2\pi t}{12} + \epsilon_t$$

Y el de una onda senoidal con 12 puntos y base armónica es:

$$x_t = a_1 + a_2 t + a_3 \sin \frac{2\pi t}{12} + a_4 \cos \frac{2\pi t}{12} + a_5 \sin \frac{\pi t}{12} + a_6 \cos \frac{\pi t}{12} + \epsilon_t$$

La inclusión de la armónica permite diseñar una gran variedad de modelos estacionarios. En general, se tiene:

$$x_t = \sum_{i=1}^n a_i f_i(t) + \epsilon_t, \quad t=1, 2, \dots, T \quad (6.1)$$

donde  $f_i(t)$  son variables independientes

$T$  = observaciones disponibles.

En forma matricial:

$$x = (x_1, \dots, x_T)' = \epsilon, \quad t=1, 2, \dots, T$$

Los coeficientes  $a$  que minimizan a:

$$SS_e = \sum_{t=1}^T \beta^{T-t} [x_t - \hat{a}' f(t)]^2$$

son:

$$\hat{a}(T) = G(T)^{-1} g(T) \quad (6.2)$$

$$\text{donde } G_{ij}(T) = \sum_{t=1}^T \beta^{T-t} f_i(t) f_j(t), \quad i, j=1, 2, \dots, n$$

$$g_i(T) = \sum_{t=1}^T \beta^{T-t} x_t f_i(t), \quad i=1, 2, \dots, n$$

Ahora supongamos que todas las funciones  $f_i(t)$  son tales que sus valores en el tiempo  $t+1$  son una combinación lineal en el tiempo previo  $t$ , entonces:

$$f_i(t+1) = L_{i1} f_1(t) + L_{i2} f_2(t) + \dots + L_{in} f_n(t), \quad i=1, 2, \dots, n$$

Si  $L$  es la matriz de los  $L_{ij}$  de  $(n \times n)$ , se tiene:

$$f(t+1) = L f(t) \quad (6.3)$$

Por ejemplo, considerando la variación del modelo de tendencia cuadrática:

$$x_t = a_1 + a_2 t + \frac{1}{2} a_3 t(t-1) + \epsilon_t$$

entonces  $f_1(t) = 1$

$$f_2(t) = t$$

$$f_3(t) = \frac{1}{2} t(t-1)$$

y se verifica que:

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

Dando a  $L$  y  $f(0)$ , se encuentra  $f(t) = L^t f(0)$

$$\text{Si } G(T) = \sum_{j=0}^{T-1} \beta^j f(-j) f'(-j) = G(T-1) + \beta^{T-1} f(-T+1) f'(-T+1)$$

la matriz  $G(T)$  se aproxima al límite  $G$ , donde:

$$G = \lim_{T \rightarrow \infty} G(T) = \sum_{j=0}^{\infty} \beta^j f(-j) f'(-j) \quad (6.4)$$

$\therefore G^{-1}$  (calcula sólo una vez). Además, para el tiempo  $T$ , se tiene

$$\begin{aligned} g(T) &= \sum_{j=0}^{T-1} \beta^j x_{T-j} f(-j) = x_T f(0) + \sum_{j=1}^{T-1} \beta^{j-1} x_{T-j} f(-j) \\ &= x_T f(0) + \beta \sum_{j=1}^{T-1} \beta^{j-1} x_{T-j} L^{-1} f(-j+1) \\ &= x_T f(0) + \beta L^{-1} \sum_{k=0}^{T-2} \beta^k x_{T-1-k} f(-k) \end{aligned}$$

$$\therefore g(T) = x_T f(0) + \beta L^{-1} g(T-1)$$

De (6) también que

$$\hat{g}(T) = G^{-1} g(T) = G^{-1} [x_T f(0) + \beta L^{-1} g(T-1)] \quad (6.5)$$

y como  $g(T-1) = G \hat{a}(T-1)$ , se tiene

$$\hat{g}(T) = x_T G^{-1} f(0) + \beta G^{-1} L^{-1} G \hat{a}(T-1)$$

que es de la forma:

$$\hat{g}(T) = h x_T + H \hat{a}(T-1) \quad (6.6)$$

donde

$$\begin{aligned} h &= G^{-1} f(0) \\ H &= \beta G^{-1} L^{-1} G \end{aligned}$$

Notese que:

$$\begin{aligned} L^{-1} G &= L^{-1} G (L')^{-1} L' \\ &= \sum_{j=0}^{\infty} \beta^j L^{-1} f(-j) f'(-j) (L')^{-1} L' = \sum_{j=0}^{\infty} \beta^j [L^{-1} f(-j)] [L' f'(-j)] L' \\ &= \sum_{j=0}^{\infty} \beta^j f(-j+1) f'(-j+1) L' \end{aligned}$$

para  $k=j+1$ , se tiene:

$$L^{-1} G = \beta^{-1} \left[ \sum_{k=1}^{\infty} \beta^k f(-k) f'(-k) \right] L' = \beta^{-1} [G - f(0) f'(0)] L'$$

Así

$$H = \beta G^{-1} L^{-1} G = G^{-1} [G - f(0) f'(0)] L' = [I - G^{-1} f(0) f'(0)] L'$$

como

$h = G^{-1} f(0)$ , entonces:

$$H = L' - h f'(0) L' = L' - h [L f'(0)]' = L' - h f'(1)$$

$$\begin{aligned} \therefore \hat{a}(1) &= h x_T + [L' - h f'(1)] \hat{a}(T-1) = \\ &= L' \hat{a}(T-1) + h [x_T - f'(1) \hat{a}(T-1)] \quad (6.7) \end{aligned}$$

Si  $\hat{x}_T$  es el pronóstico para el período  $T$ , hecho al final del período  $T-1$  y como  $\hat{x}_T = f'(1) \hat{a}(T-1)$ , entonces (6.7) queda:

$$\hat{a}(T) = L' \hat{a}(T-1) + h [x_T - \hat{x}_T] \quad (6.8)$$

donde  $h[x_T - \hat{x}_T]$  es el error para un solo pronóstico, es decir:  $e_1(T) = x_T - \hat{x}_T$ .

Finalmente, la estimación de los coeficientes en el período  $T$ , es una combinación lineal de los calculados en el período previo, es decir:

$$\hat{a}(T) = L' \hat{a}(T-1) + h e_1(T) \quad (6.9)$$

a  $h$  se le llama vector de suavización.

El pronóstico para el período  $T+1$ , hecho al final del período  $T$ , es:

$$(6.10) \quad \hat{x}_{T+1} = \hat{a}(T) f(1) = \sum_{i=1}^n \hat{a}_i(T) f_i(1), \quad i=1, 2, \dots$$

Este procedimiento requiere de la estimación inicial de los coeficientes  $\hat{a}(0)$ . Éstos se pueden obtener subjetivamente o a través del análisis de datos históricos.

Ejemplo: Para ilustrar las ideas anteriores se analizará un modelo armónico de seis parámetros

Las variables independientes, los coeficientes en el período  $T$ , y el vector de suavización, son:

$$f(t) = \begin{bmatrix} t \\ \sin \frac{2\pi t}{12} \\ \cos \frac{2\pi t}{12} \\ \sin \frac{4\pi t}{12} \\ \cos \frac{4\pi t}{12} \end{bmatrix}, \quad a(T) = \begin{bmatrix} a_1(T) \\ a_2(T) \\ a_3(T) \\ a_4(T) \\ a_5(T) \end{bmatrix}, \quad h = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \end{bmatrix}$$

La matriz de transición es

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.866 & 0.5 & 0 & 0 \\ 0 & 0 & -0.5 & 0.866 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0.866 \\ 0 & 0 & 0 & 0 & -0.866 & 0.5 \end{bmatrix}$$

Para obtener las ecuaciones de suavización se sustituye a L en  $\hat{z}(T) = L'(\hat{\theta})(T-1) + h_0 C(T)$  ∴ se tiene:

$$\begin{aligned} \hat{\theta}_1(T) &= \hat{\theta}_1(T-1) + \hat{\theta}_2(T-1) + h_1 e_1(T) \\ \hat{\theta}_2(T) &= \hat{\theta}_2(T-1) + h_2 e_2(T) \\ \hat{\theta}_3(T) &= 0.866 \hat{\theta}_3(T-1) + 0.5 \hat{\theta}_4(T-1) + h_3 C(T) \\ \hat{\theta}_4(T) &= 0.5 \hat{\theta}_3(T-1) + 0.866 \hat{\theta}_4(T-1) + h_4 C(T) \\ \hat{\theta}_5(T) &= 0.5 \hat{\theta}_5(T-1) - 0.866 \hat{\theta}_6(T-1) + h_5 C(T) \\ \hat{\theta}_6(T) &= 0.866 \hat{\theta}_5(T-1) + 0.5 \hat{\theta}_6(T-1) + h_6 C(T) \end{aligned}$$

donde  $e_i(T) = z_T - \hat{z}_T = z_T - f'(1) \hat{\theta}(T-1)$   
y el pronóstico para un período  $T+1$ , al final del período  $T$  es:

$$\hat{z}_{T+1} = f'(1) \hat{\theta}(T)$$

Ejemplo: Demostrar los cálculos necesarios para obtener los elementos del vector de suavización para un modelo específico. Para cualquier modelo, el vector de suavización se encuentra de:

$$h = G^{-1} f(0)$$

Cálculo de la matriz G.

$$G = \lim_{T \rightarrow \infty} G(T) = \sum_{j=0}^{\infty} \beta^j f(-j) f'(-j)$$

$$\therefore G = \sum_{j=0}^{\infty} \beta^j f(-j) f'(-j)$$

La matriz simétrica cuyos elementos de la diagonal es la suma de los cuadrados de los pesos de las variables independientes.

El modelo armónico de seis parámetros es:

$$G = \begin{bmatrix} A & B(1) & B(2) \\ B(1) & c(1,1) & c(1,2) \\ B(2) & c(2,1) & c(2,2) \end{bmatrix}$$

donde las submatrices  $A$ ,  $B(u)$  y  $c(u,v)$  de orden  $2 \times 2$ , son:

$$A = \begin{bmatrix} \sum_{j=0}^{\infty} \beta^j & -\sum_{j=0}^{\infty} j \beta^j \\ -\sum_{j=0}^{\infty} j \beta^j & \sum_{j=0}^{\infty} j^2 \beta^j \end{bmatrix}$$

$$B(u) = \begin{bmatrix} -\sum_{j=0}^{\infty} \beta^j \sin v \omega_j & \sum_{j=0}^{\infty} \beta^j \cos v \omega_j \\ \sum_{j=0}^{\infty} j \beta^j \sin v \omega_j & -\sum_{j=0}^{\infty} j \beta^j \cos v \omega_j \end{bmatrix}$$

para  $v = 1, 2$

$$c(u,v) = \begin{bmatrix} \sum_{j=0}^{\infty} \beta^j \sin u \omega_j \sin v \omega_j & -\sum_{j=0}^{\infty} \beta^j \sin u \omega_j \cos v \omega_j \\ -\sum_{j=0}^{\infty} \beta^j \cos u \omega_j \sin v \omega_j & \sum_{j=0}^{\infty} \beta^j \cos u \omega_j \cos v \omega_j \end{bmatrix}$$

para  $u = 1, 2$

$v = u+1$ , donde  $\omega = 2\pi/12$ .

usando transformaciones Z, se encuentran los elementos de las submatrices A y B:

$$A = \begin{bmatrix} \frac{1}{1-\beta} & \frac{-\beta}{(1-\beta)^2} \\ \frac{-\beta}{(1-\beta)^2} & \frac{\beta(1+\beta)}{(1-\beta)^3} \end{bmatrix}$$

$$B(u, v) = \begin{bmatrix} \frac{-\beta \sin u \cos v}{1 - 2\beta \cos u \cos v + \beta^2} & \frac{1 - \beta \cos u \cos v}{1 - 2\beta \cos u \cos v + \beta^2} \\ \frac{\beta(1 - \beta^2) \sin u \sin v}{(1 - 2\beta \cos u \cos v + \beta^2)^2} & \frac{-2\beta^2 + \beta(1 + \beta^2) \cos u \cos v}{(1 - 2\beta \cos u \cos v + \beta^2)^2} \end{bmatrix}$$

Con las siguientes identidades trigonométricas:

$$\sin u \cos v = \frac{1}{2} [\cos(u-v) - \cos(u+v)]$$

$$\cos u \cos v = \frac{1}{2} [\cos(u+v) + \cos(u-v)]$$

$$\cos u \sin v = \frac{1}{2} [\sin(u+v) - \sin(u-v)]$$

$$\sin u \sin v = \frac{1}{2} [\sin(u+v) + \sin(u-v)]$$

$C(u, v)$  se expresa como:

$$C(u, v) = \frac{1}{2} \begin{bmatrix} -\left(\frac{c_1}{d_1} + \frac{c_2}{d_2}\right) & -\left(\frac{c_3}{d_1} + \frac{c_4}{d_2}\right) \\ -\left(\frac{c_3}{d_1} - \frac{c_4}{d_2}\right) & \left(\frac{c_1}{d_1} + \frac{c_2}{d_2}\right) \end{bmatrix}$$

donde

$$\begin{aligned} c_1 &= 1 - \beta \cos(u+v)w \\ c_2 &= 1 - \beta \cos(u-v)w \\ c_3 &= \beta \sin(u+v)w \\ c_4 &= \beta \sin(u-v)w \\ d_1 &= 1 - 2\beta \cos(u+v)w + \beta^2 \\ d_2 &= 1 - 2\beta \cos(u-v)w + \beta^2 \end{aligned}$$

Notese que los elementos de  $G$  se determinan completamente de la decisión

$$f(0) = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

Para  $\beta = 0.90$  y  $w = 2\pi/12$ , se tiene:

$$h = \begin{bmatrix} 0.17030 \\ 0.00953 \\ 0.05453 \\ 0.15545 \\ 0.07055 \\ 0.14285 \end{bmatrix}$$

ELECCION DE LA CONSTANTE DE SUAVIZACION.

En muchas aplicaciones de suavización exponencial es necesario especificar el valor de la constante o constante del modelo.

Se analizarán algunas bases para seleccionar a la constante, puesto que esta controla el número de realizaciones pasadas de los niveles de tiempo, que a su vez influyen en el pronóstico. En el valor de la constante es pequeño, se toma un gran número de observaciones pasadas, y el pronóstico reaccionará lentamente para cambios de los parámetros del modelo, y para grandes valores, se incluyen pocos datos históricos y reaccionará rápidamente con variaciones aleatorias aumentando más sensible.

Stowen analizó la respuesta característica de la suavización exponencial para varias señales estándares, tales como impulsos, saltos y funciones de paso.

Como regla general, para un modelo constante, la constante de suavización estará entre los valores de 0.01 a 0.3. La técnica más ampliamente usada es llevar una secuencia de pruebas de los datos históricos, dando diferentes valores a la constante, eligiendo el que optimice a la aplicación.

Por ejemplo, se se tienen datos de tres años, se toman los de los dos primeros para optimizar a la constante y se simula el pronóstico de cada mes y observar su comportamiento.

Si el resultado de las pruebas dan un valor de  $\alpha > 0.3$ , la validez del método será cuestionada.

En modelos polinomiales y trascendentes se define un valor de la constante, tal que los términos en igual sean los mismos. Es decir,

supóngase que se usa una  $\alpha$ , en una suavización constante, entonces su valor equivalente en un modelo de  $n$  términos es:

$$(1 - \alpha_n)^n = 1 - \alpha$$

Se debe elegir un valor de la constante para elegir. Para un punto en el futuro el intervalo es  $n = \infty$ . Los coeficientes cambian rápidamente cuando  $\alpha = 2.05$ .

\*. - TIEMPOS PONDERADOS DE PREDICTION.

En la mayoría de los ejemplos presentados hasta ahora, nos ha interesado el pronóstico para un simple período futuro  $T$  en un punto arbitrario. Muchas veces interesa pronosticar para varios períodos. Es decir, se desea pronosticar sobre algún tiempo ponderado o planeación horizontal. Por ejemplo, en problemas de producción y planeación de inventarios, se pronostica la demanda para 12 períodos.

Para ilustrar el proceso, supóngase que  $T$  es el período presente y se desea pronosticar la demanda de los  $w$  períodos siguientes, es decir,  $T+1, T+2, \dots, T+w$ , para  $x_1, x_2, \dots, x_{T+1}, \dots, x_{T+w}$ .

∴ el pronóstico de la demanda acumulada para el tiempo ponderado para el período  $w$  es:

$$\hat{X}_w(T) = \sum_{t=T+1}^{T+w} \hat{x}_t$$

De esta forma es muy simple pronosticar la demanda para cualquier período  $w$  de tiempo ponderado.

La variancia para cualquier punto futuro es:

$$V(\hat{x}_t) = f'(t) V f(t)$$

donde la matriz de variancia-covariancia está definida por:

$$V = E(\hat{\alpha} - \alpha)(\hat{\alpha} - \alpha)' = T^{-1} G^{-1} F W^2 (F W^2)' G^{-1}$$

Si las variables independientes incluyen un polinomio a  $n$  términos infinitos, entonces  $V(\hat{x}_t)$  depende de un punto futuro en el tiempo. Brown demostró que, en general, la variancia del pronóstico de un modelo lineal se incrementa como una función lineal de tiempo, y en un cuadrático se incrementa se incrementa como una función cuadrática de tiempo. Un modelo que no incluye polinomios infinitos tiene aproximadamente la misma variancia en cualquier punto; esto es muy importante, puesto que en esta forma se obtiene un rango de valores y por ende la probabilidad de los valores que puedan tomar futuras observaciones.

Por otro lado se tiene:

$$\text{error de pronóstico } e_t = x_t - \hat{x}_t$$

$$\text{y variancia } V(e_t) = \sigma_e^2 + f'(t) V f(t)$$

La variancia del pronóstico acumulativo es  $V\{\hat{X}_w(T)\}$

$$\text{Como: } \hat{X}_w(T) = \sum_{t=T+1}^{T+w} \hat{x}_t = \sum_{t=T+1}^{T+w} \hat{\alpha}'(T) f(t)$$

$$\text{se ve que: } V\{\hat{X}_w(T)\} = \left[ \sum_{t=T+1}^{T+w} f(t) \right]' V \left[ \sum_{t=T+1}^{T+w} f(t) \right]$$

y la variancia del error acumulativo para el período  $w$  de tiempo ponderado es:

$$w \sigma_e^2 + \left[ \sum_{t=T+1}^{T+w} f(t) \right]' V \left[ \sum_{t=T+1}^{T+w} f(t) \right]$$

\* - (1) AL DE MUESTRO Y ANALISIS DEL ERROR DE PRONOSTICO.

Matemáticamente se define al error de pronóstico como:

$$e_t = x_t - \hat{x}_t$$

Y como también existe la posibilidad de muchas características muy difíciles de observar al pasar los y sucesos de tiempo. Es necesario ser creativo para este análisis.

La variancia del error de pronóstico es:

$$V(e_t) = V(x_t) + V(\hat{x}_t)$$

$$o' \quad \sigma_e^2 = \sigma_x^2 + \sigma_{\hat{x}}^2$$

Suponiendo que  $x_t$  y  $\hat{x}_t$  son independientes, entonces para un modelo constante:

$$\sigma_{\hat{x}}^2 = \frac{\alpha}{2-\alpha} \sigma_e^2$$

$$\sigma_e^2 = \frac{2}{2-\alpha} \sigma_e^2$$

Además de la media, la variancia del error, es el tener conocimiento de la forma de su distribución de probabilidad. Gran parte de los errores están distribuidos normal y, por lo tanto, gaussianos.

La medida de control es un dispositivo para medir y controlar el error de pronóstico. Existen dos formas de señal de registro, una basada en la suma de los errores y el otro en el error de suavización.

Considerando las sumas para el periodo actual T:

$$Y(T) = \sum_{t=1}^T e_t$$

$$o' \quad Y(T) = Y(T-1) + e_T$$

Si existe un sesgo consistente en el pronóstico, los errores tendrán el mismo signo para varios periodos y la suma tendrá a crecer.

Para un modelo constante se tiene:

$$\sigma_y^2 = \frac{1}{1-\beta^2} \sigma_e^2$$

o' en general, para un modelo con  $m$  parámetros:

$$\sigma_y^2 \approx \frac{1}{1-\beta^{2m}} \sigma_e^2$$

Como  $\sigma_y^2$  no se conoce exactamente se estima.

En realidad, como  $\sigma_y^2 = [1/(1-\beta^{2m})] \sigma_e^2$ , basta en estimar a  $\sigma_e^2$ .

Suponiendo que el error está distribuido normalmente con media cero y variancia  $\sigma_e^2$ . Entonces la media de desviación absoluta  $\Delta$ , es

$$\Delta = E\{|e - E(e)|\} = 2 \int_0^{\infty} (e-0) \eta(e; 0, \sigma_e^2) de = \sqrt{\frac{2}{\pi}} \sigma_e \approx 0.8 \sigma_e$$

y al final del periodo T es:

$$\hat{\Delta}(T) = \alpha |e(T)| + (1-\alpha) \hat{\Delta}(T-1)$$

$$o' \quad \hat{\Delta}(T) = \frac{1}{N} \sum_{t=T-N+1}^T |e_t|$$

$$y \quad \sigma_{\hat{\Delta}}^2 = \left[ \frac{\hat{\Delta}(T)}{0.8} \right]^2$$

para una suavización simple:

$$\sigma_e^2 = \frac{2-\alpha}{2} \sigma_e^2$$

$$\sigma_e^2 = \left( \frac{2-\alpha}{2} \right) \left[ \frac{\hat{\Delta}(T)}{0.8} \right]^2$$

$$\therefore \sigma_y^2 = \left( \frac{1}{1-\beta^2} \right) \left( \frac{2-\alpha}{2} \right) \left[ \frac{\hat{\Delta}(T)}{0.8} \right]^2 = \frac{1}{2\alpha} \left[ \frac{\hat{\Delta}(T)}{0.8} \right]^2$$

Para una suavización exponencial simple la señal de registro, basada en la suma de los errores es:

$$\left| \frac{Y(T)}{\hat{\Delta}(T)} \right| < \frac{K}{0.8} \sqrt{\frac{1}{2\alpha}} = c_1$$

donde  $K^{-1}$  elige de las tablas de la función de distribución normal acumulativa estándar para proporcionar el nivel de aproximación de significancia. (Cuando el modelo no es constante se selecciona  $K^{-1}$  correspondiente a  $C_1$  cuyos valores, por lo común, están entre 4 y 6.

Para el período  $T$ , el error de suavización es:

$$\hat{z}(T) = \alpha e_T + (1-\alpha) \hat{z}(T-1)$$

la señal apropiada es:

$$\left| \frac{\hat{z}(T)}{\hat{\Delta}(T)} \right| < C_2$$

no debe exceder de 1.0

el valor de  $C_2$  está entre 0.2 y 0.4.

Si la señal de registro excede al límite de control en dos o tres observaciones, es símbolo de que algo está mal en el sistema. Y cuando la señal se sale de control, la cantidad  $Y(T)$  o  $Z(T)$  se restituye a cero para evitar una salida falsa en periodos futuros.

Ejemplo: Como ilustración, calcular el error acumulativo de la señal de registro, considerando los datos de la demanda, situados en la columna (1) de la tabla siguiente:

Calculos muestra para el error acumulativo de la señal de registro.

$t$	(1)	(2)	(3)	(4)	(5)	(6)
	$Y_t$	$\hat{Y}_t$	$e_t$	$Y(t)$	$Z(t)$	$Y(t)/Z(t)$
	1	1	0.00	1.00	1.00	1.00
Enero	104	100.10	3.90	4.90	2.10	1.67
Febrero	98	100.49	-2.49	2.41	2.54	1.03
Abril	110	100.24	9.76	12.17	2.90	4.20
Mayo	120	101.22	18.78	30.95	4.47	6.90
Junio	118	103.10	14.90	45.85	5.53	8.29

Suponiendo que en Enero se ha determinado algún punto exterior y que el modelo apropiado es constante y con  $\alpha = 0.1$ , y  $S_{Dec} = 100.00$ ,  $Y(Dec) = 0.00$ ,  $\hat{\Delta}(Dec) = 2.00$ .

El pronóstico para Enero es:  $\hat{X}_{Ene} = S_{Dec} = 100.00$ , y el de los meses siguientes se evolva en igual forma. (ecuaciones:  $S_T = \alpha X_T + (1-\alpha)S_{T-1}$  y  $S_T = \hat{Z}_{T-1}$ ) y se muestran en la columna 2; las (3) y (4) son el

error acumulativo para cada mes; la (5) es la desviación de la media absoluta de cada mes, (ecuación:  $\hat{\Delta}(T) = \alpha |e(T)| + (1-\alpha) \hat{\Delta}(T-1)$ ).

Por ejemplo:

$$\begin{aligned} \hat{\Delta}(Ene) &= \alpha |e_{Ene}| + (1-\alpha) \hat{\Delta}(Dec) = \\ &= 0.1 |100| + (0.9)(2.00) = 1.80 \end{aligned}$$

Finalmente, la señal de registro para cada mes se calcula dividiendo el error acumulativo entre la desviación de la media absoluta. Si se elige  $\pm 4$  como la crítica de detección de la señal, entonces la reacción de la primera señal fuera de control se genera en Abril seguida por la de mayo y junio.

El método para analizar el error acumulativo tiene alguna desventaja relativa para la interpretación del error de suavización. Si ocurre un gran error aleatorio en un periodo particular, se incrementarán el de suavización y el acumulativo, pero no tanto como para que genere una señal fuera de control. Ahora supóngase que varios periodos tienen un error promedio casi cero. El error acumulativo continúa en un valor grande, pero el de suavización decrece hacia cero. En algún punto, un segundo error aleatorio aparece con el mismo signo que el anterior. La suma acumulativa de la señal de registro generará incorrectamente una señal fuera de control, pero no así el de suavización. Como otro ejemplo, supóngase que la suma acumulativa de la señal de control fue un número que la crítica detectada al comienzo, entonces el pronóstico será correcto. Como el error es cero, la suma acumulativa no cambiará y la media de desviación absoluta será pequeña.



más importante que las observaciones individuales. Por ejemplo, en un inventario puede interesar más la probabilidad de una demanda de 140 a 150 unidades para el período siguiente. Es decir, interesa la probabilidad de pronóstico en lugar de futuras observaciones.

Si  $X_t$  es una variable aleatoria con función de distribución de probabilidad  $G$ , es decir,  $\Pr\{X_t = \theta\} = G(\theta)$ .

$G$  comúnmente es desconocida, pero se puede estimar y debe ser estacionaria o cambiar lentamente en el tiempo.

Un problema de pronóstico consiste en encontrar  $\hat{\theta}_p$ , una estimación de  $\theta_p$ , tal que  $G(\theta_p) = p$ , para  $0 < p < 1$ . Es decir, se desea encontrar un estimador de  $\theta_p$  tal que la probabilidad  $p$  de las observaciones futuras sea más pequeña que  $\theta_p$ .

Supóngase que las observaciones son medidas en una escala con  $n+1$  límites de clase, es decir:

$$X_0 < X_1 < X_2 < \dots < X_n$$

los cuales deben ser definidos, de modo que a cada observación se le asigne una y sólo una clase. Es decir, sólo existe una  $k$ , de modo que:  $X_{k-1} < X_t \leq X_k$

Los límites  $X_0$  y  $X_n$  son finitos y están entre 10 y 20 clases.  $\therefore$

Si  $p_k$  es la probabilidad de que la variable aleatoria  $X_t$  caiga en el intervalo de  $X_{k-1}$  a  $X_k$ , es decir,

$$p_k = \Pr\{X_{k-1} < X_t \leq X_k\}, \quad k=1, 2, \dots, n.$$

$$\therefore \sum_{k=1}^n p_k = 1.$$

Además como  $\sum_{j=1}^k p_j = \Pr\{X_t \leq X_k\} = G(X_k)$

Se puede estimar la función desconocida de distribución  $G$  calculando las  $n$  probabilidades  $p_k$ , las cuales se expresan como un vector columna  $\hat{p}$  de  $(n \times 1)$ , por tanto:

en el tiempo  $t$  se denotan por:

$$\hat{P}(t) = \begin{bmatrix} \hat{p}_1(t) \\ \hat{p}_2(t) \\ \vdots \\ \hat{p}_n(t) \end{bmatrix}$$

La estimación de  $G(X_k)$  es  $\hat{G}(X_k) = \sum_{j=1}^k \hat{p}_j(t)$

Si la  $t$ -ésima observación  $X_t$  está asociada con el  $k$ -ésimo intervalo de clase, es decir,  $X_{k-1} < X_t \leq X_k$ .

Definiendo un vector columna  $u_t$  de  $(n \times 1)$  que tiene  $n$  ceros y un uno como la componente  $k$ -ésima.

Revisando la estimación de los períodos anteriores  $\hat{p}(t-1)$  de la información presente de acuerdo a:

$$\hat{p}(t) = \alpha u_t + (1-\alpha) \hat{p}(t-1)$$

Se demuestra fácilmente que  $\hat{p}(t)$  es insesgado [ $E\{\hat{p}(t)\} = p$ ] y la variancia de la  $k$ -ésima probabilidad es:

$$\sigma_{\hat{p}_k}^2 = \frac{\alpha}{2-\alpha} p_k (1-p_k)$$

La estimación inicial de la probabilidad  $\hat{p}(0)$  puede ser subjetiva o obtenida a través del análisis de datos históricos.

Considérese el siguiente problema de pronóstico. Supóngase que se tienen las  $n$  probabilidades  $\hat{p}(t)$  y se da  $p$ .

Se desea estimar un valor de  $\theta_p$  tal que la probabilidad de un observación futura más pequeña que  $\theta_p$  es  $p$ , es decir,  $G(\theta_p) = p$ .

Si  $p$  es tal que uno de los límites de clase satisface exactamente a  $p = \hat{G}(X_k)$ , la solución se encontrará estimando a  $\hat{\theta}_p$  con  $\hat{\theta}_p = X_k$ .

Si:  $\hat{G}(X_{k-1}) < p < \hat{G}(X_k)$ .

entonces  $\theta_p$  se estima por medio de una interpolación lineal  $\therefore$

Ejemplo: Supóngase que en el tiempo  $t-1$ , los datos de la tabla siguiente son la demanda diaria de una componente eléctrica. La observación en el período  $t$  es  $X_t = 34$ , siendo el vector  $u_t = [0, 0, 0, 1, 0]$ , con  $\alpha = 0.1$ .

DATOS DE LA DEMANDA			
k	límite de clase $Y_k$	Probabilidad $\hat{p}_k(t-1)$	$G(Y_k)$
0	0		0.00
1	10	0.60	0.60
2	20	0.15	0.75
3	30	0.15	0.90
4	40	0.05	0.95
5	50	0.05	1.00

Actualizando las probabilidades de acuerdo a la ecuación:

$$\hat{p}(t) = \alpha u_k + (1-\alpha) \hat{p}(t-1)$$

$$\hat{p}(t) = 0.1 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + 0.9 \begin{bmatrix} 0.60 \\ 0.15 \\ 0.15 \\ 0.05 \\ 0.05 \end{bmatrix} = \begin{bmatrix} 0.540 \\ 0.135 \\ 0.135 \\ 0.145 \\ 0.045 \end{bmatrix}$$

La nueva función de distribución estimada es:

$$\begin{aligned} \bar{G}(0) &= 0.000 & \bar{G}(30) &= 0.910 \\ \bar{G}(10) &= 0.540 & \bar{G}(40) &= 0.955 \\ \bar{G}(20) &= 0.675 & \bar{G}(50) &= 1.00 \end{aligned}$$

Supóngase que se quiere encontrar  $\theta_p$ , tal que la probabilidad de demanda menor que  $\theta_p$  es 0.9. Este valor es:

$$\hat{\theta}_{0.9} = \frac{[0.955 - 0.9]30 + [0.9 - 0.810]40}{0.955 - 0.810} = \underline{\underline{36.2}}$$

El procedimiento de control más automático. Se han de hallar un gran número de diseños para monitorizar y modificar automáticamente el valor de la constante en la suavización exponencial. Estas técnicas se llaman modelos de control adaptable, debido a que la constante se adapta por sí misma a cambiar en las series de tiempo.

(How describió un procedimiento de control adaptable de una sola constante de suavización exponencial. Este método tiene tres valores espaciados o distancias iguales para la constante: un valor superior ( $\alpha_1$ ), un valor medio ( $\alpha_0$ ) y un valor inferior ( $\alpha_2$ ).

$$\begin{aligned} \text{Así: } \alpha_1 &= \alpha_0 + \delta \\ \alpha_2 &= \alpha_0 - \delta \end{aligned} \quad (8.1)$$

donde  $\delta$  es una constante elegida arbitrariamente, con valor aproximado de 0.05.

Para cada período se calculan tres pronósticos, uno con cada valor de la constante. La desviación de la media absoluta se denota por  $\Delta(\alpha_0)$ ,  $\Delta(\alpha_1)$ ,  $\Delta(\alpha_2)$ , y se calcula para cada período. La regla para cambiar el valor de la constante es, si  $\Delta(\alpha_0)$  es menor que  $\Delta(\alpha_1)$  y  $\Delta(\alpha_2)$  no se hace cambio.

Si  $\Delta(\alpha_1) < \Delta(\alpha_0)$ , entonces se hace  $\alpha_0 = \alpha_1$ .

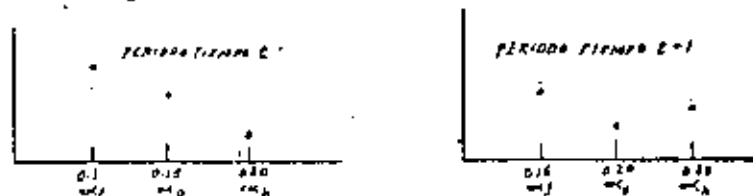
Los nuevos valores superior e inferior se eligen de acuerdo a la ecuación (8.1) usando la nueva  $\alpha_0$ .

Si  $\Delta(\alpha_2) < \Delta(\alpha_0)$ , se hace  $\alpha_0 = \alpha_2$ , y se revisan los valores inferior y superior.

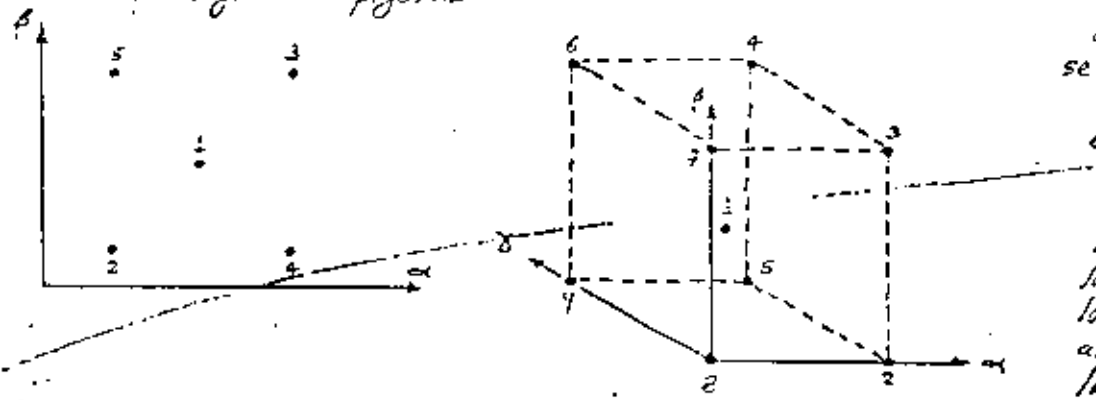
Cuando  $\Delta(\alpha_1)$  y  $\Delta(\alpha_2)$  son menores que  $\Delta(\alpha_0)$ , se ajusta  $\alpha_0$  a la dirección de la media de desviación absoluta menor.

En cada revisión la desviación media absoluta se restaura en cero y el proceso converge de nuevo.

El procedimiento de Chow, suavización exponencial de control adaptable, para el caso de  $\Delta(\alpha_2) < \Delta(\alpha_0)$  en el tiempo  $t$ , se ilustra en las siguientes figuras:



... para el control de la constante de suavización exponencial. Su método es esencialmente una extensión del de Chow para el caso de varias constantes. El procedimiento de R. y R. usa un diseño factorial de dos niveles en el cual cada constante está situada en un nivel superior e inferior. Todas las posibles combinaciones de este nivel, así como un punto central, están incluidas. El diseño para modelos de dos y tres parámetros se muestran en las siguientes figuras:



La repetición completa para cada punto de diseño se llama ciclo. Si los  $k$  parámetros de suavización están bajo control, entonces se calculan  $2^k + 1$  predicciones para cada período. La medida de efectividad del pronóstico es el cuadrado de su error. Si  $ES_{ij}$  es el cuadrado del error en el punto de diseño del ciclo  $j$ . Entonces después de  $n$  ciclos, su promedio es:

$$\overline{ES}_k = \frac{1}{n} \sum_{j=1}^n ES_{kj}$$

El procedimiento de control consiste en el cálculo de los "efectos" de cada parámetro de suavización en la misma forma que los efectos principales en un análisis de variancia, se estiman y se ajustan aquellos.

$$\bar{E}_\alpha = \frac{1}{2} (\overline{SS}_1 + \overline{SS}_4 - \overline{SS}_2 - \overline{SS}_5)$$

$$\bar{E}_\beta = \frac{1}{2} (\overline{SS}_3 + \overline{SS}_5 - \overline{SS}_2 - \overline{SS}_4)$$

Los parámetros se cambian si  $\bar{E}_\alpha, \bar{E}_\beta$  están fuera de la aproximación del 99% del error.

$$\pm 3 \sqrt{\frac{1}{n}} S$$

donde  $S$  es la estimación de la desviación estándar de los errores, y se calcula por el método de rango.

Como ilustración de la técnica, supóngase que  $\bar{E}_\alpha$  está bajo el límite inferior de error, para un ciclo particular, es decir,

$$\bar{E}_\alpha < -3 \sqrt{\frac{1}{n}} S.$$

lo que indica que  $\overline{SS}_2 + \overline{SS}_5$  es más grande  $\overline{SS}_3 + \overline{SS}_4$ , por lo tanto el diseño debe desplazarse hacia la derecha, tomando el valor superior de  $\alpha$  como el nuevo central y dejando los de  $\beta$  sin alterar. Se descartan los antiguos de  $\overline{SS}_2$  y los nuevos se calculan sobre varios ciclos:

Para el caso de tres parámetros los efectos son:

$$\bar{E}_\alpha = \frac{1}{4} (\overline{SS}_1 + \overline{SS}_3 + \overline{SS}_4 + \overline{SS}_5 - \overline{SS}_6 - \overline{SS}_7 - \overline{SS}_8 - \overline{SS}_9)$$

$$\bar{E}_\beta = \frac{1}{4} (\overline{SS}_3 + \overline{SS}_4 + \overline{SS}_6 + \overline{SS}_7 - \overline{SS}_2 - \overline{SS}_5 - \overline{SS}_8 - \overline{SS}_9)$$

$$\bar{E}_\gamma = \frac{1}{4} (\overline{SS}_4 + \overline{SS}_5 + \overline{SS}_6 + \overline{SS}_7 - \overline{SS}_1 - \overline{SS}_2 - \overline{SS}_3 - \overline{SS}_8)$$

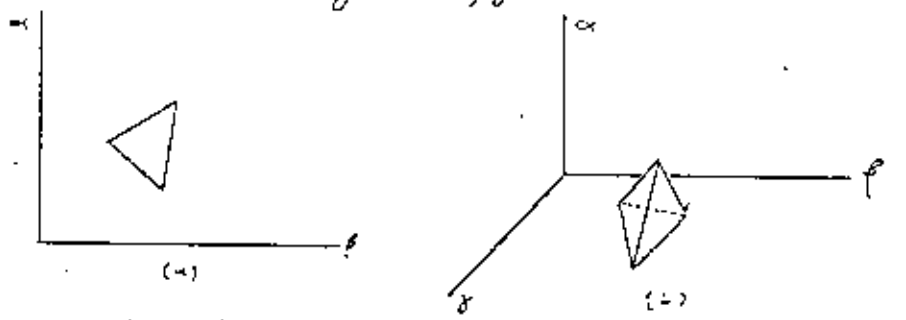
y la aproximación del 99% en los límites de control son:

$$\pm 3 \sqrt{\frac{1}{n}} S$$

La técnica de control para 3 parámetros es similar a la anterior.

Otra técnica de control adaptable es la de Montgomery. Su esquema está basado en el simplex el cual, como el factorial de dos niveles, es un diseño experimental ortogonal de primer orden.

Si  $k$  son los parámetros controlados y  $N$  las observaciones tomadas en los  $N$  vértices del Simplex, el cual para dos parámetros ( $k=2$ ) es un triángulo equilátero, para tres es un tetraedro. Estos diseños se muestran en las siguientes figuras:



La matriz de diseño  $D$  para un Simplex de orientación arbitraria se construye de  $N \times k$ , donde  $H$  es una matriz ortogonal de  $(N \times N)$ ,  $d_i$  son las filas de  $D$ .

La ventaja de este diseño es que en cada período sólo  $k$  pronósticos se calculan.

Para aplicar esta técnica a un modelo exponencial de dos o tres parámetros, se aplican las siguientes reglas:

1.- Se denota por  $E_i$  al error absoluto del pronóstico presente en el punto de diseño  $i$ th, es decir,

$$E_i = |Z_i - X_i|, \quad i = 1, 2, \dots, N$$

suponiendo que el valor máximo de  $E_i$  es en el punto  $d_j$ .

Se forma un nuevo Simplex suprimiendo a  $d_j$  de  $D$ , y se substituye el nuevo punto de diseño:

$$d_j^* = 2^{k-1} (d_1 + d_2 + \dots + d_{j-1} + d_{j+1} + \dots + d_N) - d_j$$

calcular el pronóstico para el período siguiente usando los parámetros de suavización, los cuales son elementos de  $d_j^*$

2.- Aplicar la regla 1 excepto en el punto de diseño que se ha eliminado en  $N$  sucesivos Simplexes sin ser eliminado. Esta situación surgiría para el punto  $i$ th de diseño, desvirtuando  $E_i$

y calcular el pronóstico para el período siguiente usando los parámetros de suavización en  $d_i$ . Entonces aplicar la regla

3.-  $E_i$  será el error máximo absoluto en el  $n$ -ésimo Simplex y  $E_i^*$  es el error máximo absoluto en el  $n+1$  Simplex, no regresar al  $n$ th diseño. En su lugar de oscilación mover los  $n+1$  diseños por medio de descartar al segundo error absoluto más grande.

Computacionalmente, este método es más eficiente que el de Robert y Reed, pues únicamente  $k$  pronósticos se requieren para cada período (el de R. y R. requiere de  $2^k + 1$ ).

## 8.- TÉCNICAS DE BAYES PARA PRONÓSTICO.

Los usos de Bayes que a continuación se presentan son útiles cuando no se tiene información histórica disponible en el momento de iniciar el pronóstico. Esta situación se presenta con mucha frecuencia en la práctica, por lo que al efectuar un pronóstico inicial este se realiza en bases subjetivas exclusivamente. Sin embargo, a medida que se va obteniendo información histórica de la serie de tiempo, se deberán modificar nuestros estimadores subjetivos.

El método de Bayes, proporciona un criterio para ir modificando nuestros estimadores subjetivos en términos de la información que se va obteniendo. La técnica de Bayes en pronósticos consiste en aplicar el criterio de la teoría de decisiones de Bayes, la cual a su vez, está basada en el teorema de Bayes. Se presentará primero este teorema, y después su aplicación a pronósticos para el caso de una serie en el tiempo que siga un modelo constante y de tendencia lineal.

### \* CRITERIO DE BAYES EN ESTADÍSTICA

Cuando se tiene una variable aleatoria  $X$ , que tiene una función de densidad  $f$ , la cual está caracterizada por un parámetro  $\theta$ , entonces se usa la notación  $f(X; \theta)$  o  $f(X|\theta)$ , para hacer incógnita en que la función de densidad  $f$ , depende del parámetro  $\theta$ . Por ejemplo, cuando  $f$  es una densidad de probabilidad binomial, entonces la notación  $f(X; \theta)$  significa:

$$f(X; \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \quad x=0,1,2,\dots,n$$

Si  $f$  es una distribución normal, entonces la notación  $f(X; \theta_1, \theta_2)$  significa:

$$f(X; \theta_1, \theta_2) = \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{1}{2}\left(\frac{X-\theta_1}{\theta_2}\right)^2}$$

En este ejemplo el parámetro  $\theta_1$  representa la media y  $\theta_2$  la variancia. Si una variable aleatoria  $X$  tiene una función de densidad normal con  $\theta_1$  y variancia  $\theta_2$ , se dice que  $X$  es  $N(\theta_1, \theta_2)$  o también se escribe:

$$X \sim N(\theta_1, \theta_2).$$

OBSERVACION: A una función de densidad de probabilidad también se le llama función de densidad.

CRITERIO DE BAYES: Sea  $X$  una variable aleatoria con función de densidad de probabilidad  $f(X; \theta)$ , donde el parámetro  $\theta$  es desconocido. El criterio de Bayes en teoría de decisiones consiste en considerar al parámetro desconocido  $\theta$  como una variable aleatoria. Esta consideración es la diferencia importante entre el criterio de Bayes y el criterio clásico en teoría de decisiones, ya que el criterio clásico considera que el parámetro  $\theta$  desconocido, es una constante y no una variable aleatoria.

### TEOREMA DE BAYES.

Sea  $X$  una variable aleatoria con función de densidad de probabilidad  $f(X; \theta)$  donde el parámetro desconocido  $\theta$  es considerado como una variable aleatoria. Sea  $f(\theta)$  la función de densidad de probabilidad de la variable  $\theta$ . A  $f(\theta)$  se le llama la función de densidad de probabilidad a priori de la variable aleatoria  $\theta$ .

Sean  $X_1, X_2, \dots, X_n$  observaciones de  $X$ , y sea  $f(X_1, X_2, \dots, X_n | \theta)$  la función de densidad de probabilidad condicional de  $X_1, X_2, \dots, X_n$ , dado  $\theta$ . La función de densidad de  $\theta$  dadas las observaciones  $X_1, X_2, \dots, X_n$ , está dada por:

$$f(\theta | X_1, X_2, \dots, X_n) = \frac{f(X_1, X_2, \dots, X_n | \theta) f(\theta)}{\int_0 f(X_1, X_2, \dots, X_n | \theta) f(\theta) d\theta} \quad (8.1)$$

A  $f(\theta | X_1, X_2, \dots, X_n)$  se le llama la función de densidad a posteriori de  $\theta$ . A la fórmula (8.1) se le llama Teorema de Bayes.

### NOTAS.

- 1- Para utilizar la fórmula (8.1) generalmente se efectúan los siguientes pasos:
  - a) Subjetivamente se establecen los posibles valores que puede tomar el parámetro  $\theta$ , junto con la probabilidad de que tome los valores supuestos de  $\theta$ . La función de probabilidad de estos valores también se da subjetivamente, en base a la experiencia, o de situaciones similares encontradas en otros

prol ya resueltos.

- ii) Se obtienen los valores históricos (o experimentales),  $x_1, \dots, x_n$ .
- iii) Se calcula la condicional  $f(x_1, \dots, x_n | \theta)$ , para cada uno de los valores supuestos de  $\theta$ .
- iv) Se asignan nuevas probabilidades a los valores de  $\theta$ , usando la ecuación (8.1), en base de la información proporcionada por las observaciones históricas  $x_1, \dots, x_n$ .
- v) Con las nuevas probabilidades  $f(\theta | x_1, \dots, x_n)$  para los posibles valores de  $\theta$ , es posible hacer estimaciones óptimas de  $\theta$  sobre  $\theta$ , como se verá en el teorema que aparece después.

2: En muchas ocasiones, en lugar de considerar directamente las observaciones  $x_1, \dots, x_n$ , se calcula una función  $y = u(x_1, \dots, x_n)$  de las observaciones  $x_1, \dots, x_n$ , para concentrar toda la información que proporcionan  $x_1, \dots, x_n$  en un solo valor  $y = u(x_1, \dots, x_n)$ , y en base a esta información concentrada se estiman las nuevas probabilidades de  $\theta$ , i.e. se calcula:

$$f(\theta | y) = \frac{f(y | \theta) f(\theta)}{\int_0 f(y | \theta) f(\theta) d\theta} \quad (8.2)$$

Para obtener esta probabilidad a posteriori, primero es necesario encontrar la condicional de  $f(y | \theta)$  la cual depende de la función  $f(x_1, \dots, x_n | \theta)$ , que a su vez depende de  $f(x)$ . En general, es fácil encontrar la función  $f(y | \theta)$  en términos de  $f(x)$ , aunque depende del tipo de la función  $y = u(x_1, \dots, x_n)$  que se haya elegido. Por ejemplo, si:

$$y = u(x_1, \dots, x_n) = \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

se sabe en estadística que  $y = \bar{x}$  tiene una distribución igual a la distribución  $x$ , con la misma media, pero con variancia igual a la variancia de  $x$  sobre  $n$ . En el caso de que  $x$  sea  $N(\mu, \sigma^2)$ , entonces  $y = \bar{x}$  es:

$$N\left(\mu, \frac{\sigma^2}{n}\right).$$

### CRITERIO DE BAYES EN TEORIA DE DECISIONES.

TEOREMA.- Sea  $f(x; \theta)$  una función de densidad de probabilidad, donde  $\theta$  es un parámetro desconocido considerado como una variable aleatoria. Sean  $x_1, x_2, \dots, x_n$  observaciones de  $X$ , y si  $y = u(x_1, \dots, x_n)$  una función de  $x_1, x_2, \dots, x_n$ . El estimador Bayesiano de  $\theta$ , indicado por  $\theta^*$ , usando el criterio del menor error cuadrado medio, está dado por:

$$\theta^* = \int \theta f(\theta | y) dy \quad (8.3)$$

Interpretación:

El método de Bayes asociado al criterio del menor error cuadrado medio, consiste en elegir un estimador  $\theta^*$ , que minimize el valor esperado condicional de  $(\theta - \theta^*)^2$ , dadas las observaciones  $x_1, x_2, \dots, x_n$ .

Observación.

En probabilidad se sabe que  $\int \theta f(\theta | y) dy$  es por definición la media condicional de  $\theta$  dado  $y$ . Por lo tanto, observando la ecuación (8.3) se dice que el estimador Bayesiano  $\theta^*$  de  $\theta$ , basado en el criterio del menor error cuadrado medio, es la media condicional de  $\theta$  dado  $y$ , o también se dice que  $\theta^*$  es la media a posteriori de  $\theta$ , o la media de la distribución a posteriori  $f(\theta | y)$ .

### METODO DE COHEN.

TEOREMA. Sea  $\{X_t : t \in A\}$ ,  $A = \{1, 2, \dots, n\}$ , una serie en el tiempo, y considerando que se puede representar por el modelo constante:

$$X_t = a + E_t \quad (8.4)$$

donde  $a$  es un parámetro desconocido que representa la media del proceso, y  $E_t$  es el error aleatorio con distribución  $N(0, \sigma^2)$ , donde  $\sigma^2$  es conocido. Suponga que el parámetro desconocido  $a$ , es una variable aleatoria; cuya distribución a priori es  $N(a_0, \tau^2)$ , i.e. nuestra consideración subjetiva es que  $a$  es una variable aleatoria  $N(a_0, \tau^2)$ . Sea  $y = \bar{x}$ , la función que concentra toda la información histórica  $x_1, x_2, \dots, x_n$ .

$$\hat{\alpha}^* = \frac{\bar{x} \sqrt{v_0^2} + u(\sqrt{v_e^2}/n)}{\sqrt{v_0^2} + (\sqrt{v_e^2}/n)} = \hat{\alpha}^*(n) \quad \therefore \quad \hat{\alpha}^*(n) = \frac{\bar{x} \sqrt{v_0^2} + \hat{\alpha}_0(\sqrt{v_e^2}/n)}{\sqrt{v_0^2} + (\sqrt{v_e^2}/n)} \quad (8.4)$$

$$\hat{\alpha}^* = \frac{n}{u+n} \bar{x} + \frac{u}{u+n} \hat{\alpha}_0 \quad \therefore \quad \hat{\alpha}^*(n) = \frac{n}{u+n} \bar{x} + \frac{u}{u+n} \hat{\alpha}_0 \quad (8.5)$$

donde  $u = \sqrt{v_e^2}/\sqrt{v_0^2}$ , y  $\hat{\alpha}^*$  se puede expresar como

$$\hat{\alpha}^* = \alpha x_n + (1-\alpha) \hat{\alpha}^*(n-1) \quad (8.6)$$

forma no diferente de la suavización exponencial.  
donde  $\alpha = 1/(u+n)$

ii) El pronóstico para cualquier período  $n+1$ , es

$$\hat{x}_{n+1} = \hat{\alpha}^*(n) \quad (8.7)$$

NOTA: La Técnica de Bayes deberá dejarse de usar cuando en algún tiempo ya se disponga de datos suficientes, y en ese tiempo deberá adoptarse otra técnica de pronóstico, quizás suavización exponencial u otra.

Ejemplo: Suponga que deseamos pronosticar la demanda para un producto nuevo. Sospechamos que la demanda está normalmente distribuida y que un modelo constante es apropiado, pero no se dispone de información histórica. Se supone que una razonable densidad a priori para  $\hat{\alpha}$  es  $N(50, 4)$ , y que  $\sqrt{v_e^2} = 7$ . Para el período 1, el pronóstico es:

$$\hat{x}_1 = 50$$

Se supone que la demanda real en el período 1 es  $x_1 = 56$ . Ahora el estimador de Bayes de  $\hat{\alpha}$  es:

$$\hat{\alpha}^*(1) = \frac{1}{(7/4)+1} (56) + \frac{(9/4)}{(7/4)+1} (50) \approx 52.$$

Por tanto, el pronóstico para el período 2 es:

$$\hat{x}_2 = 52$$

$$\hat{\alpha}^*(2) = \frac{1}{(7/4)+2} (58) + \frac{(13/4)}{(7/4)+2} (52) \approx 53$$

$$\text{y } \hat{x}_3 = 53$$

Este procedimiento se continúa hasta que haya suficientes datos acumulados disponibles, para que en ese momento se desarrolle un sistema de pronóstico permanente.

TEOREMA.

Considerando un modelo lineal de la forma  $x_t = \hat{\alpha} + bt + E_t$  el cual se transcribe como

$$x_t = \hat{\alpha} + b(t - \bar{t}) + E_t$$

$$\text{donde } \hat{\alpha} = \hat{\alpha} + bt \quad \text{y} \quad \bar{t} = \sum_{t=1}^n t/n = n(n+1)/2$$

Suponiendo que las densidades a priori de  $b$  y  $\hat{\alpha}$  son normales, i.e.  $b \sim N(b_0, \sqrt{v_b^2})$  y  $\hat{\alpha} \sim N(\hat{\alpha}_0, \sqrt{v_{\hat{\alpha}}^2})$ , y además que  $E_t \sim N(0, \sqrt{v_e^2})$ .

Suponiendo que  $\sqrt{v_e^2}$  es conocida.

$$\text{Sean } \hat{\alpha} = \sum_{t=1}^n x_t/n \quad \text{y} \quad \hat{b} = \sum_{t=1}^n x_t(t - \bar{t})/SS_{tt}$$

estimadores de  $\hat{\alpha}$  y  $b$ , donde  $SS_{tt} = \sum_{t=1}^n (t - \bar{t})^2$ .

Los estimadores de Bayes son:

$$b^* = \frac{w}{w + \sqrt{v_e^2}} \hat{b} + \frac{\sqrt{v_e^2}}{w + \sqrt{v_e^2}} b_0 \quad (8.8)$$

$$\text{donde } w = SS_{tt} \sqrt{v_e^2}$$

$$(\hat{\alpha})^* = \frac{z}{z + \sqrt{v_e^2}} \hat{\alpha} + \frac{\sqrt{v_e^2}}{z + \sqrt{v_e^2}} \hat{\alpha}_0 \quad (8.9)$$

$$\text{donde } z = n \sqrt{v_e^2}$$

- 1- Los coeficientes  $\hat{a}$  y  $\hat{b}$  que resultan de la regresión de  $Z_t, \dots, Z_n$ , corresponden a los estimadores obtenidos por mínimos cuadrados.
- 2- Observando (8.8) y (8.9) se nota que los estimadores de Bayes ( $\hat{a}$ ) y  $\hat{b}$  son promedios ponderados (o con prioridades) de los estimadores de mínimos cuadrados y de las medias de las a priori.
- 3- Las fórmulas (8.8) y (8.9) se deberían usar para que sucesivamente se vayan combinando los estimadores subjetivos con datos observados hasta que se tenga suficiente experiencia para desarrollar un sistema de pronóstico permanente.

La combinación de pronósticos de esta última naturaleza en la suavización exponencial la cual se deriva del método de mínimos cuadrados. En muchas series de tiempo, las observaciones sucesivas son altamente dependientes. Si este es el caso, entonces los modelos anteriormente presentados son inadecuados. Esto no quiere decir que los métodos de suavización exponencial no funcionen para este tipo de observaciones. Se puede trabajar bastante bien con ellos. Sin embargo, los métodos de pronóstico explotan esta dependencia y producen buenos resultados. Se verán algunos de estos modelos.

A muchos de estos modelos alternativos se les puede diseñar como una combinación lineal de variables aleatorias independientes, es decir,  $\epsilon_1, \epsilon_2, \dots, \epsilon_t, \dots$ , a esta sucesión se le llama un proceso "white noise" y es extraída de una distribución estable, supuestamente normal, con media cero y variancia  $\sigma^2$ . La combinación lineal de las  $\{\epsilon_t\}$  es:

$$X_t = \mu + \psi_0 \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \dots \quad (9.1)$$

donde  $\psi_i (i=0, 1, \dots)$  son los pesos y  $\mu$  determina el nivel del proceso. A esta ecuación se le llama "linear filter". Las observaciones sucesivas  $X_t$  son dependientes, puesto que son generadas de las  $\{\epsilon_t\}$ . Desde el punto de vista de (9.1) a un modelo de series de tiempo se le define como una función que transforma las series de tiempo en un proceso de "white noise".

Box y Jenkins han unificado esta metodología y desarrollado toda una filosofía; por esta razón, los modelos de series de tiempo (9.1) se les llama "Box-Jenkins".

UNA CLASE DE MODELOS DE SERIES DE TIEMPO.

El linear filter no es muy práctico por el número infinito de parámetros desconocidos que contiene (los pesos  $\psi$ ).

Al operador  $B$  de desplazamiento hacia atrás se le define como:

$$B\epsilon_t = \epsilon_{t-1}$$

$$\text{y } B^j \epsilon_t = \epsilon_{t-j}$$



Una forma alternativa de la ecuación (9.1) en términos del operador B

$$x_t - \mu = \sum_{j=0}^{\infty} \psi_j (x_{t-j} - \mu) + \epsilon_t$$

$$\text{ó } x_t - \mu = Y(B) \epsilon_t$$

$$\text{donde } Y(B) = \psi_0 B^0 + \psi_1 B^1 + \psi_2 B^2 + \dots$$

Los modelos derivados de (9.1) pueden representar series de tiempo estacionarias y no estacionarias. Por estacionarias, se quiere decir que las series de tiempo oscilan aleatoriamente alrededor de una media constante, y por no estacionarias las que no tienen una media normal. Generalmente, si la secuencia de los pesos  $\psi_0, \psi_1, \dots$  es finita o infinita y convergente, la serie es estacionaria con media  $\mu$ . Si  $\psi_0, \psi_1, \dots$  es infinita y divergente, la serie es no estacionaria y  $\mu$  es únicamente un punto de referencia para el nivel de la serie.

MODELOS AUTORREGRESIVOS.

Es conveniente trabajar las series de tiempo en términos de la desviación  $\mu$ . Por tanto, suponiendo que  $\tilde{x}_t = x_t - \mu$ , para toda  $t$ . Un caso especial importante es el modelo:

$$\tilde{x}_t = \phi_1 \tilde{x}_{t-1} + \phi_2 \tilde{x}_{t-2} + \dots + \phi_p \tilde{x}_{t-p} + \epsilon_t \quad (9.2)$$

y se le llama proceso autorregresivo (AR) y contiene p parámetros desconocidos  $\phi_1, \phi_2, \dots, \phi_p$ .

Dos importantes procesos autorregresivos son:

$$\text{AR}(1): \tilde{x}_t = \phi_1 \tilde{x}_{t-1} + \epsilon_t \quad \text{y}$$

$$\text{AR}(2): \tilde{x}_t = \phi_1 \tilde{x}_{t-1} + \phi_2 \tilde{x}_{t-2} + \epsilon_t$$

La ecuación (9.2), en términos del operador B, es:

$$\begin{aligned} \tilde{x}_t &= (\phi_1 B + \phi_2 B^2 + \dots + \phi_p B^p) \tilde{x}_t + \epsilon_t \\ \text{ó } (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) \tilde{x}_t &= \epsilon_t \end{aligned} \quad (9.3)$$

$$\text{si } \phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

$$\phi_p(B) \tilde{x}_t = \epsilon_t \quad (9.4)$$

El modelo autorregresivo se utiliza para representar series de

tiempo estacionarias y no estacionarias. Si las raíces del polinomio  $\phi_p(B) = 0$  están fuera del círculo unitario, el proceso es estacionario. Esto equivale a decir que, por ejemplo, en el AR(1) resulta tener:

$$|\phi_1| < 1$$

y en el AR(2):  $\phi_1 + \phi_2 < 1$

$$\phi_2 - \phi_1 < 1$$

$$|\phi_2| < 1$$

MODELOS DE PROMEDIO MOVIL.

Considerando el caso especial de hacer filter (9.1) con sólo el primer peso  $\neq$  no cero, se tiene:

$$\begin{aligned} \tilde{x}_t &= \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} \\ &= (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \epsilon_t \\ &= \Theta_q(B) \epsilon_t \end{aligned} \quad (9.5)$$

donde  $-\theta_1, -\theta_2, \dots, -\theta_q$  son el conjunto finito de pesos de (9.1). Al modelo (9.5) se le llama "proceso del promedio móvil" de orden q: MA(q).

Los dos casos más útiles e importantes son:

$$\text{MA}(1): \tilde{x}_t = \epsilon_t - \theta_1 \epsilon_{t-1}$$

$$\text{MA}(2): \tilde{x}_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2}$$

Existe una interesante dualidad entre el promedio móvil y el proceso autorregresivo. Por ejemplo, el proceso MA(1):

$$\tilde{x}_t = \epsilon_t - \theta_1 \epsilon_{t-1} = (1 - \theta_1 B) \epsilon_t$$

cuya solución es:  $\epsilon_t = (1 - \theta_1 B)^{-1} \tilde{x}_t$

$$\text{si } |\theta_1| < 1 \therefore \epsilon_t = \left( \sum_{j=0}^{\infty} \theta_1^j B^j \right) \tilde{x}_t = (1 + \theta_1 B + \theta_1^2 B^2 + \dots) \tilde{x}_t$$

el cual es un proceso infinito autorregresivo con pesos  $\phi_j = \theta_1^j$ . Es decir, se ha invertido el proceso MA(1) para obtener el AR( $\infty$ ).

A  $|\phi_1| < 1$ , se le llama condición de inversión para el proceso MA(1). En general, para efectuar la inversión del modelo MA(q)

al proceso AR(∞) se requiere que las raíces de  $\Theta_p(B)$  estén fuera del círculo unitario. Sin embargo, el proceso del promedio móvil es estacionario cualesquiera que sean los pesos elegidos. Esta dualidad es verdadera para procesos autorregresivos, es decir, el proceso finito AR(p) se puede invertir para obtener uno de promedio móvil de orden infinito.

MODELOS MIXTOS DE PROMEDIO MÓVIL Y AUTORREGRESIVOS.

La construcción de un modelo empírico de series de tiempo puede conducir a un modelo muy lento, si se incluyen parámetros del promedio móvil y autorregresivo, es decir, modelos mixtos de orden (p, q):

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} + \epsilon_t$$

ó  $\Phi_p(B) x_t = \Theta(B) \epsilon_t$  (1.6)

en forma abreviada: ARMA(p, q)

un modelo particularmente útil es ARMA(1, 1):

$$x_t - \phi_1 x_{t-1} = \epsilon_t - \theta_1 \epsilon_{t-1}$$

ó  $x_t = \phi_1 x_{t-1} + \epsilon_t - \theta_1 \epsilon_{t-1}$

Las condiciones de estacionaridad e invertibilidad para los procesos AR(p) y MA(q) establecen estas propiedades para ARMA(p, q). Por ejemplo, el proceso ARMA(1, 1) es estacionario si  $|\phi_1| < 1$ , e invertible si  $|\theta_1| < 1$ .

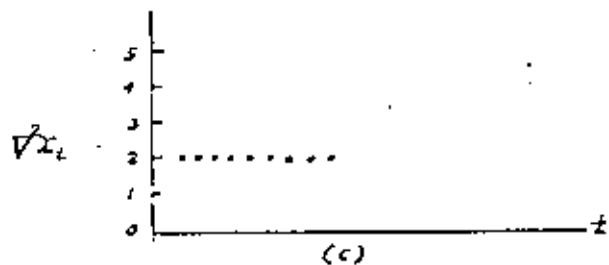
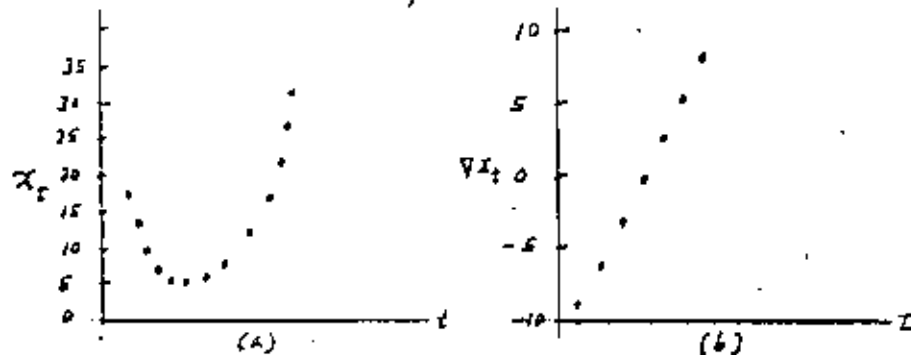
PROCESOS NO ESTACIONARIOS

Cuando las series de tiempo se comportan como si no tuvieran una media constante, se les llama no estacionarios en la media.

Para ilustrar el funcionamiento de las no estacionarias, considere el modelo discreto y determinístico que se muestran en las figuras: (a), (b), (c).

La fig. (a) muestra el comportamiento de una función no estacionaria en media y sesgo. La primera diferencia se representa en la

fig. (b), que sólo es no estacionaria en la media. La segunda diferencia se muestra en (c), es estacionaria.



Por lo tanto, para reducir una serie de tiempo estocástica no estacionaria a una estacionaria se usa diferenciación.

Definiendo al operador diferencia (hacia atrás)  $\nabla$ , como:

$$\nabla x_t = x_t - x_{t-1} \quad (1.7)$$

$\nabla$  se puede expresar en términos del operador B, como:

$$\nabla = (1 - B)$$

$$\nabla^2 = (1 - B)^2$$

$$\nabla^3 = (1 - B)^3$$

$$\nabla^4 = (1 - B)^4$$

Por ejemplo:  $\nabla^2 x_t = (1 - B)^2 x_t = x_t - 2x_{t-1} + x_{t-2}$

El operador diferencia siempre actúa sobre las observaciones originales.

Diferenciando se producen nuevas series  $w_t = \nabla^d x_t = \nabla^d \bar{x}_t$ . Estas series  $\{w_t\}$  puede tener media no cero.

El modelo matemático para el proceso del promedio móvil autorregresivo integrado ARIMA(p, d, q) de orden (p, d, q) es

$$\Phi_p(B) \nabla^d x_t = \Theta_q(B) \epsilon_t \quad (9.8)$$

cuya característica principal es su flexibilidad y su forma alternativa es  $\Phi(B) w_t = \Theta(B) \epsilon_t$

Si las series  $w_t$  tienen media no cero, es decir  $\mu_w$ , entonces el proceso ARIMA es:

$$\Phi_p(B) (w_t - \mu_w) = \Theta_q(B) \epsilon_t$$

La suposición de que  $\mu_w$  es no cero es para introducir un polinomio determinístico en una función eventual de pronóstico, y por esta razón se dice que  $\mu_w = 0$ , salvo que los datos indiquen otra cosa.

En la práctica muchas series de tiempo se pueden describir por un proceso ARIMA en el cual p, d, q no exceden de 2. Por ejemplo, si ARIMA(1,1,1):

$$(1 - \phi_1 B) \nabla x_t = (1 - \theta_1 B) \epsilon_t$$

$$\text{ó } x_t = (1 + \phi_1) x_{t-1} - \phi_1 x_{t-2} + \epsilon_t - \theta_1 \epsilon_{t-1}$$

Si p, q, d = 0, el ARIMA incluye modelos autorregresivos, promedio móvil y mixtos.

### \* CALCULO DE LOS PESOS $\psi$

En la práctica es muy útil expresar cualquier modelo ARIMA(p, d, q) en términos de los pesos  $\psi$  del linear filter. Para efectuar esto, primero recordemos que el linear filter es:

$$\bar{x}_t = \Psi(B) \epsilon_t \quad \text{ó } \epsilon_t = \frac{\bar{x}_t}{\Psi(B)}$$

sustituyendo en (9.8) con la  $x_t$  corregida para  $\mu$  se tiene

$$\Phi_p(B) \nabla^d x_t = \Theta_q(B) \frac{\bar{x}_t}{\Psi(B)}$$

$$\text{ó } \Psi(B) \Phi_p(B) (1-B)^d \bar{x}_t = \Theta_q(B) \bar{x}_t$$

por lo tanto los pesos  $\psi$  se pueden obtener de la ecuación de coeficientes de  $B$  del desarrollo:

$$(\psi_0 + \psi_1 B + \psi_2 B^2 + \dots) (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) (1-B)^d = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$$

Por ejemplo, considerando el modelo ARMA(1,1):

$$(1 - \phi_1 B) \bar{x}_t = (1 - \theta_1 B) \epsilon_t$$

Para encontrar los pesos  $\psi$ , se tiene:

$$(\psi_0 + \psi_1 B + \dots) (1 - \phi_1 B) = (1 - \theta_1 B)$$

Para  $B^0$ :  $\psi_0 = 1$

para  $B^1$ :  $\psi_1 = \psi_0 - \theta_1$

para  $B^2$  los coeficientes son  $\psi_2 - \phi_1 \psi_1 = 0$

lo cual da  $\psi_2 = \phi_1 (\psi_1 - \theta_1)$

$$\therefore \psi_j = \phi_1^{j-1} (\psi_1 - \theta_1)$$

NOTA: como conclusión se podría agregar que, en general, los modelos presentados aquí, no son los apropiados para datos estacionarios.

### 10. PRONOSTICO CON MODELOS DE BOX-JENKINS.

Box-Jenkins elaboraron una filosofía general para el desarrollo de un modelo adecuado de series de tiempo y su uso en pronósticos. Su aproximación consistió de un método iterativo de tres pasos. Primero, se construye un modelo tentativo de acuerdo a los datos actuales. Entonces los parámetros desconocidos se estiman, y finalmente se verifica el diagnóstico para determinar el modelo adecuado.

#### \* IDENTIFICACION

La identificación tentativa del modelo se efectua por medio de un

análisis de los datos históricos. Se requieren de 50 observaciones como mínimo para lograr resultados satisfactorios. El principal instrumento en este análisis es la función de autocorrelación.

Considerando la serie de tiempo estacionaria  $X_1, X_2, \dots, X_N$ , la función teórica de autocorrelación es:

$$\rho_k = \frac{E[(X_t - \mu)(X_{t-k} - \mu)]}{\sigma_x^2}, \quad k=0, 1, \dots \quad (10.1)$$

donde  $\sigma_x^2$  es la variancia,  $\rho_k$  es la autocorrelación en la última  $k$ . Obviamente  $\rho_0 = 1$ ; pero como nunca se le conoce con certeza se le asume. Una estimación satisfactoria de  $\rho_k$  es la función muestra de autocorrelación:

$$\hat{\rho}_k = \frac{\frac{1}{N-k} \sum_{t=1}^{N-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{\frac{1}{N} \sum_{t=1}^N (X_t - \bar{X})^2}, \quad k=0, 1, \dots, k \quad (10.2)$$

Para obtener los resultados, se calculan las primeras  $k = N/4$  autocorrelaciones.

El coeficiente de autocorrelación parcial (o correlación condicional)  $\phi_k$  se lo define como al coeficiente  $k$ -ésimo del proceso autorregresivo de orden  $k$ , y satisface la siguiente ecuación de Yule-Walker:

$$\rho_j = \phi_{j1}\rho_{j-1} + \phi_{j2}\rho_{j-2} + \dots + \phi_{jk}\rho_{j-k}, \quad j=1, 2, \dots, k.$$

substituyendo  $\hat{\rho}_j$  por  $\rho_j$ , se estiman los coeficientes de autocorrelación parcial:

$$\hat{\phi}_j = \phi_{j1}\hat{\rho}_{j-1} + \phi_{j2}\hat{\rho}_{j-2} + \dots + \phi_{jk}\hat{\rho}_{j-k}, \quad j=1, 2, \dots, k$$

cuya solución para  $k=1, 2, \dots, k$ , nos da la función muestra de autocorrelación parcial:  $\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_k$

Por "tailing off" se quiere significar que la función tiene una aproximación exponencial, senoidal o decaimiento geométrica, con un número relativamente grande de valores no cero. Por "cutting off" se da a entender que la función se trunca bruscamente con pocos valores no cero. Obsérvese la dualidad entre los procesos autorregresivo y promedio móvil (evidente en la tabla I).

Según Box-Jenkins, el error estándar de los  $k$ -ésimos coeficientes de autocorrelación parcial es:

$$S(\hat{\rho}_k) \approx N^{-1/2} \left[ 1 + 2 \sum_{j=1}^{k-1} \rho_j \right] \quad (10.3)$$

donde  $\rho_j = \begin{cases} \hat{\rho}_j & \text{si } \rho_j \neq 0 \\ 0 & \text{si } \rho_j = 0 \end{cases}$

el error estándar de los  $k$ -ésimos coeficientes de la autocorrelación parcial es:  $S(\hat{\phi}_k) \approx N^{-1/2} \quad (10.4)$

Generalmente, se supone a la autocorrelación, ya sea parcial o no, como cero, si el valor absoluto de su estimación es menor dos veces que su error estándar. Siempre es conveniente graficar directamente los límites  $\pm 2S(\hat{\rho}_k)$  y  $\pm 2S(\hat{\phi}_k)$  en las gráficas de las funciones de las series de tiempo no estacionarias, para muy fuertemente de un valor a otro. Si este es el caso, se calculan las funciones para la primera diferencia de las series. Si las funciones se comportan de acuerdo a la tabla I, entonces solo un grado de diferencia se necesita para lograr la estacionariedad. Si no es así, se ensayan sucesivamente diferencias de orden superior hasta que se produzca la estacionariedad.

Desde luego, esto requiere habilidad y práctica.

TABLA I

COMPORTAMIENTO DE LAS FUNCIONES DE AUTOCORRELACION TEORICA Y AUTOCORRELACION PARCIAL PARA MODELOS ESTACIONARIOS

MODELO	FUNCION DE AUTOCORRELACION	FUNCION DE AUTOCORRELACION PARCIAL
AR(p)	Tails off	Cuts off después de los últimos p.
MA(q)	Cuts off, después de los últimos q.	Tails off.
ARMA(p,q)	Tails off, muestra el amortiguamiento de la onda senoidal después de los últimos (q-p)	Tails off, muestra el amortiguamiento de la onda senoidal después de los últimos (p-q).

ESTIMACION.

Una vez que las series de tiempo han sido identificadas tentativamente, el siguiente paso es estimar los parámetros del modelo por medio de mínimos cuadrados. Anteriormente se ha definido al modelo lineal como una relación lineal con parámetros desconocidos.

Esto quiere decir que, si el modelo es lineal, entonces la derivada parcial de  $E_t$  con respecto a cualquier parámetro no es función de éstos. Por ejemplo, considerar el proceso AR(p):

$$\hat{x}_t = \phi_1 \hat{x}_{t-1} + \phi_2 \hat{x}_{t-2} + \dots + \phi_p \hat{x}_{t-p} + E_t$$

$$\text{ó } E_t = \hat{x}_t - \phi_1 \hat{x}_{t-1} - \phi_2 \hat{x}_{t-2} - \dots - \phi_p \hat{x}_{t-p}$$

Como  $\frac{\partial E_t}{\partial \phi_i} = \hat{x}_{t-i}$ ,  $i = 1, 2, \dots, p$

no es función de los  $\phi_i$ , entonces por medio de "mínimos cuadrados lineales" se estiman los parámetros de AR(p).

Ejemplo:

Si usamos las ideas anteriores obteniendo los estimadores de AR(2), por medio de mínimos cuadrados lineales.

$$\hat{x}_t = \phi_1 \hat{x}_{t-1} + \phi_2 \hat{x}_{t-2} + E_t$$

Suponiendo que se tienen N observaciones. Entonces se escriben los datos, en forma matricial, para el modelo AR(2) como:

$$\begin{bmatrix} \hat{x}_2 \\ \hat{x}_3 \\ \vdots \\ \hat{x}_N \end{bmatrix} = \begin{bmatrix} \hat{x}_2 & \hat{x}_1 \\ \hat{x}_3 & \hat{x}_2 \\ \vdots & \vdots \\ \hat{x}_{N-2} & \hat{x}_{N-3} \\ \hat{x}_{N-1} & \hat{x}_{N-2} \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} + \begin{bmatrix} E_2 \\ E_3 \\ \vdots \\ E_{N-1} \\ E_N \end{bmatrix}$$

$$\text{ó } X = Z\phi + E$$

∴ la estimación de los parámetros autorregresivos con:

$$\hat{\phi} = (Z'Z)^{-1} Z'X$$

Notar que únicamente se han utilizado las últimas N-2 observaciones, puesto que las  $\hat{x}_0$  y  $\hat{x}_1$  requeridos para el AR(2) en  $t=1$ , no existen. Una solución alternativa para un valor inicial es hacer  $\hat{x}_0 = \hat{x}_1 = 0$  y utilizar todo el conjunto de las N observaciones.

La estimación de los parámetros de un modelo de promedio móvil no es tan simple. Por ejemplo, considerando al proceso MA(1)

$$\hat{x}_t = E_t - \theta_1 E_{t-1}$$

$$\therefore E_t = (1 - \theta_1 B)^{-1} \hat{x}_t$$

La primera derivada  $\frac{\partial E_t}{\partial \theta_1} = \theta_1 (1 - \theta_1 B)^{-2}$  es una función de los parámetros desconocidos  $\theta_1$ . Por lo tanto, estos no pueden ser estimados por mínimos cuadrados lineales. Una estimación aproximada se obtiene aplicando la técnica de la suma de la función cuadrada:

$$SS_e = \sum_{t=1}^N (1 - \hat{x}_t)^2$$

VERIFICACION DEL DIAGNOSTICO

Si el modelo de ajustado es adecuado, las observaciones deberán transformarse a un proceso white noise. Así, un método lógico de verificar el diagnóstico es calcular los residuales, es decir,

$$e_t = \hat{x}_t - \hat{x}_t$$

se construye y se examina su función de autocorrelación. Si el modelo es el adecuado esta función no diferirá significativamente de cero para todos los últimos más grandes que cero. Un análisis cuidadoso de los residuales indicará a menudo un mejoramiento potencial del modelo.

PRONOSTICO

Una vez identificado el modelo correcto de las series de tiempo, estimados sus parámetros y verificado su diagnóstico, entonces, el modelo ya puede ser utilizado para generar pronósticos.

Denotamos por T al período presente, y suponiendo que se desea pronosticar la señal para el período T+t. Suponiendo además que  $\hat{x}_{T+1}$  (ó  $\hat{x}_{T+1}$  si la serie está corregida por la media) repre-

usando el pronóstico para el período  $T+t$

El pronóstico se obtiene tomando la esperanza en el origen  $T$  del modelo para el tiempo  $T+t$ . Como regla general, el pronóstico para el período  $T+t$  se construye a partir del pronóstico para los períodos  $T+1, T+2, \dots, T+t-1$ .

En este procedimiento, los  $\varepsilon_{T+j}$  que no han ocurrido en el tiempo  $T$  se reemplazan por su pronóstico  $\hat{\varepsilon}_{T+j}$ ; los  $\varepsilon_{T-j}$  se estiman de  $\varepsilon_{T-j} = x_{T-j} - \hat{x}_{T-j}$ , y los  $\varepsilon_{T-j}$  que no han ocurrido se reemplazan por ceros. Por ejemplo, considerando el modelo ARIMA(1,1,1) al final del tiempo  $T+t$ :

$$x_{T+t} = (1 + \phi_1)x_{T+t-1} - \phi_1 x_{T+t-2} + \varepsilon_{T+t} - \theta_1 \varepsilon_{T+t-1}$$

Tomando la esperanza en el tiempo  $T$  se encuentra para  $t=1$

$$E[x_{T+1}] = \hat{x}_{T+1} = (1 + \phi_1)x_T - \phi_1 x_{T-1} - \theta_1 \varepsilon_T$$

Como  $E[\varepsilon_{T+1}] = 0$ , y al final del tiempo  $T$ ,  $\varepsilon_T$  se puede estimar por medio del residual en el tiempo  $T$ ,  $\varepsilon_T = x_T - \hat{x}_T$ .

Para  $t \geq 2$ , una forma similar se obtiene:

$$E[x_{T+t}] = \hat{x}_{T+t} = (1 + \phi_1)\hat{x}_{T+t-1} - \phi_1 \hat{x}_{T+t-2}$$

También se puede generar un pronóstico usando un modelo expresado en términos de los polos  $\psi$ . En general, en el origen  $T$ , se tiene:

$$x_{T+t} = \psi \varepsilon_{T+t-1} + \psi_{t-1} \varepsilon_{T+t-2} + \psi_{t-2} \varepsilon_{T+t-3} + \dots + \varepsilon_{T+t} \quad (10.5)$$

Sin embargo, en el tiempo  $t > T$  se puede reemplazar los correspondientes  $\varepsilon_t$  por ceros y en el tiempo  $t \leq T$  se puede reemplazar  $\varepsilon_t$  por  $\varepsilon_t$ , así el pronóstico es:

$$\hat{x}_{T+t} = \psi_t \varepsilon_T + \psi_{t-1} \varepsilon_{T+1} + \dots \quad (10.6)$$

La estimación de los pesos  $\psi_T, \psi_{T+1}, \dots$  se obtiene de  $\{\hat{\psi}_i\}$  y  $\{\hat{\theta}_i\}$ . La variancia del peso  $t$ , al error de pronóstico es:

$$V(t) = \left\{ 1 + \sum_{j=1}^{t-1} \psi_j^2 \right\} \sigma_\varepsilon^2$$

Así, la aproximación 100(1- $\alpha$ ) del porcentaje de los límites de probabilidad del pronóstico para el período  $T+t$  se calculan de

$$\hat{x}_{T+t} \pm z_{\alpha/2} \left\{ 1 + \sum_{j=1}^{t-1} \psi_j^2 \right\}^{1/2} \sigma_\varepsilon \quad (10.7)$$

donde  $\sigma_\varepsilon$  es una estimación de  $\sigma_\varepsilon$  y  $z_{\alpha/2}$  es un porcentaje de la densidad normal estándar, tal que:

$$Pr\{Z \geq z_{\alpha/2}\} = \alpha/2$$

### 13-SISTEMA INTERACTIVO DE PRONÓSTICO.

Con el sistema de tiempo compartido de computadora, se ha abierto una nueva dimensión en la aplicación de modelos estadísticos y matemáticos a problemas secuenciales de decisión.

Certo la función de predicción implica un proceso secuencial, se lo puede manipular satisfactoriamente con el sistema de tiempo compartido. En seguida se describe brevemente este sistema que permite aplicar un análisis preliminar de datos, para identificar la técnica de pronóstico más apropiada para una situación dada, y aplicarla en el desarrollo de aquél.

El sistema interactivo de pronóstico ha dado excelentes resultados, tanto en la existencia, como para hacer decisiones en los negocios. Actualmente se aplica en las técnicas de administración.

En la teoría y práctica ha respondido ampliamente a los requerimientos de compañías de todos tamaños, en las que es esencial proporcionar para un número de artículos las inventas que las afectan en su decisión y ejecución. Pero mientras la mayoría de los administradores están conscientes de la necesidad de emplear métodos de pronóstico, pocos son los que están familiarizados con las técnicas y sus características, que han sido desarrolladas, y así poder seleccionar la más apropiada a su problema.

Desafortunadamente existen pocos trabajos que traten directamente el lado administrativo de problemas de pronóstico; por el contrario, la literatura sobre este tema es extensa, y consiste de un excelente número de libros artículos.

Como ejemplos de técnicas de pronóstico, se tiene el método de Forster-Spinkins para aproximación de series de tiempo, el de Johnston sobre técnicas de regresión, y el de Brown para suavización de series de tiempo. Estos métodos han encontrado amplia aceptación en investigación de operaciones, por lo tanto, es necesario describir sus aspectos técnicos y características, para trasladarlas a la práctica administrativa.

Los factores primarios concernientes a la administración, incluyen la experiencia práctica que otros han tenido al usar el método, su costo, su limitación y la exactitud en los resultados del pronóstico.

Todos aquellos que han tratado con métodos de pronóstico, saben de un número de requerimientos que son difíciles de satisfacer con los sistemas y técnicas existentes. Tres de los más importantes son los siguientes:

1.- La dificultad de mantener la flexibilidad de aproximación en nuevas situaciones.

Los administradores han encontrado que es difícil desarrollar un método de pronóstico de su predicción, sobre todos los demás, para utilizarlo en cualquier nueva situación. Sin embargo, reconocen la necesidad de considerar un rango de técnicas alternativas para tales casos. Una fuente de esta dificultad es la confianza puesta en el criterio de un solo técnico. Es difícil esperar imparcialidad de parte de éste hacia cada alternativa del método, sencillamente a causa de la magnitud intelectual que dicha tarea requiere. Lo correcto sería que un sistema sustentase explícitamente la consideración de varias técnicas alternativas para cada de los problemas de pronóstico.

2.- Consideración de todos los factores relevantes al seleccionar una técnica de pronóstico.

Los administradores están conscientes de la necesidad de considerar no solo la exactitud, sino también otros factores al seleccionar la técnica de pronóstico. Como el intercambio y criterio se deben hacer respecto a estos factores, el administrador necesita estar involucrado activamente para aplicar cualquier sistema de pronóstico.

3.- Un procedimiento para seleccionar rápidamente una técnica alternativa.

Los profesores que han enseñado problemas a muchos administradores al preverlos, es que ante una nueva situación, se ven en la necesidad de turnarla directamente a algún técnico de su personal, por no sentirse competentes para hacer el análisis preliminar. Sin embargo, si se realiza una selección preliminar de técnicas alternativas para un problema dado, a la larga se ahorra tiempo y esfuerzo, y también se estimulará la adopción de procedimientos formales de pronósticos en casos en donde no sea necesario consultar a un especialista.

Salidas de los problemas que han encontrado los administradores, los que enseñan programas de administración también han identificado la tendencia en clase hablar de detalles técnicos excluyendo algunas consideraciones prácticas. Esto es natural, dado que la mayor parte de la literatura sobre pronósticos es técnica en su orientación, y existe poco escrito sobre la experiencia actual.

Otro problema al que se enfrentan los que enseñan pronósticos, es que toma un tiempo considerable al estudiante aplicar una técnica de manera completa a una simple situación. Puesto que la mayoría de los cursos son relativamente cortos, no hay tiempo suficiente para aplicar un número de estas técnicas a varias situaciones del problema.

Finalmente, los profesores se encuentran con la dificultad particular para enseñar las suposiciones inherentes en cada método alternativo de pronósticos y su implicación en la práctica. Una solución obvia para esto último, sería dar al estudiante alguna experiencia práctica para aplicar los métodos. Pero sin más, por falta de tiempo esto no se puede llevar a cabo.

Basándose, en gran parte, sobre un reconocimiento de los problemas anteriores, los autores han desarrollado un sistema interactivo tanto para la enseñanza como para su aplicación

práctica. Este sistema ha sido instalado en computadoras de tiempo compartido.

En seguida se describe este sistema interactivo, como manejarlo, la experiencia práctica de los autores en la enseñanza, identificando y explicando las técnicas para cada problema particular de administración.

A la fecha, los resultados han sido optimistas, y parece ser se ha alcanzado el objetivo para problemas previamente planeados.

DESCRIPCION DEL SISTEMA COMPUTARIZADO.

El sistema interactivo de pronóstico está dividido en dos segmentos secuenciales.

El primer segmento llamado SIMYL está dirigido para permitir al usuario ejecutar un análisis preliminar de sus datos, a fin de identificar dos o tres técnicas de pronóstico que pudieran ser las más apropiadas para su caso.

Como se muestra en el diagrama de flujo (fig. 1), se comienza por introducir los datos que se desean examinar para usarlos como base del pronóstico. Este segmento hace una serie de preguntas del administrador respecto a su criterio acerca de los datos y las características del problema, más importantes en la selección del método. Entre los factores que necesariamente se deben considerar, están los siguientes:

- i) El tiempo horizontal para efectuar una decisión: periodo inmediato, periodo corto, periodo medio y periodo largo.
- ii) El esquema de datos: estacional, horizontal, tendencia, cíclico o aleatorio.
- iii) El tipo de modelo deseado: series de tiempo, causal, estadístico o no estadístico.
- iv) El valor del pronóstico, así como la cantidad que puede ser gastada en su obtención.



v) La exactitud requerida y su justificación.

vi) La complejidad tolerable.

vii) La disponibilidad de datos históricos.

Algunos de estos factores se los puede analizar mejor a través de pruebas estadísticas, mientras que otros implican el criterio del usuario. Y otros más requieren de información acerca de la técnica de pronóstico en sí misma, lo cual puede ser proporcionado por anteriores usuarios. Todos estos factores son importantes y su consideración específica se debe hacer antes de seleccionar el método más apropiado.

Para poder elegir el mejor método se debe tener, por un lado, conocimiento de todas las técnicas de pronóstico; y por el otro, se deben evaluar todos los factores que influyen en la selección. Esta tarea no es fácil, aun por expertos en este campo. Para los novatos o para muchos dificultades, una de las más importantes es la falta de experiencia al evaluar los factores relevantes que influirán en la elección del método de pronóstico.

La primera característica del segmento SIBYL, es que considera a todos los siete factores mencionados anteriormente, y da apoyo al usuario al manejarlos. La base lógica de esta consideración se muestra en la fig. 2. Esta figura ha sido desarrollada como base para comparar técnicas disponibles sobre diferentes factores. La influencia de esta estructura en el diseño del sistema, está reflejada en la secuencia de preguntas y respuestas. En este segmento (SIBYL), la pregunta inicial trata con la serie de datos que se está usando como base para el pronóstico; y también se grafican para obtener distintas estadísticas.

La siguiente sección de preguntas trata con el uso de la autocorrelación para identificar el esquema básico de los datos. En esta parte de SIBYL, la autocorrelación se calcula para varios tiempos de demora, y se grafica o se presenta en forma

de resumen estadístico. Una vez calculada la autocorrelación, SIBYL ayuda al usuario a identificar el esquema que mejor representa sus datos.

A través de una serie de preguntas el programa obtiene información sobre los factores importantes y necesarios para seleccionar el método de pronóstico y suministra una lista de tres o cuatro técnicas que parecen ser las más adecuadas para el problema. También en este punto le da al usuario, un número de estadísticas comparativas sobre aquellas técnicas sugeridas y pregunta para seleccionar una para desarrollarla en el pronóstico presente. Una vez que se ha elegido el método, el control pasa al segundo segmento mayor del sistema: RUNNER.

Antes de describir a RUNNER, hay un par de características de este sistema de pronóstico que facilitan su uso, y merecen especial atención. Una de estas, es la opción que tiene el usuario de responder cualquier pregunta desde el sistema con la palabra HELP. Cuando el usuario contesta de esta manera recibe información adicional no sólo en forma de respuesta, sino también una explicación más general del factor y su relación con el pronóstico.

La segunda consiste de un manual suplementario exhaustivo, ilustrado con ejemplos que involucran los principios importantes de pronóstico, y el uso de técnicas específicas.

Al segundo segmento principal del sistema se le llama RUNNER y se muestra en la figura 3. Este segmento está compuesto de un número de subrutinas, en las que cada una es la versión computarizada de cada técnica especificada de pronóstico.

El programa RUNNER permite al usuario seleccionar de un conjunto de subrutinas el método de pronóstico adecuado a sus datos. Se debe recordar que cuando el usuario llega a este punto, SIBYL ya ha identificado dos o tres técnicas apropiadas al problema. La mayoría de preguntas y respuestas involucradas en este segmento del programa, tratan con un conjunto de parámetros para un

método, especificando previamente. Por ejemplo, si el método es un simple ajuste lineal, este requiere que el usuario defina tres parámetros: los pesos, la constante de adaptación y el número de iteraciones.

Una vez que el método de predicción está fijo, se determina la exactitud de las iteraciones. Esto se hace en términos del error y su porcentaje. Finalmente, el RUNNER se lo utiliza para comparar diferentes técnicas para el mismo conjunto de datos históricos.

Después de "correr" la combinación SIBYL-RUNNER, el usuario tendrá:

- 1- Una análisis general de los datos.
- 2- Una selección de las técnicas disponibles de predicción.
- 3- Un examen detallado de las técnicas más apropiadas.
- 4- La selección final de una técnica para su situación.

#### EXPERIENCIA EN EL USO DEL SISTEMA INTERACTIVO DE PROLSTICO.

El uso de este sistema de pronóstico, basado en tiempo compartido, ha sido bien recibido en la enseñanza. Los estudiantes han sido más motivados, usando esta aproximación, que con la de los libros de texto tradicionales; y científicos que han guiado en la aplicación práctica de estas técnicas alternativas.

Una de las características más atractivas de este sistema, es que se puede usar eficientemente por alguien familiarizado con éste, y con predicción empírica; ya que permite expresar mucho de la información descriptiva. Así, los estudiantes han encontrado que es muy útil, no sólo en su aprendizaje inicial, sino que también, cada vez que han estudiado un número de técnicas, les facilita su aplicación a situaciones específicas.

Aunque el sistema SIBYL-RUNNER no ha sido cabalmente probado en los negocios como en la enseñanza, ha proporcionado una ayuda de gran utilidad en pronósticos en los que ha sido aplicado. Una de las ventajas de este sistema en

administración, es que permite obtener en forma rápida y a bajo costo un pronóstico aproximado. En resumen, este método es adecuado como guía para examinar un amplio rango de alternativas para cualquier nueva situación, además, ayuda a considerar todos aquellos factores que son importantes para seleccionar la técnica más apropiada para un problema dado. La conclusión de los autores, es que este sistema satisface muchas de las necesidades existentes en el área de pronósticos.

Naturalmente, este sistema tiene limitaciones. En primer lugar, es que tanto administradores como estudiantes, deben adentrarse en pronósticos, así como en el rango de técnicas disponibles antes de empezar a utilizar al SIBYL-RUNNER.

En segundo lugar, las preguntas y respuestas pueden ser rovinas fuera de tiempo, y por lo tanto fallar el entusiasmo del usuario. Tercero, la simplicidad del sistema y su uso fácil puede llevar a aplicarlo, en casos en donde es necesario un análisis más sofisticado, lo cual debe ser efectuado por un especialista en pronósticos.

Este paquete está diseñado para adaptarse a la aplicación de métodos sistematizados de pronóstico, no para reemplazarlos.

Actualmente, un simple programa de computadora escribe al lenguaje de programación BASIC incorporando las características del sistema SIBYL-RUNNER. Además, existe un programa BASIC que representa a cada una de las técnicas de pronóstico incluyendo este sistema. Por ejemplo, existe un programa para suavización exponencial, otro para regresión simple, y así sucesivamente. Estos programas, que representan cada método de pronóstico, son llamados y controlados directamente por el programa SIBYL-RUNNER. Sin embargo, también se pueden correr individualmente.

El sistema SIBYL-RUNNER está programado en FORTRAN.



CONCLUSIONES:

Los efectos de los series de tiempo han tenido un desarrollo considerable a partir de su introducción. Como ya se vió, existe una gran variedad de métodos que se pueden utilizar para adaptarse a la mayoría de estimación, pronóstico o control, que necesariamente implican series de tiempo, obteniéndose un alto grado de exactitud con su empleo.

Sin embargo, existe un problema que no es técnico, sino más bien consistente en la falta de comunicación entre los expertos en métodos de serie de tiempo y los usuarios potenciales. Aunque estos métodos son conocidos ampliamente, gracias a Box y Jenkins, Watts y Kalman, subsisten problemas de tipo práctico.

No existe una documentación adecuada para la descripción verbal de cómo debe usarse cada método y cómo interpretar los resultados en términos de programas de computadora para personas no capacitadas técnicamente. La descripción de los métodos es demasiado matemática y los diferentes programas existentes no están estandarizados.

El interés por el método no está compensado de los beneficios de éstos. Los métodos de series de tiempo de filtro, pronóstico o control. Existe una carencia de información para permitir el avance de costos, de exactitud, y desde luego en el análisis final, de los beneficios de usar tales técnicas. En fin, la crítica es simi-

lar a aquella que hizo Grayson a la ciencia de la administración, principalmente que el uso de métodos de series de tiempo es un arte que requiere de conocimientos prácticos que sólo unos cuantos especialistas saben aplicar.

En opinión de los autores, los inconvenientes antes mencionados se deben corregir. Existe la necesidad de varias prioridades de investigación.

Primero, los hallazgos de la investigación deberán establecer el nivel de exactitud, preferiblemente expresado en promedio de porcentaje de error de cada método individual. Segundo, una representación detallada del costo actual del desarrollo y funcionamiento de cada técnica. Tal información sería inestimable para cualquier usuario potencial. El trabajo, aunque no es fácil, dado el gran número de factores que influyen la exactitud y el costo, no es tan difícil como las tareas llevadas a cabo por físicos y otros científicos.

Otra área de investigación debería ser dirigida hacia el desarrollo de filtros que incluyan esquemas de movimiento medio. Esto incrementaría la eficiencia con poca carga computacional extra. La aplicación de los filtros, fuera de las situaciones de ingeniería, parece prometer mucho.

Los filtros combinan las características, tanto de los métodos de suavización (en términos de costos), como las del esquema autorregresivo, en términos de rigorismo teórico, sin ser tan difícil — o complejo su uso. Por lo tanto, su utilización en aplicaciones fuera del campo de la ingeniería podría ser sumamente útil.

Otra área que requiere atención, es la de predecir el ciclo de negocios de una manera más satisfactoria. Esta es una área de extrema importancia en pronósticos de negocios, en la cual, a la fecha, ningún método de series de tiempo ha tenido éxito. Existe, claro, algunos intentos aislados, tales como los de McLaughlin y Boyle. Sin embargo, los resultados han sido mediocres. También debe de haber un trabajo creativo de programación para elaborar paquetes de pronóstico que sean fáciles de entender, de usar, y que presenten sus resultados de tal forma que personas no especializadas, puedan utilizarlos para sus necesidades de estimación, de planeación, de pronóstico y de control.

Finalmente, el trabajo que queda, y el más importante, es dar a conocer los métodos de series de tiempo a los no expertos. Esto crearía un gran número de usuarios, y por lo tanto, aumentaría el interés y la actividad, en el área de las series de tiempo.

En opinión del autor, la tarea más urgente es la de educar al usu-

rio potencial sobre los méritos de la aproximación de series de tiempo, aplicadas a datos cronológicos.

## BIBLIOGRAFIA

- \* Operations research in production planning, scheduling, and inventory control.  
de L.A. Johnson.  
D.C. Montgomery.  
Edit. John Wiley and sons, Inc. London.
- \* The statistical analysis of Time Series.  
de T.W. Anderson.  
Edit. John Wiley and sons, Inc. New York.
- \* Elements of Applied stochastic Processes.  
de U. Narayan Bhat.  
Edit. John Wiley and sons, Inc.
- \* Interactive forecasting. Volume II.  
de Spyros Makridakis.  
Edit. Christine Doerr, Hewlett. Packard.



DIVISION DE EDUCACIÓN CONTINUA  
FACULTAD DE INGENIERIA U.N.A.M.

COMPUTACION APLICADA A LA PLANEACION URBANA

PROGRAMACION LINEAL

Dr Sergio Fuentes Maya

Agosto, 1981

CONTENIDO	Pag.
1. INTRODUCCION	1.1
2. BASES METODOLOGICAS DE P.L.	2.1
2.1 Conceptos y definiciones	2.2
2.2 Problemas lineales duales	2.5
2.3 El método Simplex revisado	2.11
2.4 Análisis Post-óptimo	2.19
2.5 Parametrización	2.36
3. EJEMPLOS DE APLICACION	3.1.

## 1. Introducción

La programación lineal es la parte de la programación matemática más popular y ampliamente desarrollada. Su aplicación a problemas reales como a diversas ramas de la ciencia tiene como propósito resolver problemas de asignación de recursos escasos. El auge de la programación lineal se debe en gran parte a su facilidad para plantear y resolver problemas reales ó una simplificación de los mismos, así como a la sencillez que existe para la interpretación de sus resultados. Una ventaja importante de la programación lineal es que dispone de métodos especiales para el análisis de un mismo problema cuando existen cambios en los datos. Esto permite establecer cotas sobre los resultados de un problema lineal cuando la información sobre algunos de sus parámetros es incierta o bien varía en intervalos definidos. Para ello, en los últimos veinte años se han desarrollado las técnicas de análisis post-óptimo y de parametrización que permiten determinar los efectos de cambios en los datos.

En la práctica también son importantes los métodos de solución de la programación lineal, pues aún problemas pequeños involucran una gran cantidad de operaciones que hacen indispensable el uso de las computadoras. Es por ello que se han desarrollado "paquetes comerciales" eficientes y sofisticados que inclusive permiten la solución de problemas de gran escala. La importancia de estos paquetes no sólo está en el hecho de que resuelve el problema dando una solución óptima, sino que



además, proporcionan al usuario una gran cantidad de información adicional que le permite hacer un análisis post-óptimo, esto es, le permiten determinar los rangos de variación de diversos elementos del problema sin que cambie la solución óptima. Otros procesos que es posible realizar con estos paquetes son la parametrización, almacenamiento de la base y cambio arbitrario de un parámetro cualquiera del problema de P.L.

En esta ponencia desarrollamos las bases metodológicas de la programación lineal y presentaremos algunos ejemplos de aplicación.

## 2. Bases metodológicas de la P.L.

El propósito de este capítulo es presentar de manera uniforme los principales conceptos y resultados de la programación lineal. La idea de establecer de manera formal algunos de estos resultados se justifica de la necesidad de familiarizar al lector con las implicaciones de los mismos, las cuales, son frecuentemente usadas en la interpretación de los reportes de salida de los paquetes comerciales de la programación lineal. Los resultados fundamentales son enunciados y el lector puede recurrir a cualquier libro de programación lineal para su demostración o profundización en el tema.

Este capítulo se divide en cinco secciones, en la primera se dan algunas definiciones y resultados importantes introduciendo con ellos la notación que se maneja más a menudo. En la segunda sección se define el problema dual, cómo se obtiene y los resultados que caracterizan las soluciones óptimas de la programación lineal: los teoremas de dualidad y complementaridad. En la tercera sección se describe el método Simplex Revisado y se resuelven algunos ejemplos. La sección 4 trata el análisis post-óptimo y finalmente, la sección 5 introduce mediante un ejemplo, el concepto de parametrización.

2.1 Conceptos y definiciones de P.L.

Un problema lineal consiste en la maximización o minimización de una función lineal de varias variables sujeta a restricciones lineales en estas mismas variables. Una forma particular, del problema lineal, es la forma estándar:

$$\text{minimice } c_1x_1 + c_2x_2 + \dots + c_nx_n$$

sujeto a

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \dots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \dots + a_{2n}x_n = b_2$$

..... (P)

$$a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \dots + a_{mn}x_n = b_m$$

$$x_1 \geq 0 ; x_2 \geq 0 ; \dots ; x_n \geq 0$$

donde los coeficientes  $a_{ij}$ ,  $b_i$  y  $c_j$  son números reales, y  $x_i$ ,  $i=1, \dots, n$  son las variables a determinar. Una forma compacta y usual de escribir (P) es:

$$\text{minimice } cx$$

$$Ax = b$$

$$x \geq 0$$

donde A, es una matriz  $m \times n$ ; c, vector hilerá de n componentes; b, un vector columna de m componentes y x, un vector columna de n variables.

Considere el sistema de ecuaciones lineales

$$Ax = b$$

donde A es una matriz  $m \times n$ ; b, vector columna de m componentes; y x, vector columna de n incógnitas. Sea B una submatriz de A de orden  $m \times m$  que no es singular y suponga que las  $n-m$  componentes del vector x no asociadas a las columnas de B se hacen igual a cero. La solución del conjunto de ecuaciones resultante se denomina una solución básica con respecto a la base B. Las componentes de x asociadas a las columnas de B se denominan variables básicas. La solución básica es degenerada si una o más de las variables básicas tiene valor cero.

La idea al definir la solución básica del sistema  $Ax=b$  es que si podemos escribir  $A = [B, R]$ , donde B es una matriz no singular, entonces una solución de este sistema de ecuaciones se puede determinar observando que, si hacemos  $Bx_B=b$  el vector  $x = \begin{bmatrix} x_B \\ 0 \end{bmatrix} = \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix}$  es la solución que deseamos. Note que en una solución básica no degenerada es inmediata la identificación de las columnas de A que forman la matriz no singular B. Sin embargo, en una solución degenerada existe cierta ambigüedad para identificar B, pues, las variables básicas con valor cero pueden ser confundidas con las variables no básicas cuyo valor es cero también.

Sea el problema lineal en forma estándar

$$\begin{aligned} & \text{minimice } cx \\ \text{sujeto a } & Ax = b \\ & x \geq 0 \end{aligned}$$

Se dice que  $x$  es una solución factible si satisface las restricciones de este problema. Si la solución factible es también básica se dice que es una solución factible básica. De manera semejante se define la solución factible básica degenerada. Una solución factible que adquiere el valor mínimo de la función objetivo en el problema lineal se denomina solución óptima. La base  $B$  asociada a la solución óptima se denomina base óptima.

TEOREMA FUNDAMENTAL DE LA P.L.

Dado un problema lineal en la forma estándar en donde  $A$  es una matriz  $m \times n$  con rango  $m$ :

- a. Si existe una solución factible, existe una solución factible básica;
- b. Si existe una solución factible óptima, existe una solución factible básica que es óptima.

Este resultado permite reducir el problema a programación lineal a la búsqueda de soluciones básicas cuyo número es finito.

2.2 Problemas lineales duales

Considere los problemas lineales

minimice $cx$	maximice $\lambda b$
(P) $Ax \geq b$	$\lambda A \leq c$ (D)
$x \geq 0$	$\lambda \geq 0$

donde  $A$  es una matriz  $m \times n$ ;  $b$ , un vector columna de  $m$  componentes;  $c$ , un vector hilera de  $n$  componentes;  $x$ , un vector columna de  $n$  incógnitas; y  $\lambda$ , un vector hilera de  $m$  variables. Estos problemas se denominan problemas lineales duales y se dice que (P) es el problema primal y (D) el correspondiente problema dual.

Esta definición de problemas lineales duales permite determinar el problema dual de un problema lineal cualquiera. Esto se obtiene, básicamente, mediante la transformación del problema lineal original a la forma del problema (P).

Ejemplo 1.

Considere el problema lineal en forma estándar

$$\begin{aligned} & \text{minimice } cx \\ \text{(P')} & \quad Ax = b \\ & \quad x \geq 0 \end{aligned}$$

que puede escribirse en forma equivalente como:

minimice  $cx$

$$\begin{aligned} Ax &\geq b \\ -Ax &\geq -b \\ x &\geq 0 \end{aligned}$$

y cuyo correspondiente problema dual es:

maximice  $ub-vb$

$$\begin{aligned} uA-vA &\leq c \\ u &\geq 0 \quad v \geq 0 \end{aligned}$$

donde  $u$  y  $v$  son vectores hilera de  $m$  componentes. Si  $\lambda = u-v$ , el correspondiente par de problemas duales asociados es

minimice $cx$	maximice $\lambda b$
(P') $Ax = b$	$\lambda A \leq c$ (D')
$x \geq 0$	$\lambda$ no restringida

denominada la forma asimétrica, pues en un problema, el vector de variables es restringido y en el otro es no restringido.

En general, se cumple que si alguna de las restricciones del problema primal es una igualdad, la componente correspondiente del vector  $\lambda$  en el problema dual será una variable no restringida. Recíprocamente, si alguna de las componentes del vector  $x$  en el problema primal es no restringida, la desigualdad correspondiente en el problema dual será igualdad.

Ejemplo 2. Determine el dual del problema lineal

$$\begin{aligned} \text{minimice } & cx \\ & Ax \leq b \\ & x \geq 0 \end{aligned}$$

Primero conviene reformular este problema como

$$\begin{aligned} \text{minimice } & cx \\ & -Ax \geq -b \\ & x \geq 0 \end{aligned}$$

cuyo dual asociado es dado por

$$\begin{aligned} \text{maximice } & -\lambda b \\ & -\lambda A \leq c \\ & \lambda \geq 0 \end{aligned}$$

o si se prefiere, (i.e., el caso usado en paquetes comerciales

$$\begin{aligned} \text{maximice } & \lambda b \\ & \lambda A \leq c \\ & \lambda \leq 0 \end{aligned}$$

Utilizando estas ideas podemos establecer las formas generales de problemas lineales duales:

PRIMAL	DUAL
Minimice $qy + cx$ $Q_1y + A_1x \geq b_1$ $Q_2y + A_2x = b_2$ $Q_3y + A_3x \leq b_3$ y no restringida ; $x \geq 0$	Maximice $\lambda_1b_1 + \lambda_2b_2 + \lambda_3b_3$ $\lambda_1Q_1 + \lambda_2Q_2 - \lambda_3Q_3 = q$ $\lambda_1A_1 + \lambda_2A_2 - \lambda_3A_3 \leq c$ $\lambda_1 \geq 0 ; \lambda_2$ no restringida ; $\lambda_3 \geq 0$ .
Maximice $qy + cx$ $Q_1y + A_1x \geq b_1$ $Q_2y + A_2x = b_2$ $Q_3y + A_3x \leq b_3$ y no restringida ; $x \geq 0$	Minimice $\lambda_1b_1 + \lambda_2b_2 + \lambda_3b_3$ $-\lambda_1Q_1 + \lambda_2Q_2 + \lambda_3Q_3 = q$ $-\lambda_1A_1 + \lambda_2A_2 + \lambda_3A_3 \geq c$ $\lambda_1 \geq 0 ; \lambda_2$ no restringida ; $\lambda_3 \geq 0$ .

La relación más importante y significativa de los problemas lineales duales queda resumida en el teorema de dualidad, el cual presenta la relación que existe entre los valores de las funciones objetivo de estos problemas y las relaciones que gobiernan a las restricciones de un problema con las variables de decisión de su dual. Antes de establecer este resultado conviene señalar que dados los problemas lineales duales

$$\begin{array}{ll}
 \min z = cx & \max z = \lambda b \\
 \text{(P)} \quad Ax \geq b & \lambda A \leq c \quad \text{(D)} \\
 x \geq 0 & \lambda \geq 0
 \end{array}$$

Se cumple que si  $x_0$  y  $\lambda_0$  (arbitrarias) satisfacen las restricciones de estos problemas se tiene que  $cx_0 \geq \lambda_0 b$ , pues

$$cx_0 \geq (\lambda A)x_0 = \lambda(Ax_0) \geq \lambda_0 b$$

Asimismo, se tiene que  $cx_0 = \lambda_0 b$  implica el par  $(x_0, \lambda_0)$  es la solución de los problemas lineales duales.

TEOREMA DE DUALIDAD

Si alguno de los problemas lineales duales (P) ó (D) tiene solución óptima, lo mismo es cierto del otro problema y el correspondiente valor de la función objetivo es el mismo. Por otra parte, si uno de los problemas tiene función objetivo no acotada, el otro problema, no tiene solución factible.

## TEOREMA DE COMPLEMENTARIDAD

Considere los problemas lineales duales

$$\begin{array}{ll} \min z = cx & \max w = b' \\ (P) \quad Ax \geq b & \lambda A \leq c \quad (D) \\ \quad x \geq 0 & \lambda \geq 0 \end{array}$$

Sean  $x^*$  y  $\lambda^*$  soluciones factibles de los problemas respectivos.

Entonces, una condición necesaria y suficiente para que  $x^*$  y  $\lambda^*$  sean soluciones óptimas es que satisfagan las relaciones

$$\begin{array}{ll} a. \quad x_i^* > 0 & \text{implica} \quad \lambda^* a_{ij} = c_j \\ b. \quad x_i^* = 0 & \text{si} \quad \lambda^* a_{ij} < c_j \\ c. \quad \lambda_j^* > 0 & \text{implica} \quad a^j x^* = b_j \\ d. \quad \lambda_j^* = 0 & \text{si} \quad a^j x^* > b_j \end{array}$$

donde  $a_j$  ( $a^j$ ) es el  $i$ -ésimo ( $j$ -ésimo) vector columna (hílera) de la matriz  $A$ .

Una interpretación económica de este resultado es como sigue:

Si en el problema (P) se tiene una restricción activa, esto es,  $a^j x^* = b_j$ , entonces el precio a que se compraría una unidad adicional del recurso  $j$  es igual a  $\lambda_j^*$ . Además, si la restricción es no activa, esto es  $a^j x^* > b_j$ , el precio a que se compraría la unidad adicional de recurso es igual a cero.

De este resultado se justifica que los elementos  $\lambda_j^*$   $j=1, \dots, m$ , sean denominados los precios sombra o precios de oportunidad.

## 2.3 El método Simplex Revisado

Existen distintos métodos de solución del problema de programación lineal, el más frecuentemente utilizado para la solución de problemas de programación lineal es el llamado Simplex Revisado. Este método consiste en pasar de una solución factible básica a otra, de tal manera, que en cada paso el valor de la función objetivo disminuya. La terminación del proceso se garantiza en un número finito de pasos, ya que, el número de soluciones factibles es finito. Este método además proporciona elementos suficientes para calcular la solución del problema dual.

Dada la importancia del método simplex revisado conviene describir las bases del mismo:

Considere el problema lineal

$$\begin{array}{l} \text{minimice } z = cx \\ Ax = b \\ x \geq 0 \end{array}$$

donde  $A$  es matriz  $m \times n$ ;  $b$ , vector columna de  $m$  componentes;  $c$ , vector hileras de  $n$  componentes;  $y$ ,  $x$  vector de  $n$  variables.

Suponga que la base  $B$  consiste de las primeras  $m$  columnas de  $A$ .

Suponga que particionamos  $A$ ,  $x$  y  $c$  como sigue:

$$A = [B, R] \quad ; \quad x^t = [x_B^t, x_R^t] \quad ; \quad c = [c_B, c_R]$$

Entonces, el problema lineal es equivalente a:

$$\text{minimiza } z = c_B x_B + c_R x_R$$

$$Bx_B + Rx_R = b$$

$$x_B \geq 0 ; x_R \geq 0$$

y una solución factible básica asociada con la base B es  $x^t = (x_B, x_R) = (B^{-1}b, 0)$  con valor de la función objetivo  $z_0 = c_B x_B = c_B B^{-1}b$ . Sin embargo, para valores de  $x_R \neq 0$ , el vector  $x_B$  está dado por  $x_B = B^{-1}b - B^{-1}Rx_R$ . De donde, sustituyendo  $x_B$  en  $z$  tenemos

$$z = c_B (B^{-1}b - B^{-1}Rx_R) + c_R x_R$$

$$= c_B B^{-1}b + (c_R - c_B B^{-1}R)x_R$$

$$= z_0 + (c_R - c_B B^{-1}R)x_R$$

y se observa que, si el vector de costos relativos o reducidos  $\bar{c}_R = c_R - c_B B^{-1}R \geq 0$ , entonces

$$z_0 \leq c_B B^{-1}b + (c_R - c_B B^{-1}R)x_R = z$$

y por lo tanto, la solución básica original  $(x_B, 0)$  es óptima. Este resultado se conoce como: Principio de Optimalidad.

Al expresar el valor de la función objetivo en términos de las variables no básicas  $x_R$ , podemos determinar si alguna de ellas debe entrar a la base, o, si el proceso de búsqueda de una solución óptima ha llegado a su fin. Esto es, si la

$i$ -ésima componente del vector de costos relativos es  $\bar{c}_{R_j} < 0$  entonces, para valores de  $x_j$  positivos tenemos que  $\bar{c}_{R_j} x_j < 0$ . De aquí  $z_0 > z$  y por lo tanto, la solución básica anterior no es la mínima y debe considerarse una nueva solución donde  $x_j$  esté en la base.

### Descripción del Método Simplex Revisado

Suponga que B es una base del problema lineal y que la solución  $x = (x_B, 0) = (B^{-1}b, 0)$  es una solución factible.

1. Calcule  $\lambda = c_B B^{-1}$ . Calcule el vector de costos relativos  $\bar{c}_R = c_R - \lambda R$ . Si  $\bar{c}_R \geq 0$  la solución es óptima.
2. Determine un vector  $a_j$  que entre a la base. Este será aquél con costo relativo más negativo. Calcule  $y = B^{-1}a_j$  que expresa el vector  $a_j$  en términos de la base actual.
3. Determine el vector que sale de la base, esto es, si  $\bar{b} = B^{-1}b$ ,  $y = B^{-1}a_j$  determine el índice  $k$  tal que

$$b_k / y_k = \min \{ b_i / y_i ; \text{si } y_i > 0 \}$$

donde  $b_i$  ( $y_i$ ) denota el  $i$ -ésimo elemento del vector  $\bar{b}$  ( $y$ ).

Finalmente, actualice la base (y su inversa) y regrese a 1.

Ejemplo 1. Considere el problema lineal

$$\text{minimice } z = -20x_1 - 10x_2 - x_3$$

$$3x_1 + 2x_2 + 10x_3 \leq 10$$

$$2x_1 + 4x_2 + 20x_3 \leq 15$$

$$x_1 \geq 0 ; x_2 \geq 0 ; x_3 \geq 0$$

Para resolver este problema usando el método simplex revisado, se introducen variables de holgura y se obtiene

$$\text{minimice } z = -20x_1 - 10x_2 - x_3$$

$$3x_1 + 2x_2 + 10x_3 + x_4 = 10$$

$$2x_1 + 4x_2 + 20x_3 + x_5 = 15$$

$$x_1 \geq 0 ; x_2 \geq 0 ; x_3 \geq 0 ; x_4 \geq 0 ; x_5 \geq 0$$

cuya tableau inicial es:

$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$b$
3	2	10	1	0	10
2	4	20	0	1	15
-20	-10	-1	0	0	0

Una base factible inicial puede identificarse inmediatamente del tableau simplex. Por lo tanto:

Iteración 1.  $B = [a_4, a_5]$  y  $c_B = [0, 0]$ . Asimismo,

	$B^{-1}$	$B^{-1}b$	$B^{-1}a_1$
$x_4$	1	0	10
$x_5$	0	1	15

Entonces  $\lambda = c_B B^{-1} = [0, 0]$  y puesto que  $R = [a_1, a_2, a_3]$  se tiene que  $c_R = [-20, -10, -1]$ . Asimismo,

$$\bar{c}_R = c_R - \lambda R = [-20, -10, -1]$$

Seleccionando al vector  $a_1$  para entrar a la base se tiene de la comparación de  $B^{-1}a_1$  con  $B^{-1}b$  que el vector que sale de la base es  $a_4$ , pues  $\min \{ 10/3, 15/2 \} = 10/3$ .

Iteración 2. Usando  $B = [a_1, a_5]$  se tiene que  $c_B = [-20, 0]$  y podemos determinar la inversa de la base B, usando el elemento pivote indicado en la iteración anterior.

	$B^{-1}$	$B^{-1}b$
$x_1$	1/3	10/3
$x_5$	-2/3	25/3

de donde  $\lambda = c_B B^{-1} = [-20/3, 0]$  y dado que  $R = [a_2, a_3]$  se tiene que  $c_R = [0, -10, -1]$ . Asimismo,

$$\bar{c}_R = c_R - \lambda R = [20/3, 10/3, 197/3].$$

Puesto que  $\bar{c}_R \geq 0$ , la solución óptima es  $x_1^* = 10/3$ ;  $x_5^* = 25/3$ ;  $x_2^* = x_3^* = x_4^* = 0$ , y la función objetivo es  $z^* = -200/3$ .



Ejemplo 2. Considere el problema lineal\*

$$\text{maximizar } z = 2x_1 + x_2 - 3x_3 + 5x_4$$

$$x_1 + 2x_2 + 4x_3 - x_4 \leq 6$$

$$2x_1 + 3x_2 - x_3 + x_4 \leq 12$$

$$x_1 + x_3 + x_4 \leq 4$$

$$x_1 \geq 0 ; x_2 \geq 0 ; x_3 \geq 0 ; x_4 \geq 0$$

Resolveremos este problema usando el criterio de optimalidad correspondiente, esto es, la solución óptima deberá satisfacer  $\bar{c}_R = c_R - \lambda R \leq 0$ . Para ello note que el tableau inicial es:

$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$b$
1	2	4	-1	1	0	0	6
2	3	-1	1	0	1	0	12
1	0	1	1	0	0	1	4
2	1	-3	5	0	0	0	0

y que una base inicial se tiene en forma inmediata, específicamente  $B = [a_5, a_6, a_7]$ . Por lo que podemos empezar a resolver el problema con el método simplex revisado.

Iteración 1. Sea  $B = [a_5, a_6, a_7]$  y  $c_B = [0, 0, 0]$

	$B^{-1}$			$B^{-1}b$	$B^{-1}a_4$
$x_5$	1	0	0	6	-1
$x_6$	0	1	0	12	1
$x_7$	0	0	1	4	1

De donde,  $\lambda = c_B B^{-1} = [0, 0, 0]$  y  $R = [a_1, a_2, a_3, a_4]$ . De aquí que  $c_R = [2, 1, -3, 5]$  y  $\bar{c}_R = c_R - \lambda R = [2, 1, -3, 5]$ .

Por lo tanto, se selecciona el vector  $a_4$  para entrar a la base y de la comparación de  $B^{-1}b$  con  $B^{-1}a_4$  se obtiene que el vector que sale de la base es  $a_7$  pues:

$$\min\{12/1, 4/1\} = 4$$

Iteración 2. Usando  $B = [a_5, a_6, a_4]$  se tiene que  $c_B = [0, 0, 5]$  y la inversa de la nueva base  $B$  puede obtenerse usando el elemento pivote indicado en la iteración anterior.

	$B^{-1}$			$B^{-1}b$	$B^{-1}a_2$
$x_5$	1	0	1	10	2
$x_6$	0	1	-1	8	3
$x_4$	0	0	1	4	0

de donde  $\lambda = c_B B^{-1} = [0, 0, 5]$  y  $R = [a_1, a_2, a_3, a_7]$ . Entonces  $c_R = [2, 1, -3, 0]$  y  $\bar{c}_R = c_R - \lambda R = [-3, 1, -8, -5]$ .

Por lo tanto, se selecciona el vector  $a_2$  para entrar a la base y de la comparación  $B^{-1}b$  con  $B^{-1}a_2$  se obtiene que  $a_6$  sale de la base ya que

$$\min \{10/2, 8/3\} = 8/3$$

Iteración 3. La nueva base es  $B = [a_5, a_2, a_4]$  y  $c_B = [0, 1, 5]$ . Se tiene la inversa de esta nueva base usando el elemento pivota indicado en la iteración anterior.

	$B^{-1}$			$B^{-1}b$
$x_5$	1	-2/3	5/3	14/3
$x_2$	0	1/3	-1/3	8/3
$x_4$	0	0	1	4

de donde  $\lambda = c_B B^{-1} = [0, 1/3, 14/3]$  y  $R = [a_1, a_6, a_3, a_7]$ .

Entonces  $c_R = [2, 0, -3, 0]$  y  $\bar{c}_R = c_R - \lambda R = [-10/3, -1/3, -22/3, -14/3]$

Puesto que  $\bar{c}_R \leq 0$  la solución óptima es

$x_2^* = 8/3$ ,  $x_4^* = 4$ ,  $x_5^* = 14/3$ ,  $x_1^* = x_3^* = x_6^* = x_7^* = 0$  y el valor de la función objetivo es  $z^* = 68/3$ . Los valores duales asociados a la solución óptima son:  $\lambda_1 = 0$ ,  $\lambda_2 = 1/3$  y  $\lambda_3 = 14/3$ .

### 2.4 Análisis post-óptimo

Este problema surge del interés por conocer la forma en que varía la solución óptima de un problema de programación lineal, al cambiar algunos de los parámetros usados. El aspecto importante de esta problemática es que, frecuentemente, se requiere un mínimo de cálculos adicionales para obtener la nueva solución óptima sin tener que resolver el problema totalmente. Las operaciones requeridas para este análisis se conocen con el nombre de análisis de sensibilidad ó post-óptimo.

El interés de este análisis se enfoca a establecer las condiciones bajo las cuales la base óptima se conserva al variar el vector de recursos y el vector de costos. Asimismo, se desea establecer el costo que representa para la solución óptima la introducción de variables no-básicas. Con el propósito de tener un marco para el análisis del comportamiento del problema lineal considere:

$$\begin{aligned} \text{minimize } z &= cx \\ \text{(P)} \quad Ax &= b \\ x &\geq 0 \end{aligned}$$

Suponga que  $A = [B, R]$ , donde  $B$  es una matriz invertible,  $c = [c_B, c_R]$  y  $x^t = [x_B^t, x_R^t]$ . Entonces usando la solución básica  $x$  asociada con  $B$  podemos transformar (P) en:

$$\text{minimice } z = c_B x_B + c_R x_R$$

$$Bx_B + Rx_R = b$$

$$x_B \geq 0 ; x_R \geq 0$$

Suponiendo que la solución básica  $x = [x_B, x_R] = [B^{-1}b, 0]$  es óptima, las condiciones de optimalidad son:

$$a. \quad x_B = B^{-1}b \geq 0$$

$$b. \quad c_R - c_R B^{-1} \geq 0$$

donde se observa que la factibilidad del problema primal no depende del vector de costos  $c$ ; de la misma manera, la factibilidad del problema dual no depende del vector de requerimientos  $b$ . Estas observaciones son la base del análisis del problema de variación de datos, pues lo que se buscará es satisfacer las mismas condiciones de optimalidad al cambiar parámetros.

#### a. Cambio en el vector de requerimientos

Suponga que en el problema original se cambia  $b$  por el vector  $b + ad$  donde  $a$  es un escalar y  $d$  es un vector columna de  $m$  componentes. Analizaremos las condiciones bajo las cuales la base óptima  $B$  sigue siendo óptima. Para la discusión supondremos que la solución óptima original es no degenerada, esto es,  $x_B = B^{-1}b > 0$ . Note que la condición necesaria y suficiente para que  $B$  siga siendo óptima es que

$$B^{-1}(b + ad) \geq 0$$

Sin embargo esto se cumple, al menos, para valores de  $a$  pequeños pues  $B^{-1}b > 0$ . En este caso el cambio en la función objetivo es:

$$\Delta z = c_B B^{-1} [b + \Delta b] - c_B B^{-1} b = c_B B^{-1} \Delta b = \lambda \Delta b$$

y se concluye que la solución del problema dual  $\lambda$  es una medida de la rapidez de variación de los valores óptimos de la función objetivo del primal respecto a los requerimientos.

Si definimos  $y = B^{-1}d$  se puede verificar que los valores de  $a$  para los cuales la base óptima  $B$  se conserva están dados por el intervalo  $[\underline{a}, \bar{a}]$  en que

$$\bar{a} = \min_i \{-x_{Bi}/y_i ; y_i < 0\}$$

$$\underline{a} = \max_i \{-x_{Bi}/y_i ; y_i > 0\}$$

donde  $x_{Bi}$  es la componente  $i$ -ésima del vector  $x_B$ . Esto significa que para valores fuera de este intervalo la solución básica dada por  $x_B = B^{-1}b + ay$ , deja de ser factible (y óptima) del problema modificado. En particular, si se tiene que  $y = B^{-1}d \geq 0$  la solución básica  $x_B = B^{-1}b + ay$  es óptima para toda  $a \geq 0$ . Análogamente, si  $y \leq 0$  la solución  $x_B$  es óptima para toda  $a \leq 0$ .

Ejemplo 1. Considere el problema

$$\begin{aligned} \text{minimice } z &= -20x_1 - 10x_2 - x_3 \\ 3x_1 + 2x_2 + 10x_3 + x_4 &= 10 \\ 2x_1 + 4x_2 + 20x_3 + x_5 &= 15 \\ x_1 \geq 0; x_2 \geq 0; x_3 \geq 0; x_4 \geq 0; x_5 \geq 0. \end{aligned}$$

cuyo tableau simplex inicial es

$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_b$
3	2	10	1	0	10
2	4	20	0	1	15
-20	-10	-1	0	0	0

y el tableau simplex final es

$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$b$
1	2/3	10/3	1/3	0	10/3
0	8/3	40/3	-2/3	1	25/3
0	10/3	197/3	-20/3	0	200/3

Suponga que deseamos encontrar el intervalo de variación de la primera componente del vector  $b^t = [10, 5]$  sin que cambie la base óptima ya determinada. Asimismo, calcule los valores de la función objetivo para los límites de ese intervalo.

La inversa de la base óptima es

$$B^{-1} = \begin{bmatrix} 1/3 & 0 \\ -2/3 & 1 \end{bmatrix}$$

Considere ahora el cambio del vector  $b$  por  $b + \alpha e_1$  donde  $\alpha$  es un escalar y  $e_1 = [1, 0]$ . Para determinar el intervalo de variación de  $\alpha$  hagamos

$$y = B^{-1}e_1 = \begin{bmatrix} 1/3 & 0 \\ -2/3 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/3 \\ -2/3 \end{bmatrix}$$

El valor máximo de  $\alpha$ , denotado  $\bar{\alpha}$ , es tal que

$$x_B = B^{-1}b + \alpha y = \begin{bmatrix} 10/3 \\ 25/3 \end{bmatrix} + \alpha \begin{bmatrix} 1/3 \\ -2/3 \end{bmatrix} \geq 0$$

de donde  $\bar{\alpha} = (25/3)/(2/3) = 25/2$ . El valor mínimo de  $\alpha$  es  $\underline{\alpha} = (-10/3)/(1/3) = -10$ . Por lo tanto la base asociada a la solución óptima original no cambia si el valor de la primera componente de  $b$  está en el intervalo  $[0, 45/2]$ . Los valores de la función objetivo cuando  $b_1$  toma estos valores extremos son:

$$z(0) = 0 \quad ; \quad z(45/2) = -150$$

b. Cambio en el vector de costos

Suponga que en el problema original se cambia el vector de costos  $c$  por  $c + \beta h$ , donde  $\beta$  es un escalar y  $h$  un vector hilera de  $n$  componentes. Analizaremos las condiciones bajo las cuales la base óptima  $B$  sigue siendo óptima. Supondremos para el análisis que la solución óptima del problema dual es no-degenerada, esto es, el sistema  $\lambda A \leq c$ , donde  $A = [B, R]$  y  $c = [c_B, c_R]$  puede expresarse como:

$$\lambda B = c_B \quad \text{y} \quad \lambda R < c_R$$

que indica que las primeras  $m$  desigualdades del sistema se satisfacen con igualdad y las  $n-m$  restantes con estricta desigualdad. Sin embargo, puesto que  $\lambda = c_B B^{-1}$ , la solución óptima no-degenerada implica que

$$c_R - c_B B^{-1} R = c_R - \lambda R > 0$$

Por otra parte, la condición necesaria y suficiente para que la base  $B$  siga siendo óptima es que se cumpla

$$[c_R + \Delta c_R] - [c_B + \Delta c_B] B^{-1} R \geq 0$$

donde  $[\Delta c_B, \Delta c_R] = \Delta c = \beta h$  es la variación propuesta al vector de costos. Esta expresión equivale a

$$[c_R - c_B B^{-1} R] + [\Delta c_R - \Delta c_B B^{-1} R] \geq 0$$

y se satisface para el caso que se analiza con valores de  $\beta$  suficientemente pequeños, debido a que la solución óptima del

problema dual es no-degenerada. Si la base  $B$  sigue siendo óptima el vector  $\lambda = [c_B + \Delta c_B] B^{-1}$  es la solución óptima del dual y el cambio en la función objetivo es

$$\Delta w = [c_B + \Delta c_B] B^{-1} b - c_B B^{-1} b = \Delta c_B B^{-1} b$$

que equivale a  $\Delta w = \Delta c_B x_B + \Delta c_R x_R = \Delta c x$  pues  $x_R = 0$  y  $x_B = B^{-1} b$ . De donde se concluye que la solución óptima del problema primal representa la rapidez de cambio de la función objetivo del problema dual respecto al vector de costos.

Si  $\bar{c}_R = [c_R - c_B B^{-1} R] > 0$  y  $s = h_R - h_B B^{-1} R$  puede verificarse que los límites de  $\beta$  para los cuales la base  $B$  no cambia quedan dados por el intervalo  $[\underline{\beta}, \bar{\beta}]$  donde

$$\bar{\beta} = \min \{ -\bar{c}_{Rj} / s_j \ ; \ s_j < 0 \}$$

$$\underline{\beta} = \max \{ -\bar{c}_{Rj} / s_j \ ; \ s_j > 0 \}$$

quiere  $\bar{c}_{Rj}$  es la componente  $i$ -ésima de  $\bar{c}_R$  (y lo mismo sucede con  $s_j$ ). Finalmente note que si  $s \geq 0$  el vector solución  $\lambda = [c_B + \beta h_B] B^{-1}$  es óptimo para toda  $\beta \geq 0$ . Análogamente, si  $s \leq 0$ , la solución  $\lambda$  es óptima para toda  $\beta \leq 0$ .

**Ejemplo 2.** Determine el intervalo de variación de la primera componente del vector de costos del ejemplo 1 (de esta sección) sin que cambie la base óptima original.

En el ejemplo 1 se estableció que

$$B^{-1} = \begin{bmatrix} 1/3 & 0 \\ -2/3 & 1 \end{bmatrix}$$

donde  $B = [a_1, a_5]$  y  $R = [a_2, a_3, a_4]$ . Asimismo, se verifica que  $c_B = [-20, 0]$  y  $c_R = [-10, -1, 0]$ . De donde para determinar el intervalo de variación de la primer componente de  $c_B$ , que coincide con la primer componente de  $c$ , define el vector

$h = [1, 0, 0, 0, 0]$ . Entonces  $h_B = [1, 0]$ ,  $h_R = [0, 0, 0]$  y

$$\bar{c}_R = c_R - c_B B^{-1} R = [10/3, 197/3, 20/3]$$

$$s = h_R - h_B B^{-1} R = [-2/3, -10/3, -1/5]$$

Por lo tanto la cota superior de  $\theta$  es dada por

$$\bar{\theta} = \min \left\{ \frac{-10/3}{-2/3}, \frac{-197/3}{-10/3}, \frac{-20/3}{-1/5} \right\} = 5$$

Por otra parte  $\theta \leq 0$  que implica  $\theta_{\min} = -\infty$ . De aquí se concluye que el intervalo de variación de la primera componente de  $c$ , sin que cambie la base, es  $(-\infty, -15)$ . De donde tenemos que:

$$z(-\infty) = -\infty, \quad z(-15) = -50$$

**Ejemplo 3.** Considere el problema

$$\text{maximizar } z = 2x_1 + x_2 - 3x_3 + 5x_4$$

$$\text{sujeto a } x_1 + 2x_2 + 4x_3 - x_4 \leq 6$$

$$2x_1 + 3x_2 - x_3 + x_4 \leq 12$$

$$x_1 + x_3 + x_4 \leq 4$$

$$x_1 \geq 0; x_2 \geq 0; x_3 \geq 0; x_4 \geq 0; x_5 \geq 0; x_6 \geq 0; x_7 \geq 0.$$

cuyo tableau inicial, con variables de holgura es:

$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$b$
1	2	4	-1	1	0	0	6
2	3	-1	1	0	1	0	12
1	0	1	1	0	0	1	4
2	1	-3	5	0	0	0	0

La base óptima  $B = [a_5, a_2, a_4]$  y su inversa son:

$$B = \begin{bmatrix} 1 & 2 & -1 \\ 0 & 3 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad B^{-1} = \begin{bmatrix} 1 & -2/3 & 5/3 \\ 0 & 1/3 & -1/3 \\ 0 & 0 & 1 \end{bmatrix}$$

Se desea determinar el intervalo de variación de cada una de las componentes de los vectores de requerimientos y de costos, sin que cambie la base óptima original. Asimismo, calcule los valores de la función objetivo para los puntos límites de dicho intervalo.

Primero observe que la solución es dada por

$$x_B^* = [x_5, x_2, x_4] = [14/3, 8/3, 4]$$

y que  $x_R^* = [x_1, x_3, x_6, x_7] = [0, 0, 0, 0]$ . Asimismo, la solución del problema dual es  $\lambda^* = [0, 1/3, 14/3]$ . Si deseamos determinar el intervalo de variación de  $b_1$ , la primera componente de  $b$ , sin que cambie la base, sea:

$$\underline{b} = b + \alpha e_1^t = \begin{bmatrix} 6 \\ 12 \\ 4 \end{bmatrix} + \alpha \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

Para que  $B$  siga siendo óptima se debe cumplir que:

$$B^{-1}\underline{b} = \begin{bmatrix} 14/3 \\ 8/3 \\ 4 \end{bmatrix} + \alpha \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \geq 0$$

esto es,  $B^{-1}\underline{b} \geq 0$ . Sin embargo, esto es cierto cuando

$\alpha \geq -14/3$ , que equivale a tener  $b_1 \geq 6 - 14/3 = 4/3$ . Asimismo,

$$x_B = B^{-1}\underline{b} = \begin{bmatrix} x_5 \\ x_2 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 & -2/3 & 5/3 \\ 0 & 1/3 & -1/3 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ 12 \\ 4 \end{bmatrix} = \begin{bmatrix} b_1 - 4/3 \\ 8/3 \\ 4 \end{bmatrix}$$

En el valor mínimo,  $x_B^t = [0, 8/3, 4]$ , por lo tanto  $z = 68/3$ .

Procediendo de manera semejante se tiene que el intervalo de variación de la segunda componente de  $b$  es:

$$4 \leq b_2 \leq 19$$

y además los valores de  $x_B$  en función de  $b_2$  son:

$$x_B = \begin{bmatrix} x_5 \\ x_2 \\ x_4 \end{bmatrix} = \begin{bmatrix} 38/3 - 2/3 b_2 \\ b_2/3 - 4/3 \\ 4 \end{bmatrix}$$

por lo tanto  $z(b_2) = b_2/3 + 56/3$ ;  $z(4) = 20$  y  $z(19) = 25$

Análogamente se obtiene que el intervalo de variación de  $b_3$  es:

$$6/5 \leq b_3 \leq 12$$

y los valores  $x_B$  en función de  $b_3$  son:

$$x_B = \begin{bmatrix} x_5 \\ x_2 \\ x_4 \end{bmatrix} = \begin{bmatrix} 5/3 b_3 - 2 \\ 4 - b_3/3 \\ b_3 \end{bmatrix}$$

de donde  $z(b_3) = \frac{14}{3} b_3 + 4$ ;  $z(6/5) = 9.6$  y  $z(12) = 60$ .

Para analizar la variación de los coeficientes del vector de costos debemos especificar, primeramente, si son elementos de

$$c_B = [c_5, c_2, c_4] \quad \text{o} \quad c_R = [c_1, c_6, c_3, c_7]$$

Suponga que se desea variar un elemento de  $c_R$ . Entonces la condición de optimalidad que debe satisfacerse es:

$$\bar{c}_R = c_R - c_B B^{-1} R \leq 0$$

donde  $c_R = c_R + \alpha e_i$ ,  $i=1,3,6,7$ . Note que  $c_R$  es el nuevo vector de costos no-básicos. En particular, si  $i=1$ , tenemos que:

$$\bar{c}_R = (\alpha - 10/3, -1/3, -22/3, -14/3) \leq 0$$

de donde  $\alpha \leq 10/3$ , lo cual es equivalente a  $c_1 \leq 16/3$ . Además, el valor de la función objetivo cuando  $c_1$  toma este valor extremo no se altera porque la variable  $x_1$  no está en la base. De manera análoga, se obtiene que  $c_3 \leq 13/3$  y para los coeficientes  $c_6$  y  $c_7$ , correspondientes a las variables de holgura, resulta que  $c_6 \leq 1/3$  y  $c_7 \leq 14/3$ .

Si la componente a variar es un elemento de  $c_B$  entonces la condición de optimalidad que se debe satisfacer es:

$$\bar{c}_R = c_R - c_B B^{-1} R \leq 0$$

donde  $c_B = c_B + \alpha e_i$ ,  $i=2,4,5$ . En particular, si  $i=2$  se tiene que  $c_B = (0, 1+\alpha, 5)$  y que

$$0 \geq \bar{c}_R = -\frac{1}{3} [10 + \alpha, 1 + \alpha, 22 - 2\alpha, 14 - \alpha]$$

que implica  $-1 \leq \alpha \leq 11$  y equivalentemente  $0 \leq c_2 \leq 12$ . Además, si se sustituye el coeficiente  $c_2=1$  de la función objetivo por estos valores extremos se obtiene en cada caso:

$$z(0) = 20.0 \quad \text{y} \quad z(12) = 52.0$$

De manera análoga, resulta que  $c_4 \geq 5/3$  y sustituyendo este valor extremo en la función objetivo nos queda:

$$z(5/3) = 28/3$$

Para el coeficiente  $c_5$  que corresponde a la variable de holgura básica  $x_5$ , el proceso es el mismo y se obtiene que:

$$-\frac{22}{19} \leq c_5 \leq \frac{1}{2}$$

los valores de la función objetivo al sustituir estos extremos son:

$$z(-22/19) = 17.26316 \quad \text{y} \quad z(1/2) = 25.$$



c. Introducción de variables no-básicas

Otro aspecto importante del análisis post-óptimo es la evaluación de los cambios en la función objetivo debidos a la introducción de variables no-básicas en la solución óptima del problema de programación lineal. Específicamente, sea B la base óptima del problema lineal cuya solución general es:

$$x_B = B^{-1}b - B^{-1}R x_R \geq 0 \quad (1)$$

donde  $x_B$  es óptima cuando  $x_R = 0$ . Suponga que la variable no-básica  $x_1$  puede tener valores distintos de cero y que las otras variables no-básicas permanecen con valor igual a cero. Se desea determinar el intervalo de variación de dicha variable no-básica sin que cambie la base óptima. La consecuencia de que la variable no básica tenga valores fuera de este intervalo es que al menos una variable básica se vuelve no factible. El proceso de determinación de la primera variable básica que se hace no factible en cada extremo se denomina "LIMITING PROCESS" en los listados de computadora. Si la variable básica en cuestión es de holgura, la pérdida de factibilidad es equivalente a la no factibilidad de la restricción en que se encuentra dicha variable de holgura.

Finalmente se desea determinar el costo unitario de disminución de beneficios debidos a la introducción de esta variable y los correspondientes valores de la función objetivo en los extremos de dicho intervalo.

Ejemplo 4. Considere el problema:

$$\begin{aligned} \text{Maximizar } z &= 2x_1 + x_2 - 3x_3 + 5x_4 \\ \text{sujeto a} \quad &x_1 + 2x_2 + 4x_3 - x_4 \leq 6 \\ &2x_1 + 3x_2 - x_3 + x_4 \leq 12 \\ &x_1 + x_3 + x_4 \leq 4 \\ &x_1 \geq 0 ; x_2 \geq 0 ; x_3 \geq 0 ; x_4 \geq 0 \end{aligned}$$

cuyo tableau inicial con variables de holgura es:

$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	b
1	2	4	-1	1	0	0	6
2	3	-1	1	0	1	0	12
1	0	1	1	0	0	1	4
2	1	-3	5	0	0	0	0

y cuyo tableau óptimo es:

$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$B^{-1}b$
4/3	0	19/3	0	1	-2/3	5/3	14/3
1/3	1	-2/3	0	0	1/3	-1/3	8/3
1	0	1	1	0	0	1	4

Determine los intervalos de variación de las variables no-básicas sin que cambie la base. Asimismo, determine los costos unitarios de disminución de beneficios debidos a la introducción de las variables no-básicas en la solución óptima y los correspondientes valores de la función objetivo en los extremos de dichos intervalos.

Para ello, suponga que la variable  $x_1$  puede tener valores distintos de cero y que las otras variables no-básicas permanecen con valor igual a cero. Entonces, sustituyendo valores en la relación (1) se tiene que:

$$x_B = \begin{bmatrix} x_5 \\ x_2 \\ x_4 \end{bmatrix} = \begin{bmatrix} 14/3 \\ 8/3 \\ 4 \end{bmatrix} - \begin{bmatrix} 4/3 \\ 1/3 \\ 1 \end{bmatrix} x_1$$

Los valores de  $x_1$  para los cuales  $x_B \geq 0$  son dados por  $x_1 \leq 7/2$ . Note que cuando  $x_1 > 7/2$  se tiene  $x_5 < 0$  que es equivalente a que la restricción uno no se cumpla pues

$$x_1 + 2x_2 + 4x_3 - x_4 = 6 - x_5$$

En este caso se dice que la restricción uno es la que limita el que la variable  $x_1$  pueda tener valores mayores que  $7/2$ .

Por otra parte observe que el valor de la función objetivo al considerarse  $x_1$  en la solución óptima es

$$Z = 2x_1 + (8/3 - 1/3x_1) + 5(4 - x_1) = 68/3 - 10/3x_1$$

De donde el costo unitario debido a la introducción de  $x_1$  es igual a  $10/3$  y los límites de la función objetivo para los valores extremos de  $x_1$  son  $Z = -\infty$  y  $Z = 11$ .

Considere ahora la introducción de la variable no-básica  $x_3$  en la solución óptima. Primero conviene analizar el intervalo de valores que puede tener  $x_3$  sin que cambie la base. Sea

$$x_B = \begin{bmatrix} x_5 \\ x_2 \\ x_4 \end{bmatrix} = \begin{bmatrix} 14/3 \\ 8/3 \\ 4 \end{bmatrix} - \begin{bmatrix} 19/3 \\ -2/3 \\ 1 \end{bmatrix} x_3$$

De donde los valores de  $x_3$  tales que  $x_B \geq 0$  están dados por  $-4 \leq x_3 \leq 14/19$ . Note que si  $x_3 < -4$  entonces la variable básica  $x_2$  deja de ser factible; si  $x_3 > 14/19$  entonces  $x_5 < 0$  que equivale a decir que la restricción uno no se satisface. En la operación "LIMITING PROCESS" se dirá que  $x_2$  y la restricción uno limitan la variación de  $x_3$  más allá de los límites determinados para esta variable. Asimismo, se observa que el valor de la función objetivo al introducir  $x_3$  es

$$Z = 68/3 - 22/3x_3$$

que equivale a decir que el costo unitario de deterioro de la función objetivo es  $22/3$ . En particular, los valores extremos de esta función para los límites de variación de  $x_3$  son:

$$Z(-4) = 52 \quad ; \quad Z(14/19) = 17.2631$$

## 2.5 Parametrización

En el análisis post-óptimo se discute únicamente los cambios de datos del problema lineal sin que exista cambio de la base óptima. La parametrización proporciona una visión más amplia, pues nos permite analizar el comportamiento de la solución óptima y el valor de la función objetivo, aún cuando existan cambios en la base óptima. Este análisis permite conocer la influencia de cada una de las componentes del vector de recursos o de costos en la función objetivo. La curva resultante se denomina función de beneficios y es importante para estudios económicos más globales, en donde se distribuyen recursos escasos a unidades productivas. Un ejemplo típico de esto, en el caso de planeación agrícola, se tiene al determinar los beneficios económicos de un distrito de riego en la función del volumen de agua que se le asigne.

Considera los dos tipos más usuales de parametrización. El primero queda dado por

$$\begin{aligned} z(\alpha) &= \text{maximice } cx \\ Ax &= b + \alpha d \\ x &\geq 0 \end{aligned}$$

donde se desea saber la forma de la función  $z(\alpha)$  cuando varía  $\alpha$ . Se puede demostrar que la forma típica de esta función es:

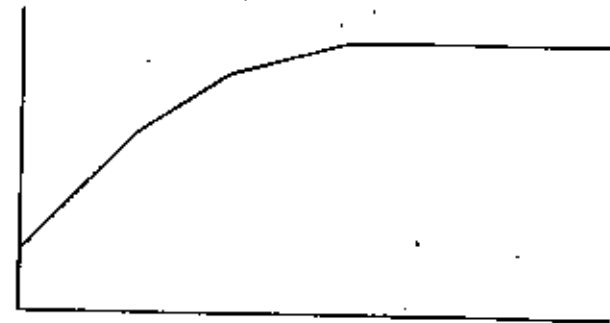


Fig. 2.1 Curva de variación del lado derecho.

El segundo tipo de problema está dado por:

$$\begin{aligned} z(\alpha) &= \text{max}(c + \alpha h)x \\ Ax &= b \\ x &\geq 0 \end{aligned}$$

En este caso puede demostrarse que la forma de la función  $z(\alpha)$  cuando varía  $\alpha$  es como sigue:

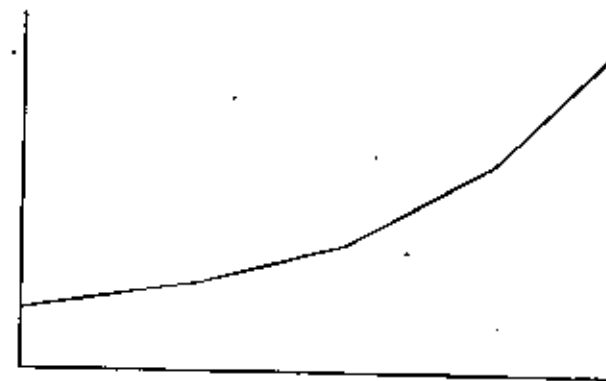


Fig. 2.2 Curva de variación de costos.

Ejemplo 1. Considere el problema:

$$\text{Maximizar } z = 2x_1 + x_2 - 3x_3 + 5x_4$$

$$\text{sujeto a } x_1 + 2x_2 + 4x_3 - x_4 \leq \alpha$$

$$2x_1 + 3x_2 - x_3 + x_4 \leq 12$$

$$x_1 + x_3 + x_4 \leq 4$$

$$x_1 \geq 0; x_2 \geq 0; x_3 \geq 0; x_4 \geq 0$$

cuyo tableau inicial con variables de holgura es:

$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$b$
1	2	4	-1	1	0	0	$\alpha$
2	3	-1	1	0	1	0	12
1	0	1	1	0	0	1	4
2	1	-3	5	0	0	0	0

Se desea determinar la curva de beneficios máximos de este problema en función del parámetro  $\alpha$ .

Iteración 1. Sean  $B = [a_5, a_6, a_7]$  y  $c_B = [0, 0, 0]$

	$B^{-1}$	$B^{-1}b$	$B^{-1}a_4$
$x_5$	1 0 0	$\alpha$	-1
$x_6$	0 1 0	12	1
$x_7$	0 0 1	4	(1)

Entonces  $\lambda = c_B B^{-1} = [0, 0, 0]$  y  $R = [a_1, a_2, a_3, a_4]$ . Asimismo, se tiene que  $c_R = [2, 1, -3, 5]$  y  $\bar{c}_R = c_R - \lambda R = [2, 1, -3, 5]$ .

Se selecciona el vector  $a_4$  para entrar a la base y el vector  $a_7$  sale de la base, ya que  $\min\{12/1, 4/1\} = 4$ .

Iteración 2. La nueva base es  $B = [a_5, a_6, a_4]$  y  $c_B = [0, 0, 5]$ .

	$B^{-1}$	$B^{-1}b$	$B^{-1}a_2$
$x_5$	1 0 1	$\alpha + 4$	(2)
$x_6$	0 1 -1	8	(3)
$x_4$	0 0 1	4	0

La inversa de  $B$  se calculó con el elemento pivote de la iteración anterior. Entonces  $\lambda = c_B B^{-1} = [0, 0, 5]$  y  $R = [a_1, a_2, a_3, a_7]$ . Asimismo,  $c_R = [2, 1, -3, 0]$  y  $\bar{c}_R = c_R - \lambda R = [-3, 1, -8, -5]$ . Se selecciona al vector  $a_2$  para entrar a la base y de la comparación de  $B^{-1}b$  y  $B^{-1}a_2$  se obtiene que:

$$\min\{(\alpha+4)/2, 8/3\} = \begin{cases} 8/3 & \text{si } \alpha \geq 4/3 \\ (\alpha+4)/2 & \text{si } \alpha \leq 4/3 \end{cases}$$

Iteración 3. ( $\alpha \leq 4/3$ ). La nueva base es  $B = [a_2, a_6, a_4]$  y  $c_B = [c_2, c_6, c_4] = [1, 0, 5]$ . Usando el elemento pivote (línea punteada) de la iteración 2 se tiene:

	$B^{-1}$			$B^{-1}b$
$x_2$	1/2	0	1/2	$2+\alpha/2$
$x_6$	-3/2	1	-5/2	$2-3\alpha/2$
$x_4$	0	0	1	4

Entonces  $\lambda = c_B B^{-1} = [1/2, 0, 11/2]$  y  $R = [a_1, a_6, a_3, a_7]$ . Asimismo  $c_R = [2, 0, -3, 0]$  y  $\bar{c}_R = c_R - \lambda R = [-4, 0, -21/2, -11/2]$ .

Por lo tanto, si  $-4 \leq \alpha \leq 4/3$ , la solución óptima es:

$x_2^* = 2+\alpha/2$ ,  $x_4^* = 4$ ,  $x_6^* = 2-3\alpha/2$ ,  $x_1^* = x_3^* = x_5^* = x_7^* = 0$  y el valor de la función objetivo es  $z^* = 22 + \alpha/2$ .

Iteración 3. ( $\alpha \geq 4/3$ ).  $B = [a_5, a_2, a_4]$  y  $c_B = [c_5, c_2, c_4]$ .

	$B^{-1}$			$B^{-1}b$
$x_5$	1	-2/3	5/3	$\alpha-4/3$
$x_2$	0	1/3	-1/3	8/3
$x_4$	0	0	1	4

de donde  $\lambda = c_B B^{-1} = [0, 1/3, 14/3]$  y  $R = [a_1, a_6, a_3, a_7]$ . Asimismo  $c_R = [2, 0, -3, 0]$  y  $\bar{c}_R = c_R - \lambda R = [-10/3, -1/3, -22/3, -14/3]$ .

Puesto que  $\bar{c}_R \leq 0$  se tiene que la solución óptima es

$x_2^* = 8/3$ ;  $x_4^* = 4$  y  $x_5^* = \alpha-4/3$ ,  $x_1^* = x_3^* = x_6^* = x_7^* = 0$  y el valor de la función objetivo es  $z^* = 68/3$ .

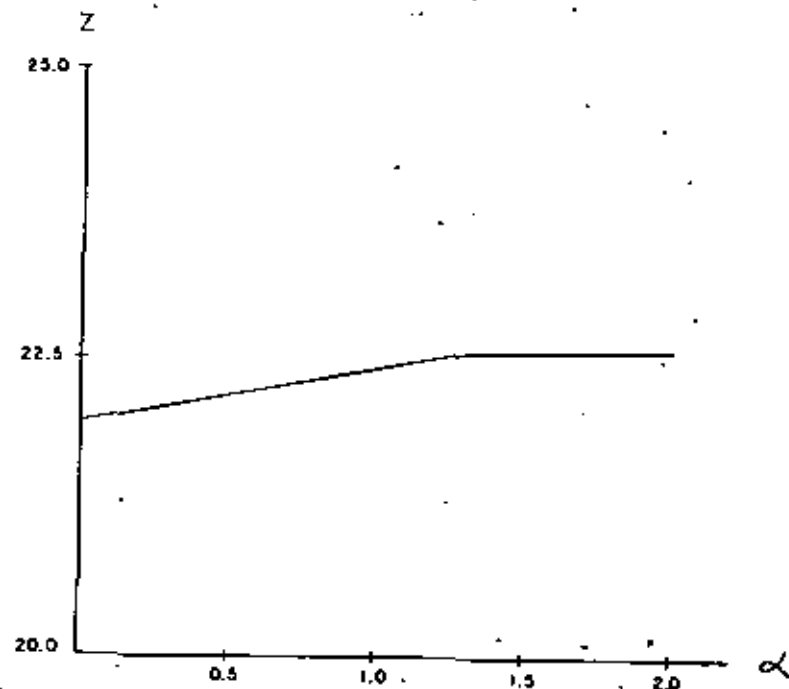


Fig. 2-3 Curva de variación del lado derecho.

Ejemplo 2. Considere el problema

$$\text{Maximizar } z = ax_1 + x_2 - 3x_3 + 5x_4$$

sujeto a

$$x_1 + 2x_2 + 4x_3 - x_4 \leq 6$$

$$2x_1 + 3x_2 - x_3 + x_4 \leq 12$$

$$x_1 + x_3 + x_4 \leq 4$$

$$x_1 \geq 0; x_2 \geq 0; x_3 \geq 0; x_4 \geq 0$$

y suponga que se desea determinar la curva de beneficios máxi-  
mos al variar el parámetro  $a$ . El tableau inicial con variables  
de holgura es

$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$b$
1	2	4	-1	1	0	0	6
2	3	-1	1	0	1	0	12
1	0	1	1	0	0	1	4
$a$	1	-3	5	0	0	0	0

Iteración 1: Sea  $B = [a_5, a_6, a_7]$  y  $c_B = [0, 0, 0]$

	$B^{-1}$	$B^{-1}b$	$B^{-1}a_4$
$x_5$	1 0 0	6	-1
$x_6$	0 1 0	12	1
$x_7$	0 0 1	4	1

de donde,  $\lambda = c_B B^{-1} = [0, 0, 0]$  y  $R = [a_1, a_2, a_3, a_4]$ . Asimismo,  
 $c_R = [a, 1, -3, 5]$  y  $\bar{c}_R = c_R - \lambda R = [a, 1, -3, 5]$ . Se selecciona el  
vector  $a_4$  para entrar a la base y sale de la base el vector  $a_7$ .

Iteración 2. Sean  $B = [a_5, a_6, a_4]$  y  $c_B = [0, 0, 5]$ .

	$B^{-1}$	$B^{-1}b$	$B^{-1}a_2$
$x_5$	1 0 1	10	2
$x_6$	0 1 -1	8	3
$x_4$	0 0 1	4	0

donde la inversa de  $B$  se obtuvo con el elemento pivote indica-  
do en la iteración anterior. Se tiene  $\lambda = c_B B^{-1} = [0, 0, 5]$  y  
 $R = [a_1, a_2, a_3, a_7]$ . Asimismo  $c_R = [a, 1, -3, 0]$  y tenemos que  
 $\bar{c}_R = c_R - \lambda R = [a-5, 1, -8, -5]$ . Se selecciona el vector  $a_2$  para  
entrar a la base y el vector  $a_6$  sale de la base.

Iteración 3. Sea  $B = [a_5, a_2, a_4]$  y  $c_B = [0, 1, 5]$

	$B^{-1}$	$B^{-1}b$	$B^{-1}a_1$
$x_5$	1 -2/3 5/3	14/3	4/3
$x_2$	0 1/3 -1/3	8/3	1/3
$x_4$	0 0 1	4	1

De donde  $\lambda = c_B B^{-1} = [0, 1/3, 14/3]$  y  $R = [a_1, a_6, a_3, a_7]$ .

Entonces,  $c_R = [a, 0, -3, 0]$  y

$$\bar{c}_R = c_R - \lambda R = [a - 16/3, -1/3, -22/3, -14/3]$$

Si  $a \leq 16/3$  la solución es óptima y es:

$$x_2^* = 8/3 ; x_4^* = 4 ; x_5^* = 14/3 ; x_1^* = x_3^* = x_6^* = x_7^* = 0$$

con valor de la función objetivo  $z(a) = 68/3$ .

Iteración 4. Si  $a \geq 16/3$  entonces el vector  $a_1$  entra a la base

y sale el vector  $a_5$ . Sean  $B = [a_1, a_2, a_4]$  y  $c_B = [a, 1, 5]$ .

	$B^{-1}$			$B^{-1}b$	$B^{-1}a_D$
$x_1$	3/4	-1/2	5/4	7/2	-1/2
$x_2$	-1/4	1/2	-3/4	3/2	1/2
$x_4$	-3/2	1/2	-1/4	1/2	1/2

de donde  $\lambda = c_B B^{-1} = [-4 + 3a/4, 3 - a/2, -2 + 5a/4]$  y se tiene que

$R = [a_5, a_6, a_3, a_7]$ . Entonces  $c_R = [0, 0, -3, 0]$  y

$$\bar{c}_R = c_R - \lambda R = [4 - 3a/4, a/2 - 3, 22/3 - 11a/4, 2 - 5a/4]$$

Si  $16/3 < a < 6$ , la solución es óptima y es:

$$x_1^* = 7/2 ; x_2^* = 3/2 ; x_4^* = 1/2 ; x_3^* = x_5^* = x_6^* = x_7^* = 0$$

cuyo valor de la función objetivo es  $z(a) = 7a/2 + 4$ .

Iteración 5. Si  $a \geq 6$  entonces el vector  $a_6$  entra a la base y

sale el vector  $a_4$ . Sean  $B = [a_1, a_2, a_6]$  y  $c_B = [a, 1, 0]$ .

	$B^{-1}$		$B^{-1}b$	
$x_1$	0	0	1	1
$x_2$	1/2	0	-1/2	1
$x_6$	-3/2	1	-1/2	1

Entonces  $\lambda = c_B B^{-1} = [1/2, 0, a - 1/2]$  y  $R = [a_5, a_4, a_3, a_7]$ .

Asimismo  $c_R = [0, 5, -3, 0]$  y

$$\bar{c}_R = c_R - \lambda R = [-1/2, 6 - a, -a - 9/2, -a + 1/2]$$

Si  $a \geq 6$  tenemos que  $\bar{c}_R \leq 0$  y la solución actual es óptima:

$$x_1^* = 4 ; x_2^* = 1 ; x_6^* = 1 ; x_3^* = x_4^* = x_5^* = x_7^* = 0$$

con valor de la función objetivo igual a  $z(a) = 4a + 1$ .

La gráfica de la función de beneficios máximos resultante al variar el parámetro  $a$  en el problema resuelto se proporciona en la siguiente hoja.

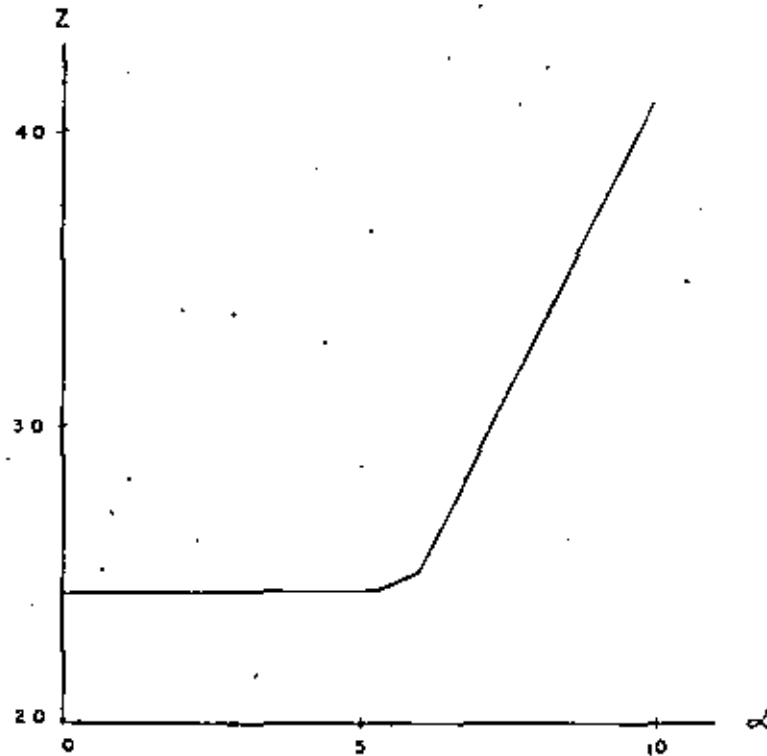


Fig. 2.4 Curva de variación del vector de costos.

### 1. PROBLEMA DE CONTRATACION.

La compañía Aeronaves del Pacífico necesita determinar cuántas aeromozas contratar y adiestrar en los próximos seis meses. Las necesidades de la compañía expresadas como horas-vuelo-aeromoza son:

Enero	febrero	marzo	abril	mayo	junio
8000	9000	7000	10000	9000	11000

El entrenamiento necesario para que una aeromoza dé servicio en un vuelo dura un mes; por lo que cada muchacha debe contratarse por lo menos un mes antes.

El entrenamiento necesita 100 horas de supervisión de aeromoza ya entrenadas, por lo que se dispone 100 horas-vuelo-aeromoza menos, durante un mes por cada aeromoza en entrenamiento. Cada aeromoza entrenada puede trabajar 150 horas en un mes y la compañía áerea tiene 60 aeromozas entrenadas al principio de enero.

Por razones sindicales, si el tiempo máximo disponible de las aeromozas entrenadas excede al requerido por la compañía en el mes (horas-vuelo y supervisión), éstas trabajarán menos de 150 horas y no se despide a nadie. Sin embargo, en cada mes aproximadamente 10% de las aeromozas con experiencia dejan el trabajo por razones de matrimonio u otras. Además, en cada mes, 5% de las personas que se contratan (y terminan su entrenamiento), son rechazadas por varias razones.



Si se consideran los salarios y otros beneficios, cada aeromoza adiestrada cuesta a la compañía \$8,000.00 mensuales, y cada aeromoza en entrenamiento, \$4,000.00.

La compañía desea determinar el plan de contratación y adiestramiento de aeromozas a costo mínimo.

Formulación:

Sea  $x_t$  número de personas contratadas al inicio del mes  $t$

$y_t$  número de aeromozas con entrenamiento al inicio del mes  $t$ .

$D_t$  número de horas-vuelo-aeromoza necesarias en el mes  $t$ .

En este caso el problema consiste en

$$\text{minimizar } z = 800 \sum_{t=1}^6 y_t + 4000 \sum_{t=1}^5 x_t$$

sujeto a

$$y_{t+1} = 0.95x_t + 0.9y_t \quad t=1,2,\dots,5$$

$$150y_t \geq D_t + 100x_t \quad t=1,2,\dots,6$$

$$y_1 = 50$$

$$x_t \geq 0 ; y_{t+1} \geq 0 \quad t=1,2,\dots,5.$$

## 2. PROBLEMA DE INVERSIONES

Un inversionista dispone de \$40,000.00 y desea establecer un plan de inversiones que maximice la cantidad de dinero que puede acumular al final de los próximos cinco años.

El inversionista dispone de varias actividades financieras.

En la actividad A, cada peso invertido al comienzo de un año produce \$ 1.50 (una ganancia de 0.50) dos años más tarde (en el momento preciso para una reinversión). En la actividad B, cada peso invertido al principio de un año le produce \$ 1.80 tres años después. Se tienen además dos actividades financieras: C y D, que estarán disponibles solamente una vez en el futuro. Cada peso invertido en C, en el comienzo del segundo año, le produce \$ 2.25 cuatro años más tarde. Finalmente, en la actividad D, cada peso invertido al principio del quinto año le produce \$ 1.30 un año más tarde.

Formulación:

Sea  $x_{ij}$  la cantidad de dinero invertido al principio del año  $i$  en la cantidad  $j$  ( $i=1,2,3,4,5$ ) ; ( $j=A,B,C$ ).

$w_i$ , la cantidad de dinero que no se invierte al principio del año  $i$ .  $i=1,2,3,4,5$ .

En este caso se desea maximizar el dinero acumulado al final del quinto año, esto es, se desea

$$\text{maximizar } z = 1.7x_{4A} + 1.8x_{4B} + 2.25x_{2C} + 1.30x_{5D}$$

sujeto a

$$x_{1A} + x_{1B} + w_1 = 40000$$

$$x_{2A} + x_{2B} + x_{2C} + w_2 = w_1$$

$$x_{3A} + x_{3B} + w_3 = w_2 + 1.5x_{1A}$$

$$x_{4A} + w_4 = w_3 + 1.5x_{2A} + 1.8x_{1B}$$

$$x_{5D} = w_4 + w_4 + 1.5x_{3A} + 1.8x_{2B}$$

$$x_{ij} \geq 0 \quad i=1,2,3,4,5 \quad j=A,B,C,D.$$

$$w_{ij} \geq 0 \quad i=1,2,3,4.$$

### 3. PROBLEMAS DE REGRESION

Un problema importante en el campo de la estadística es el denominado problema de regresión lineal. A grandes rasgos, éste consiste en determinar la línea recta que mejor representa un conjunto de datos estadísticos como  $(x_1, y_1), \dots, (x_n, y_n)$ . Suponiendo que la línea recta es dada por la ecuación  $y = a + bx$ , el propósito es determinar las constantes  $a$  y  $b$  que proporcionan la mejor representación de los datos de acuerdo a un criterio especificado. Algunos de los criterios, en que la programación lineal es útil, son:

a. minimice  $\sum_{i=1}^n |y_i - a - bx_i|$ .

b. minimice { máximo  $|y_i - a - bx_i|$  ;  $i=1,2,3, \dots, n$  }.

En cada uno de estos criterios se minimizan las desviaciones entre los valores observados y aquéllos que son calculados por medio de la ecuación de la recta.

Una forma de reformular el problema dado por el criterio a es como sigue. Observe que éste es equivalente a

$$\text{minimizar } \sum_{i=1}^n z_i$$

$$z_i \geq |y_i - a - b_i x_i| \quad i=1, \dots, n$$

$$z_i \geq 0 \quad i=1, \dots, n$$

Asimismo, usando la definición de valor absoluto, se tiene que el problema es equivalente a

$$\begin{aligned} & \text{minimizar } \sum_{i=1}^n z_i \\ & z_i \geq y_i - a - bx_i \quad i=1, \dots, n \\ & z_i \geq -y_i + a + bx_i \quad i=1, \dots, n \\ & z_i \geq 0 \quad i=1, \dots, n \end{aligned}$$

que es un problema de programación lineal. Por otra parte, considere el problema de regresión con criterio b y observe que éste es equivalente a

$$\begin{aligned} & \text{minimizar } z \\ & z \geq |y_i - a - bx_i| \quad i=1, \dots, n \end{aligned}$$

De donde podemos concluir que el problema de regresión es equivalente a

$$\begin{aligned} & \text{minimizar } z \\ & z \geq y_i - a - bx_i \quad i=1, \dots, n \\ & z \geq -y_i + a + bx_i \quad i=1, \dots, n \\ & z \geq 0 \end{aligned}$$

que es un problema de programación lineal.

COMPUTACION APLICADA A LA PLANEACION URBANA

AGOSTO, 1981

FECHA	HORA	T E M A	PROFESOR
Lunes 10	9:00 - 11:45	Comptadoras en Planeación Urbana Panorama del Curso	Dr Willard B Hansen Arq Alejandro Villanueva
	11:45 - 11:30	Descanso	
	11:30 - 13:00	Geo-procesamiento	Dr Adolfo Guzmán Arenas
	13:00 - 14:00	Comida	
	14:00 - 15:30	Modelos Digitales de Terrenos	Act Dora Luz Gómez
	15:30 - 15:45	Descanso	
	15:45 - 18:00	Computación Gráfica y Mapas por Computado- ra I SYMAP	Arq Alejandro Villanueva
Martes 11	9:00 - 10:45	Archivos Geográficos Urbanos	Dr Willard B Hansen
	10:45 - 11:00	Descanso	
	11:00 - 12:00	Sistemas Interactivos de Geo-procesamien- to	Gte Eelipe de Teresa y Polignac
	12:00 - 13:00	Visita a SPR	
	13:00 - 14:00	Comida	
	14:00 - 16:00	Percepción Remota	M en C Armando Jinich
	16:00 - 16:15	Descanso	
	16:15 - 18:00	Sistemas Interactivos para la Toma de Decisiones con Objetivos Multiples en Sistemas Urbanos	M en C Fausto Riveros Acosta
Miércoles 12	9:00 - 10:45	Sistemas de Información Urbana	Dr Willar B Hansen
	10:45 - 10:55	Descanso	
	10:55 - 11:55	Sistema de Información para el Desarrollo Urbano, caso de estudio DDF	Arq Alfonso Goveia
	11:55 - 15:00	Descanso	

Miércoles 12	12:00 - 13:00	Insumo/Producto en Sistemas de Ciudades	Dr José Villanueva Ledesma
	13:00 - 14:00	Comida	
	14:00 - 16:00	Minicomputadoras y Microprocesadores	Ing Luis Cordero Borboa
	16:00 - 16:15	Descanso	
	16:15 - 18:00	Computación Gráfica y Mapas por Computadora II Calcomp. Visita CENIA	Gte Lou Meier
Jueves 13	9:00 - 10:45	Análisis Estadístico	Dr Willard B Hansen
	10:45 - 10:55	Descanso	
	10:55 - 11:25	Ejercicios SPSS	Act Juan Manuel Padilla
	11:25 - 11:30	Descanso	
	11:30 - 13:00	Métodos de Pronóstico	M en C Rubén Téllez Sánchez
	13:00 - 14:00	Comida	
	14:00 - 15:30	Optimización en Sistemas Urbanos Tempo	Dr Sergio Fuentes Maya
	15:30 - 16:00	Ejercicios Tempo	Act Juan Manuel Padilla
16:00 - 18:00	Computación Gráfica y Mapas por Computadora III, Visita a Soluciones Gráficas	Ing Douglas Thorson	
Viernes	9:00 - 10:45	Modelos de Simulación, Uso de Suelo-Transporte	Dr Willard B Hansen
	10:45 - 11:00	Descanso	
	11:00 - 12:00	Modelos Urbanos, Casos de Estudio	Dr Cuauthémoc Rodríguez
	12:00 - 12:05	Descanso	
	12:05 - 13:00	Simulación Mediante Juegos. METRO, APEX, CLUG	Arq Alejandro Villanueva
	13:00 - 14:00	Comida	
	14:00 - 16:15	Conceptos Generales y Lenguajes de la Simulación de Sistemas	M en C Arturo García del Busto
	16:15 - 16:30	Descanso	
16:30 - 18:00	Simulación Urbana con Dynamo	M en I Francisco Alvarez-Caso	



DIVISION DE EDUCACION CONTINUA  
FACULTAD DE INGENIERIA U.N.A.M.

COMPUTACION APLICADA A LA PLANEACION URBANA

How to describe pure forma and how to measure  
differences in shapes using shape numbers

Agosto, 1981

*Reprinted from*

*PATTERN RECOGNITION*

*Vol. 12, No. 2, pp. 101-112*

HOW TO DESCRIBE PURE FORM AND HOW TO  
MEASURE DIFFERENCES IN SHAPES USING  
SHAPE NUMBERS

ERNESTO BRIBIESCA and ADOLFO GUZMAN

Research Department, DETENAL-SPP (Mexico), Calzada San Antonio Abad 124, Mexico 8,  
D.F. and Computer Science Dept., IIMAS, National University of Mexico

PERGAMON PRESS  
OXFORD · NEW YORK · FRANKFURT · PARIS  
1980

## HOW TO DESCRIBE PURE FORM AND HOW TO MEASURE DIFFERENCES IN SHAPES USING SHAPE NUMBERS

ERNESTO BRITESCA and ADOLFO GUZMAN

Research Department, DETENAL-SPP (Mexico), Calzada San Antonio Abad 124,  
Mexico 8, D.F. and Computer Science Dept., IIMAS, National University of Mexico

(Received 5 April 1979; in revised form 26 September 1979; received for  
publication 15 October 1979)

**Abstract**—The shape number of a curve is derived for two-dimensional non-intersecting closed curves that are the boundary of simply connected regions. This description is independent of their size, orientation and position, but it depends on their shape. Each curve carries "within it" its own shape number. The order of the shape number indicates the precision with which that number describes the shape of the curve. For a curve, the order of its shape number is the length of the perimeter of a 'discrete shape' (a closed curve formed by vertical and horizontal segments, all of equal length) closely corresponding to the curve. A procedure is given that deduces, without table look-up, string matching or correlations, the shape number of any order for an arbitrary curve. To find out how close in shape two curves are, the degree of similarity between them is introduced; dissimilar regions will have a low degree of similarity, while analogous shapes will have a high degree of similarity. Informally speaking, the degree of similarity between the shapes of two curves tells how deep it is necessary to descend into a list of shapes, before being able to differentiate between the shape of those two curves. Again, a procedure is given to compute it, without need for such list or grammatical parsing or least square curve or area fitting. The degree of similarity maps the universe of curves into a tree or hierarchy of shapes. The distance between the shapes of any two curves, defined as the inverse of their degree of similarity, is found to be an ultradistance over this tree. The shape number is a description that changes with skewing, anisotropic dilation and mirror images, as the intuitive psychological concept of "shape" demands. Nevertheless, at the end of the paper a related Theory "B" of shapes is introduced that allows anisotropic changes of scale, thus permitting for instance a rectangle and a square to have the same B shape. These definitions and procedures may facilitate a quantitative study of shape.

Curve description  
Form similarity  
Image processing

Chain encoding  
Shape comparison

Shape code  
Measure of shape difference

Silhouettes

Shape numbers  
Binary picture

### INTRODUCTION

The study of shape is an important part of the field of Pattern Recognition.

As pointed out by Lord Kelvin, a science begins to emerge when it is possible to make measurements of the phenomena that such science seeks to understand, allowing quantitative comparison and mathematical relations among them.

This paper gives a procedure to measure (i.e. to assign a number to) the resemblance between any two shapes.

With the help of procedures like this, a quantitative study of shape may be possible.

#### *Precious work on shape*

Shape extraction is an active field. Sequential extraction of shape features<sup>(1)</sup> can be performed making only one pass over the image. For global shape analysis, several authors have used Freeman chains, medial axis transforms, decomposition into primary convex subsets, polar co-ordinates,<sup>(2)</sup> decomposition at concave

vertices; decomposition by clustering, mirroring axes<sup>(3)</sup> and stroke detectors. These and other methods are reviewed by Pavlidis.<sup>(4)</sup>

### 1. THE SHAPE NUMBERS

#### 1.1. *What is a shape*

A region is a simply connected portion of a plane limited by a curve boundary. That is, no holes; no self-intersecting boundary. It is a closed boundary. A given region has a size, a position, and an orientation in the plane. This defines a flat region, which is uniquely defined by the curve it has as boundary. This paper deals with shapes of regions, but the shape numbers used here can also be applied to open curves. In addition, Section IV describes regions with holes.

A shape is what remains of a region after disregarding its size, position and orientation in the plane. That is, two regions have the same shape if we can make them coincide exactly by translation and rotation in the plane, in addition to a uniform change of scale (the



$x$  and  $y$  co-ordinates increase by the same factor).

A region and its mirror images will not have the same shape, in general.

This definition coincides with the intuitive psychological definition of "shape".

If a notation is going to be used to represent the shape of a region, it has to be independent of the position, orientation and size of such region. It should be reproducible: a region, when translated, magnified and rotated should still give the same description as when untransformed. Two regions with different shapes should produce different descriptions. Finally, the shape number should be unique for a given region; for instance, it should not depend on an arbitrary starting point or a particular co-ordinate system.

If the notation can be deduced exclusively from the region, without comparison with a table of canonical shapes or shape descriptors, for instance, then we can expect savings in memory and computer time for the procedure that computes the shape description.

### 1.2 Continuous and discrete shapes

A shape is *discrete* if the boundary of the region is formed by segments of a square grid. Otherwise, the shape is *continuous*.

### 1.3 Mating a continuous shape into a discrete shape

A square grid may be overlaid on top of a continuous shape to obtain a discrete shape. The quantization of the shape is as follows: a square of the grid is "black" (inside the discrete shape) if more than 50% of it is covered by the continuous shape; otherwise it is "white" or outside (Fig. 1). The size, orientation and position of this grid will influence the resulting discrete shape.

A discrete shape, obtained from a continuous shape in the above manner, can not be a shape descriptor of the continuous shape, because it depends on the size and orientation of the grid. This will be solved in Section 1.6.

Now, some shape descriptors will be given.

### 1.4 Eccentricity

The *eccentricity* (ratio of the major to minor axis, Fig. 2) of a region is a descriptor that depends only on its shape.

The *major axis* of a region is the line joining the two perimeter points furthest away from each other. The *minor axis* is perpendicular to the major axis, and of length such that a box could be formed that just

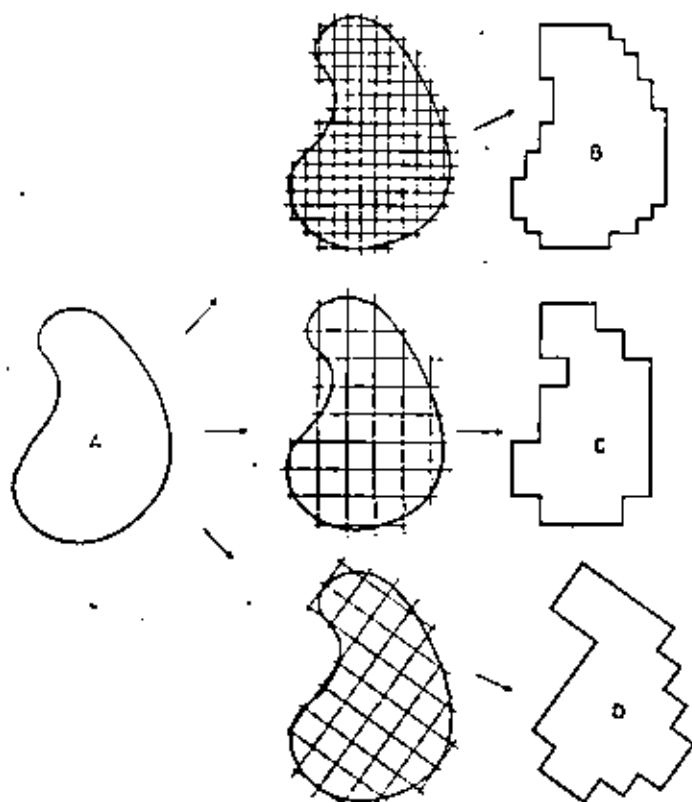


Fig. 1. Continuous and discrete shapes. Continuous shape  $A$  gives rise to several discrete shapes  $B$ ,  $C$ ,  $D$ . If it is desired to have a *unique* discrete shape derived from  $A$ , then it is necessary to specify the grid size (related to the *order* of the discrete shape), as well as its orientation and position with respect to the continuous curve  $A$ . In this manner, for a given order  $n$ , the discrete shape corresponding to  $A$  will be unique. This is accomplished in Section 1.6.



Fig. 2. Basic rectangle. (a) Minor axis of  $c$ . (b) Major axis of  $c$ . (c) Region. (d) Basic rectangle of  $c$ . The eccentricity  $e = b/a$  is always greater than or equal to 1. It is a shape descriptor, although not a good one.

encloses the region. This box is called the *basic rectangle* (Fig. 2).

Occasionally, there will be more than one major axis in a region. In that case, select that which gives the shorter minor axis; if necessary, add additional criteria to make the choice of major axis a unique choice.

1.5 *Freeman chain and its derivative*

*Freeman chain in four directions.* For a given region and a given square grid of fixed orientation and size, the Freeman chain in four directions is the curve obtained by walking clockwise on the grid (on its "wires") around and outside the squares that are more than half contained by the region (Fig. 3).

*Derivative of Freeman Chains.* It is the chain number obtained by clockwise replacing each convex corner of the Freeman chain by a 1, each straight corner by a 2, and each concave corner by a 3, as Fig. 3 suggests. The number obtained ("E" in Fig. 3) will be different if we change the size or orientation of the grid. In the next section a method appears that makes the "derivative of Freeman chain" independent of these changes. This new derivative will be called the shape number.

1.6 *The shape numbers*

This section tells how to obtain our proposed description for the shape of *shapes* and *regions*. The procedure to find the shape number of a region is as follows:

1. A grid of arbitrary cell size is overlaid on top of the region. A "black" region is formed with all the cells

that fall 50 per cent or more inside the region.

2. The boundary of such a black region is the chain sought after. This chain is denoted by its derivative notation (q.v.). We collect these numbers travelling clockwise. Refer to Fig. 4.

Observe that there are several strings of digits 1, 2 and 3 corresponding to the above chain, depending on the starting point (see Fig. 4):

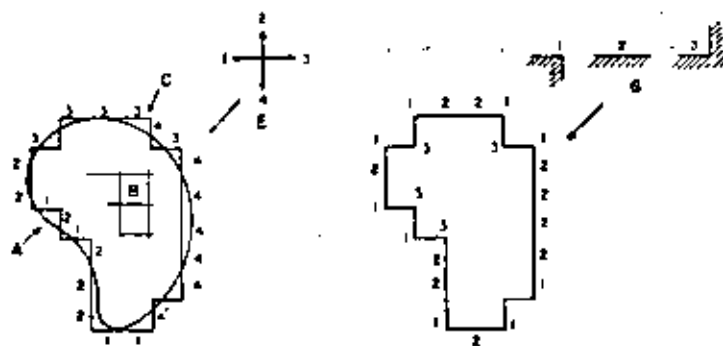
- 12131131213113 (A)
- 21311312131131 (B)
- 13113121311312 (C)
- 31131213113121 (D)
- 11312131131213 (E)
- 13121311312131 (F)
- 31213113121311 (G)
- 12131131213113 (H)
- 21311312131131 (I)
- 13113121311312 (J)
- 31131213113121 (K)
- 11312131131213 (L)
- 13121311312131 (M)
- 31213113121311 (N)

Observe also that one of them is a minimum, when regarded as a number in base 3: (E) in the above example.

3. Select the chain that is minimum as the chain that represents the region. In the example, it is 11312131131213. Observe that the minimum chain always starts with a 1, since every discrete shape contains at least four 1's.

*What size of grid? What orientation?* Unless a procedure is given that normalizes these questions and provides unique answers, a region will have several shape numbers.

The adopted posture is that the orientation of the grid will be normalized, but its size will be a parameter that will allow us to vary the precision of the shape number. Nevertheless, although the size of the cell of the grid varies according to the precision, the number of segments of the grid (sides of each cell) into which the region will be mapped is no longer at user's will, but



D: 12223354344444411222121

F: 113122131222131122313

Fig. 3. Chains. A: the region. B: The grid. C: The Freeman chain in four directions. D: Its chain number. E: The four directions of (B) used to code (C) into (D). F: the derivative of (C). G: The three types of corners used to code (C) into (F).

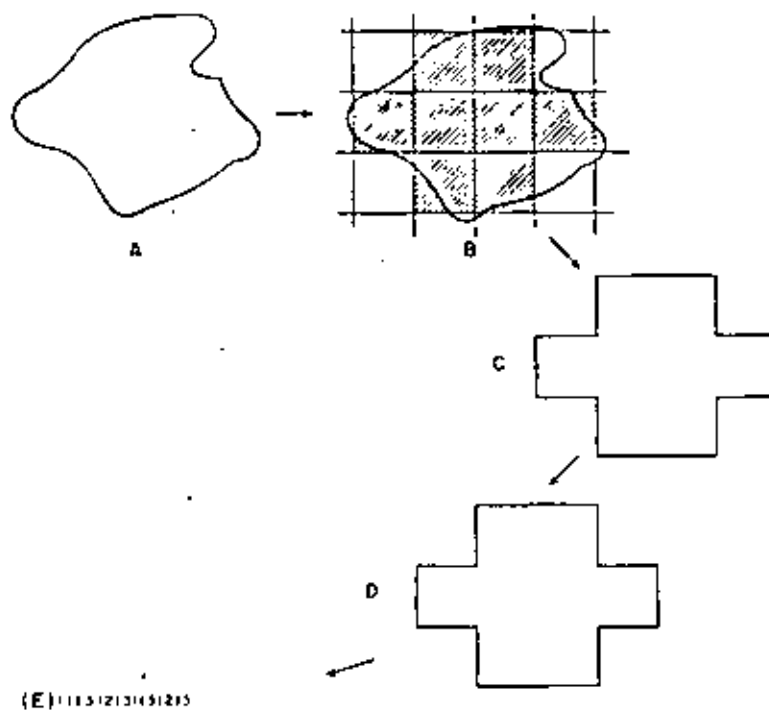


Fig. 4. Shape number. (A) the continuous shape. (C) The discrete shape. (E) The shape number.

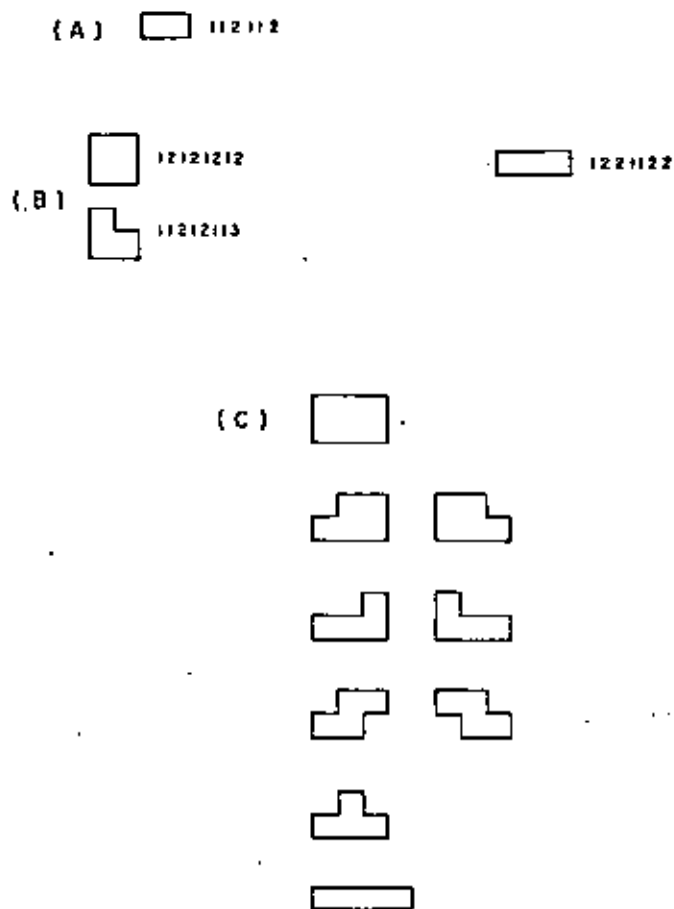


Fig. 5. All the shapes of orders 6, 8 and 10. (A) Of order 6. (B) Of order 8. (C) Of order 10.

it is dictated by the precision he specifies.

The orientation of the grid is not arbitrary, but it is made to coincide with the major axis of the region. The reason is clear: each region should carry along with it its own direction of the grid. In this manner, if the region rotates, the grid rotates the same amount and a code is obtained invariant under rotations.

*Procedure to achieve a unique shape number.* Given a region surrounded by its basic rectangle, a grid of a given (fixed) size could be placed on top of the rectangle, in order to extract the unique shape number of the region. Instead, the user is allowed to tell how many digits he wants his shape number to contain. That is known as the *order* of the shape number.

It is clear that the same shape gives rise to several shape numbers. But, given  $n$ , the shape number of order  $n$  of that shape is unique.

Shortly a procedure will be shown to find the shape number of order  $n$  of a region, for a given  $n$ . Before that, however, the families of discrete closed shapes of several orders are presented.

*All the shapes of order 4.* These are all the regions that can be formed with four sticks of the same size, when they can be placed only collinearly or at 90 degrees with respect to each other.

There is only one closed shape of order 4, the square: 1111.

This is the most primitive or fundamental shape. Imagine you are looking at things very far away; you

can not really differentiate much. All objects would look round (square, in this paper) and equal.

*All the shapes of order 5.* No shape number of odd order represents a closed figure. For a closed figure, number of corners = number of sticks = order of figure.

This paper does not deal with open figures. Not all ternary numbers with an even number of digits are shape numbers. Most of them do not close.

*All the shapes of order 8, 10 and 12.* See Figs. 5 and 6.

## 2. USING THE SHAPE NUMBERS FOR SHAPE DESCRIPTIONS

### 2.1 The order of a shape number

The order of a shape number is the number of ternary digits that the shape number contains. It is always even, because the boundary is closed.

### 2.2 How to find the shape number of order $n$ of a continuous shape

The procedure is as follows:

1. Find the basic rectangle of the region.
2. From the family of discrete shapes of order  $n$ , find the rectangle of order  $n$  with eccentricity closest to that of the region. (This is easy. For instance, for  $n=22$ , the rectangles of order 22—those with perimeter equal to 22—are of sides 6 by 5, 7 by 4, 8 by 3, 9 by 2 and 10 by 1.)

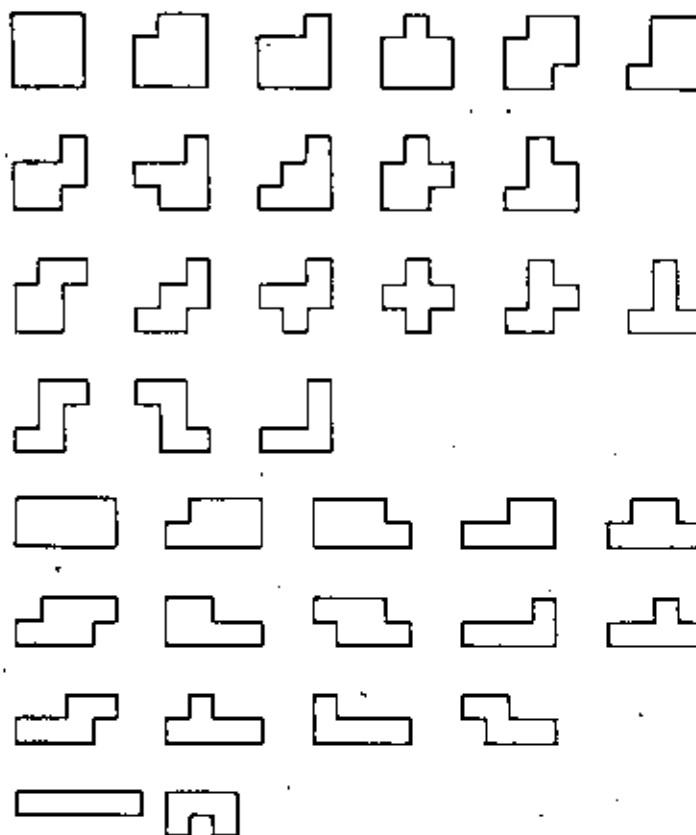


Fig. 6. All the shapes of order 12.

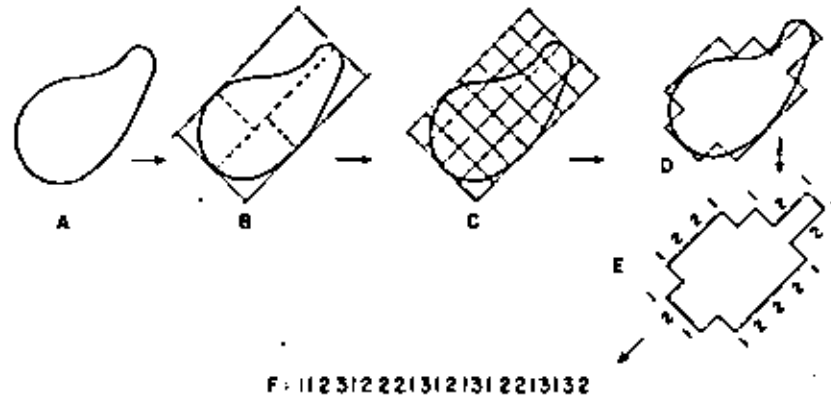


Fig. 7. Finding the shape number of order 22. The continuous (A) is encased in its basic rectangle (B). Given  $n = 22$ , a square grid of appropriate size (see text) is centered (C) on the basic rectangle. A discrete shape (D) is obtained. The "derivative notation" (E) is found. Traveling clockwise, the chain with the minimum absolute value (F) is the shape number of order 22 digits. (D) has 22 sides, as well as 22 corners.

In practice, it is better to approximate the longer side of the rectangle instead of the eccentricity. If  $e =$  eccentricity, one can deduce that the longer side is  $b = (n/2)(e/1+e)$ . Select a rectangle with longest side closest to that quantity. Lay this rectangle, centered, so as to cover the region, and make a grid of  $u$  by  $h$  square cells ("C", in Fig. 7).

3. Make black (= 1) all those cells falling more than 50% inside the region; leave white (0, outside) all others.

The boundary of this black region, expressed in the derivative notation, is the desired shape number.

Remember to write down the digits of the chain traveling clockwise, and selecting as the starting point the corner of type I that makes the chain number the smallest of the  $n$  possible chain numbers. An example is given in Fig. 7.

Notice that the resulting shape number is indeed of order  $n$ . This will not be true if the figure has depressions (concavities) in its boundary. The depression in the boundary makes the order bigger. Each depression of depth  $k$  increases the order of the shape by  $2k$ .

When, looking for a shape number of order  $n$ , if a number of order  $n + 2d$  results, try next to look for a shape number of order  $n - 2d$ . Due to the presence of the holes, the shape number  $n - 2d$  will be increased by an amount equal to the "hole excess"  $2d$ , thus yielding the desired order  $n$ . This relation holds only approx-

imately, since the size of the holes of order  $n$  is smaller than those of order  $n - 2d$ . Thus, in practice, try the basic rectangles of order  $n - 2d$ ,  $n - 2d + 2$ ,  $n - 2d + 4$ , ...,  $n - 2$ , and when we obtain a shape number of order  $n$ , that is it. See also Section 2.3.

*Properties of the shape number.* It is insensitive to orientation of the region, to its position, to its size, and to the origin of the chain. It is therefore appropriate to think that the shape number of a region indeed describes its shape (cf. Section 1.1).

Also, since it is possible to compute the shape number of a region without reference to a table of sorted shapes (canonical shapes), we avoid making correlations or comparisons of shapes or of strings. That is, the shape number of a region can be deduced solely from the region.

In addition, the precision of the resulting shape number can be varied. This is done with the order of the shape number; that is, the size of the sticks (or of the grid) that we use to find it.

### 2.3 Shapes without shape numbers

A shape with a thin isthmus (narrower than the size of the grid) will not yield one shape number, since the procedure of Section 2.2 will split the continuous shape into several discrete shapes (Part II of Fig. 8). Some shapes with depressions in their boundaries may not have a shape number. This is discussed in Section 3.3d.

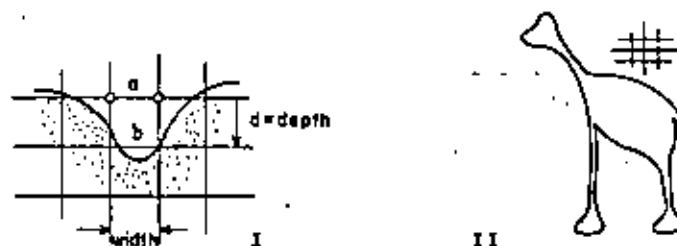


Fig. 8. Depressions and degenerate shapes. I. A depression of depth  $d$  increases the shape number by  $2d$ . II. Degenerate regions split the discrete shape but do not have a shape number.

3. USING THE SHAPE NUMBERS TO MEASURE SHAPE SIMILARITY

3.1 The degree of similarity between two shapes

The shape number of a region enables us to find instances of a given shape, even when distorted by enlargement or rotation. It answers the questions "Have these two regions the same shape?", up to an order  $n$ .

In practice, however, a shape rarely repeats itself, due to noise and the allowable variations (for instance, ten silhouettes of apples have similar but not identical shapes). The relevant questions to answer are "How much different are these two forms?", "How much do these two shapes resemble each other?", "Is region  $A$  closer in shape to  $B$ , or to  $C$ ?", This section gives a procedure to quantitatively answer these questions.

When the shapes of two regions  $A$  and  $B$  are compared, we can notice that the shape of order 4 of  $A$ ,  $s_4(A)$ , is equal to 1111 (the only shape of order 4), and is therefore equal to  $s_4(B)$ .

Also  $s_6(A) = s_6(B)$ ; probably  $s_8(A) = s_8(B)$ . It is likely that their first few shape numbers be identical. The reason is that the discrete shapes are coarse and not varied at low orders, where the "resolution" is low.

Nevertheless, most likely  $s_{100}(A) \neq s_{100}(B)$ , also  $s_{98}(A) \neq s_{98}(B)$ , etc. This is expected, because, due to the finer precision at higher orders, there exists a large variety of shapes, thus the discrimination between  $A$  and  $B$  is more demanding.

Of course, if  $A$  and  $B$  were very similar (but not identical), one might need to go up to say 170 to find

that  $s_{170}(A) \neq s_{170}(B)$ . On the other hand, if they are visibly different (not alike at all), already at order 10 we will find  $s_{10}(A) \neq s_{10}(B)$ .

Thus, as we increase the order  $n$  of the two shape numbers  $s_n(A)$  and  $s_n(B)$ , they begin equal but at some order they become different. How long they remain equal gives us an idea of the similarity between the shapes of  $A$  and  $B$ .

Degree of similarity  $k$  between the shapes of two regions  $A$  and  $B$ : it is the largest order for which their shape numbers still coincide.

That is, it is the largest  $m$  for which  $s_m(A) = s_m(B)$ , but  $s_{m+1}(A) \neq s_{m+1}(B)$  for all  $i$  greater than 0.

That is, we have  $s_4(A) = s_4(B)$ ,  $s_6(A) = s_6(B)$ ,  $s_8(A) = s_8(B)$ , ...,  $s_k(A) = s_k(B)$ ,  $s_{k+1}(A) \neq s_{k+1}(B)$ ,  $s_{k+2}(A) \neq s_{k+2}(B)$ , ...

If  $A$  and  $B$  are regions with degree  $k$  of similarity we write  $a \approx_k b$ .

Example. For the figures of Fig 9 we have for  $A$  to  $F$ :

- $s_4(A) = s_4(B) = \dots = s_4(F) = 1111$ ;
- $s_6(A) = s_6(B) = \dots = s_6(F) = 112112$ ;
- $s_8(A) = s_8(D) = s_8(E) = 12121212$ ;
- $s_8(B) = 11212113$ ;  $s_8(C) = s_8(F) = 11221122$ ;
- $s_{10}(A) = 1212212122$ ;  $s_{10}(B) = 1121221123$ ;
- $s_{10}(C) = 1122113113$ ;
- $s_{10}(D) = s_{10}(E) = 1131212122$ ;
- $s_{10}(F) = 1122121213$ ;  $s_{12}(D) = 113113121213$ ;
- $s_{12}(E) = 113121221213$ .

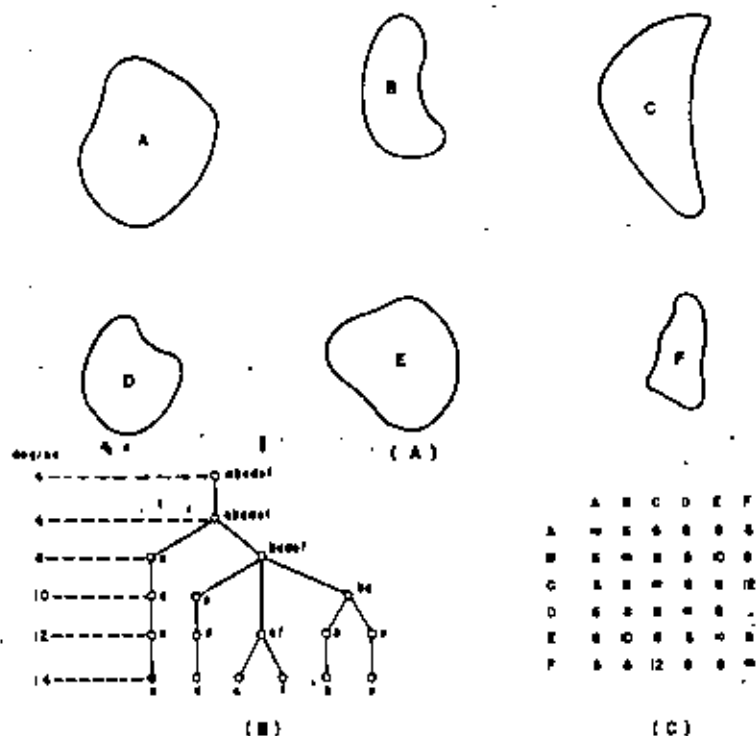


Fig 9 Degree of similarity. (A) Regions to be analyzed. (B) Similarity tree for (A). (C) Similarity matrix for regions (A). The shapes form a hierarchy, a tree with root at degree = 4

Therefore,  $A$  and  $B$  have a degree of similarity equal to 6:  $A \approx_6 B$ . Also,  $A \approx_4 E$ ;  $E \approx_8 B$ ;  $D \approx_{10} E$ ; etc.

This is represented in the figure both as a similarity tree and as a similarity matrix.

The similarity matrix is symmetrical; in fact, it is easily proved that, for arbitrary regions  $A$  and  $B$ ,

- (1) (Thm.) The relation " $A$  and  $B$  have degree  $k$  of similarity" (for a fixed  $k$ ) is not an equivalence relation.
- (2) (Thm.) The relation " $A$  and  $B$  have degree of similarity of at least  $k$ " (for a fixed  $k$ ) is an equivalence relation.

In fact, the equivalence classes of (2) for  $k=10$  are nine, since there are only nine discrete shapes of order 10.

Informally speaking, the size (power) of the magnifying lens that barely confuses two regions gives the degree of similarity between such regions.

The comparison procedure could be visualized as follows: A number (a shape number of high order) is associated to each one of two regions. If the numbers are equal, the regions have identical shape. If not, another pair of numbers (shape numbers of the next lower order) is deduced, and so on until we find that the two numbers coincide. The number of stages needed is an indication of the dissemblance of the two shapes.

### 3.2 The distance between two shapes

*Distance.* (definition) The distance between two shapes  $A$  and  $B$  is defined to be the inverse of their degree of similarity,  $d(A, B) \hat{=} 1/k$ .

Then  $d$  is an ultradistance, obeying

$$d(A, A) = 0 \quad (1)$$

$$d(A, B) \geq 0;$$

$$d(A, B) = 0 \text{ if and only if } A = B \quad (2)$$

$$d(A, C) \leq \text{Sup}\{d(A, B), d(B, C)\}. \quad (3)$$

### 3.3 Comments on this theory of shapes

a. *No parsing is necessary.* To find the degree of similarity between  $A$  and  $B$ , their shape numbers are compared for equality. Two shape numbers of different order are incommensurable. Two shape numbers of the same order are either equal or different. If different, that is it. There is no need to compare "how close in shape they are". String matching<sup>22</sup> is not needed.

To find out the degree of similarity, a binary search is used. First see whether the shape numbers at order 8 are equal or not. Then compare the shape numbers at the highest required accuracy (say, 100). Then at the middle. Then at the middle of the remaining valid half. And so on.

b. *Intuitively satisfying.* Shape numbers are not invariant under reflexions (mirror images), skewing, or unequal expansion along the  $X$  and  $Y$  axes. These transformations alter what could be considered the intuitive shape of a figure. At the end of the paper a Theory "B" of shapes is presented, where the last

transformation is allowed, i.e. a circle and an ellipse have the same Bshape number.

c. *Occasional loop in the similarity tree.* Due to noise or the 50 per cent requirement for quantization, and at low orders, a transitory divergence and then convergence in the shapes of two regions is sometimes observed, v.gr.,

$$s_4(A) = s_4(B)$$

$$s_{10}(A) \neq s_{10}(B)$$

$$s_{12}(A) = s_{12}(B)$$

$$s_{14}(A) \neq s_{14}(B)$$

$$s_{16}(A) \neq s_{16}(B)$$

i.e. they were already different at order 10, but they are again equal at order 12 (however, only to separate soon forever). This still gives a unique number for the shape of a region, but makes the definition of degree of similarity less attractive, and the procedure to find it, unreliable. Only loops of size 2 (such as the example given) have been found, infrequently. These loops disappear in theory B.

d. *Non existent shape numbers.* Shape number of order  $n$  may occasionally not exist for a given figure, due for instance to symmetrical holes of type 1 in Fig. 8. This does not bother the similarity procedure, but it is a nuisance not to have that shape number.

e. *Quantization of the eccentricity.* The basic rectangles of order 12 have eccentricities equal to 1 (the square of 3 by 3), 2 (the rectangle of 4 by 2) and 5 (the rectangle of 5 by 1). For an object of eccentricity 1.6, one of these has to be used. An error is going to be committed in any case. There seems to be no way out of this.

A theory is now presented that has none of these problems.

### 3.4 Theory "B" for Shape description

To obtain this new theory, the current theory undergoes some changes:

1. Force the eccentricity of any region to be equal to one, by performing an unequal dilation of its axes. The only discrete Bshapes that now exist are those obtained from squares. All the rectangles have disappeared.
2. Do not go into depressions (1 in Fig. 8) with width smaller than the size of the side of the cell of the grid. This avoids degenerate shapes.

That is, if a region is "scratched" by thin lines (thinner than the size of the grid) that belong to the background ignore them (act as if they were not there) or else, if they cannot be ignored, this theory "B" says that the size of the grid is inappropriate to describe such region, and that its Bshape does not exist at this order. Higher resolution is needed.

3. Let the depressions where the sticks do go in (because they are wider than Part 1 of Fig. 8)

generate Bshape numbers having a number of (ternary) digits larger than the expected order. That is, do not correct the anomaly that these depressions cause. The perimeter of the Bshapes no longer tells its order.

4. Eliminate the orders that are not powers of two. The only valid orders for Bshape numbers are 4, 8, 16, 32, ... These numbers still indicate the perimeter of the basic square of the region.

The procedure is the following:

#### How to find the Bshape number of order $n$

1. Find the basic rectangle of the region and convert it to a square. Declare that the Bshape number does not exist if the region has necks (isthms) or depressions (channels, fjords) narrower than  $4n$  or  $2^2 - n$ .
2. Make a grid by dividing the side of the basic square into  $n^2$  equal parts.
3. Mark with a 1 each cell of the grid of step 2 that is more than 50 per cent contained in the region. The collection of grid squares containing a 1 form a discrete Bshape.
4. Find the shape number of the discrete Bshape of step 3, and give that as answer (even if it has more than  $n$  ternary digits).

The order  $n$  of a Bshape number is four times the number of parts into which the side of the basic square

was divided. It is also the perimeter (measured by the number of sticks) of the basic square.

It is no longer the perimeter of the discrete Bshape, nor the number of ternary digits of the Bshape number.

Given a shape, it is easy to derive its Bshape number. An example is given in Fig. 11.

The degree of similarity between the Bshapes of two regions is obtained as before. No change in the definition.

*Downwards constructability.* Given the Bshape number of order  $n$  of a region, the Bshape number of order  $n/2$  can be deduced from it, by regrouping appropriate sets of 4 neighboring cells into a cell for the lower order. Therefore, if two regions have the same Bshape number of order  $n$ , they will continue to have equal Bshape numbers of smaller order, until they cease to exist. This gets rid of problem 3.3c of the former theory.

*Upwards existence.* If the Bshape number of order  $n$  of a region exists, the existence of numbers for higher order is guaranteed, since the channels or narrow parts that could not split the shape at order  $n$ , will also be unable to split it at higher orders. This defeats problem 3.3d of the former theory.

*Quantization of the eccentricity.* Finally, problem 3.3e of the former theory is not present in theory B because all eccentricities are now equal to 1.

Some example of similarity comparison using theory B are given in Fig. 10.

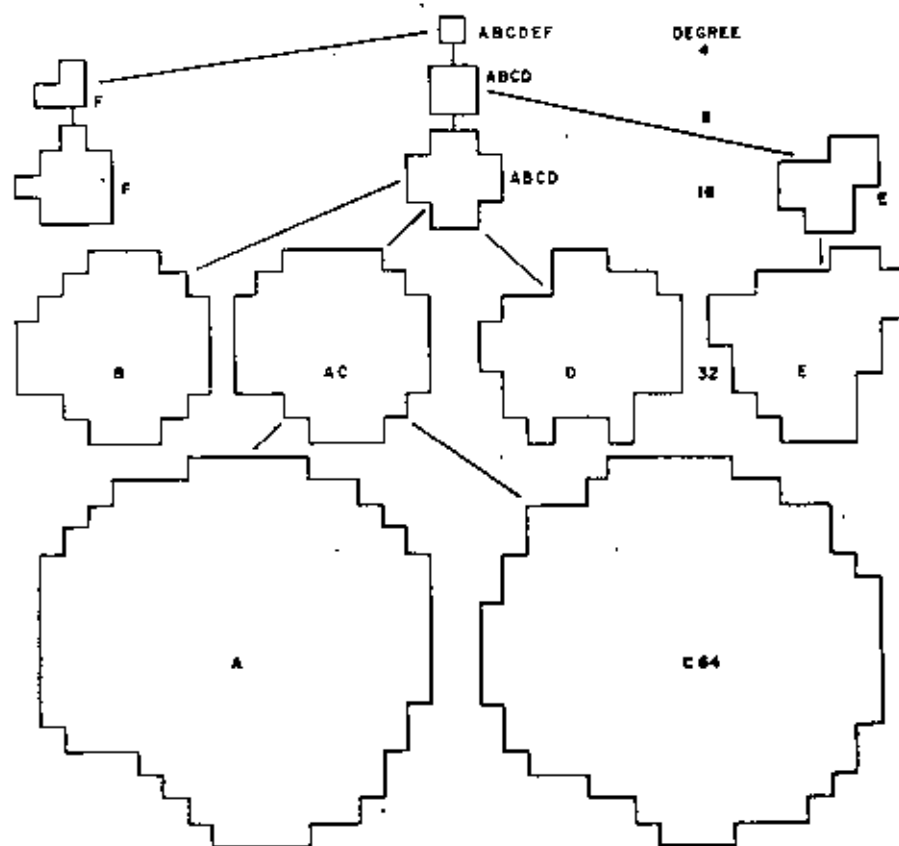
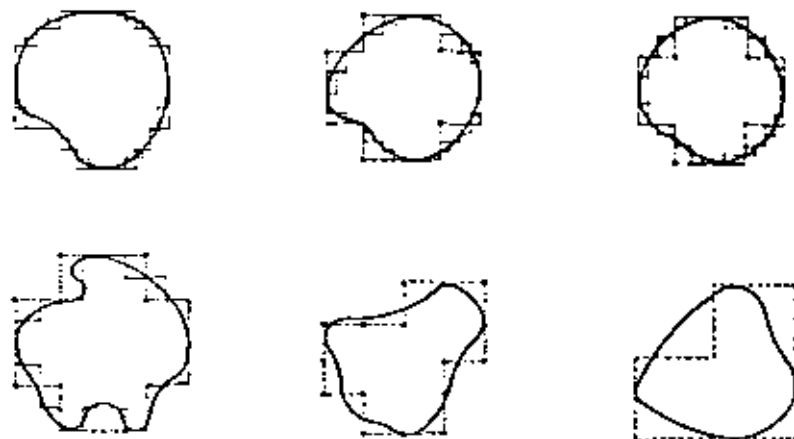


Fig. 10. Similarity tree for the Bshapes of Regions A to F. The tree shows that the degree of similarity between B and E is 8, but between B and C is 16.





	A	B	C	D	E	F
A						
B	16					
C	32	6				
D	6	6	6			
E	8	8	8	8		
F	4	4	4	4	4	

TABLE "SIMILARITY MATRIX FOR THE BSHAPES OF REGIONS A-F"

Notice that  $A \sim_{16} B$ ,  $C \sim_{32} D$ , but  $E \not\sim_{32} D$ .

	A	B	C	D	E	F
A	0	1/16	1/32	1/6	1/8	1/4
B		0	1/6	1/6	1/6	1/4
C			0	1/6	1/8	1/4
D				0	1/8	1/4
E					0	1/4
F						0

TABLE "DISTANCE MATRIX FOR BSHAPES OF REGIONS A TO F"

A and C are very together (1/32) in Bshape.  
The region F is quite dissimilar (1/4) in shape to all others.

Fig. 11. Example of Bshapes and their similarity. The arrows on the figures signal the beginning of the string of order 32 or 64.

*Disadvantage of theory "B":* Squeezing along one axis is now a valid (Bshape preserving) transformation. Thus, either the application does not care about the eccentricity or aspect ratio, or this has to be carried as another parameter, in addition to the Bshape number.

Also, more care needs to be exercised now when selecting the major and minor axis, to avoid noise perturbations. It may pay to use the rectangle suggested in.<sup>(13)</sup>

#### 4. THE SHAPE NUMBERS OF SHAPES WITH HOLES

It is possible to assign shape numbers for regions with holes, and to use them for shape comparison and shape similarity measurement. The idea is to use the basic rectangle of the outer boundary for discretization of all the boundaries (both the outer and the inner boundaries). Using the shortest possible vertical or horizontal cuts, join the boundaries among them. Each cut reduces by one the number of boundaries. Finally, a single boundary is found. Then such a boundary can be described by an ordinary shape number. Such a shape number is then associated with the original region.

Notice that no other shape with holes could also be the owner of that shape number, since the new number

has "touching edges" (those running along the cut). And since the set of cuts is unique (cf. discussion below), the resulting shape number is also unique. See Fig. 13.

The procedure is detailed now for Bshapes. To find the Bshape number of order  $n$  of a region with holes, proceed as follows:

1. Find the Bshape number of order  $n$  of the outer boundary.
2. Using the grid defined in (1), find the discrete Bshapes of the inner boundaries.
3. Let a "cut" be a sequence of purely vertical or purely horizontal segments of the grid. Find the minimum spanning tree of cuts that connects the boundaries (This tree can be found as follows: (a) find the two boundaries closest to each other; that is, the two boundaries with the shortest cut joining them. That cut belongs to the tree, and these two boundaries are now joined by such a cut. (b) Now, find the boundary closest to that new boundary. That defines another cut. This new cut belongs to the tree, and this new boundary is now joined (by such a cut) to the former collection of boundaries. (Now we have three boundaries joined by two cuts). (c) Keep iterating (b), each time adding a new boundary (the closest one) to the collection of boundaries, and its corresponding cut to the mini-

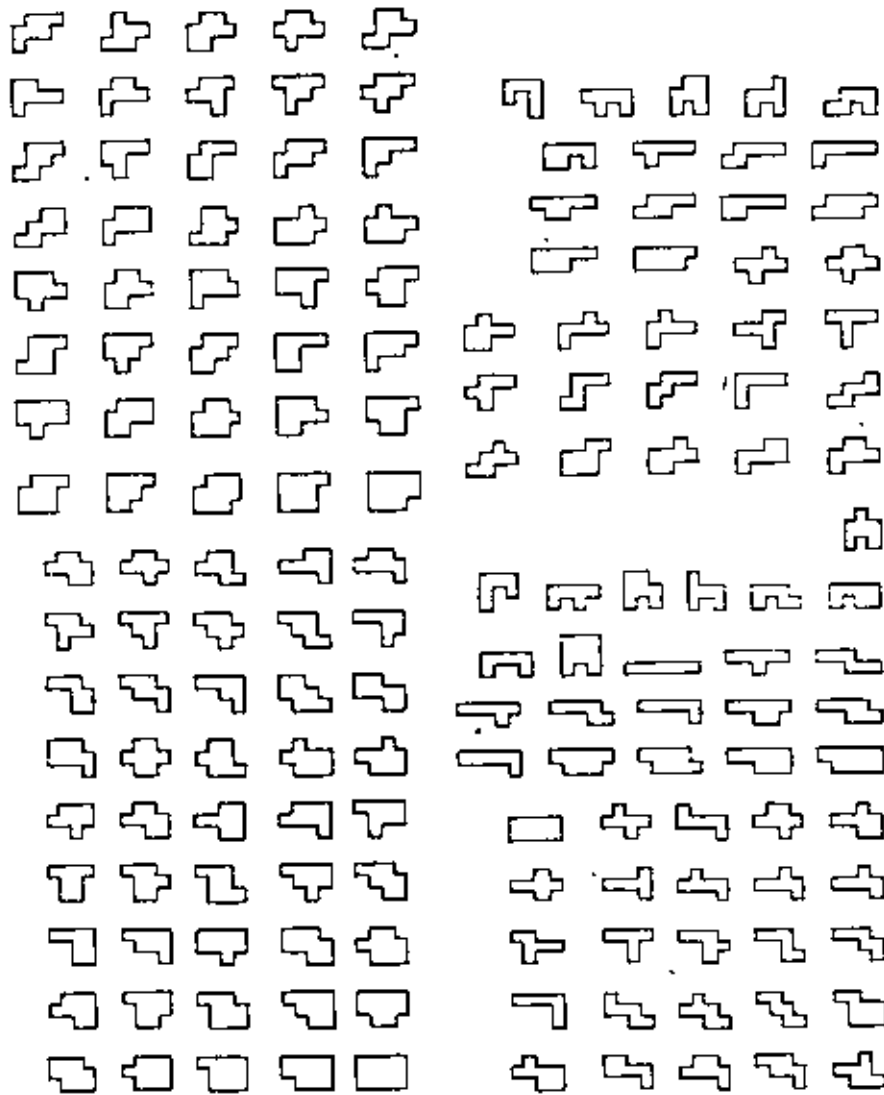


Fig. 12. All the shapes of order 14.

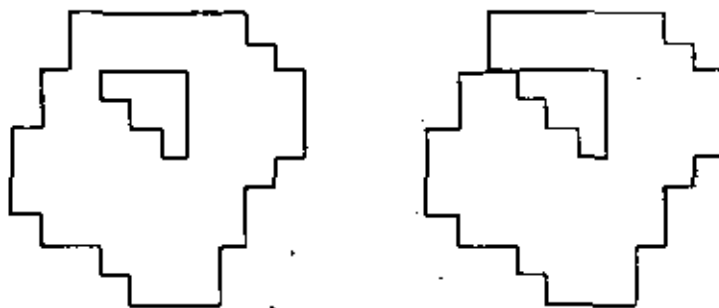


Fig. 13. Shapes with holes. To find the Bshape number of a discrete shape ( $A$ ) with a hole, cut a channel across the region, so as to have a simply connected region ( $B$ ); then return the Bshape of ( $B$ ) as the answer. The text provides more explanation.

imum spanning tree of cuts. When all boundaries are joined, stop.

The result is a simply connected boundary.

4. Find the Bshape of this simply connected boundary, and give that as answer.

If there are two cuts of equal length, use the cut that minimizes the resulting Bshape number. This favors cuts near the starting point of the Bshape number.

With this tie-breaking rule, the Bshape number is unique.

### CONCLUSIONS

For each two-dimensional region, a shape number can be derived. This number depends exclusively on the form of the region.

These shape numbers can be found without table look-up or correlation or string matching.

The shape numbers can be of different order; high orders are more accurate for shape description. Informally, the number of ternary digits of a shape number will tell its order.

The degree of similarity between two regions, introduced in this paper, permits to find out how close in shape two regions are. Two regions with shapes that look alike will have a high degree of similarity.

Informally speaking, the degree of similarity between the shapes of two regions is the highest optical resolution (power of the magnifying lens) that still confuses them.

The distance between two shapes is defined and it is found to be an ultradistance or ultrametric.

The Bshape numbers allow additional advantages and overcome some problems of the (ordinary) shape numbers.

The shape numbers of figures with holes are defined.

Suggestion for further work: apply the shape numbers to three-dimensional surfaces enclosing a volume.

**Acknowledgements** - The research herein reported was partially done under the Joint Research Agreement (NI-1976) between DETENAL and UNAM. The work was carried on at IIMAS-UNAM and at DETENAL.

The computer programs were written at the M.I.T. Artificial Intelligence Laboratory (Boston) during a summer visit; we acknowledge the hospitality of Prof. Marvin Minsky.

Some ideas developed during a winter stay at the Ecole Nationale Supérieure des Télécommunications (Paris), where we thank Prof. C. Guéguen.

Discussions at the Mathematics Dept. Centro Investigación del Politécnico and at the Computer Science Dept. IIMAS have been valuable.

### REFERENCES

1. E. Brabiesca and A. Guzmán, Shape description and shape similarity measurement for two-dimensional regions. *Proc. 4th Int. Jnt. Conf. on Pattern Recognition*, Kyoto, Japan, Nov. 1978, 608-612. Also available as a Technical Report PR-78-18, Computer Science Dept., IIMAS, National University of Mexico, 1978.
2. A. Guzmán and H. V. McIntosh, "CONVERT". *Communications of the A.C.M.*, 9, (8), 604-615 (1966).
3. H. Freeman and R. Shapira, Determining the minimum-area enclosing rectangle for an arbitrary closed curve. *Comm. Ass. Comput. Mach.*, 12, (7), 409-413 (1971).
4. A. K. Agrawala and A. V. Kulkarni, A sequential approach to the extraction of shape features. *Computer Graphics and Image Processing* 6, 538-557 (1977).
5. T. Pavlidis, Algorithms for shape analysis of contours and waveforms. *Proc. 4th Int. Jnt. Conf. on Pattern Recognition*, Kyoto, 70-85 (1978).
6. W. Perkins, A model-based vision system for industrial parts. *IEEE Trans. Computers*, C-27, 126-143 (1978).
7. H. Wechsler, A structural approach to shape analysis using mirroring axes. *Computer Graphics and Image Processing* 9, (3) 246-266 (1979).

**About the Author** - ERNESTO BRIBIESCA was born in 1952 in Mexico city. He attended the Instituto Politécnico Nacional, where he graduated as an Electrical and Electronics Engineer in 1974. His B.Sc. Thesis was a geographic data base system. He is currently working at DETENAL, (Dirección General de Estudios del Territorio Nacional) where he carries on research on reconfigurable data bases and computer studies on pure form and measures of shape resemblance.

**About of Author** - ADOLFO GUZMÁN was born in Ixtaltepec, Mexico, in 1943. He graduated from the Instituto Politécnico Nacional (Mexico) in 1965, obtaining a B.Sc. as an Electrical and Electronics Engineer, producing a thesis on "CONVERT", a pattern matching language. His graduate studies were carried on at the Massachusetts Institute of Technology, where he obtained his Ph.D. with the thesis "Decomposition of a Visual Scene into three-dimensional bodies". He stayed until 1970 as an Assistant Professor at the Electrical Engineering Department of M.I.T. After returning to Mexico, he was Director of the Centro Nacional de Cálculo of the Politécnico, Professor Titular at the Research Center (CIEA-IPN) of the Politécnico, and Director of the IBM Latin American Scientific Center. Currently, he is a Professor ("Investigador Titular") at the Applied Mathematics Institute (IIMAS-UNAM) of the National University of Mexico (UNAM), where he teaches graduate courses in Computer Architecture and Image Processing. His current interests are reconfigurable computer architectures, image analysis (measures for Pattern similarity), remote sensing and pictorial-geographic data bases (digital terrain models).



DIVISION DE EDUCACION CONTINUA  
FACULTAD DE INGENIERIA U.N.A.M.

COMPUTACION APLICADA A LA PLANEACION URBANA

Reconfigurable geographic data bases

Agosto, 1981

REPRINTED FROM:

# **PATTERN RECOGNITION IN PRACTICE**

Proceedings of an International Workshop  
held in Amsterdam, May 21-23, 1980

edited by

**Edzard S. GELSEMA**

Department of Medical Informatics  
Free University, Amsterdam

and

**Laveen N. KANAL**

Department of Computer Science  
University of Maryland, College Park, Md.



1980

NORTH-HOLLAND PUBLISHING COMPANY - AMSTERDAM • NEW YORK • OXFORD

## RECONFIGURABLE GEOGRAPHIC DATA BASES

Adolfo Guzmán

Computing Systems Dept., UNAM  
National University of Mexico

Information systems that handle geographic information usually store the data in one of these manners: (1) as a matrix of adjacent PIXELS, such as the LANDSAT Systems; (2) as polygons surrounding areas with uniform properties; (3) as shape descriptions with neighborhood relations among them. The time and difficulty used to retrieve, update, locate and transform the data, depend on the format it is stored.

In this paper a configurable data base is proposed, that dynamically changes its data representation, trying to provide best performance in view of current and past uses of it. This data base contains a cost function to be minimized, a statistics module that collects information of past usages of the data base, and a predictor that extrapolates the current user's behavior into future (possible) users behavior. Finally, according to the prediction, a reconfigurator module actually changes the format of the stored data to the most suitable format.

### INTRODUCTION

Geographic information is obtained in large quantities through the world. Airplanes and satellites such as LANDSAT and SKYLAB obtain each day vast amounts of pictorial data.

In order to efficiently use this information (see "The need for electronic paper"), it is necessary to improve the acquisition, distribution, storage and access to relevant parts of this data.

This article addresses the last two aspects of the problem: how to store the information in an economic way, and how to get it back efficiently when needed.

The first part of the paper describes and delimits the problem; the second part reviews actually available solutions. The main thrust of our proposition is in the third part, where a system is proposed that stores geographic information in several ways, and dynamically changes these storage formats, according to an efficiency criterion. The last part of the article describes other reconfigurable systems, one of which is already implemented at the University of Mexico.

The purpose of the paper is to invite criticisms and suggestions, since the proposed system will soon be implemented at the Computing Systems Department.

### LIMITATIONS OF GEOGRAPHIC INFORMATION SYSTEMS.

Memory. Pictorial and graphical information systems require large amounts of storage areas and fast access to it.

Fortunately, memory costs go down each year. Also, new memory types, between central memory and disk, have appeared (ccd's, bubble memories).

Also, minicomputers and recent microcomputers possess a large address space, which simplifies software writing.

In addition, each information level (each type of picture, map, detail and precision) has its optimum way of storing it. These ways have not been systematically combined in the same package.

Moreover, different uses of the same information have different retrieval speeds, according again to the storage format.

CPU Time. To passage large quantities of data, a great amount of computer's time is required. Processing is slow.

Fortunately, computer prices are also going down, at the same time that their speed goes up, because of faster components and better architectures (pipelines, asynchronous computers, image machines [5, 12, 13], heterarchical architectures [11]).

Data Acquisition. To acquire map information is expensive, as well as to digitize it. The same is true for image data.

Fortunately, a great part of the information is already in digital form, because it is acquired in that way, or it can be cheaply digitized if this is done at the early stages of its acquisition.

Languages. There is a need to design and use better picture-manipulation/picture-processing languages.

Some new designs [1] possess specialized data-types, and can be executed in parallel machines.

Man-machine interaction. The mutual collaboration between man and machine must be such that each part carries on the tasks best suited for it. It is also required that the common "language" for interaction should be picture related (application related).

It is relevant here to mention the efforts of Gaillat [6] to marry Detenal's geographic data bank [2] with the P.R. System [10], in order to produce a smart assistant to the photointerpreter.

#### PROBLEM DESCRIPTION

A bidimensional image or map is a set of geometric elements possessing attributes (properties). There are three types of geometric elements: points (well, ruins, airport), lines (rivers, boundaries, highways) and regions (lake, city, swamp). Examples of attributes are: population, altitude, chemical composition, slope. In addition, the adjective "geometric" indicates that each element has a position (x,y coordinates) in the plane.

#### THE NEED FOR "ELECTRONIC PAPER"

Most of the geographic information is recorded in maps and pictures, and it is accessed with the help of ruler and compass, copied by hand and ink, much in the same manner of Ptolemaic times.

These archaic methods of storing and accessing information create a strong bottleneck that delimits its uses.

Fortunately, the computer and visual/geographic information systems [10] provide new ways to store and use this information. Each way of storage has advantages and drawbacks. Why not combine these different manners of storage through an "intelligent" system that dynamically chooses the best for a given application or pattern of use?

A geographic data base is a data base containing items that are geometric elements. Since a set of geometric elements forms an image or map, it is appropriate to say that a geographic data base manipulates maps and images. This article deals with the problem "how to design geographic data bases that exhibit reconfigurability"

A data base is reconfigurable [7] if its contents can be stored in a different manner, changing the form or format in which they are kept, but keeping invariant the information contained. For instance, we can represent Lake Chapala (Figure "Lake Chapala") (a) by its boundary; (b) by a collection of pixels; (c) by its shape number [1]. From the point of view of the user, (a), (b) and (c) represent the same lake in different formats. Now, if the data base automatically changes (due to some convenient reason) the storage format of Lake Chapala from (a) to (b), it is a reconfigurable data base. The user did not order the change; he is unaware of it.

Reconfigurable data bases are able to choose the best format for the information they keep, according to the patterns of use of the data. They need to have, for each format, an access mechanism; conversion routines (from one format to another) are also needed.

An attribute is inheritable if the subelements of an element also possess the attribute. Thus, "it is a pine forest" is an inheritable attribute, since each part of an element that "is a pine forest" is also a pine forest. The property (this town) "contains 2000 inhabitants" is not inheritable.

Inheritable attributes allow savings in memory, because it is not necessary to repeat the attribute for each subelement; the attribute can be placed at a "high level" element. This will become clear later.

#### ACTUAL SOLUTIONS FOR DATA REPRESENTATION

A reconfigurable geographic data base has several manners to represent and store its data. It is now convenient to study them, in terms of occupied memory and also in terms of access methods and access time.

##### Representation of points.

X,y representation. A point is represented by a pair of coordinates. The attributes associated to that point are represented by a collection of pairs property (attribute)-value.

##### Representation of lines.

Freeman chains. A line is represented by the relative coordinates of sequential points. The coordinates of each point are expressed relative to the preceding point in the line.

The inheritable attributes are associated with the line, and the non-inheritable attributes are defined along the line.

Storage of Freeman chains is compact. Reconstruction of the line is easy, but the "geometric" properties of the represented line are difficult to obtain; for instance "What is the shortest distance between two lines?".

Polygonal lines. A line is represented by the absolute (x,y) coordinates of sequential points. That is, the line is approximated by straight-line segments, and each vertex is recorded.

The inheritable attributes are associated with the line, and the non-inheritable attributes are defined along the line. For instance, "elevation" is defined along



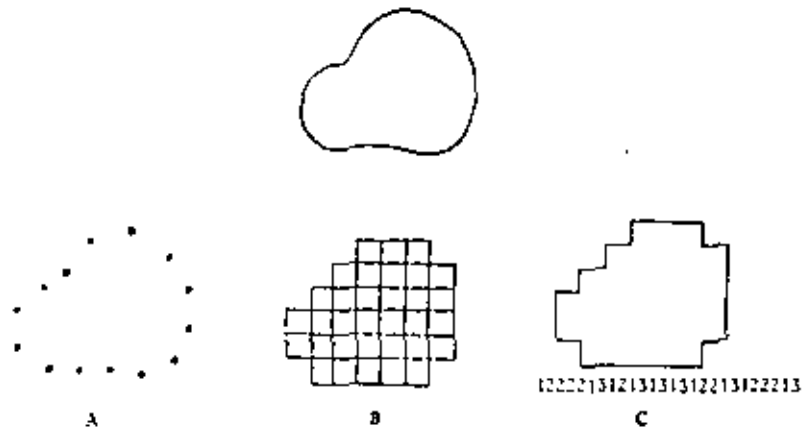


Figure "LAKE CHAPALA"

- (A) Stored as a collection of boundary points: (3.5, 4.2), (3.6, 4.7), ...  
 (B) Stored as a collection of pixels:  
 (C) Stored as a shape number: 1222213121313131312213122213

In a reconfigurable geographic data base, you could store Lake Chapala as a collection of pixels and later discover that the information system now contains it as a shape number. Under what conditions is such change desirable? Read the section "The proposed system".

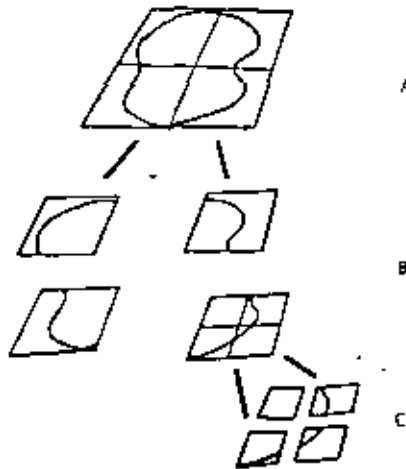


Figure "BRIBIESCA'S REPRESENTATION"

The region (A) begins to be subdivided (B) into a hierarchy of pixels (quad tree), but this subdivision is stopped prematurely (C). Each final leaf (C) then carries as attribute the Freeman chain describing its portion of the boundary. This increases the precision with which final graphs can be obtained.

the line, v. gr., at fixed intervals.

This is a popular representation. It demands more memory than the Freeman chain. Reconstruction is easy.

#### Representation of regions

Pixels. Each region is divided in squares (pixels) of size dictated by the desired precision. Attributes are assigned to each pixel, as if it were a point. For this reason, it is not possible to distribute further non-inheritable attributes within a pixel. For example, if a pixel contains "5000 inhabitants", we can not know from this information if they are located near its center, or in its SE corner. Should this information become necessary, the picture has to be resampled (in order to measure the attributes again) with a finer pixel size.

Pixel representation avoids storing the x,y coordinates of each pixel: these coordinates are implicit properties [7], because from the position of the pixel information within the record, it is possible to deduce the coordinates.

Access to the information is fast, since it is possible to compute in advance the record position of the needed information. Geometric properties are easily obtained, too, because each pixel has known coordinates.

Storage of properties associated with each pixel follows one of these two ways:

- (A) Access through the property. See figure 'Pixel Representation, (a)'. This is in widespread use. Each property is represented by a 2-d matrix of values. For instance, LANDSAT data consists of a "green image" stored in a large matrix containing numbers from 0 to 127 (the intensity in the green range), plus a "red image", plus an "infrared image", etc.

Other example: a matrix of altitudes.

This form of storage is advantageous when the values of the stored attribute exist everywhere (v.gr., altitudes). But it will be expensive to use for properties that show up only in a few places (copper deposits), since most of the data will consist of zeroes.

If the data base is going to frequently answer questions that combine several properties ("give me all the places with green=80 and altitude between 1500 and 2000"), it is disadvantageous to have the properties of the same pixel stored in different planes (matrices), since several records --one for each property-- need to be read from disk.

- (B) Access through the pixel. The properties or attributes of each pixel are placed in a list associated to the pixel. Stored in this format, a LANDSAT image will be a matrix of 4-tuples of numbers, each 4-tuple being of the form 64-15-119-23, meaning "green=64, red=15, infrared=119, etc." See Figure 'Pixel Representation, (b)'.

For attributes that exist everywhere, storage size is the same for both (A) and (B). But now suppose that it is necessary to store "rare" properties, that show up only in a few places. Then we could use (B) with explicit properties [7], meaning that together with the attribute value it is necessary to store the name (which is generally a number) of the attribute.

That is, a pixel could have the following property list:

```
GREEN - 64
COPPER - 301
ALTITUDE - 1500
```

while another pixel could have the following property list:

```
GREEN - 59
CACTUS - 22%
ALTITUDE - 1500
ARCHAEOLOGICAL MONUMENTS - 5
AIRPORTS - 1
```

Very well. Do not store inexistent attributes. Store the existing attributes as explicit properties. Nevertheless, these rules give rise to property lists of unequal length (See Figure "Pixel Representation, (c)") which is fatal for quick access to a given pixel. It is necessary to keep all the list equally long. If this is accomplished, we may as well use arrays instead of lists. I have found two ways to keep property lists of equal length:

- (1) In Detenal's geographic data bank [2], a "reasonable maximum length" is computed, and this size is assigned to each "list" (an array, in fact). If a pixel has fewer properties, space on its property list is left unused. If it has more, some properties are lost (not stored).
- (2) In the Mexican Child data base [9] popular properties (those existing everywhere) are stored implicitly [7], but "rare" properties are stored in a list of fixed size, like Detenal's. Both (1) and (2) achieve the use of fixed size property lists.

Victor Germán Sánchez [7] abounds in the advantages of implicit vs. explicit properties, when to use which, and when to reconfigure from one to another.

Quad trees. [14]. Each region is divided in an initial number of (big) pixels; to each of them properties are associated as explained under "pixels" above. Now, each of these pixels is considered under some homogeneity criterion (v. gr., that the gray level maximum difference between subportions of the pixel should not exceed 10% of the average; or that  $\max \text{height} - \min \text{height} < 10 \text{ m}$ ). If it is homogeneous, it is left alone and it does not generate sons. If not homogeneous, it produces a fixed number (four, for the quad-trees of U. of Maryland) of children. These are pixels of smaller size. All of them collectively cover the same region as their father. Attributes are also associated to these smaller pixels, thus allowing greater accuracy in the representation. Each pixel is further considered for subdivision, and so on; the process stops when all the (final) leaves of the tree satisfy the homogeneity criterion.

The pixels need not be square; the brothers need not be of the same size or shape: Dora Gómez [18] uses, for representation of terrain and 3-d surfaces (elevations), quad trees of irregularly shaped triangles of different size.

Each pixel does not need to have exactly four sons. Bribiesca [2] uses, in Detenal's geographic data bank, a different number of sons for each level, in order to have layer of pixels representing a given map scale. See Figure "Tree of Pixels in Detenal's Geographic Data Bank".

Quad trees allow easy access to the information. If we only keep the attributes of the final leaves (those pixels having no descendants), storage shrinks.

Quad trees keep fairly well most geometric and neighborhood properties.

Polygon representation. Describe each region by its boundary; describe this boundary by a polygon (see "polygonal lines" above); that is, by a number of perimeter points.

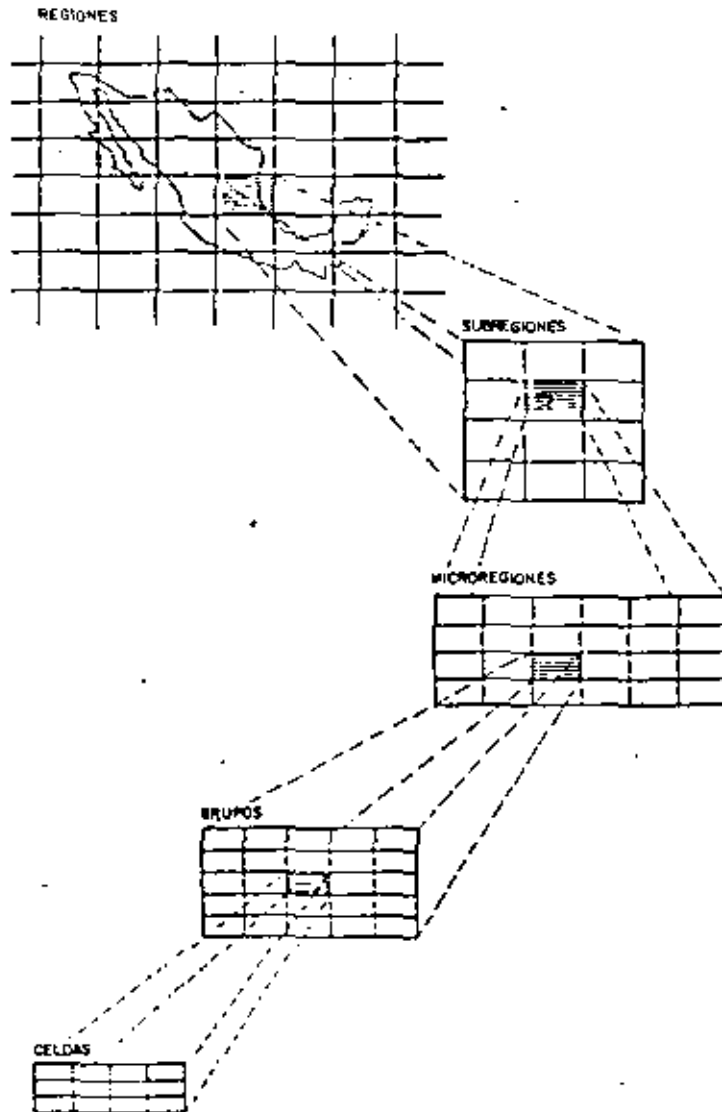


Figure "TREE OF PIXELS IN DETENAL'S GEOGRAPHIC DATA BANK"

Each layer is formed by pixels that represents maps at different scales. These scales determine the number of sons each pixel must have. At each level, all pixels have the same number of sons.

Properties are attached to each polygon.

Polygon representation occupies little memory. Nevertheless, to find region intersections ("give me all the deciduous forests that grow on top of sandy soil") is slow.

First pixels, then polygons. In a recent implementation of Detenal's geographic data bank, Fribiesca uses quad-trees up to a certain level, then stops. But at these leaves of the tree he then stores the Freeman chain of the part of the boundary of the region crossing that pixel. See Figure "Fribiesca's Representation".

This is useful when most questions asked to the data base do not need high precision in their answers (they are answered with the help of the pixels only), but the final maps need to be drawn precisely. Instead of a map with staircase-like boundaries, each step the size of a pixel, he uses the Freeman chain within a pixel and obtains a much smoother boundary. And he avoids a lot of polygon intersections.

Shape numbers. When a region is considered disregarding its size, orientation and position, it is seen as a "shape". From each shape, it is possible [5, 4] to deduce a unique shape number, that allows easy comparisons between shapes. With the help of the shape numbers it is possible not only to decide whether two shapes are identical, but (if not identical), how close in shape they are. They give a metric or distance that measures the similarity in shape.

Shape numbers occupy very little memory. Unfortunately, operations other than shape comparison become difficult on them: intersection of two shapes, for instance.

#### ACTUAL SOLUTIONS FOR DATA ACCESS

Geometric elements can be accessed (a) according to their geographic location; (b) according to the value of their attributes. It is very common to combine both mechanisms: "give me all the places within 10 kilometers of Popocatepetl Volcano having lava soil".

The access mechanisms for the attributes are precisely those used in general information systems: arrays, records, inverted lists, access vectors, hash coding, multilists, index sequential files, auxiliary keys, etc. [7].

The access mechanism for points is through their coordinates.

The access mechanism for lines is through the coordinates (or coordinates interval) of some point of it: its centroid, or a rectangle just enclosing the line.

For pixel access, to access a pixel with given x,y coordinates, these coordinates are used to compute the row (of the matrix) containing it. This record is then read from disk into memory. Then x gives us the position within the row.

To access a pixel within a quad tree, the tree is penetrated always from its root. Some interval comparisons of the coordinates of the needed pixel will reveal which of its sons contains such point. In this manner we descend towards the inner levels. To go down from one level to the next, some comparisons are required. Access is fast.

#### WHY ARE RECONFIGURABLE DATA BASES NEEDED?

The data access mechanisms described above for geographic information (that is, for purely geometric elements) greatly differ from those used in conventional information systems. When the first methods are combined with the last, a great variety results. Each combination has its ecological niche.

Nevertheless,

- (1) The needs for accessing and the subsequent modification of the information may vary, as time changes, and a different form of data representation may become more attractive.
- (2) Even if each image or map is represented in its most efficient format, it is quite common for the data base to have to answer questions that need to access information from several images, hence to have to manipulate several formats.
- (3) To transform each format to a canonical form each time a question is answered, in order to simplify the obtention of the answer, may become inefficient.
- (4) To save the same map information in different formats is very expensive in memory, as well as in updating time.

On the other hand, conventional (i.e., not geographic) information systems exhibit some diversity of formats for storing their data. Recently, there has been some interest in the implementation of reconfigurable systems [7], even in the form of reconfigurable machines [12].

#### THE PROPOSED SYSTEM

It is hereby proposed to design and to implement a geographic reconfigurable data base that:

- (1) Accepts and uses the main formats employed in cartographic/geographic/image processing systems.
- (2) Answers arbitrary questions about the information it contains, irrespective of the formats used to store it.
- (3) Obtains these answers in an optimal or near-optimal manner, with respect to the formats of the involved information. That is, "does its best" when it is required to agglutinate information kept in dissimilar forms.
- (4) Collects and extrapolates statistics about questions and updates, so as to decide to change the format of some portion of the information, if the trends show this to be desirable.

#### The parts of the system

The reconfigurable geographic data base will consist of the following models:

- A syntactic analyzer/parser for queries/updates.
- An optimizer.\*
- Access routines for different representations.
- Utility library to perform the operations.
- Reporter/graph maker.
- Statistics module.\*
- Predictor.\*
- Cost function.\*
- Reconfigurator.\*

With asterisks are signaled the modules needed to make to data base reconfigurable.

#### How the system works

The syntactic analyzer converts input from the user into calls to the appropriate access routines and utility subprograms. The optimizer is needed before these calls are made, because the richness of formats makes more desirable to chose a good way

to perform the operations. Then the access routines bring the data into memory, and the utility subprograms perform the desired operations. These operations usually end with a report or map being produced: this is the job of the report/graph maker.

While all this is being done, the statistics module counts the number of access and updates. The predictor extrapolates past use into future behavior: it predicts what future use of the data base will arise. From time to time, a cost function is invoked to ascertain whether the current formats are cost-effective. If not, this cost function suggests a reconfiguration. This reconfiguration is performed by the Reconfigurator.

We now proceed to describe each of the starred parts: those modules needed to make the data base reconfigurable.

The optimizer. When an execution command is received, the form and order of execution of the operations is important. For instance, if the command makes necessary to perform operations between a pixel image and a map stored in polygon representation (as it was the case with Gaillard [6]), the optimizer decides whether to transform the polygon to pixel representation and perform the operations in that format, or, instead, to do everything in polygon representation.

The statistics module. This part of the system keeps several counters, for retrievals, access, updates, searches and the like. This information pertains to the "recent past" behavior of the data base.

The predictor. By extrapolation of the behavior of the data base in the recent past, the predictor calculates its behavior in the immediate future, taking notice of trends.

The cost function. Each form of storage has different costs in terms of speed of access, storage required, and cpu time needed to perform several operations. The cost function contains all these parameters. From time to time, the system uses this function to compute the current cost (of the use of the data base in the immediate future). Specifically, it computes the cost of continuing using the data base with the current formats, as well as the cost of use with alternate formats. It then selects the best (least expensive) alternative, and indicates to the reconfigurator how it wants the data base to be changed.

The reconfigurator. Transforms the data base, changing the formats of storage of some of the items, but keeping unaltered its information. It basically consists of several transformation routines that go from one format type to another.

The reconfigurator works under request from the cost function. Since the task of reconfiguring also has a (high) cost, it should be taken into account in the cost function, at least in an approximate manner.

## CONCLUSIONS

Geographic information systems perform useful and necessary tasks, in view of the large amounts of data and processing required. The formats that the information systems use for storing its data have several costs in terms of speed and memory requirements. It is therefore advisable to design a reconfigurable geographic data base, which always keeps its information stored in a best or near-best set of formats.

Drawbacks. The proposed system is larger than fixed-format systems, since it handles richer varieties of formats. The cost of the Reconfigurator may be large in terms of programming effort and running time.

Advantages. The user needs not worry about how to capture its data, because it is internally converted, if necessary, into a near-optimum format. If the evolution

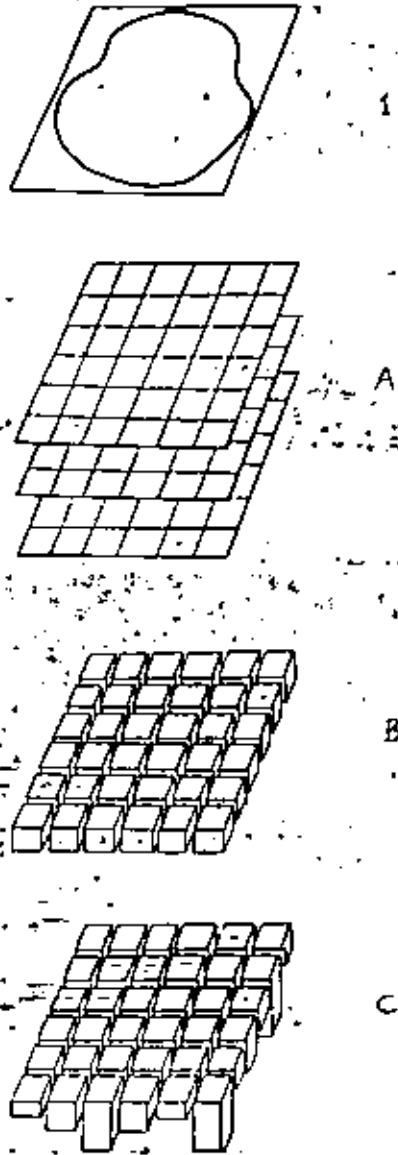


Figure - "PIXEL REPRESENTATION"

- (1) The region
- (A) Access through the property: each attribute forms a matrix of values
- (B) Access through the pixel: each pixel has a small list of attributes
- (c) Access through the pixel: same as (B), but the lists are of unequal length



function is properly chosen, the overall performance of the system will surpass that of a fixed-format system.

Future work. A reconfigurable data base is already implemented [7] at the University of Mexico. We propose to implement the system described in this paper at the Computing Systems Dept. of the University. Main difficulties felt are: the evaluation function, and the optimizer.

#### ACKNOWLEDGMENTS

My appreciation to the members of the Computing Systems Department, and particularly to Renato Barrera, who is coauthor of several ideas.

The AFR Project [12], with its reconfigurable computer, has provided a good place to work and think.

Work hereby described was partially supported by a grant (#1632) from the Consejo Nacional de Ciencia y Tecnología (Mexico).

#### REFERENCES

1. Barrera, A., Guzmán, A., Jinich, J., Radhakrishnan, T. Design of a High Level Language for Image Processing. IIMAS, 1979, PR-78-22. University of Mexico.
2. Bribiesca, E., Guzmán, A. User's manual to Exploit a Geographic Data Bank. OCAL-74-17, Scientific Center, IBM de México. Doc. 74. (in Spanish).
3. Bribiesca, E., and Guzmán, A. Shape Description and Shape Similarity measurement for two-dimensional regions. Proceedings of the 4th International Conference on Pattern Recognition Kyoto, Japan, 1978, 608-612. Also available as Comunicaciones Técnicas IIMAS-UNAM, 1978::9, 166 (PR-78-18). IIMAS, National University of Mexico. Accepted for publication in Journal of Geoprocessing.
4. Bribiesca, E. A How to describe pure form and how to measure differences in shapes using shape numbers. 1979 IEEE Conference on Pattern Recognition and Image Processing, Chicago, U.S.A. Accepted for publication in Pattern Recognition Journal. (1980).
5. Briggs, F.A., Fu, K.S., Hwang, K., Patel, J.H. IM: a reconfigurable multi-processor system for pattern recognition and image processing. TR-EE-79-11, March, 79. School of Electrical Engineering, Purdue University, U.S.A.
6. Gaillat, J.P., Méndez, R. An intelligent assistant to the photointerpreter. Comunicaciones Técnicas IIMAS, Univ. of Mexico, 1979 (in Spanish).
7. Germán-Sánchez, Víctor. A general reconfigurable information system. M. Sc. Thesis, IIMAS, National Univ. of Mexico. To appear in 1980 (in Spanish).
8. Gómez, Dora and Guzmán, A. Digital Model for three-dimensional surface representation. Journal of Geoprocessing, (1979), 53-70. Elsevier Publishing Co.
9. Guzmán, A., et al. Banco de datos del niño Mexicano. Comisión Nal. para el Año Internacional del niño. DIF. 1979. (in Spanish)
10. Guzmán, A., Seco, Rosa and Sánchez, Víctor Germán. Computer Analysis of LANDSAT images for crop identification in Mexico. Proceedings of the International Conference on Information Sciences and Systems, August 19-24, 1976. University of Patras, Patras, Greece.
11. Guzmán, A. Heterarchical architectures for parallel processing of digital images. IIMAS, 1979, AFR-79-3 (PR-79-23). Comunicaciones Técnicas, Serie Naranja. Univ.

- Nal. de México. Presented at the II Seminario Internacional sobre el uso de los sensores remotos. México, D.F., 1979.
- 12 Guzmán, A., y Segovia, Raymundo. A Parallel reconfigurable LISP machine. Proceedings of the International Conference on Information Sciences and Systems, August 19-24, 1976. University of Patras, Patras, Greece. 207-211.
  - 13 Levisaldi, S., Duff, M., Preston, K., Norgen, P., Toriwaki, J. Theoretical and practical considerations in the application of neighborhood logic to image Processing. Proc. 4th Intl. Joint Conf. on Patt. Recogn. Kyoto, 1979. 139-145
  - 14 Samet, H: Computing perimeters of images represented by quad trees. TR-755 # 368, DAAG-53-76C-0138, April 1979. Computer Science Center; Univ. of Maryland



**DIVISION DE EDUCACION CONTINUA  
FACULTAD DE INGENIERIA U.N.A.M.**

COMPUTACION APLICADA A LA PLANEACION URBANA

Digital model for three-dimensional surface representation

Agosto, 1981

## DIGITAL MODEL FOR THREE-DIMENSIONAL SURFACE REPRESENTATION

DORA GÓMEZ and ADOLFO GUZMÁN

Computer Science Department, IIMAS, National University of Mexico, Mexico 20 D.F.  
(Mexico)

## ABSTRACT

Gómez, D. and Guzmán, A., 1979. Digital model for three-dimensional representation. *Geo-Processing*, 1:53-70.

A tree of planar or spherical triangles is used to represent a 3-d surface; rather flat regions will be represented by large triangles; while abrupt zones will require further subdivision of the model into smaller triangles. Their vertices are not placed on a regular grid; they are allowed to fall at (or near) places such as ridges and peaks, where the change in slope is significant.

Starting from a collection, not necessarily good or complete, of "significant" points, the model selects five of them to form four triangles. Each triangle either matches the surface within a prespecified error tolerance, or else is further subdivided, by selecting appropriate "significant" points, into four triangular sons, which then receive in turn the same treatment. The tree stops growing when all the surface is represented within the specified tolerance. The model consists of the vertex points arranged into a table suitable for quick retrieval and interpolation.

Thus, the model guides its own construction; its components (points) are taken from the set of "significant" points, not in an arbitrary fashion but only where, and when needed. Since the model proposes the approximate location of the next point to be included in it, the set of "significant" points may be small or non-existent.

A constant signal to noise ratio and a representation thrifty in storage are achieved in this manner.

The model is being tested for use in digital representation of terrain elevation. Large savings in memory are expected, when compared to contour lines storage, for instance.

The paper concludes with some comments in favor of the use of this model to describe gray level pictures.

## INTRODUCTION

The digital representation of three-dimensional surfaces plays important roles in photogrammetry and cartography (drainage patterns, contour lines (Gómez, 1978), valleys formation, stereoscopy (Gómez), dimensions of the human body); scene analysis (occlusion of bodies, explanation of regions (Guzmán, 1971) and objects, range finding, shape from shading (Horn, 1970); computer graphics (hidden lines, shadows, coloring, specular reflexions); image processing; remote sensing (thickness of ice (Jensen, 1976), underground geology), and other disciplines.

The surfaces to be considered are of the following types:

$$z = f(x,y) \quad (I)$$

$$g(x,y,z) = 0 \quad (II)$$

Surfaces of type (I) give a single  $z$  value ("height") at each pair of coordinates  $x,y$ ; surfaces of type (II) can represent more general surfaces in the space, for instance the skin of a hand (Figure 1). The examples of the paper refer only to surfaces of type (I), but the model hereby proposed can be used for both types.

Our model for a surface consists of a minimum complete cover of triangles; that is, a mutually exclusive, collectively exhaustive finite set of triangles such that (1) each point in the surface is represented by exactly one triangle and (2) every triangle represents at least one point of the surface. To this we add the important restriction: (3) the difference between the coordinates  $x,y,z$  of a point in the surface and those  $x',y',z'$  of its representative (as given or computed from the model) is less than a pre-specified error or tolerance  $\epsilon$ . Thus, for two tolerances  $\epsilon_1 > \epsilon_2$ , the same surface will be represented by two models  $M_1$  and  $M_2$ , where  $M_2$  is a refinement of  $M_1$ .

For surfaces of type (I), we make  $x'=x$ ,  $y'=y$  and the distance between the real point and the model point is just the difference in heights.

#### Non uniqueness of the model

Several complete covers for a surface exist; in order to save memory, it is preferable to use a cover with fewer triangles (hence, each triangle covers in the average a larger area). As long as they obey restriction (3) above. The idea is: (4) not to subdivide a triangle unless it fails to represent reality within the  $\epsilon$  tolerance.

Even so, a 3-d surface can be represented by several models complying with (4) above. There is no unique model; by changing the position of the original rectangular frame a new representation is obtained. A different starting rectangle in the procedure or a different orientation of the grid will produce a different cover of triangles.

The width of the sidewalk is a constant fraction  $k_1$  (say, 10%) of the corresponding median:  $w_{SI} = k_1 \cdot s - r$ . To incorporate these sidewalks to the model, it was only necessary to modify the definition of the function *inside*, so that a point is "inside" triangle  $\Delta d$  if it falls inside it or at the sidewalk. (Refer to fig. 7). If not significant point  $r$  is found near enough  $r''$  so as to meet the 10% requirement, the model orders to find (to fabricate) a new significant point with coordinates  $x,y$  closer to  $x'',y''$ ; the model only accepts significant points close enough to  $r''$  ( $r''$  is the midpoint of  $(c_x,c_y,0)-(d_x,d_y,0)$ ) to ensure that the resultant flaps will not be wider than  $k_1$ . An optimal way (suggestion 1) is needed to compute  $k_1$ .

The sidewalks slightly contradict the assumption (1); uniqueness of representation for a point, of the introduction. This is of no importance.

This is not a problem for 3-d surface representation, as used for instance in applications to cartography and computer graphics. For surface comparison it is

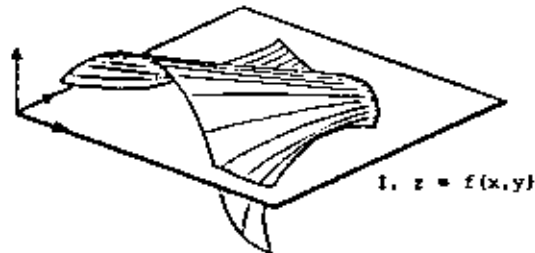


Fig. 1. TYPES OF SURFACES. The model described in this paper is able to represent either (I) single height surfaces, or (II) more general surfaces.

much better to have a unique (canonical) model, perhaps through a normalization procedure.

**The set of significant points.** Using a method external to the model, for instance stereoscopy (Gómez), gradient extraction (Signor and Kadler, 1978), river following, or others, an initial set of "significant" points is chosen on the 3-d surface that we want to represent. A point is called "significant" if in its neighborhood the change in slope is large.

The model begins by using some of these points; if it later finds necessary to grow, it indicates the approximate place ( $x,y$  coordinates) where a new "significant" point should be added to the model.

The model thus consists of a subset of "significant" points, defining a triangular irregular mesh; if the original set of "significant" points is too small, the model will suggest where to add one; if too many, most of them will be ignored (not included in the model); if the procedure that implements "significance" is noisy or unreliable, the model still guarantees the  $\epsilon$  tolerance, but storage economy suffers.

Therefore, in a computer implementation, it is not necessary to obtain first the set of significant points and then to pick the model from them; instead, the

model can begin to grow as soon as five or six are found, and the procedure that extracts significant points is called by the model as it deems necessary.

#### Obtention of the three-dimensional surface

It is assumed that the surface to be modelled already was obtained and exists available in some suitable representation, v.gr., a 2-d matrix containing height values. This data could have been obtained by stereocorrelation (Gómez) of a pair of pictures, by interpolation of digitized contour lines (Bribiesca and Avilés, 1974) or by other means.

#### CONSTRUCTION OF THE MODEL

In order to describe the model, it is necessary to explain

- (1) its constituent parts. In this case, they are vertices ("significant" points from the 3-d surface to be represented) that form planar, but tilted, triangles.
- (2) how the model is stored; the data structure used to keep the model in memory (primary or secondary storage). A tree of triangles, each with none or four sons, is used.
- (3) the use of the model: the procedure to follow for reconstruction of the 3-d surface from the model; the way to obtain the coordinates of a point in the surface from the cover of triangles. Here a directed access is used to the correct triangle starting from the top of the tree of (2), and falling down the appropriate chain of triangle sons, using little search and no backtracking.
- (4) the construction of the model, i.e., the obtention of its parts from the 3-d surface. A recursive procedure will be presented, where the model guides its own construction, by suggesting places (x,y coordinates) where to incorporate into itself points from the 2-d surface that are also "significant" with respect to changes in slope.

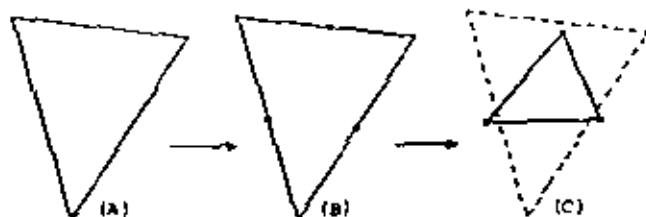


Fig. 2. TRIANGLE REFINEMENT. If it is necessary to refine triangle (A), three new vertices are proposed at the mid-points (B) of the sides; "significant" points are located near those mid-points; once they are found (C), four new triangles stand instead of the original (A).

## The parts of the model

To represent a surface  $z = f(x,y)$ , the model uses a collection of planar tilted triangles; each of them is defined by its three vertices, chosen to lie on the surface  $z = f(x,y)$  to be represented.

All the points inside the triangle are interpolated linearly: the surface inside the triangle is considered flat (but not horizontal, in general). Since the real surface  $z = f(x,y)$  is not flat, an error is introduced by this assumption. If everywhere in the triangle this error (height difference) does not exceed a tolerance  $\epsilon$ , the planar triangle is considered to be a good (and final or "terminal") representative for that region of the surface, and it is included in the model. If the error is larger, the triangle is discarded by dividing it into four smaller triangles, each of which in turn undergoes the same treatment.

Initially the surface is divided into a small set of arbitrarily chosen large triangles: if the surface is bound by a rectangle (as it is frequently the case in maps), four triangles are chosen as shown in part C of Figure 3.

The final model contains triangles (of different sizes) that represent the surface  $z = f(x,y)$  with a tolerance  $\epsilon$ . Each of the vertices of these triangles was proposed by the model by dividing a triangle in four through inclusion of new vertices near the middle points of the sides (Figure 2).

Once every triangle is refined, the vertices ( $E$  in Fig. 3) are stored in an appropriate way, suitable for quick data retrieval for surface reconstruction.

## When to stop refining

A triangle such as in Figure 2A is refined further, unless

- 1) the difference between the real height  $z = f(\bar{x}, \bar{y})$  and the computed height  $Z$  at the center of mass  $(\bar{x}, \bar{y}, E)$  of the triangle is smaller than  $\epsilon$ , and
- 2) every point in a grid of points spaced at most  $K$  units apart and inside  $A$  is within  $\epsilon$  of the real point on the surface  $z = f(x,y)$ .

Test (1) is a quick test; test (2) is applied only if (1) does not find a difference exceeding  $\epsilon$ .

$K$ , the distance between two points in the grid of (2), is a function of  $\epsilon$ ; normally,  $K = \min(\epsilon/m, K_0)$ , where  $m$  is the mean slope of the surface at the triangle ( $A$ ), and  $K_0$  is the diameter of the smallest topographic feature (hill, ravine) that it is necessary to represent in the model. Generally  $K_0$  is given by the user of the model: "be sure to check the model every 500 horizontal meters for accuracy"; then  $K_0 = 500$ .

Flow diagram. The procedure for construction of the model could be summarized as:

- Let  $T$  be the set of triangles that are candidates to be included in the model.

Initialize  $T$  with the four triangles of (C), Fig. 3.



- Mark every triangle of  $T$  as "terminal" if it passes tests (1) and (2) of the Section "When to stop refining". If these tests fail for a triangle, mark it "non-terminal", divide it into four sons (cf. Fig. 2) and add them to  $T$ .
- Exit when all triangles of  $T$  (including all the additions to  $T$ ) are marked (either "terminal" or "non-terminal"). Then  $T$  is the model.

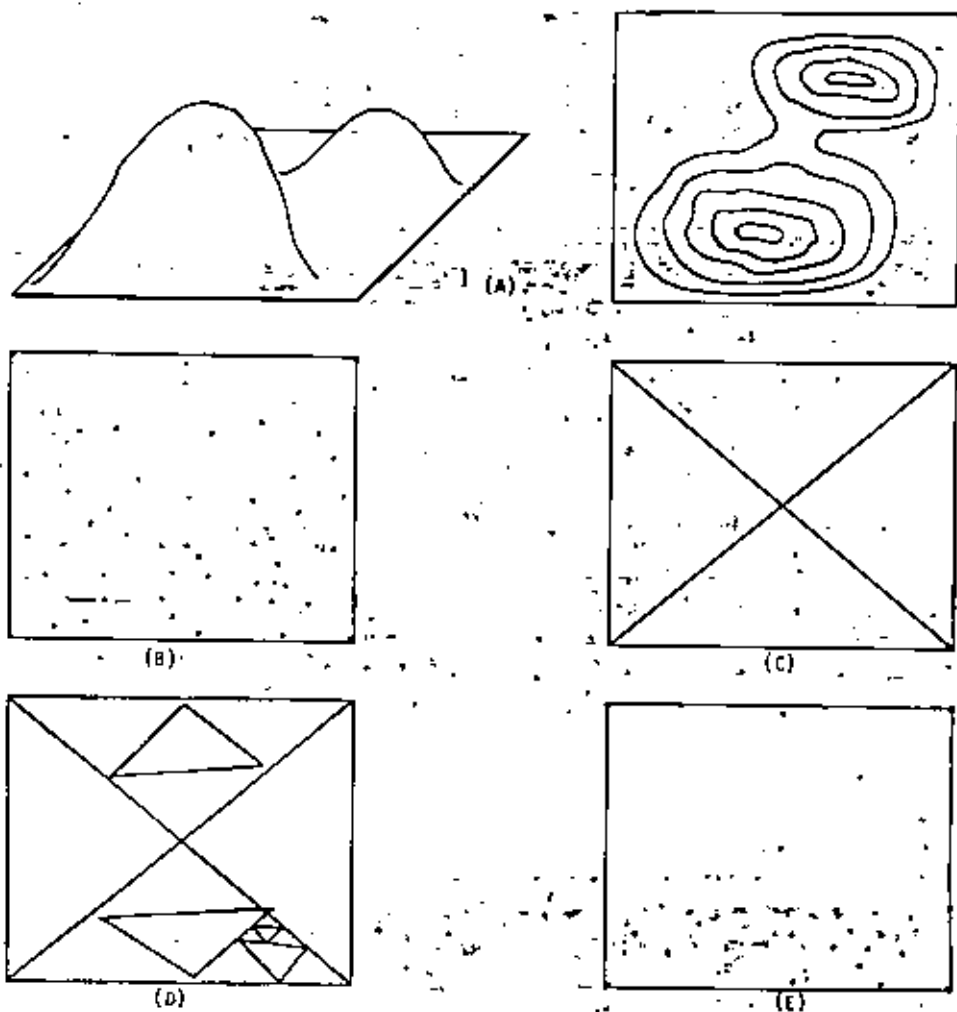


Fig. 3. MODEL BUILDING: (a) the surface, (b) the "significant" points, (c) the four initial triangles, (d) the final triangles, (e) the points of the model, usually a subset of (b). Triangles (d) are in the space (they project out of the paper); similarly, points (e) have three coordinates.

TABLE I. FLOW DIAGRAM FOR MODEL BUILDING.

This simple program constructs surface models such as that shown in fig. 6.

BEGIN

$T$  the four initial triangles of fig. 3C;

For every triangle in  $T$

if it passes tests (1) and (2) of section "When to stop refining"  
 then mark it 'terminal'  
 else mark it non-terminal and  
 add its four sons to  $T$ ;

END.

A non-terminal triangle is not needed in the model, since

- (1) its accuracy is worse than  $\epsilon$ , and
- (2) some of its descendants are a *finite* terminal triangles, hence suitable for modelling.

Thus, the model could be just the collection of terminal triangles.

This is advisable when the cover is made of similar triangles (q.v.), where it is easy to pick up the correct triangle for surface reconstruction. If the triangles are not similar, it is preferable to retain the non-terminal triangles into the model. This facilitates the addressing of the correct terminal triangle that gives the height  $z$  of a point  $(x,y)$  (i.e., the point  $(x,y,z)$  that represents the point  $(x,y,z)$  of the 3-d surface). More of this in the section IV 'Data Retrieval for Surface Reconstruction'.

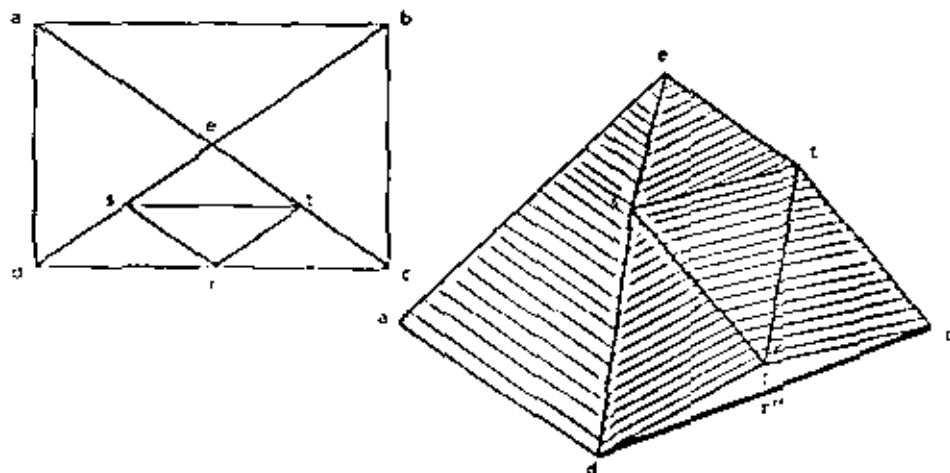


Fig. 4. MODEL VISUALIZATION. We try to give an isometric view of the appearance of the triangular model in 3-d space. Point  $E=(s_x, s_y, z)$  does not lie on the 3-d line  $d - c$ ; but point  $(s_x, s_y, 0)$  does lie on the line  $(d_x, d_y, 0) - (c_x, c_y, 0)$ . Triangles such as  $d - c - s$  or  $d - t - c$  are not part of the model; they represent no part of the real 3-d surface because they lie vertical. They are useless.

### Cover of similar triangles

Two polygons are similar if the corresponding angles are equal, the sides parallel and their length proportional.

If in Figure 2 we stop the refinement at (B), choosing the midpoints as new vertices to include in the model, the final cover of the model is composed of two families of similar triangles, because a line joining the middle points of two sides is parallel to the third side.

A word of caution: the triangles are not similar as they lie in the 3-d space. Their projections on the plane  $x,y$  do form a family of similar 2-d triangles (for triangles  $a b c$  and  $d e f$  of Fig. 4, and all their descendants) another family of similar 2-d triangles for triangles  $d a c$  and  $e c f$ , and all their descendants.

The advantages of the cover of similar triangles are:

- (a) storage of these triangles is easy. (Klinger and Nikitas) stores a hierarchy of squares.
- (b) reconstruction of the surface from the model becomes simplified.
- (c) a set of "significant" points (B in Fig. 3) is not needed.

The disadvantage comes from (c):

- (d) the model might contain more points, since they are not special or significant: they are not the best to choose for interpolation of planes.

### DATA STRUCTURE FOR MODEL STORAGE

This section describes the way to organize the storage of the model. Essentially, the storage consists of a collection of triangles. Each triangle is stored in a "frame"; each of them contains

- three internal vertices
- a "terminal" or non-terminal mark for each son.

The terminal mark (zero) indicates that a triangle son already fulfills the accuracy, hence it (the son) has no sons of its own --need not be further subdivided.-- The non-terminal mark, an integer different from zero, indicates the location (frame) in the model matrix occupied by this triangle son. Thus, when a node is marked as non-terminal, the mark itself also says where (in what frame) that son is stored. See Fig. 5 and Table II. Slightly different conventions were used in IIMAS-UNAM (Gómez, 1978).

The model is stored in a matrix (C, Fig. 5) which is a collection of frames. A non-terminal triangle occupies a frame; it stores clockwise (B, Fig. 5) its three central vertices and a mark specifying for each son whether it is terminal or not. A terminal triangle does not use a frame, since it has no sons. But a non-terminal triangle could very well have four terminal sons. That is the case of frames 4 to 9 of Fig. 6.

The initial frame, frame 1, is stored in a slightly different manner (part A of Fig. 5, because it describes a rectangle.

A more complicated example is given in Fig. 6.

Storage of vertices. When describing a non-terminal triangle (v.gr., triangle 1 2 3 of B, Fig. 5, only vertices 4, 5 and 6 are stored in the frame belonging to that triangle 1 2 3, since vertices 1, 2 and 3 were undoubtedly stored in the ancestors of triangle 1 2 3. This avoids multiple storage of vertices, and exploits the fact that in order to examine whether a point (x,y) falls inside the

TABLE II. NAMING CONVENTIONS.

These conventions are important for correct storage of vertices (such as 4 of the internal triangle P), and its subsequent appropriate retrieval for reconstruction of the 3-d surface. For a use, see definition of procedure 'altitude' in section "Data retrieval for surface reconstruction".

CONVENTIONS I. - Refer to part (A) of Fig. 5.

Vertices of triangles which are sons of the rectangle are named as shown. The correct names for (A) are:

rectangle: a b c d

M = triangle a b e

N = triangle b c e

O = triangle c d e

P = triangle d a e

CONVENTIONS II. - Refer to part (B) of Fig. 5.

Vertices of triangles that are sons of triangles are named clockwise, starting with the vertex that also belongs to the father.

If the triangle to be named is the internal triangle (P), then start with the vertex that falls near the middle point of line 1-2, where 1 is the first of the vertices that belong to the father, and 2 is the second of them.

The correct names for triangles of (B) are:

triangle 1 2 3      (first vertex is 1)

M = triangle 1 4 6

N = triangle 2 5 4

O = triangle 3 6 5

P = triangle 4 5 6

2-d<sup>2</sup> triangle 1 2 3 or not, we already asked a similar question to the ancestors of 1 2 3. In this way the coordinates to vertices 1, 2 and 3 are already known when triangle 1 2 3 is accessed (cf. Section 'Data Retrieval for Surface Reconstruction').

A vertex is stored by storing its three coordinates  $x, y, z$ . Some duplication (not triplication or multiplication) occurs when a vertex such as 5, 20 or 15 in Fig. 6 gets stored by two non-terminal brother triangles. For instance, vertex 5 is stored at frame 4 that describes triangle 3 1 9, and also at frame 3 that

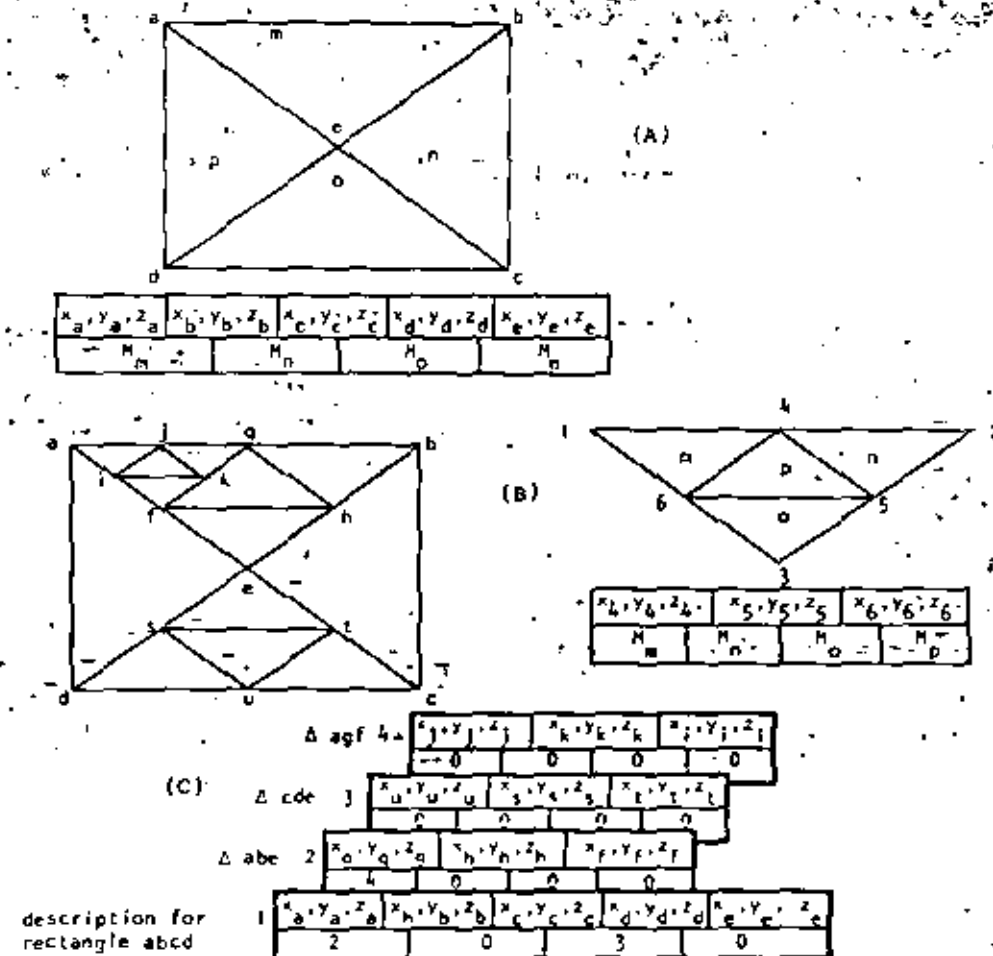


Fig. 5. DATA STRUCTURE. (A) Storage conventions for the initial rectangle. (B) Storage conventions for a non-terminal triangle. Its sons are M, N, O, P. (C) Example of a model and its data structure. Only a non-terminal triangle uses up a frame. The 2 0 3 0 marks of frame 1 mean that son M is non-terminal and it is described in frame 2, son N is terminal, son O is non-terminal and it is described in frame 3, and son P is terminal (mark = 0 means terminal). The model is stored in a matrix (C) which is a collection of frames.

describes triangle 2 5 3 9. The trivial cure will be to keep a table of vertices, and to store in the frame pointers to the table, instead of the three coordinates  $x, y, z$ .

This table of vertices is not used in our model because it saves little storage:

- (1) if both a pointer and a vertex coordinate occupy a word of memory, then to use the table requires 2 pointers + 3 coordinates = 5 words; not to use the table requires 3 coordinates + the same 3 coordinates = 6 words;
- (2) if for some reason triangle 2 5 3 9 selects vertex 5 as the "significant" point near the mid-point of side 3-9 (Refer to Fig. 6), but triangle 3 1 9 selects vertex 5' (a different vertex, near vertex 5 but not the same) as the "significant" point near the mid-point of side 9-3, then the table wastes memory.

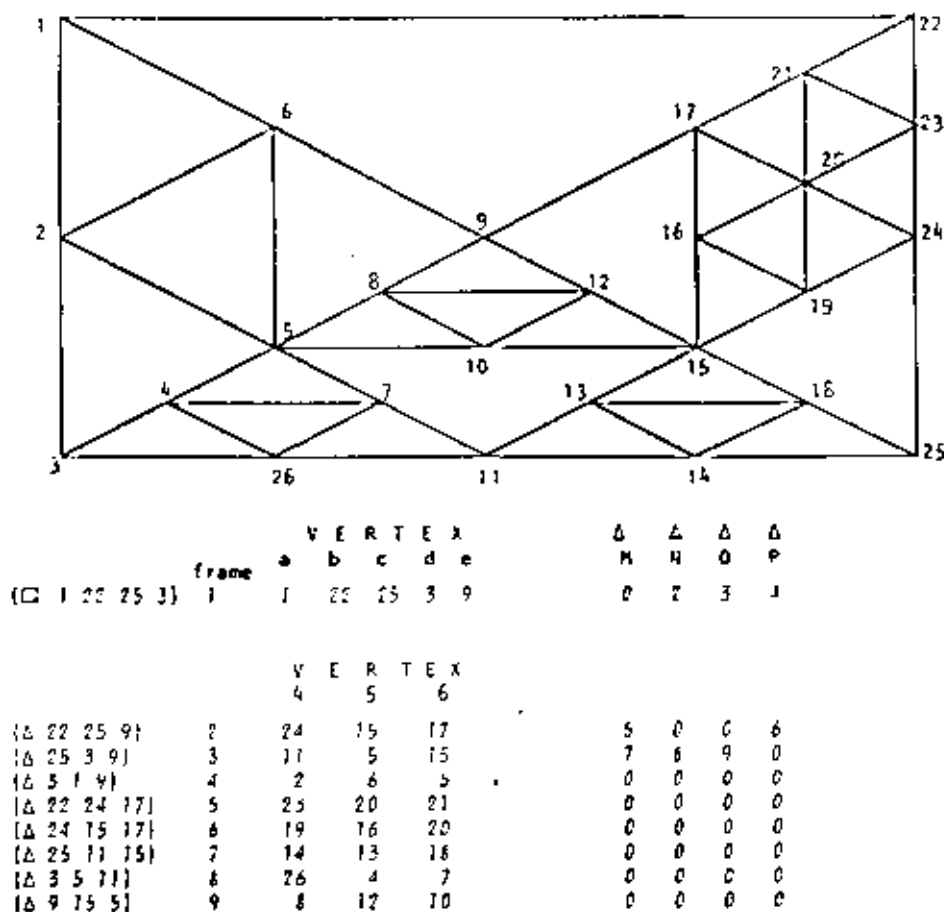


Fig. 6. MODEL EXAMPLE. This example was constructed using the rules (A) and (B) of Fig. 5 and Table II. Each frame consists of vertices and pointers to other frames. Only non-terminal triangles occupy a frame of the matrix. This matrix is the model.

## Simplified storage for cover of similar triangles

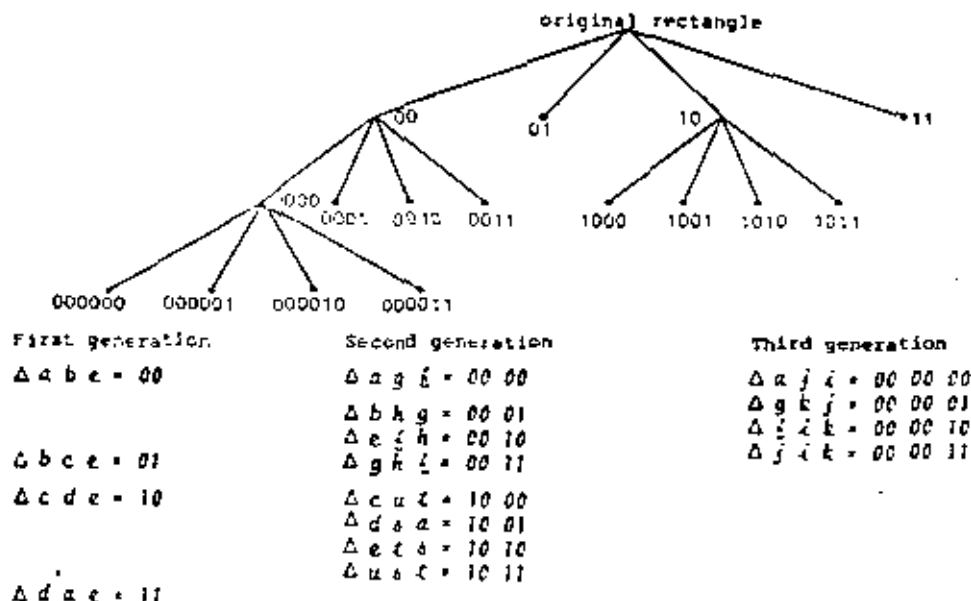
If we assume that the rectangle  $a b c d$  (Fig. 5) is a square and that the "significant" points are exactly at the mid-points of the sides of the triangles, instead of near them, then all the two-dimensional triangles<sup>2</sup> are similar (in fact, they are isosceles right angled triangles) and the  $(x,y)$  coordinates of any vertex need not be stored, since they are the average of the  $(x,y)$  coordinates of the vertices of an appropriate side.

The new representation for square  $a b c d$  of Fig. 5 is:

	Frame #	vertex a	vertex b	vertex c	vertex d	vertex e	$\Delta$ M N O P
(rectangle $a b c d$ )	<u>1</u>	$Z_a$	$Z_b$	$Z_c$	$Z_d$	$Z_e$	<u>2</u> 0 <u>3</u> 0
			vertex 4	vertex 5	vertex 6		
(triangle $a b c$ )	<u>2</u>		$Z_g$	$Z_h$	$Z_i$		<u>3</u> 0 0 0
(triangle $c d e$ )	<u>3</u>		$Z_u$	$Z_v$	$Z_w$		0 0 0 0
(triangle $a g h$ )	<u>4</u>		$Z_j$	$Z_k$	$Z_l$		0 0 0 0

If the original area is not an square but a rectangle, we will have two families of similar two-dimensional triangles.

If we denote the sons M, N, O and P by 00, 01, 10 and 11, then we could form from Figure 5 the following tree:



These codes could be combined with the  $z$  values to render a compact model. We do not pursue this further. In a similar manner, a tree of squares can be represented (Klinger and Nikitas).

#### DATA RETRIEVAL FOR SURFACE RECONSTRUCTION

In order to recover the 3-d surface, it is sufficient to ask the model what is the  $z$  value for any pair  $x, y$ . This is realized by the function ALTITUDE.

```

ALTITUDE (x, y)  % returns the height z of the point (x, y) as obtained from
                % the model. It is defined as:

a := MODEL [ 1, 1]; % first vertex of frame 1. Frame 1 is the rectangle.
b := MODEL [ 2, 1]; % MODEL [*, 1] is the frame 1, a non-terminal triangle.
c := -MODEL [ 3, 1]; % MODEL [*, *] is the matrix containing the whole model.
d := MODEL [ 4, 1];
e := MODEL [ 5, 1];
m := MODEL [ 6, 1]; n := MODEL [ 7, 1]; o := MODEL [ 8, 1];
p := MODEL [ 9, 1]; % retrieving the pointers to the sons.
error := -1;

ALTITUDE := if inside (a, b, e, x, y)
then      if m=0 then height (a, b, e, x, y)
          else ZETA (a, b, e, x, y, m)
else if inside (b, c, e, x, y)
then      if n=0 then height (b, c, e, x, y)
          else ZETA (b, c, e, x, y, n)
else if inside (c, d, e, x, y)
then      if o=0 then height (c, d, e, x, y)
          else ZETA (c, d, e, x, y, o)
else if inside (d, a, e, x, y)
then      if p=0 then height (d, a, e, x, y)
          else ZETA (d, a, e, x, y, p)
else error;

END ALTITUDE.

```

Function INSIDE (a, b, c, x, y) is true if the point  $(x, y, 0)$  is inside the triangle  $(a_x, a_y, 0)$ ,  $(b_x, b_y, 0)$ ,  $(c_x, c_y, 0)$  with sidewalks (see Fig. 7).

A point  $p$  is inside triangle  $a b c$  if  $p$  and  $c$  fall on the same side of  $a b$  and  $p$  and  $b$  lie on the same side of  $a c$ , and  $p$  and  $a$  rest on the same side of  $b c$ . A thesis (Gómez) contains listings and results.



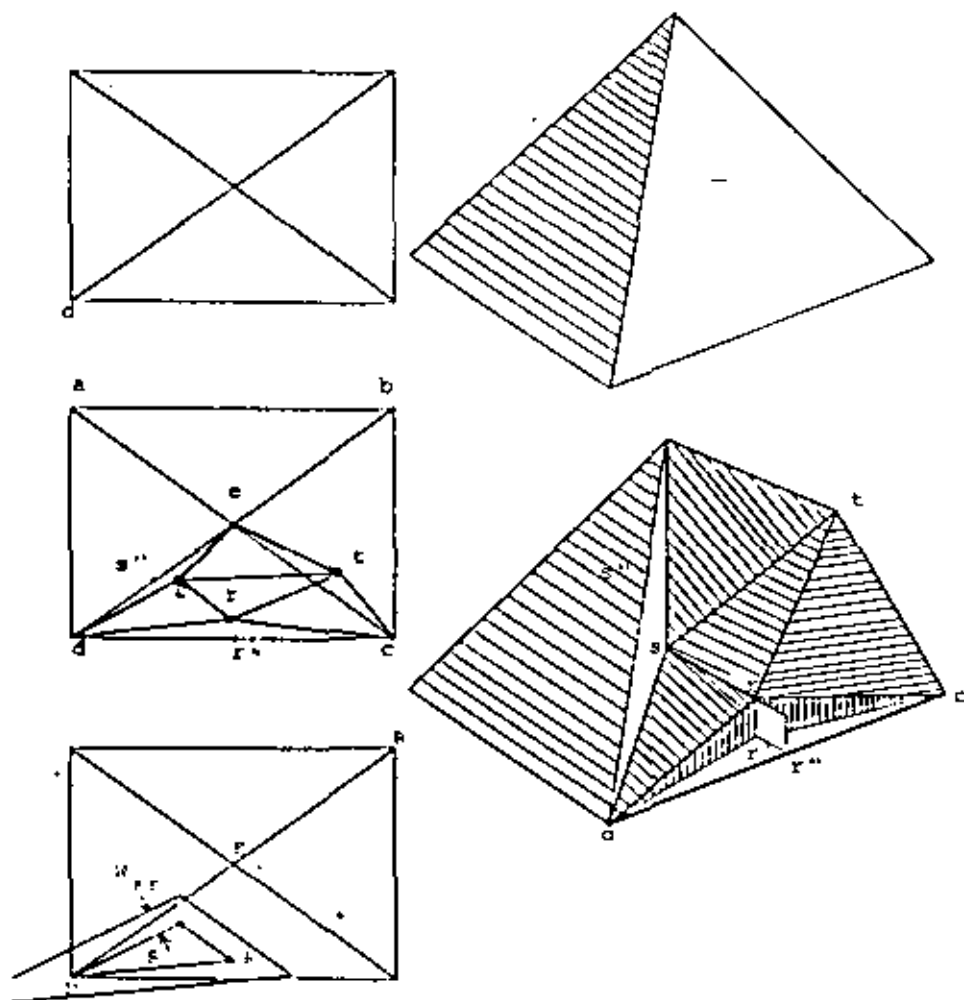


Fig. 7. COPLANAR SIDEWALKS. Compare with Figure 4. If point  $r = (r_x, r_y, 0)$  does not fall on line  $(c_x, c_y, 0) \rightarrow (d_x, d_y, 0)$ , an horizontal area  $c d t$  will be without coverage by the triangles; a corresponding part of the 3-d surface will fail to be represented. The cure for this is to give "flaps" to the triangles, so that triangle  $t d s$  (and its other three brothers) are enlarged by a coplanar sidewalk that covers up to  $t$ .

Procedure ZETA ( $v1, v2, v3, x, y, m$ ) is defined as

```

v4 := model [1,m];
v5 := model [2,m];
v6 := model [3,m];
mm := model [4,m];
n := model [5,m];
o := model [6,m];
p := model [7,m];

ZETA := if inside (v1,v4,v6,x,y)
then      if mm=0 then height (v1,v4,v6,x,y)
           else ZETA (v1,v4,v6,x,y,mm)  † see Table II 'Naming Conventions'
else if inside (v2,v5,v4,x,y)
then      if n=0 then height (v2,v5,v4,x,y)
           else ZETA (v2,v5,v4,x,y,n)
else if inside (v3,v6,v5,x,y)
then      if o=0 then height (v3,v6,v5,x,y)
           else ZETA (v3,v6,v5,x,y,o)
else if p=0
then      height (v1,v4,v6,x,y)
else      zeta (v1,v4,v6,x,y,p);

END ZETA.

```

The search for the correct triangle that represents a point generates no backtracking. At each level of the tree of triangles, we simply go down to the next level through the appropriate son (that son containing the point), until we hit a terminal triangle, where we compute the height by a planar interpolation.

#### CONCLUDING REMARKS

Since a gray level picture can be seen as a surface in three dimensions,  $z$  being the gray level value, it is in principle possible to use the models described here to represent them. This could have use for shape comparison of these surfaces, but the authors have not experimented with this. The idea, anyway, is to use models with large  $\epsilon$  (large error tolerance, coarse representation) to compare two surfaces; if the models are equal (in some appropriate sense, for instance, the quantized  $z$  values agree) then we could afford comparison with a smaller  $\epsilon$  (more accurate representation). In this way the shape similarity between any two 3-dimensional surfaces (or any two gray level pictures) can be ascertained. A related paper (Bribiesca and Guzmán, 1978a) develops this idea fully for two-dimensional flat regions (binary pictures). The largest problem

with this approach is to find a normalization procedure (the basic rectangle of (Bribiesca and Guzmán, 1978b)) that will produce a unique model for the 3-d case: it is easier to compare canonical models.

The method described in this paper is currently being implemented and tested for representation of topographic surfaces formerly described by their contour lines.

Merging of models into a larger model. If four adjacent surfaces  $a, b, c, d$  are represented by models  $\underline{a}, \underline{b}, \underline{c}, \underline{d}$ , the model of the joint surface  $[a, b, c, d]$  is formed by creating a new frame 1 (cf. Fig. 5) which has as non-terminal pointers  $M_n, M_R, M_O$  and  $M_p$ , pointers to the frames 1 of  $\underline{a}, \underline{b}, \underline{c}$  and  $\underline{d}$ .

Significant points vs. correlation points. The significant points (also called surface-specific points (Fauker, et. al., 1976) are those points of the terrain where slope changes in an important way. The points that a correlation routine finds in an easy manner, based for instance in the two pictures of a stereo pair, are called "correlation points;" they are points that are easy to correlate in the pictures, because the gray levels in their neighborhood are quite different from others, hence they can be identified rapidly and unmistakably. But they will not necessarily fall on top of "significant" points.

The components of the model. The model so far described and its construction can be seen as formed by:

- a tessellation of polygons (Gómez, 1978) (triangles in this case);
- an accuracy criteria, which tells whether a polygon of the model needs further refinement (in our case, comparison of modelled vs. real heights, cf. section "when to stop refining");
- a procedure to refine the model (in our model, select a significant point near the middle point of a side);
- a manner to store the model (as exemplified in Fig. 6);
- a way to access the model (as seen in section "Data Retrieval for Surface Reconstruction");
- a method to reconstruct the surface from the model (this is given by the procedure *height*  $(a, b, c, x, y)$  evaluated at the appropriate triangle  $a b c$  which contains the point  $(x, y, 0)$ ; the appropriate definition of containment is embodied in procedure *inside*  $(a, b, c, x, y)$ , which takes into account, for instance, the "flaps" of Fig. 7).

#### Suggestions for further work

1. Refer to Fig. 7. Do not use  $k_1 = 10\%$  for the width of the sidewalks. Compute instead the maximum distance that  $(x_x, x_y, 0)$  can be from  $r'$  for the enlarged

triangle  $i, d, s$  to meet still the error tolerance  $\epsilon$ . This has to do with average slopes of the triangles.

2. Refer to section "Simplified storage for cover of similar triangles". Fully develop the model that uses the representation of each triangle as a string of pairs of binary digits, v.gr., triangle  $g, k, j = 00\ 00\ 01$  (the son  $N$  of the son  $M$  of the son  $N$  of the rectangle).
3. Do not retrieve the triangles from the root of the tree (cf. section "Data retrieval for surface reconstruction") but store them so as to access them by a double binary search on the coordinates of the vertices (Gómez).
4. Consider the methods of this paper and of (Bribiesca and Avilés, 1974; Bribiesca and Guzmán, 1976a) as similar procedures that address data representation at arbitrary accuracy levels, and use them for shape comparison.

#### ACKNOWLEDGEMENTS

Andrew Clement and T. Peucker gave the triangular idea; Renato Barrera contributed to the concept of a hierarchy of triangles and other good advices. Abel Carreño and Angel García Amaro, of CETENAL, gave good photogrammetric advice. T. Radhakrishnan kindly revised the manuscript.

Work herein reported was partially done under the Joint Research Agreement (IX-1976) between CETENAL and LIAM.

#### REFERENCES

- Bribiesca, E. and Avilés, R. 1974. Codificación en cadenas y técnicas de reducción de información para mapas y dibujos lineales. IBM Latin American Scientific Center (Mexico City). Informe CCAL-74.
- Bribiesca, E. and Guzmán, A. 1978a. Shape description and shape similarity measurement for two-dimensional regions. Submitted to Fourth International Joint Conference on Pattern Recognition. Kyoto, Japan.
- Bribiesca, E. and Guzmán, A. 1978b. Shape numbers: a notation to describe pure form and to measure resemblance and difference in shape. Computer Science Dept., LIMAS, National University of Mexico. Report PR-78-18.
- Gómez, D. Modelos digitales del terreno de precisión variable. B.S. Thesis, Facultad de Ciencias, Universidad Nacional de México (in preparation).
- Gómez, D. 1978. Tessellation of triangles of variable precision as an economical representation for DTM's. Proceedings of the Digital Terrain Models Symposium St. Louis, Mo. Available from American Society of Photogrammetry.
- Guzmán, A. 1971. Analysis of curved line drawings using context and global information. In Machine Intelligence VI, (D. Michie and B. Meltzer, eds) University of Edinburgh Press. Chapter 20.
- Horn, S.K.P. 1970. Shape from shading: a method for obtaining the shape of a smooth opaque object from one view. Ph.D. Thesis, E.E. Dept., M.I.T. Project MAC Technical Report MAC-TR-79.
- Jensen, H.H. 1976. Collaboration in Physics within the Nordic countries. Europhysics news 7, 5, pp 1-4.

- Klinger, A. and Nikitas, A. Picture decomposition, tree data structures, and identifying directional symmetries as node combinations. *Computer Graphics and Image Processing* (to appear).
- Peucker, T., Fowler, R.J., Little, J.J. and Mark, D.M. 1976. Triangulated irregular networks for representing three-dimensional surfaces. Simon Fraser University, Burnaby, Canada. Technical Report # 10.
- Signor, G. and Radler, M. 1978. Une application de la corrélation numérique d'images: la stéréophotogrammétrie automatique. *Congres AFCET/IRIA, Reconnaissance des Formes et Traitement des Images*. Paris.



**DIVISION DE EDUCACION CONTINUA  
FACULTAD DE INGENIERIA U.N.A.M.**

COMPUTACION APLICADA A LA PLANEACION URBANA

Simap

Mapas por Computadora

Arq. Alejandro Villanueva Egan

Agosto, 1981



DIVISION DE ESTUDIOS SUPERIORES  
FACULTAD DE INGENIERIA  
SUBJEFATURA DE SISTEMAS

COMPUTACION APLICADA A LA PLANEACION URBANA

DECFI. AGOSTO 10-14



SYMAP

MAPAS POR COMPUTADORA

ARQ. ALEJANDRO VILLANUEVA EGAN.

1: DESCRIPCION DEL SISTEMA

SYMAP es un programa para la producción de mapas que fue desarrollado por el Laboratorio de Computación Gráfica y Análisis Espacial de la Universidad de Harvard. En la actualidad existen numerosos usuarios de este programa en todo el mundo. Su popularidad se basa en el hecho de que todos los centros de computación tienen los dos ingredientes que necesita para elaborar mapas: una computadora y una impresora de línea estándar. Como se mencionó anteriormente las impresoras de línea produce dibujos tomando ventaja de los diferentes tonos de gris logrados mediante la sobrepresión de diferentes caracteres alfanuméricos, utilizando generalmente 132 símbolos por línea. Esto no es una restricción al tamaño del mapa, ya que se puede producir un mapa en franjas que pueden ser ensambladas posteriormente.

La forma más común de alimentar con datos al programa es mediante tarjetas perforadas, pero la información también puede ser leída a partir de cintas o discos magnéticos. Este proceso normalmente es resultado de alguna computación previa, tal como la utilización de datos estándar, análisis de regresión, análisis factorial y otros métodos similares.

Hay un formato estándar para las tarjetas de datos, pero puede cambiarse bajo el control del usuario.

El programa está escrito en FORTRAN IV nivel G; necesita 128 K bytes de memoria y consta aproximadamente de 6000 instrucciones contenidas en un programa principal y 49 subrutinas. Para producir un mapa deben conjuntarse los diversos paquetes y opciones que se describen en los capítulos siguientes.

Entre las aplicaciones más frecuentes del programa se encuentran las siguientes:

- a. Mapeo estadístico
- b. Análisis de Mercados.  
Definición de patrones de eficiencia, y el carácter socio económico para la penetración de mercados potenciales.
- c. Planeación urbana.  
Localización de servicios e identificación de áreas problema.
- d. Selección de terrenos óptimos.  
Gratificación de factores como costo, pendiente, tipo de suelo, vegetación, profundidad de aguas freáticas, accesos a vías de comunicación, etc.
- e. Mapeo de recursos naturales.
- f. Análisis de la calidad del aire y del agua.  
Para identificar patrones espaciales y temporales relacionados a diferentes concentraciones de partículas en suspensión.
- g. Demografía.  
Para definir patrones pasados, presentes o proyecciones futuras de características de la población.
- h. Geología.  
Simulación de fallas en los estratos geológicos.

## II: TIPO DE MAPAS QUE SE PUEDEN PRODUCIR CON SYMAP

### 1. MAPAS CON INFORMACION POR ZONAS. (CROQUETAS).



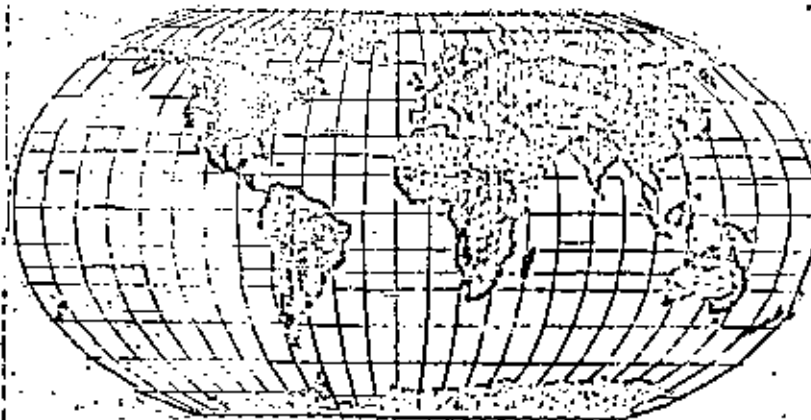
Las zonas pueden ser parcelas, manzanas, delegaciones, municipios, distritos, regiones, países, etc.

### 2. MAPAS CON INFORMACION MUESTRAL



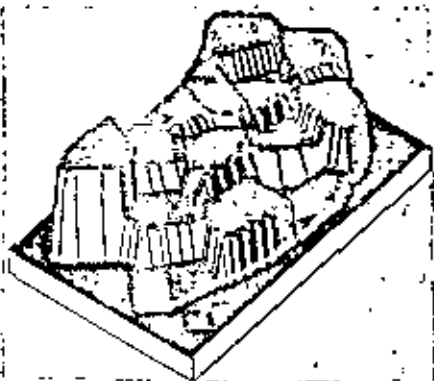
Mediante procedimientos estadísticos, el programa permite estimar el comportamiento de un fenómeno en la totalidad del área bajo estudio, basándose solamente en información tomada en puntos muestrales.

La información puede estar basada en encuestas, estaciones de monitoreo, fotografías aéreas o imágenes de satélite.

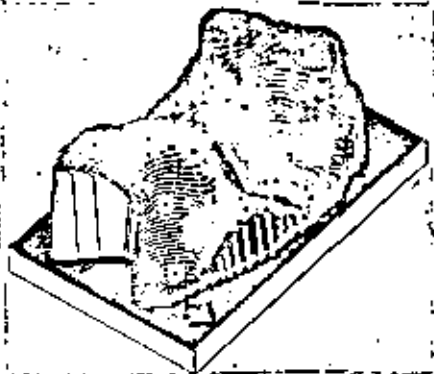


A continuación se muestran algunos ejemplos de los mapas que se pueden producir en otros programas similares y en algunos casos, compatibles con el programa SYMAP.

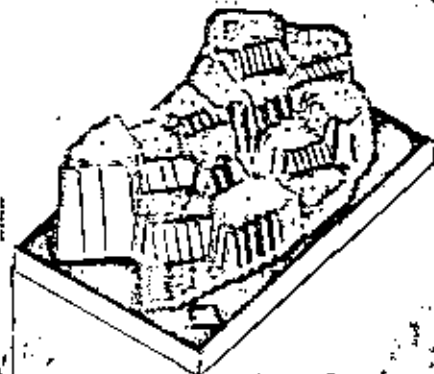




Mapa Isoplata



Mapa Próximo.



### III. OPCIONES DEL PAQUETE F-MAPA

1. Dimensiones del mapa.
2. Ventana del mapa.
3. Número de niveles ó intervalos de clase.
4. Valor mínimo de rango.
5. Valor máximo del rango.
6. Tamaño de los intervalos de clase ó niveles.
7. Simbolismo.
8. Eliminación de líneas de contorno ó isolinias.
9. Eliminación del histograma.
10. Texto explicativo.
11. Impresión del valor de los datos en el mapa.
12. Repetición múltiple de opciones.
13. Escala del mapa.
14. Márgenes del mapa.
15. Número de caracteres por pulgada.
16. Mapas grandes.
17. Eliminación de los resultados tabulares.
18. Invalidación de valores faltantes.
19. Valor mínimo para datos inválidos.
20. Valor máximo para datos inválidos.
21. Grabación del mapa en cinta.
22. Contornos continuos.
23. Eliminación del simbolismo de los puntos inválidos.
24. Eliminación de la interpretación numérica.
25. Eliminación de los símbolos de los puntos dato.
26. Alineamiento de sobreimpresión.
27. Tipo de mapa.
28. Extrapolación fraccional relativa.
29. Mínimo absoluto en la extrapolación.
30. Máximo absoluto en la extrapolación.
31. Radio inicial de búsqueda.
32. Radio máximo de búsqueda.
33. Número de puntos dato para la interpolación.
34. Independencia.

Opción 1.- DIMENSIONES DEL MAPA (1 tarjeta)

Col. 5	1 Para identificar la opción
Col. 11-20	Dimensión vertical del mapa, dada en pulgadas y como número decimal.
Col. 21-30	Dimensión horizontal del mapa dada en pulgadas y como número decimal.

Si no se incluye esta opción, el programa determina -- las dimensiones del mapa asignado 13 pulgadas a la dimensión mayor, siempre y cuando no se utilice la opción 13, sola ó en combinación con las opciones 2 y 14. Si la dimensión horizontal excede las 13 pulgadas (ancho del papel de la computadora), el mapa es impreso en secciones que deberán ensamblarse. El tamaño del mapa está limitado a 72 pulgadas en cualquier dirección, para mapas mayores, debe especificarse la opción 16.

Opción 2.- VENTANA DEL MAPA (1 tarjeta)

Col. 5	2 Para identificar la opción.
Col. 11-20 y 21-30	Coordenadas vertical y horizontal del punto superior izquierdo de la ventana.
Col. 31-40 y 41-50	Coordenadas vertical y horizontal del punto inferior derecho de la ventana.

La ventana es una región rectangular que el usuario desea observar dentro del área de estudio. Se especifica

ca mediante los puntos extremos del rectángulo, el cual deberá tener sus lados paralelos a los márgenes del mapa. Si no se especifica esta opción, se logra el mismo efecto mediante la opción 14.

Opción 3.- NUMERO DE NIVELES O INTERVALOS DE CLASE (1 tarjeta)

Col. 5	3 Para identificar la opción.
Col. 11-20	Número deseado de niveles, desde 1 hasta 10, perforado como número decimal.

Esta opción sirve para especificar el número de niveles ó intervalos de clase en que se desea dividir el rango total de los valores contenidos en los datos, con el propósito de asignar un simbolismo particular a todos los datos contenidos en el mismo intervalo. Si no se especifica esta opción, el programa divide el rango total de valores en cinco intervalos iguales. Puede usarse en combinación con las opciones 4, 5 y 6.

Opción 4.- VALOR MINIMO DEL RANGO (1 tarjeta).

Col. 5	4 Para identificar la opción.
Col. 11-20	El valor mínimo deseado, perforado como número decimal.

Todos los valores menores que este aparecerán con un simbolismo de 'L', que puede ser modificado mediante la opción 7. Si no se especifica esta opción, el programa toma como valor mínimo al menor de los valores contenidos en el paquete E-VALORES y que no esté declarado inválido por las opciones 18, 19 y/o 20.

Opción 5.- VALOR MÁXIMO DEL RANTO (1 tarjeta)

Col. 5                    5 Para identificar la opción.  
 Col. 11-20                El valor máximo deseado, perforado como un número decimal.

Todos los valores mayores que este aparecerán con un simbolismo especial de 'H', que puede modificarse mediante la opción 7. Si no se especifica esta opción, el programa toma como valor máximo al mayor de los valores contenidos en el paquete E-VALORES y que no está declarado inválido por las opciones 18, 19 y 20.

Opción 6. TAMAÑO DE LOS INTERVALOS DE CLASE O NIVELES (1 tarjeta).

Col. 5                    6 Para identificar la opción.

Especificando solo este número, el programa distribuirá equitativamente los valores de manera que haya aproximadamente la misma frecuencia en cada uno de los niveles. En caso de que se deseen intervalos diferentes, el tamaño de estos deberá estar indicado en la misma tarjeta de la siguiente manera:

Col. 11-20                Tamaño del primer intervalo, perforado como un número decimal.  
 Col. 21-30                Tamaño del segundo intervalo, perforado como un número decimal.

Así se continúa en campos de 10 columnas hasta la columna 70. En caso de más de 6 niveles, se usa otra

tarjeta. Esta opción puede usarse en combinación con las opciones 3, 4 y 5. Si no se especifica, el programa calcula 5 intervalos de igual tamaño ó intervalos iguales de acuerdo al mismo número especificado en la opción 3.

El tamaño de los intervalos puede especificarse de diferentes maneras:

Ejemplo 1:                Si existen 5 niveles y el tamaño de cada nivel es el doble del tamaño del nivel previo, se deben perforar los siguientes números:

Columna	5	11 - 20	21 - 30	31 - 40	41 - 50	51 - 60
	6	1.	2.	4.	8.	16.

Ejemplo 2:                Si se desea dividir los datos en cuatro grupos el menor 10%, el siguiente 25%, el siguiente 35% y 30% el restante, se deben perforar los siguientes números:

Columna	5	11 - 20	21 - 30	31 - 40	41 - 50
	6	10.	25.	35.	30.

Ejemplo 3.                Para especificar los siguientes intervalos

0	150	200	271.5	500	750	889	1200
---	-----	-----	-------	-----	-----	-----	------

Se deben perforar los números siguientes:

Columna	5	11 - 20	21 - 30	31 - 40	41 - 50	51 - 60	61 - 70
Tarjeta 1	6	150.	50.	71.5	228.5	250.	139.
Tarjeta 2		111.					

Cualquier valor que cae en el límite entre dos intervalos, es asignado al nivel mayor, con la excepción de valores que caigan en el límite del último intervalo.

#### Opción 7. SIMBOLISMOS (5 tarjetas)

En esta opción se especifican los caracteres que se de sean imprimir y sobre-imprimir para formar los símbolos representativos de cada uno de los niveles. Cada símbolo puede estar compuesto por un máximo de cuatro caracteres, perforados en la columna correspondiente al nivel, en estas tarjetas:

##### Tarjeta 1

Col. 3            7 para identificar la opción.

##### Tarjetas 2, 3, 4 y 5

Col. 1-10        Símbolo para cada uno de los intervalos dados en orden ascendente. Se debe utilizar solo las columnas necesarias para los niveles especificados.

Col. 11-20      Simbolismo para los puntos señal de los intervalos 1-10 en ese orden. Solo deben usarse las columnas necesarias para el número de intervalos especificado, dejándolas demás en blanco.

Col. 21          Símbolo para aquellas áreas del mapa cuyo

valor interpolado es menor que el especificado en la opción 4.

Col. 22          Símbolo para las posiciones de los puntos dato cuyos valores son menores que el mínimo establecido en la opción 4.

Col. 23          Símbolo para aquellas áreas del mapa cuyo valor interpolado es mayor que el especificado en la opción 5.

Col. 24          Símbolo para las posiciones de los puntos dato cuyos valores son mayores que el máximo establecido en la opción 5.

Col. 25          Símbolo para el fondo que aparece entre el área de estudio y los bordes del mapa, así como en las zonas declaradas inválidas por las opciones 18, 19 y/o 20.

Col. 26          Símbolo para las isoclinas y contornos de las zonas.

Col. 27          Símbolo para aquellas áreas del mapa donde no hay interpolación, debido al uso de la opción 15 y/o barreras impermeables.

Col. 28          Símbolo para indicar la presencia de dos o más puntos dato en la misma localización de la impresora.

Col. 29          Símbolo para indicar las posiciones de los puntos dato declarados inválidos en las opciones 18, 19 y/o 20.

Cuando no se especifica esta opción el programa tiene un conjunto de simbolismos estándar que se muestran a continuación:

Simbolismo general Simbolismo de los puntos dato

Columna:	1 2 3 4 5 6 7 8 9 0	1 1 1 1 1 1 1 1 1 2
Niveles		
1	.	1
2	- X	1 2
3	. 0 X	1 2 3
4	. + 0 X	1 2 3 4
5	. + 0 0 X	1 2 3 4 5
6	. + X 0 0 X	1 2 3 4 5 6
7	. ' + X 0 0 X	1 2 3 4 5 6 7
8	. ' + X 0 0 0 X	1 2 3 4 5 6 7 8
9	. ' = + X 0 0 0 X	1 2 3 4 5 6 7 8 9
10	. ' - = + X 0 0 0 X	1 2 3 4 5 6 7 8 9 0

A continuación se muestran ejemplos de simbolismo:

Ejemplo 1.

Se tienen 6 niveles y se desea invertir el orden del simbolismo estándar:

Columna	1 2 3 4 5 6 7 8 9 0	1 1 1 1 1 1 1 1 1 2 2 2 2 2 2
Tarjeta 1	7	
Tarjeta 2	0 0 0 X + .	1 2 3 4 5 6
Tarjeta 3	X X -	
Tarjeta 4	A	
Tarjeta 5	V	

Ejemplo 2:

Se desea suprimir todo el simbolismo dejando las isolíneas y los contornos de las zonas en negro y con un fondo de puntos:

Columna	1 2 3 4 5 6 7 8 9 0	1 1 1 1 1 1 1 1 1 2 2 2 2 2 2
Tarjeta 1	7	
Tarjeta 2	1 2 3 4 5	- 0
Tarjeta 3		X
Tarjeta 4		A
Tarjeta 5		V

Ejemplo 3:

Se desea usar los símbolos A, B, C, D y E para las clases de datos en un mapa de proximidad

Columna	1 2 3 4 5 6 7 8 9 0	1 1 1 1 1 1 1 1 1 2 2 2 2 2 2
Tarjeta 1	7	
Tarjeta 2	A B C D E	A B C D E
Tarjeta 3		/ / / / /
Tarjeta 4		
Tarjeta 5		

Opción 8. ELIMINACION DE LAS LINEAS DE LOS CONTORNOS Y LAS ISOLINEAS (1 tarjeta)

Col. 5 8 Para identificar la opción.

Estas líneas generalmente aparecen como blancas, a menos que sea especificado otro simbolismo para ellas en la opción 7. Si no se especifica esta opción, estas líneas aparecerán en el mapa.

Opción 9. ELIMINACION DEL HISTOGRAMA (1 tarjeta)

Col. 5 9 Para identificar la opción.

Si no se especifica esta opción, aparecerá impreso debajo del mapa un histograma que muestra la distribución de frecuencias de los datos en los intervalos de clase.

Opción 10. TEXTO EXPLICATIVO (12 tarjetas máximo)

El usuario dispone de esta opción para agregar a sus mapas información complementaria de cualquier tipo. El texto aparecerá inmediatamente debajo del mapa.

Primer tarjeta:

Col. 4-5 10 Para identificar la opción.

Tarjetas intermedias:

Col. 1-72 El texto explicativo deseado; pueden incluirse hasta 30 tarjetas.

Ultima tarjeta:

Col. 1-4 9999 para terminar el texto.

Opción 11. IMPRESION DEL VALOR DE LOS DATOS EN SUS LOCALIZACIONES CORRESPONDIENTES. (1 tarjeta)

Col. 4-5 11 para identificar la opción.

Si no se especifica esta opción aparecerá impreso el número del nivel ó intervalo en el que se encuentre el valor del punto dato.

Opción 12. REPETICION MULTIPLE DE OPCIONES (1 tarjeta).

Esta opción permite repetir todas las opciones usadas en el mapa previo.

Col. 4-5 12 Para identificar la opción.

Opción 13. ESCALA DEL MAPA (1 tarjeta)

Col. 4-5 13 Para identificar la opción.

Col. 11-20 La escala deseada, perforada como número decimal.

Si el mapa está medido en un sistema coordenado cualquiera, la escala especifica el número de pulgadas -- que representará a cada unidad de medida del mapa base. Por ejemplo, 2.0 producirá un mapa en el que cada unidad del mapa base está representada por 2 pulgadas. Si el mapa está medido en octavos y décimos de pulgada, la escala especifica que tantas veces se desea aumentar o disminuir el mapa. Por ejemplo, 1.0 - producirá un mapa a la misma escala del mapa base, 2.0 uno al doble, etc.. Si no se especifica esta opción, la escala quedará determinada por las opciones 1, 2 - y/ó 14.

Opción 14. MARGENES DEL MAPA (1 tarjeta)

Col. 4-5 14 Para identificar la opción.

Col. 11-20 Número de pulgadas deseadas como margen a partir del borde superior del mapa.

Col. 21-30 Número de pulgadas deseadas como margen a partir del borde izquierdo del mapa.

- Col. 31-40 Número de pulgadas deseadas como margen a partir del límite inferior del mapa.
- Col. 41-50 Número de pulgadas deseadas como margen a partir del límite derecho del mapa. Todos los márgenes deben perforarse como números decimales.

Se pueden especificar valores positivos y negativos para los márgenes. Los valores positivos añaden área a la ventana del mapa, mientras que los negativos le restan.

Opción 15. NÚMERO DE CARACTERES POR PULGADA (1 tarjeta)

- Col. 4-5 15 Para identificar la opción.
- Col. 11-20 Número de renglones por pulgada en los -- que la computadora ha sido ajustada para imprimir, dado como número decimal.
- Col. 21-30 Número de columnas por pulgada.

En muchas computadoras la impresora de línea ha sido ajustada para imprimir 6 renglones y 10 columnas de caracteres por pulgada. SYMAP supone que los resultados estarán dados en 8 renglones y 10 columnas de caracteres por pulgada y así es como produce mejores resultados. Si la impresora no está ajustada de esta manera, se deberá usar esta opción para evitar deformaciones en el mapa.

Opción 16. MAPAS GRANDES (1 tarjeta)

- Col. 4-5 16 Para identificar la opción.

Esta opción permite al usuario producir mapas mayores de 72 pulgadas (183 cm) en cualquier sentido. Si el mapa pasa de esta medida y esta opción no es usada, el programa reducirá la dimensión mayor a 33 pulgadas.

Opción 17. ELIMINACION DE LOS RESULTADOS TABULARES (1 tarjeta)

- Col. 4-5 17 Para identificar la opción.

Mediante esta opción se eliminan los resultados tabulares que aparecen antes del mapa y que contienen información sobre la localización, valor y nivel asignado a cada zona ó punto dato.

Opción 18. INVALIDACION DE VALORES FALTANTES (1 tarjeta)

- Col. 4-5 18 Para identificar la opción.

Con esta opción se hacen inválidos los valores de --- ó 0.0 y en blanco. El símbolo 'M' aparecerá en su localización. Si no se especifica, todos los valores serán considerados válidos.

Opción 19. VALOR MINIMO PARA DATOS INVALIDOS (1 tarjeta)

- Col. 4-5 19 Para identificar la opción.
- Col. 11-20 El valor que se declara inválido, perforado como número decimal. Este y todos los valores mayores serán declarados inválidos.

Puede usarse en combinación con la opción 20.

Opción 20. VALOR MAXIMO PARA DATOS INVALIDOS (1 tarjeta)

Col. 4-5 20 Para identificar la opción.  
 Col. 11-20 El valor que se declara inválido, perforado como número decimal. Este y todos los valores menores quedarán declarados inválidos.

Opción 21. GRABACION DEL MAPA EN CINTA (1 tarjeta)

Col. 4-5 21 Para identificar la opción.  
 Col. 19-21 1.0 Para obtener un listado de estos valores.

Al producirse un mapa, el programa calcula un valor para cada localización impresa en el mapa; en el caso de mapas isopletas y de proximidad lo hace mediante interpolación. Esta opción almacena estos valores en cinta para su utilización posterior por otros programas (SYMVU, por ejemplo).

Opción 22. CONTORNOS CONTINUOS (1 tarjeta)

Col. 4-5 22 Para identificar la opción.

Esta opción asegura la aparición de líneas de contorno que podrían ser suprimidas para permitir la representación de simbolismo descriptivo cuando el espacio entre puntos dato ó zonas es inadecuado para representar simbolismo y líneas de contorno a la vez.

Opción 23. ELIMINACION DEL SIMBOLISMO DE LOS PUNTOS INVALIDOS (1 tarjeta).

Col. 4-5 23 Para identificar la opción.

Elimina el simbolismo que aparece en cualquier punto - dato cuyo valor asociado está declarado inválido por las opciones 18, 19 y/ó 20.

Opción 24. ELIMINACION DE LA INTERPRETACION NUMERICA (1 tarjeta).

Col. 4-5 24 para identificar la opción.

Después del mapa y el texto explicativo, el programa - imprime información sobre los valores extremos, valores inválidos, los límites de los intervalos de clase y el porcentaje que estos representan en el rango, de la variable. Si no se desea esta información, debe especificarse esta opción.

Opción 25. ELIMINACION DE LOS SIMBOLOS DE LOS PUNTOS DATO. (1 tarjeta)

Col. 4-5 25 Para identificar la opción.

Elimina la aparición del simbolismo de los puntos dato poniendo en su lugar el simbolismo del valor interpolado en esa localización.

Opción 26. ALINEAMIENTO DE SOBREPRESION (1 tarjeta)

Col. 4-5 26 Para identificar la opción.

La sobreimpresión es realizada con sistemas diferentes en diversas computadoras, que no siempre coinciden con



de SYMAP. Esta opción sustituye este método en el caso de que el alineamiento sea incorrecto.

Opción 27. TIPO DE MAPA (1 tarjeta)

Col. 4-5 27 Para identificar la opción.

Esta opción permite al programa identificar el mapa como isopleta cuando se incluyen mapas coropletas en la misma corrida.

Opción 31. EXTRAPOLACION FRACCIONAL RELATIVA.

(Necesaria para mapas de proximidad) (1 tarjeta)

Col. 4-5 31 Para identificar la opción.

Col. 11-20 Límite deseado para la extrapolación expresada como una fracción del rango de la variable, perforado como número decimal.

El programa extrapola esta fracción del rango total de valores arriba ó abajo de un extremo local. Si no se especifica esta opción, el programa supone una extrapolación fracción de 0.1; en los mapas de proximidad dejar en blanco las columnas 11-20.

Opción 32. MÍNIMO ABSOLUTO EN LA EXTRAPOLACION (1 tarjeta)

Col. 4-5 32 Para identificar la opción.

Col. 11-20 Mínimo deseado, perforado como número decimal.

Esta opción fija un valor mínimo, abajo del cuál la

computadora no extrapolará.

Se recomienda su empleo cuando el usuario conoce un mínimo lógico para sus datos (0.0 para densidades de población, por ejemplo).

Si no se especifica, el programa tendrá un mínimo de interpolación igual al valor mínimo de los datos válidos.

Opción 33. MÁXIMO ABSOLUTO EN LA EXTRAPOLACION

(1 tarjeta)

Col. 4-5 33 Para identificar la opción.

Col. 11-20 Máximo deseado, perforado como número decimal.

Esta opción fija un valor máximo, arriba del cuál la computadora no extrapolará. Se recomienda su empleo cuando el usuario conoce un máximo lógico para sus datos (100.0 para datos porcentuales, por ejemplo). Si no se especifica, el programa tendrá un máximo de extrapolación igual al valor máximo de los datos válidos.

Opción 34. RADIO INICIAL DE BÚSQUEDA (1 tarjeta)

Col. 4-5 34 Para identificar la opción.

Col. 11-20 Radio inicial de búsqueda deseado, perforado como un número decimal.

El radio de búsqueda es la distancia sobre la cuál el programa busca puntos de datos para usarlos como base en-

la interpolación. Si no se especifica, el radio de búsqueda está basado en el número y dispersión de los puntos dato, utilizando en promedio 7 puntos.

las opciones 35 y 36. Si no se especifica esta opción, el programa calcula valores a cada 2 caracteres en sentido vertical y a cada 1 en sentido horizontal.

Opción 35. RADIO MAXIMO DE BUSQUEDA. (1 tarjeta)

Col. 4-5 35 Para identificar la opción.  
Col. 11-20 Radio máximo de búsqueda perforado como número decimal.

La computadora buscará puntos dato con los cuales interpolara a una distancia no mayor de este radio, que no deberá ser menor que el inicial.

Opción 36. NUMERO DE PUNTOS DATO PARA LA INTERPOLACION.  
(Necesaria para mapas de proximidad) (1 tarjeta).

Col. 4-5 36 Para identificar la opción.  
Col. 11-20 Número deseado de puntos (no más de 10), perforado como número decimal.

Si no se especifica esta opción, se tiene un mínimo de 4 y un máximo de 10, teniendo como promedio 7 puntos para la interpolación. Para mapas de proximidad dejar en blanco las columnas 11-20.

Opción 37. INDEPENDENCIA (Necesaria para mapas de proximidad)  
(1 tarjeta)

Col. 4-5 37 Para identificar la opción.

Esta opción previene el suavizamiento de las líneas creadas por las barreras C mediante la utilización de

IV. EL ALGORITMO DE INTERPOLACION DE SYMAP

IV.1 Introducción

Su objetivo es crear una superficie que cumpla ciertas propiedades para poder representar espacialmente un fenómeno a partir de la información proporcionada en un cierto número de puntos dato. Puede aplicarse a diferentes campos tales como demografía, meteorología, planeación urbana, contaminación ambiental, etc.

El programa toma las coordenadas de los puntos dato y sus valores asociados para construir una superficie - diferenciable en forma continua, que pasa por los puntos dato y representa las tendencias que éstos muestran.

El método consiste en obtener para cada localización de impresión en el mapa un promedio ponderado de las pendientes y los valores de los puntos dato cercanos, calculándolo mediante un modelo de tipo gravitacional.

El valor para cada localización se estima a partir de la fórmula:

$$z_p = \frac{\sum_i w_i z_i}{\sum_i w_i} \tag{1}$$

donde:

- $w_i$  = la ponderación del punto dato i.
- $z_i$  = el valor en el punto dato i, modificado por la pendiente en i y su desplazamiento con respecto al punto p.

IV.2 MODELO BASICO

De acuerdo al modelo básico, el valor en el punto P debe ser el promedio ponderado de los valores en los puntos dato 1, 2, ..., n, considerando la ponderación como el inverso de la distancia al cuadrado.

- Sea:  $\overline{P1}$  = distancia del punto P al punto dato 1
- $\overline{P2}$  = distancia del punto P al punto dato 2
- $z_1$  = valor en el punto dato 1
- $z_2$  = valor en el punto dato 2
- $z_p$  = valor que se va a calcular para el punto P.

$$z_p = \frac{\frac{1}{(\overline{P1})^2} z_1 + \frac{1}{(\overline{P2})^2} z_2 + \frac{1}{(\overline{P3})^2} z_3 + \dots + \frac{1}{(\overline{Pn})^2} z_n}{\frac{1}{(\overline{P1})^2} + \frac{1}{(\overline{P2})^2} + \frac{1}{(\overline{P3})^2} + \dots + \frac{1}{(\overline{Pn})^2}}$$

O bien:

$$z_p = \frac{\sum_{i=1}^n \frac{1}{(\overline{P_i})^2} z_i}{\sum_{i=1}^n \frac{1}{(\overline{P_i})^2}}$$

Si el punto P está muy cerca del punto 1, por ejemplo, entonces  $\overline{P1}$  es pequeña comparada con  $\overline{P2}$  y  $\overline{P3}$ , etc., por lo tanto el peso  $\frac{1}{(\overline{P1})^2}$  es grande comparada con  $\frac{1}{(\overline{P2})^2}$  y  $\frac{1}{(\overline{P3})^2}$  etc.

IV.3 MODIFICACIONES

Se deben hacer las siguientes modificaciones al método para hacerlo más eficiente:

a. Radio de búsqueda.

Por razones prácticas, tiene que limitarse el número de puntos dato que se consideren para interpolar el valor en una localización dada. Tomando en cuenta el número de puntos dato y el área sobre la que se extienden, el algoritmo determina un radio inicial de búsqueda R, en tal forma que un círculo con ese radio generalmente tendrá el número promedio de puntos en los que se basará la interpolación.

Para cualquier localización P, se eligen aquellos puntos dato cuya distancia efectiva a P, sea menor que R. Si dentro del círculo hay mas puntos que el número máximo especificado, se contrae el radio de búsqueda hasta que queden exactamente el número de puntos dato permitidos. Si dentro del radio inicial quedan menos puntos que el número mínimo permitido, el radio es expandido hasta que se encuentren todos los puntos especificado por el usuario ó hasta que sean usados todos los puntos que no estén bloqueados por una barrera impermeable.

Las ponderaciones consistentes en la inversa de la distancia al cuadrado, se usan para puntos dato cercanos a P. Cuando la distancia a un punto dato se aproxima al radio final de búsqueda R', la ponderación en ese punto tiende a cero.

b. Dirección

Para que las localizaciones relativas de los puntos dato entren en el cálculo, debe encontrarse el "aislamiento direccional" del punto dato i mediante la fórmula:

$$Q_i = \frac{1}{P_1} [1 - \cos(i P_1)] + \frac{1}{P_2} [1 - \cos(i P_2)] + \dots + \frac{1}{P_n} [1 - \cos(i P_n)]$$

Si los otros puntos dato 1, 2, 3, ... están en la misma dirección que i con respecto a P, entonces los ángulos i P1, i P2, i P3, son pequeños. Las cantidades (1 - cos(i P)) son también pequeñas y la suma Q<sub>i</sub> es cercana a cero. Dado que i no es el único punto en una dirección particular, se le da una ponderación reducida definida de la siguiente manera:

Si P<sub>j</sub> es la distancia de P al punto dato j, la distancia ponderada es

$$\frac{1}{(P_j)^2} \text{ para } 0 < P_j \leq R'/3$$

S<sub>j</sub> =

$$\frac{1}{(P_j)^2} \text{ para } R'/3 < P_j \leq R'$$

Sea N =  $\sum_j S_j$  y sea

$$T_i = \sum_{j \neq i} S_j \times [1 - \cos(i P_j)]$$

los puntos  $j$ , son puntos dato vecinos dentro del radio de búsqueda  $R'$ . La ponderación total en el punto dato  $i$  para la localización  $P$  es:

$$w_i = (S_i)^2 \times (H+T_i)$$

Entre mayor sea el aislamiento direccional de un punto dato y menor su distancia al punto considerado, mayor es su ponderación y por tanto su influencia en la determinación del valor interpolado.

c. Pendientes

Para evitar que la superficie presente niveles en los puntos dato, se calcula un gradiente bi-dimensional (pendiente)  $\left[ \frac{\partial Z}{\partial X} \right]_i + \left[ \frac{\partial Z}{\partial Y} \right]_i$  en cada punto dato, tomando un promedio ponderado de las pendientes de varios planos secantes. Cada plano contiene al punto  $P$  y a uno de sus puntos dato vecinos  $i$ ; el plano es tan horizontal como sea posible; la línea que va de  $P$  al punto dato debe ser la línea de pendiente más pronunciada en ese plano. El valor de la superficie en el punto  $P$  se aproxima a:

$$Z_i^1 = Z_i + \Delta Z_i$$

donde:

$$\Delta Z_i = \left\{ \left[ \frac{\partial Z}{\partial X} \right]_i \Delta X + \left[ \frac{\partial Z}{\partial Y} \right]_i \Delta Y \right\} \times k_i$$

$\Delta_x$  y  $\Delta_y$  son las diferencias  $x$  y  $y$  tomadas a partir de  $i$ .

El factor  $k_i = \frac{a}{a + P_i}$  se introduce en tal forma, que el efecto de considerar la pendiente es pequeño a grandes distancias. El parámetro  $a$  se escoge en tal forma que aunque  $i$  fuera el punto dato con la pendiente más pronunciada,  $\Delta Z$  pueda ser menor en magnitud que una fracción especificada del rango total de  $Z$ .

Si el valor en un punto dato  $i$  es mayor que los valores de los puntos dato circundantes, la pendiente en ese punto no será cero a menos que las pendientes ponderadas de los planos secantes se cancelen a sí mismas. Así, la superficie continuará creciendo en alguna dirección y por tanto los máximos y mínimos relativos no ocurren con frecuencia en los puntos dato.

d. Barreras

El algoritmo se modifica si se incluyen barreras a la interpolación. Las distancias usadas en los cálculos son "distancias efectivas". El cuadrado de la distancia efectiva es igual al cuadrado de la distancia real más la suma de los cuadrados de las resistencias de todas las barreras atravesadas un número  $n$  de veces. A las barreras impermeables se les asigna una resistencia tan grande que la distancia efectiva resultante hace que no sea considerado un punto como punto dato cercano, y por tanto se elimina del proceso de interpolación.

IV.4 EJEMPLO DE INTERPOLACION USANDO EL ALGORITMO DE SYMAP.

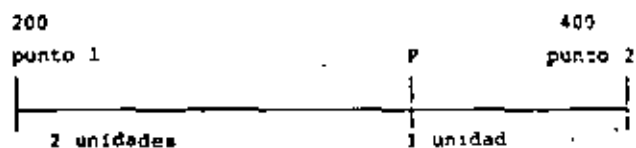
punto 1 = 200 hab/km<sup>2</sup>

Densidad de población

punto 2 = 400 hab/km<sup>2</sup>

$$z_p = \frac{(217.4) (1.875) + (384.6) (2.5)}{3.375} = 341$$

Densidad de población en el punto P = 341 hab/km<sup>2</sup>



i	S <sub>i</sub>	H	T <sub>i</sub>	W <sub>i</sub>
1	.5	1.5	2.0	.375
2	1.0	1.5	1.0	2.5

i	Z <sub>i</sub>	$\frac{\partial Z}{\partial x}  _{x_i}$	K <sub>i</sub>	$\partial Z_i$	Z <sub>i</sub>
1	200	66.7	.261	17.4	217.4
2	400	66.7	-.211	-15.4	384.6

$$W_i = 3.375$$

El valor en el punto P, es:

$$z_p = \frac{\sum_1^i W_1 + \sum_2^i W_2}{W_i}$$

sustituyendo:



DIVISION DE EDUCACION CONTINUA  
FACULTAD DE INGENIERIA U.N.A.M.

COMPUTACION APLICADA A LA PLANEACION URBANA

Imgrid

Uso del suelo e impacto ambiental

Arq. Alejandro Villanueva Egan

Agosto, 1981

MEMPLC DE ESTADISTICAS BASICAS  
 ESTADISTICAS BASICAS  
 FILE ESTALIST (CREATION DATE = 06/02/81) ARCHIVO CON LA INFORMACION  
 SUBFILE SUEA SUEB SUEC

06/02/81

PAGE

POSITIONAL INDEX  
 VARIABLE PAGE VARIABLE PAGE VARIABLE PAGE VARIABLE PAGE  
 V2 0 V3 0 V4 0 V8 0

MEMPLC DE ESTADISTICAS BASICAS  
 ESTADISTICAS BASICAS  
 FILE ESTALIST (CREATION DATE = 06/02/81) ARCHIVO CON LA INFORMACION  
 SUBFILE SUEA SUEB SUEC

06/02/81

PAGE 5

ALPHABETIC INDEX  
 VARIABLE PAGE VARIABLE PAGE VARIABLE PAGE VARIABLE PAGE  
 2 0 V3 0 V4 0 V8 0

061

MEMPLC DE ESTADISTICAS BASICAS

06/02/81

PAGE

DATA TRANSFORMATION DONE UP TO THIS POINT..

NO OF TRANSFORMATIONS	0
NO OF RECORD VALUES	0
NO OF ARITHM. OR LOG. OPERATIONS	0

AVE FILE

FILE ESTALIST HAS BEEN SAVED WITH 9 VARIABLES..

EQNLP SUBFILE CASNGT V1 V2 V3 V4 V8 V9

THE SUBFILES ARE..

NAME	NO OF CASES
SUEA	15
SUEB	13
SUEC	

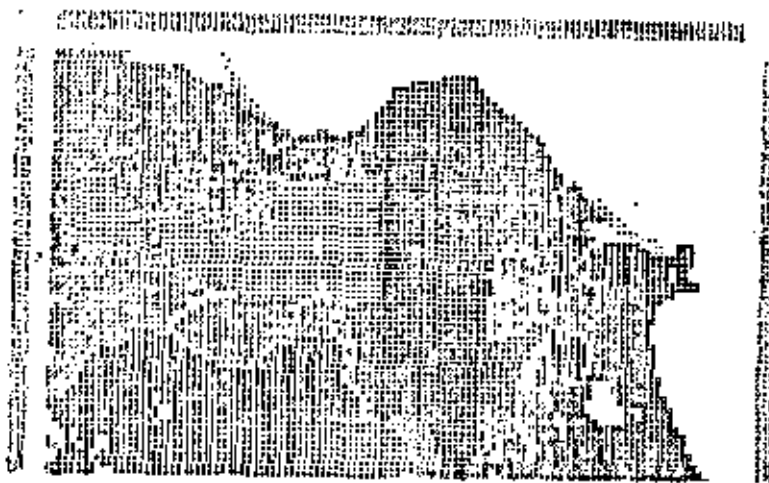
FINISH





COMPUTACION APLICADA A LA PLANEACION URBANA

DECFI, AGOSTO 10-14



IMGRID

USO DEL SUELO E IMPACTO AMBIENTAL

ARG. ALEJANDRO VILLANUEVA EGAN.

1: DESCRIPCION DEL SISTEMA

IMGRID es un conjunto de programas para computadora que procesa y mapea la información generada dentro de un proceso de planeación ecológica del uso del suelo. Los seis programas que lo componen constituyen técnicas avanzadas de inventarios y análisis de recursos. El sistema permite que los programas interactúen entre sí, ya que los resultados de cada etapa son almacenados en archivos de disco magnético para que puedan ser utilizados por los demás módulos o programas en el proceso.

Los programas están escritos en el lenguaje de programación FORTRAN IV y en el sistema UNAM/B-6700 se tienen dos subrutinas en ALGOL para el cambio de nombre en los archivos.

En este sistema, el usuario es un factor esencial para el éxito del proceso, ya que los resultados dependen de su conocimiento y la información de que disponga sobre el proyecto en cuestión.

Las ventajas que ofrece el sistema son las siguientes:

- a. Un proceso de planeación bien estructurado  
El sistema ofrece un marco de referencia para analizar en forma explícita la forma en que se toman las decisiones y se asignan criterios de evaluación.
- b. Conocimiento de las soluciones propuestas y sus repercusiones ambientales.  
Permite predecir los impactos probables de la urbanización sobre los eco-sistemas existentes, dándole así una capacidad mayor de previsión a los planes.
- c. Flexibilidad  
Debido a la generalidad en su diseño, este sistema se puede aplicar a cualquier proyecto, independientemente de su localización y escala.
- d. Mayor conocimiento de los eco-sistemas existentes en el área  
Al utilizar este sistema, el grupo planificador va adquiriendo conocimientos acerca de las componentes ambientales relevantes, su funcionamiento, su interacción, los factores que las controlan, como afectan y sus afectadas por el uso del suelo, etc.

e. Facilidad de utilización.

El sistema está diseñado para ser utilizado por personas sin conocimientos en programación, dado que se controla mediante instrucciones sencillas que realizan operaciones específicas. Una vez que el usuario entiende la naturaleza de las instrucciones, está en posibilidad de manejar el sistema sin mayor conocimiento del mismo.

f. Adaptabilidad con otros sistemas

INGRID puede formar parte como un sub-sistema dentro de un sistema mayor de recopilación, almacenamiento, análisis y graficación de información geográfica.

g. Ahorro en tiempo y costo mediante la computadora.

La facilidad de manipulación y manejo de grandes cantidades de información por las computadoras le da a INGRID una ventaja de eficiencia económica sobre los métodos tradicionales de análisis subjetivo y dibujo manual.

## II: COMPONENTES DEL SISTEMA

Los programas para computadora que componen el sistema INGRID son los siguientes:

a. BANCO DE DATOS

Programa ECODATOS/INGRID.

Permite almacenar en forma digital la información relevante de los recursos del suelo registrada a partir de los mapas fuente. El programa crea un archivo en disco magnético para cada variable, pudiendo contener hasta 50 variables diferentes subdividida cada una de ellas en un número máximo de 10 categorías.

Estos archivos son utilizados por los programas correspondientes a la producción de mapas y modelos analíticos.

b. MODELOS DE ATRACTIVO

Programa ATRACTIVO/INGRID

Mediante este programa se pueden especificar modelos para encontrar la localización más adecuada de los usos del suelo propuestos.

Es posible considerar hasta 20 diferentes usos del suelo, cada uno con su modelo correspondiente. El método utilizado por los modelos consiste en obtener para cada una de las celdillas en el área de estudio, un índice de "atractivo" que refleja su adecuación para localizar un uso del suelo específico.

El usuario determina las variables del banco de datos que intervienen en el modelo, las jerarquiza y asigna preferencias a sus categorías, expresando así, los factores de localización de cada uso del suelo que sean importantes a su criterio.

El programa utiliza categorías que toman las variables en cada celdilla y clasifica el índice como un promedio ponderado de las asignaciones del usuario.

El atractivo de cada celdilla puede tomar un valor entre 0 y 9, considerándose al 0 como "nada atractivo" y al 9 como "lo más atractivo".

Al efectuarse una corrida, se crea un archivo en disco magnético para los resultados de cada modelo. Estos archivos son utilizados por los programas correspondientes a la producción de mapas y evaluación de planes.

c. MODELOS DE VULNERABILIDAD.

Programa IMPACTO/INGRID

Mediante este programa se pueden especificar modelos para estimar el daño probable ocasionado por los usos del suelo propuestos sobre los eco-sistemas existentes en el área de estudio.

Es posible considerar hasta 20 sistemas diferentes, cada uno con su modelo correspondiente. El método utilizado por los modelos consiste en obtener para cada celdilla en el área de estudio un índice de vulnerabilidad o grado de impacto negativo producido por un uso del suelo específico. El usuario determina las variables del banco de datos que intervienen en el modelo y pondera sus categorías en forma individual y combinada para expresar su sensibilidad al impacto considerado.

El programa accesa el banco de datos para registrar las categorías que toman las variables en cada celdilla y determina el grado de impacto negativo producido de acuerdo a las asignaciones hechas por el usuario.

El impacto puede tomar un valor entre 1 y 4, de acuerdo a la siguiente escala:

- 1 = competitiva
- 2 = moderada
- 3 = severa
- 4 = terminal

Al efectuarse una corrida, se crea un Archivo en disco magnético para los resultados de cada modelo.

Estos archivos son utilizados por los programas correspondientes a la producción de mapas y evaluación de planes.

#### d. EVALUACION DE PLANES

Programa PLANES/INGRID

Programa EVALUACION/INGRID

Mediante estos dos programas es posible almacenar en disco magnético los diversos planes de uso del suelo generados por el usuario y evaluarlos con respecto a los criterios de atractivo y vulnerabilidad formulados en los modelos. El resultado de la evaluación consiste en dos tablas de resumen.

La primera, expresa el atractivo ó adecuación lograda en la localización de los usos del suelo. La segunda, indica el grado de impacto negativo causado por el plan sobre cada uno de los sistemas descritos por el usuario.

Estos resultados son almacenados en archivos de disco magnético para ser accedidos por el programa de producción de mapas.

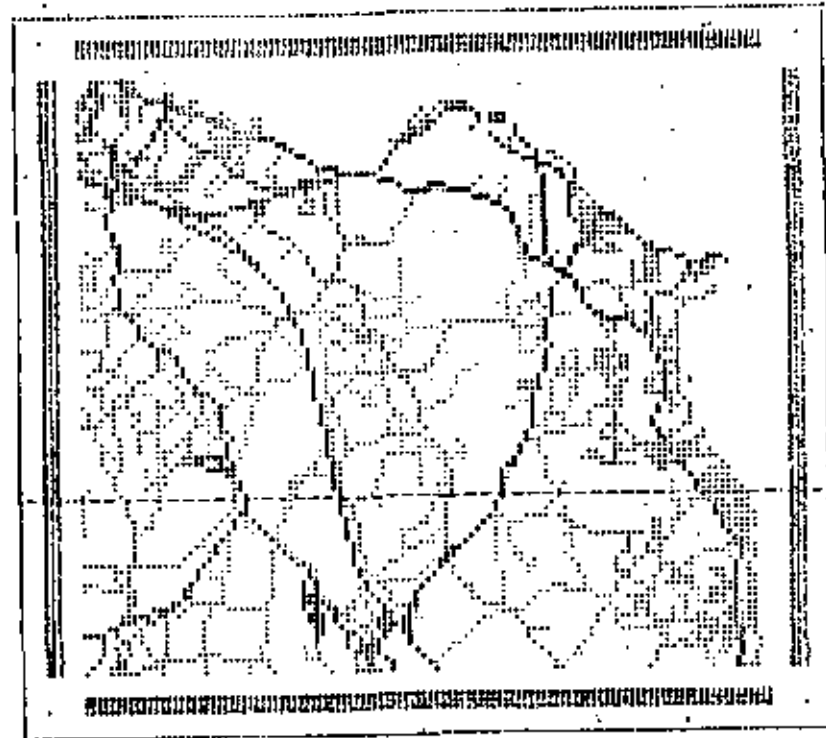
En una corrida pueden evaluarse hasta 30 diferentes planes de uso del suelo.

#### 5. PRODUCCION DE MAPAS

Programa MAPAS/INGRID

Mediante este programa se utiliza la impresora de líneas de la computadora para producir mapas de la información contenida en el banco de datos y la generada por los modelos y la evaluación de planes.

El usuario puede controlar aspectos del mapa tales como tamaño, simbolismo, numeración de referencia, textos explicativos y obtención de histogramas. La figura muestra un ejemplo de estos mapas.





DIVISION DE EDUCACION CONTINUA  
FACULTAD DE INGENIERIA U.N.A.M.

COMPUTACION APLICADA A LA PLANEACION URBANA

Imgrid

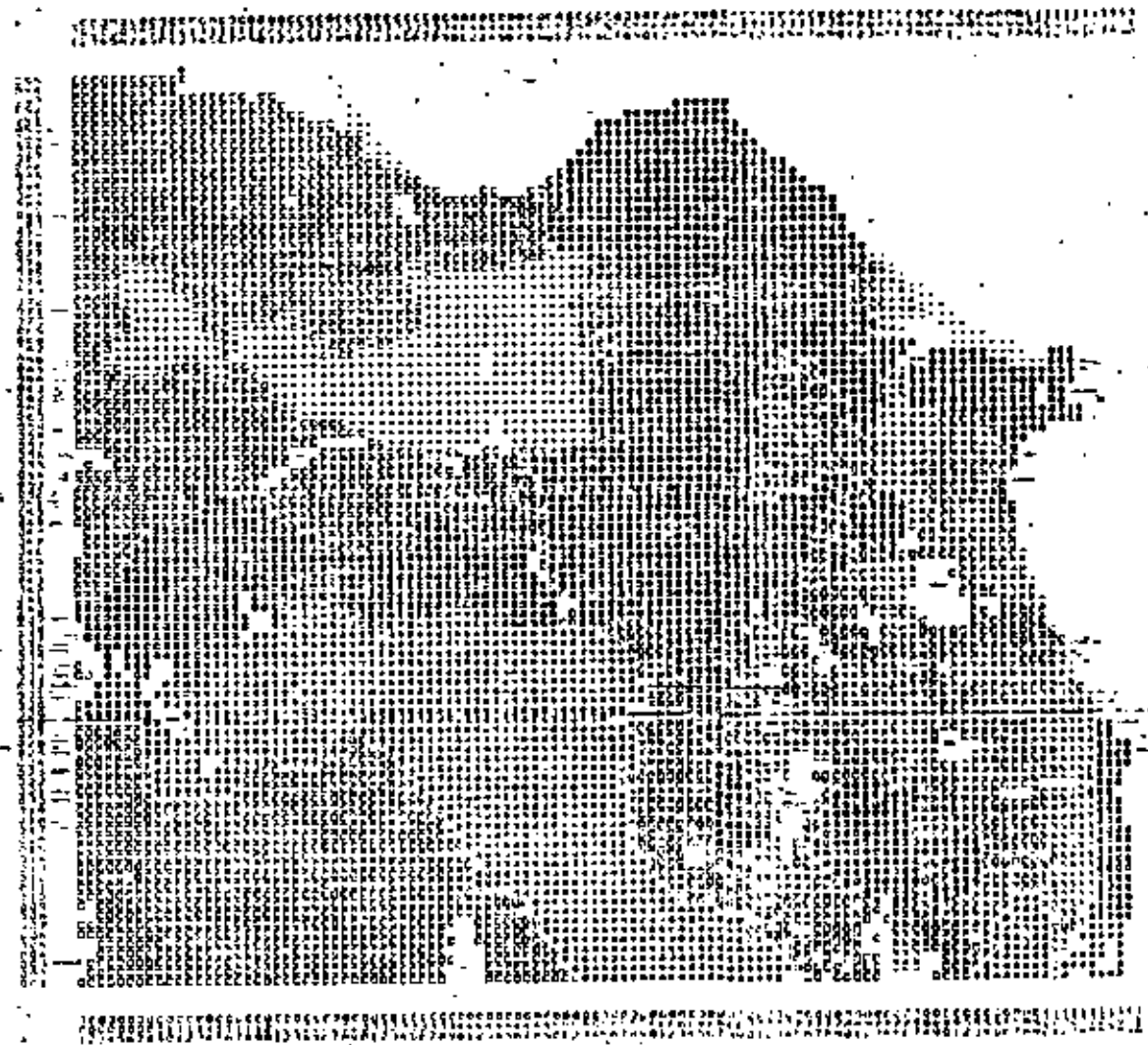
Sistema programado de planeación urbano-ambiental

Agosto, 1981

DIVISION DE ESTUDIOS DE POSGRADO  
FACULTAD DE INGENIERIA  
SUBJEFATURA DEL AREA DE INGENIERIA DE SISTEMAS  
PROGRAMOTECA



UNIVERSIDAD NACIONAL  
AUTONOMA



INGRID

SISTEMA PROGRAMADO DE PLANEACION URBANO-AMBIENTAL

2

# INGRID

Versión 1.0/DEPFI/B-6700

MANUAL DEL USUARIO.

Preparado por:

ARQ. ALEJANDRO VILLANUEVA EGAN

basado en la información y los programas proporcionados  
por el DR. CARL STEINITZ y DAVID SINTON de la Universidad  
de Harvard, así como por el DR. RICHARD TOTH y LARRY  
WEGKAMP de la Universidad Estatal de Uta.

CONTENIDO

I. INTRODUCCION.

- 1. Descripción del sistema.
- 2. Componentes.
- 3. Forma de utilización.
- 4. Aplicaciones.
- 5. Descripción del ejemplo utilizado.

II. BANCO DE DATOS.

- 1. Selección de variables y definición de sus categorías.
- 2. Determinación del tamaño de la celdilla.
- 3. Geo-codificación del contorno del área de estudio.
- 4. Geo-codificación de las variables.
- 5. Creación de archivos del banco de datos.
- 6. Mapeo del banco de datos.
- 7. Ejemplo de banco de datos.

III. MODELOS DE ATRACTIVO.

- 1. Selección de variables.
- 2. Jerarquización de las variables.
- 3. Preferencia ó deseabilidad de las categorías.
- 4. Codificación de modelos.
- 5. Creación de archivos de atractivo.
- 6. Mapeo de los resultados.
- 7. Ejemplo de modelos de atractivo.

IV. MODELOS DE VULNERABILIDAD.

- 1. Selección de variables.
- 2. Estimación de la sensibilidad de las categorías.
- 3. Estimación de la sensibilidad combinada de la 2a. y 3a. variables.
- 4. Agrupación de los usos del suelo.

- 5. Estimación de impacto negativo.
- 6. Codificación de los modelos.
- 7. Creación de archivos de vulnerabilidad.
- 8. Mapeo de los resultados.
- 9. Ejemplo de modelos de vulnerabilidad.

V. EVALUACION DE PLANES DE USO DEL SUELO.

- 1. Localización de usos del suelo.
- 2. Codificación de los planes.
- 3. Creación de los archivos de planes.
- 4. Evaluación de los Planes.
- 5. Mapeo de los resultados.
- 6. Análisis de resultados y generación de nuevos planes.
- 7. Ejemplo de evaluación de mapas.

VI. PRODUCCION DE MAPAS.

- 1. Forma de alimentar al programa.
- 2. El paquete MAP.
- 3. Opciones del paquete MAP.
- 4. Ejemplos.

VII. BIBLIOGRAFIA

Reconocimientos.



Planeación ecológica del uso del suelo.

El deterioro medio-ambiental es actualmente uno de los problemas más urgentes a resolver por la sociedad moderna. La preocupación pública a este respecto ha motivado el desarrollo de métodos avanzados para hacer más eficiente la toma de decisiones que involucran la calidad de los recursos ambientales.

Dentro de estos métodos se encuentra el sistema INGRID. Este sistema combina las facilidades de manipulación y graficación e información de las computadoras con métodos que permiten estructurar la opinión de grupos interdisciplinarios de expertos.

INGRID está orientado para desarrollar un enfoque ecológico en la planeación del uso del suelo, en el cual se diseñan y evalúan los planes de acuerdo a las componentes ambientales del terreno.

## I.1 Descripción del sistema.

INGRID es un conjunto de programas para computadora que procesa y mapea la información generada dentro de un proceso de planeación ecológica del uso del suelo. Los seis programas que lo componen constituyen técnicas avanzadas de inventario y análisis de recursos. El sistema permite que los programas interactúen entre sí, ya que los resultados de cada etapa son almacenados en archivos de disco magnético para que puedan ser utilizados por los demás módulos ó programas en el proceso. Los programas están escritos en el lenguaje de programación FORTRAN IV y en el sistema UNAM/B-6700 se tienen dos subrutinas en ALGOL para el cambio de nombre en los archivos.

En este sistema, el usuario es un factor esencial para el éxito del proceso, ya que los resultados dependen de su conocimiento y la información de que disponga sobre el proyecto en cuestión.

Las ventajas que ofrece el sistema son las siguientes:

a. Un proceso de planeación bien estructurado.

El sistema ofrece un marco de referencia para analizar en forma explícita la forma en que se toman las decisiones y se asignan criterios de evaluación.

b. Conocimiento de las soluciones propuestas y sus repercusiones ambientales.

Permite proceder los impactos probables de la urbanización sobre los eco-sistemas existentes, dándole así una capacidad mayor de previsión a los planes.

c. Flexibilidad.

Debido a la generalidad en su diseño, este sistema se -- puede aplicar a cualquier proyecto, independientemente -- de su localización y escala.

d. Mayor conocimiento de los eco-sistemas existentes en el -- area.

Al utilizar este sistema, el grupo planificador va adquiriendo conocimientos acerca de las componentes ambientales relevantes, su funcionamiento, su interacción, los -- factores que las controlan, como afectan y sus afectadas por el uso del suelo, etc.

e. Aumento de la creatividad.

El procedimiento desarrollado por IMGRID permite aumentar la habilidad creativa del planificador mediante un -- proceso de búsqueda sistemático que permite evaluar y -- comparar una gran cantidad de conceptos alternativos de -- diseño y desarrollo espacial de los planes.

f. Facilidad de utilización.

El sistema está diseñado para ser utilizado por personas sin conocimientos en programación, dado que se controla mediante instrucciones sencillas que realizan operaciones específicas. Una vez que el usuario entiende la naturaleza de las instrucciones, está en posibilidad de manejar el sistema sin mayor conocimiento del mismo.

g. Adaptabilidad con otros sistemas.

IMGRID puede formar parte como un sub-sistema dentro de un sistema mayor de recopilación, almacenamiento, análisis

7  
sis y graficación de información geográfica.

h. Ahorro en tiempo y costo mediante la computadora.

La facilidad de manipulación y mapeo de grandes cantidades de información por las computadoras le dá a IMGRID una ventaja de eficiencia económica sobre los métodos tradicionales.

## I.2 COMPONENTES DEL SISTEMA

Los programas para computadora que contiene el sistema IMGRID son los siguientes:

### a. BANCO DE DATOS. Programa ECODATOS/IMGRID.

Permite almacenar en forma digital la información relevante de los recursos del suelo registrada a partir de los mapas fuente. El programa crea un archivo en disco magnético para cada variable, pudiendo contener hasta 50 variables diferentes subdividida cada una de ellas en un número máximo de 10 categorías.

Estos archivos son accedidos por los programas correspondientes a la producción de mapas y modelos analíticos.

### b. MODELOS DE ATRACTIVO Programa ATRACTIVO/IMGRID

Mediante este programa se pueden especificar modelos para encontrar la localización más adecuada de los usos del suelo propuestos. Es posible considerar hasta 20 usos del suelo diferentes, cada uno con su modelo correspondiente. El método utilizado por los modelos consiste en obtener para cada una de las celdillas en el área de estudio, un índice de "atractivo" que refleja su adecuación para localizar un uso del suelo específico.

El usuario determina las variables del banco de datos que intervienen en el modelo, las jerarquiza y asigna preferencias a sus categorías, expresando así, de una manera formal, los factores de localización de cada uso del suelo que sean importantes a su criterio.

El programa accesa el banco de datos para registrar las categorías que toman las variables en cada celdilla y calcula el índice como un promedio ponderado de las asignaciones del usuario.

El atractivo de cada celdilla puede tomar un valor entre 0 y 9, considerándose el 0 como "nada atractivo" y el 9 como "lo más atractivo".

Al efectuarse una corrida, se crea un archivo en disco magnético para los resultados de cada modelo. Estos archivos son utilizados por los programas correspondientes a la producción de mapas y evaluación de planes.

### c. MODELOS DE VULNERABILIDAD. Programa IMPACTO/IMGRID

Mediante este programa se pueden especificar modelos para estimar el daño probable ocasionado por los usos del suelo propuestos sobre los eco-sistemas existentes en el área de estudio.

Es posible considerar hasta 30 eco-sistemas diferentes, cada uno con su modelo correspondiente. El método utilizado por los modelos consiste en obtener para cada celdilla en el área de estudio un índice de vulnerabilidad ó grado de impacto negativo producido por un uso del suelo específico. El usuario determina las variables del banco de datos que intervienen en el modelo y pondera sus categorías en forma individual y combinada para expresar su sensibilidad al impacto considerado.

El programa accesa el banco de datos para registrar las categorías que toman las variables en cada celdilla y determina el grado de impacto negativo producido de acuerdo a las asignaciones hechas por el usuario. El impacto puede tomar un valor entre 1 y 4, de acuerdo a la siguiente escala:

1 = Compatible.

2 = Moderado.

3 = Severo.

4 = Terminal.

Al efectuarse una corrida, se crea un archivo en disco magnético para los resultados de cada modelo.

Estos archivos son utilizados por los programas correspondientes a la producción de mapas y evaluación de planes.

#### d. EVALUACION DE PLANES

Programa PLANES/IMGRID

Programa EVALUACION/IMGRID

Mediante estos dos programas es posible almacenar en disco magnético los diversos planes de uso del suelo generados por el usuario y evaluarlos con respecto a los criterios de atractivo y vulnerabilidad formulados en los modelos. El resultado de la evaluación consiste en dos tablas sumario.

La primera, expresa el atractivo ó adecuación lograda en la localización de los usos del suelo. La segunda, indica el grado de impacto negativo causado por el plan sobre cada uno de los sistemas descritos por el usuario.

Estos resultados son almacenados en archivos de disco magnético para ser accedados por el programa de producción de mapas.

En una corrida pueden evaluarse hasta 30 diferentes planes de uso del suelo.

5. PRODUCCION DE MAPAS.  
Programa MAPAS/IMGRID

Mediante este programa se utiliza la impresora de líneas de la computadora para producir mapas de la información contenida en el banco de datos y la generada por los modelos y la evaluación de planes.

El usuario puede controlar aspectos del mapa tales como tamaño, simbolismo, numeración de referencia, textos explicativos y obtención de histogramas (Ver figura ).

### 1.3 FORMA DE UTILIZACION

El conjunto de programas para computadora que contiene el sistema, es utilizado para procesar la información generada por el grupo planificador en las diferentes etapas del proceso de planeación.

En cada una de estas etapas, se requiere realizar un análisis previo que permita hacer estimaciones y tomar decisiones acerca de la información y los criterios con que se va a alimentar a la computadora. Es decir, en el proceso de planeación establecido por IMGRID, una parte se realiza "fuera de la computadora" y la otra "dentro de la computadora".

De manera general, el proceso sigue el diagrama ilustrado en la figura , pero debe entenderse que no es un procedimiento lineal sino que al efectuar cualquiera de las etapas debetenerse en mente el proceso en su totalidad.

Un diagrama detallado del proceso, indicando las decisiones del grupo planificador; los programas utilizados y los resultados obtenidos se muestra en la figura .

## 1. PRE-ANALISIS.

- Tipo de proyecto
- Definición de sitio, programa y contexto

### BANCO DE DATOS

## 2. CRITERIOS DE LOCALIZACION

- Análisis de Usos del suelo

MODELOS DE ATRACTIVO

MAPEO

MODELOS DE VULNERABILIDAD

## 3. CRITERIOS DE IMPACTO AMBIENTAL.

- Análisis de componentes medio-ambientales.

### EVALUACION DE PLANES.

## 4. GENERACION DE PLANES

- Desarrollo de conceptos espaciales alternativos
- Análisis de resultados.

Figura . Diagrama del proceso de utilización de IMGRID.



## II. BANCO DE DATOS.

El banco de datos consiste en un conjunto de archivos en disco magnético, que pueden ser accedidos por los programas correspondientes a los modelos y producción de mapas. Se cuenta con un archivo para cada variable y pueden considerarse hasta 50 diferentes variables en el banco.

Las variables están subdivididas en categorías identificadas por un dígito entre 0 y 9. La información es registrada en forma digital mediante una retícula regular que se superpone a los mapas fuente. El registro es celdilla por celdilla e hilera por hilera. El escoger el tamaño adecuado de cada celdilla ó sea, el área de la unidad básica de análisis, es una de las decisiones más importantes en esta etapa del proceso.

Para la elaboración del banco de datos se deben llevar a cabo los siguientes pasos:

### 17 ENLISTAR LAS VARIABLES Y DEFINIR SUS CATEGORIAS:

Basándose en los análisis de la etapa anterior y la disponibilidad de información y recursos, se deben seleccionar las variables que estarán contenidas en el banco de datos, así como las categorías en que se subdivide cada una de ellas. Cada variable debe tener un número en orden creciente a partir del 1 y las categorías deben ser identificadas por un dígito entre 0 y 9. Por ejemplo:

Número de la variable: 1

Nombre de la variable: USO DEL SUELO EXISTENTE.

Categorías:

- 0 = No hay datos.
- 1 = Campamentos.
- 2 = Uso diurno.
- 3 = Servicio forestal.
- 4 = Escuela.
- 5 = Estacionamiento.
- 6 = Pasto.
- 7 = Caminos.
- 8 = Reserva forestal.
- 9 = Océano.

Número de la variable: 2

Nombre de la variable: PORCENTAJE DE PENDIENTE.

Categorías:

- 0 = No hay datos.
- 1 = 100% agua.
- 3 = 0 a 9%.
- 5 = 10 a 15%.
- 7 = 16 al 25%.
- 8 = Más del 25%.
- 9 = Océano.

2. DETERMINAR EL TAMAÑO DE LA CELDILLA.

La información contenida en los mapas fuente es registrada en forma digital (geo-codificación), mediante una retícula regular sobrepuesta. La forma de las celdillas es cuadrada y su área debe ser la misma en toda el área de estudio. Su tamaño depende de los siguientes factores:

- Exactitud y tipo de los datos disponibles.
- Tamaño del menor rasgo que se desee registrar.
- Propósito para el cual van a usarse los datos.
- Tamaño del área de estudio.
- Limitaciones en los recursos para geo-codificar la información.

Las figuras muestran ejemplos de retículas sobrepuestas para el registro de información geográfica.

3. GEO-CODIFICAR EL CONTORNO DEL AREA DE ESTUDIO.

Contorno regular:

1a. tarjeta:

Col. 1-5	Número de celdillas de la retícula en sentido vertical. Perforado como un número entero, justificado a la derecha.
Col. 10	0
Col. 15	0

2a. tarjeta:

Col. 1-5            99999 para indicar que termina el contorno.

Contorno irregular:

1a. tarjeta:

Col. 15            Número de hileras de la retícula en el sentido vertical que tienen el mismo desplazamiento hacia la derecha y la izquierda. Perforado como un número entero y justificado a la derecha.

Col. 6-10            Número de celdillas que se desplazan a la derecha del margen del contorno regular. Perforado como número entero y justificado a la derecha.

Col. 10-15            Número de celdillas que se desplazan a la izquierda del contorno regular. Perforado como número entero y justificado a la derecha.

Debe repetirse este procedimiento en tantas tarjetas como sea necesario para especificar todas las fronteras del área de estudio.

Última tarjeta:

Col. 1-5            99999 para indicar que termina el contorno.

La figura ilustra la forma de geo-codificar el contorno del área de estudio.

4. GEO-CODIFICAR LAS VARIABLES.

La geo-codificación se hace superponiendo la retícula al área de estudio y registrando el valor ó categoría que toma la variable en cada una de las celdillas, anotando el número correspondiente de acuerdo a la subdivisión especificada.

Los datos pueden ser registrados como:

- a) Datos de punto (vgr. una cascada, un pozo, etc.).
- b) Porcentaje de la celdilla con una actividad determinada.
- c) Tipo predominante de uso del suelo.
- d) Datos de línea (vgr. una carretera, un río, etc.).

Cada columna en una tarjeta representa una celdilla de la retícula. Cada tarjeta representa una hilera de celdillas. Si

ción 10 de Texto explicativo en cada mapa, deberá contener una clave con el número y nombre de la variable y las categorías - en que se subdivide.

Una vez revisados, analizados y corregidos los mapas, es posible pasar a la siguiente etapa del proceso de IMGRID.

se tienen más de 80 celdillas en una hilera se deberán usar - dos o más tarjetas.

1a. tarjeta:

Col. 1-2                    Número de la variable, perforado como número entero justificado a la derecha.

Tarjetas siguientes:

Col. 1-80                    Valor o categoría que toma la variable en cada una de las celdillas. Deben perforarse como números enteros de una cifra (0-9) en las columnas correspondientes. Se debe utilizar solamente el número de columnas necesarias. Para registrar todas las celdillas de una hilera de la retícula. Se deben perforar tantas tarjetas como sea necesario para registrar todas las hileras de la retícula.

El programa limita el número de variables a 50 y el de categorías para cada variables a 10, numerándolas del 0 al 9.

La figura                    muestra un ejemplo de geo-codificación de una variable.

5. CREACION DE ARCHIVOS DEL BANCO DE DATOS.

El almacenamiento de las variables en archivos en disco magnético para su posterior utilización en la producción de mapas y modelos, se realiza mediante la alimentación de la información geo-codificada en el paso anterior al programa ARCHIVOS/IMGRID.

En el sistema UNAM/B-6700 esta información debe ir precedida por tres tarjetas de control y por el contorno del área de estudio geo-codificado en el paso 3. Después de las tarjetas de la última variable debe colocarse una tarjeta de control final para la terminación del trabajo.

La forma de preparar el paquete de datos se ilustra en la figura

6. MAPEAR EL BANCO DE DATOS.

Debe producirse un mapa para cada una de las variables incluidas en el banco de datos utilizando el programa MAPAS/IMGRID, cuya descripción es proporcionada en el capítulo VI. La op--

. | 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

52  
51  
50  
49  
48  
47  
46  
45  
44  
43  
42  
41  
40  
39  
38  
37  
36  
35  
34  
33  
32  
31  
30  
29  
28  
27  
26  
25  
24  
23  
22  
21  
20  
19  
18  
17  
16  
15  
14  
13  
12  
11  
10  
9  
8  
7  
6  
5  
4  
3  
2  
1

# VALLE DE PACIFICO

19

CAYO GAVIOTAS

BOCA ANDREA

BAHIA BLANCA

PUNTA CHINUC

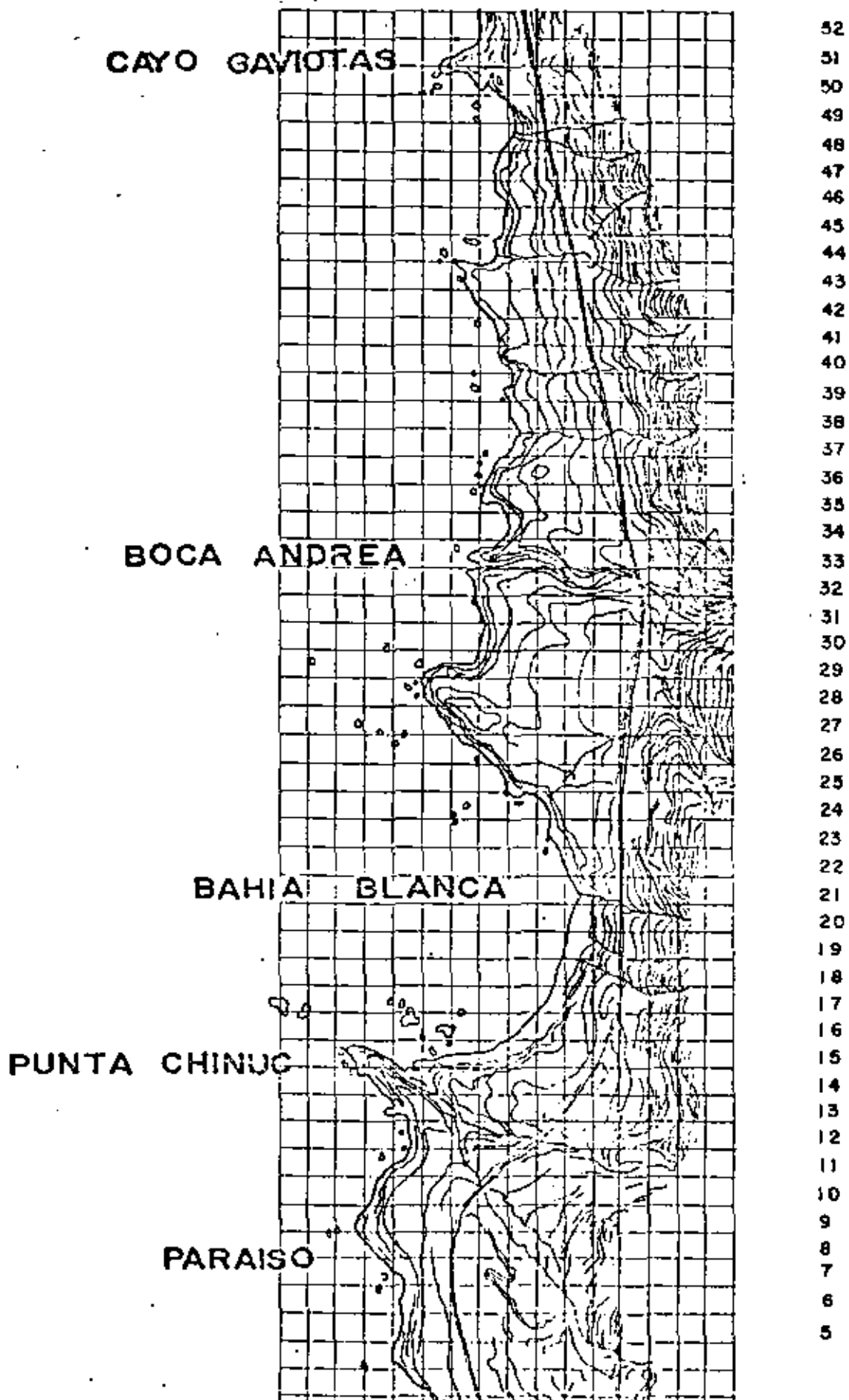
PARAISO



# VALLE DE PACIFICO

20

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16



CAYO GAVIOTAS

BOCA ANDREA

BAHIA BLANCA

PUNTA CHINUC

PARAISO

52  
51  
50  
49  
48  
47  
46  
45  
44  
43  
42  
41  
40  
39  
38  
37  
36  
35  
34  
33  
32  
31  
30  
29  
28  
27  
26  
25  
24  
23  
22  
21  
20  
19  
18  
17  
16  
15  
14  
13  
12  
11  
10  
9  
8  
7  
6  
5



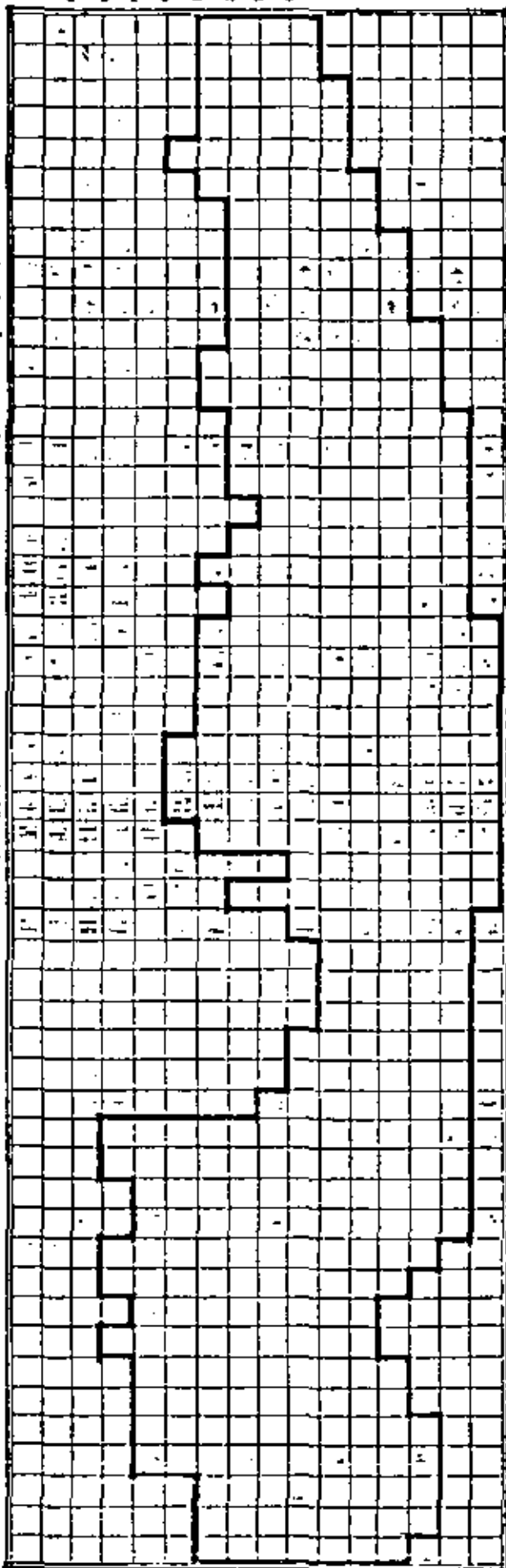
# CONTORNO IRREGULAR

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

C O L U M N A S

21

NUMERO DE MEDIDAS	NO. DE CELULAS QUE SE DESPLAZAN A LA DERECHA DEL CONTORNO NEGRO	NO. DE CELULAS QUE SE DESPLAZAN A LA IZQUIERDA DEL CONTORNO NEGRO
2	6	6
2	8	5
1	5	8
1	5	4
1	2	4
3	7	3
1	7	2
2	6	2
3	7	1
1	8	1
1	7	1
1	6	1
1	7	1
4	6	0
3	5	0
1	6	0
1	9	0
1	7	0
1	9	1
3	10	1
2	9	1
1	8	1
2	3	1
2	4	1
1	3	2
1	3	3
1	4	4
1	3	4
2	4	3
2	4	2
2	6	2
1	6	3



62  
61  
60  
59  
49  
48  
47  
46  
45  
44  
43  
42  
41  
40  
39  
38  
37  
36  
35  
34  
33  
32  
31  
30  
29  
28  
27  
26  
25  
24  
23  
22  
21  
20  
19  
18  
17  
16  
15  
14  
13  
12  
11  
10  
9  
8  
7  
6  
5  
4  
3  
2  
1

# CODIFICACION

## VARIABLE I

### USO DE SUELO EXISTENTE

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

- 0 = No hay datos
- 1 = Campamentos
- 2 = Uso diurno
- 3 = Administración del servicio forestal.
- 4 = Escuela
- 6 = Pasa
- 7 = Carreteras
- 8 = Reserva forestal
- 9 = Oceano

9	9	9	9	9	9	8	6	7	8	0	0	0	0	0	0	52
9	9	9	9	9	9	8	6	7	8	0	0	0	0	0	0	51
9	9	9	9	9	9	8	6	7	8	8	0	0	0	0	0	50
9	9	9	9	9	9	8	6	7	8	8	0	0	0	0	0	49
9	9	9	9	9	8	6	6	7	8	8	0	0	0	0	0	48
9	9	9	9	9	8	8	7	8	8	8	0	0	0	0	0	47
9	9	9	9	9	9	8	6	7	8	8	0	0	0	0	0	46
9	9	9	9	9	9	8	6	7	8	8	8	0	0	0	0	45
9	9	9	9	9	9	8	6	7	8	8	8	0	0	0	0	44
9	9	9	9	9	9	8	6	7	5	8	8	0	0	0	0	43
9	9	9	9	9	9	8	6	7	5	8	8	8	0	0	0	42
9	9	9	9	9	9	8	6	6	6	7	8	8	8	0	0	41
9	9	9	9	9	9	8	6	6	6	7	8	8	8	0	0	40
9	9	9	9	9	9	8	6	6	7	8	8	8	8	0	0	39
9	9	9	9	9	9	8	6	6	7	8	8	8	8	0	0	38
9	9	9	9	9	9	8	6	6	6	7	8	8	8	0	0	37
9	9	9	9	9	9	8	6	6	6	7	8	8	8	0	0	36
9	9	9	9	9	9	8	6	6	6	7	8	8	8	0	0	35
9	9	9	9	9	9	8	6	6	6	6	7	8	8	8	0	34
9	9	9	9	9	9	8	6	6	6	6	7	8	8	8	0	33
9	9	9	9	9	9	8	8	6	6	6	6	7	8	8	8	32
9	9	9	9	9	9	8	6	6	6	6	6	7	8	8	8	31
9	9	9	9	9	9	8	6	6	6	6	6	7	8	8	8	30
9	9	9	9	9	9	8	6	6	6	6	6	7	8	8	8	29
9	9	9	9	9	8	8	6	6	6	6	6	3	3	8	8	28
9	9	9	9	9	8	6	6	6	6	6	6	3	3	6	8	27
9	9	9	9	9	2	2	6	6	6	6	6	3	3	8	8	26
9	9	9	9	9	9	2	2	6	6	6	7	8	7	8	8	25
9	9	9	9	9	9	9	9	9	2	6	7	7	8	8	8	24
9	9	9	9	9	9	2	2	6	6	7	7	8	7	7	7	23
9	9	9	9	9	9	9	9	9	2	2	7	7	8	8	0	22
9	9	9	9	9	9	9	9	9	9	2	7	7	8	8	0	21
9	9	9	9	9	9	9	9	9	9	2	7	7	8	8	0	20
9	9	9	9	9	9	9	9	9	9	2	7	8	8	8	0	19
9	9	9	9	9	9	9	9	9	9	2	2	7	8	8	8	18
9	9	9	9	9	9	9	9	9	9	2	2	7	8	8	8	17
9	9	9	9	9	9	9	9	9	9	2	2	7	8	8	8	16
9	9	9	9	9	9	9	9	9	9	2	2	6	7	8	8	15
9	9	9	2	2	2	2	2	2	6	6	7	8	8	8	0	14
9	9	9	8	8	8	6	6	6	2	7	8	8	8	8	0	13
9	9	9	9	8	6	6	6	7	7	8	8	8	8	8	0	12
9	9	9	8	8	7	7	7	7	1	8	8	8	8	0	0	11
9	9	9	8	7	6	7	7	1	1	8	8	8	0	0	0	10
9	9	9	9	8	6	7	7	7	8	8	8	0	0	0	0	9
9	9	9	8	6	6	7	7	8	8	8	8	0	0	0	0	8
9	9	9	9	8	6	7	7	7	8	8	8	8	0	0	0	7
9	9	9	9	8	6	7	8	7	7	7	8	8	0	0	0	6
9	9	9	9	8	6	7	8	8	8	7	7	7	7	0	0	5
9	9	9	9	2	2	2	7	8	8	8	8	8	8	0	0	4
9	9	9	9	9	9	2	7	8	8	6	8	8	8	0	0	3
9	9	9	9	9	9	8	7	8	8	8	8	8	8	0	0	2
9	9	9	9	9	9	8	8	7	8	8	8	8	0	0	0	1

MAPA DE LA VARIABLE 1 DEL INVENTARIO DE DATOS USOS DEL SUELO ACTUALES

AREA DE ESTUDIO VALLE PACIFICO LABORATORIO DE PLANEACION URBANA

00000000000000000000  
000000000011111111  
1234567890123456

SUBVARIABLES. DATO

- 0 = NO HAY DATOS
- 1 = CAMPAMENTOS
- 2 = USO DIRECTO
- 3 = ADMINISTRACION DE SERVICIO FORESTAL
- 4 = ESCUELA
- 6 = PASTO
- 7 = CACHETERAS
- 8 = RESERVA FORESTAL
- 9 = OCEANO

EL TAMAÑO DE LA CELULA DE LA MALLA ES DE 200 METROS CUADRADOS

NIVELES

SIEMBOLOS

FRECUENCIA

3	4	5	6	7	8	9
XXXXXX	XXXXXX	XXXXXX	XXXXXX	XXXXXX	XXXXXX	XXXXXX
UUUUUUUU	UUUUUUUU	UUUUUUUU	UUUUUUUU	UUUUUUUU	UUUUUUUU	UUUUUUUU
IIIIIIII	IIIIIIII	IIIIIIII	IIIIIIII	IIIIIIII	IIIIIIII	IIIIIIII

6	1	2	3	4	5	6	7	8	9
UUUUUUUU	UUUUUUUU	UUUUUUUU	UUUUUUUU	UUUUUUUU	UUUUUUUU	UUUUUUUU	UUUUUUUU	UUUUUUUU	UUUUUUUU
IIIIIIII	IIIIIIII	IIIIIIII	IIIIIIII	IIIIIIII	IIIIIIII	IIIIIIII	IIIIIIII	IIIIIIII	IIIIIIII
XXXXXX	XXXXXX	XXXXXX	XXXXXX	XXXXXX	XXXXXX	XXXXXX	XXXXXX	XXXXXX	XXXXXX

76 205 517



00000000000000000000  
000000000011111111  
1234567890123456

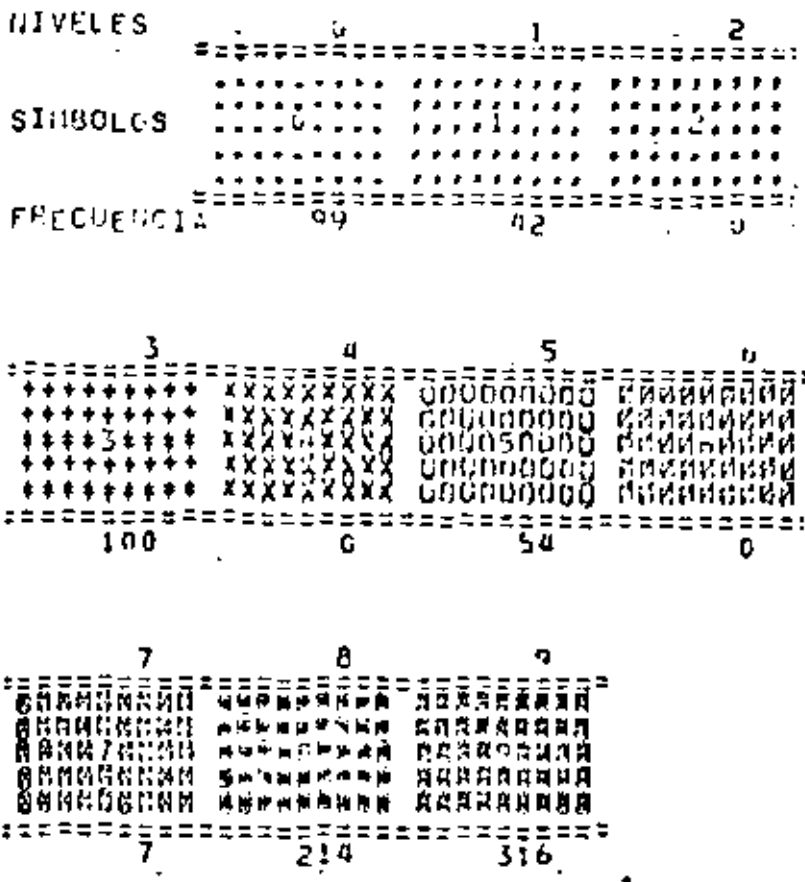
MAPA DE LA VARIABLE 2 DEL INVENTARIO DE DATOS PORCIENTO DE PENDIENTE  
AREA DE ESTUDIO VALLE DEL PACIFICO LABORATORIO DE PLANEACION URBANA

0000000000000000  
0000000001111111  
1234567890123456

SUBVARIABLES DATO

- 0 = NO HAY DATOS
- 1 = 0-9%
- 2 = 10-15%
- 3 = 16-25%
- 4 = 25%+
- 5 = OCEANO

EL TAMAÑO DE LA CELULA DE LA MALLA ES DE 2,5 ACRES



052  
051  
050  
049  
048  
047  
046  
045  
044  
043  
042  
041  
040  
039  
038  
037  
036  
035  
034  
033  
032  
031  
030  
029  
028  
027  
026  
025  
024  
023  
022  
021  
020  
019  
018  
017  
016  
015  
014  
013  
012  
011  
010  
009  
008  
007  
006  
005  
004  
003  
002  
001



0000000000000000  
0000000001111111  
1234567890123456



MAPA DE LA VARIABLE 4 DEL INVENTARIO DE DATOS VEGETACION POR ZONA

AREA DE ESTUDIO VALLE DEL PACIFICO LABORATORIO DE PLANEACION URRUTIA

0000000000000000  
0000000011111111  
1234567890123456

SUBVARIABLES DATO

- 0 = NO HAY DATOS
- 1 = ZONA COSTERA
- 2 = SUELO PASTADO
- 3 = VEGETACION CHAPARRAL
- 4 = ZONA DE BOSQUE
- 9 = OCEANO

EL TAMAÑO DE LA CELULA DE LA MALLA ES DE 2.5 ACRES

NIVELES	0	1	2
SÍMBOLOS	0	1	2
FRECUENCIA	101	0	0

	3	4	5	6
FRECUENCIA	200	106	39	0

	7	8	9
FRECUENCIA	0	0	517

052  
051  
050  
049  
048  
047  
046  
045  
044  
043  
042  
041  
040  
039  
038  
037  
036  
035  
034  
033  
032  
031  
030  
029  
028  
027  
026  
025  
024  
023  
022  
021  
020  
019  
018  
017  
016  
015  
014  
013  
012  
011  
010  
009  
008  
007  
006  
005  
004  
003  
002  
001



0000000000000000  
0000000011111111  
1234567890123456

MAPA DE LA VARIABLE 5 DEL INVENTARIO DE DATOS DE SIDA EN ARGENTINA

AREA DE ESTUDIO VALLE DEL PACIFICO LABORATORIO DE PLANEACION URBANA

0000000000000000  
00000000111111  
1234567890123456

SUBVARIABLES

- 0 = NO HAY DATOS
- 2 = MFA
- 3 = 1-25%
- 4 = 26-50%
- 5 = 51-75%
- 7 = 76-100%
- 9 = BOCALADO

EL TAMAÑO DE LA CÉLULA DE LA MALLA ES DE 25 METROS

NIVELES: 0 1 2

SÍMBOLOS

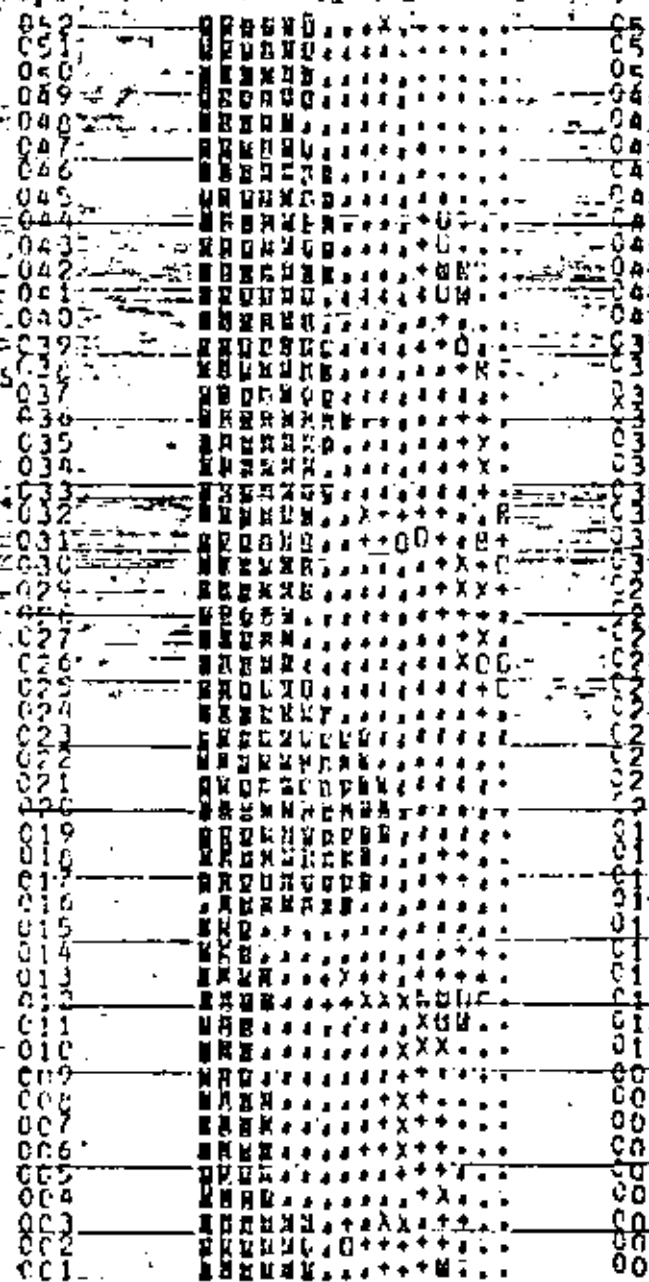
FRECUENCIA 98 0 300

	3	4	5	6
+++++	XXXXXXXXXX	OOOOOOOO	KKKKKKKK	MMMMMMMM
+++++	XXXXXXXXXX	OOOOOOOO	KKKKKKKK	MMMMMMMM
+++++	XXXXXXXXXX	OOOOOOOO	KKKKKKKK	MMMMMMMM
+++++	XXXXXXXXXX	OOOOOOOO	KKKKKKKK	MMMMMMMM
+++++	XXXXXXXXXX	OOOOOOOO	KKKKKKKK	MMMMMMMM

70 23 11 0

	7	8	9
MMMMMMMM	MMMMMMMM	KKKKKKKK	MMMMMMMM
MMMMMMMM	MMMMMMMM	KKKKKKKK	MMMMMMMM
MMMMMMMM	MMMMMMMM	KKKKKKKK	MMMMMMMM
MMMMMMMM	MMMMMMMM	KKKKKKKK	MMMMMMMM
MMMMMMMM	MMMMMMMM	KKKKKKKK	MMMMMMMM

13 0 317



0000000000000000  
00000000111111  
1234567890123456







MAPA DE LA VARIABLE B DEL INVENTARIO DE DATOS SUCCIOS

AREA DE ESTUDIO VALLE DEL PACIFICO

LABORATORIO DE PLANEACION URBANA

0000000000000000  
00000000111111  
1234567890123456

SU VARIABLES DATO

- 0 = NO HAY DATOS
- 1 = NFG2
- 2 = ST
- 3 = TBC
- 4 = CNE
- 5 = RIGH
- 6 = CZT
- 7 = GPCH
- 8 = CZE
- 9 = OCLAL

EL TAMAÑO DE LA CELULA DE LA MALLA ES DE

2.5 ACRES

NIVELES	0	1	2
SIMBOLOS	.....	.....	.....
FRECUENCIA	90	29	17

	3	4	5	6
SIMBOLOS	.....	.....	.....	.....
FRECUENCIA	96	19	169	10

	7	8	9
SIMBOLOS	.....	.....	.....
FRECUENCIA	32	14	317

042	.....	001
051	.....	005
050	.....	005
049	.....	002
048	.....	002
047	.....	004
046	.....	004
045	.....	004
044	.....	006
043	.....	004
042	.....	004
041	.....	004
040	.....	004
039	.....	000
038	.....	000
037	.....	000
036	.....	000
035	.....	000
034	.....	000
033	.....	000
032	.....	000
031	.....	000
030	.....	000
029	.....	000
028	.....	000
027	.....	000
026	.....	000
025	.....	000
024	.....	000
023	.....	000
022	.....	000
021	.....	000
020	.....	000
019	.....	000
018	.....	000
017	.....	000
016	.....	000
015	.....	000
014	.....	000
013	.....	000
012	.....	000
011	.....	000
010	.....	000
009	.....	000
008	.....	000
007	.....	000
006	.....	000
005	.....	000
004	.....	000
003	.....	000
002	.....	000
001	.....	000

0000000000000000  
00000000111111  
1234567890123456

MAPA DE LA VARIABLE 9 DEL INVENTARIO DE DATOS PROXIMIDAD AL AGUA

AREA DE ESTUDIO VALLE DEL PACIFICO

LABORATORIO DE PLANIFICACION URBANA

0000000000000000  
0000000001111111  
1234567890123456

SUBVARIABLES (ATG)

- 0 = NO HAY DATOS
- 2 = COMPONENTES DENTRO DE LA CELULA
- 3 = AGUA EN LAS CELULAS ADYACENTES
- 4 = AGUA EN UNA CELULA DE PROXIMIDAD
- 5 = AGUA EN DOS CELULAS DE PROXIMIDAD
- 6 = ZONA DE AREA
- 9 = OCEANO

EL TAMAÑO DE LA CELULA DE LA PALLA ES DE

225 ACRES

NIVELES

0 1 2

SÍMBOLOS

0 1 2 3

FRECUENCIA

90 0 3

3	4	5	6
+++++ - Y Y Y Y Y X X X	00000000	IIIIIIII	RIIR
+++++ - Y Y Y Y Y X X X	00000000	IIIIIIII	RIIR
+++++ - Y Y Y Y Y X X X	00000000	RIIR	RIIR
+++++ - Y Y Y Y Y X X X	00000000	IIIIIIII	RIIR
+++++ - X X X X X X X X	00000000	IIIIIIII	RIIR
117	87	70	81

7 8 9

IIIIIIII	IIIIIIII	XXXXXX
IIIIIIII	IIIIIIII	XXXXXX
IIIIIIII	IIIIIIII	XXXXXX
IIIIIIII	IIIIIIII	XXXXXX
IIIIIIII	IIIIIIII	XXXXXX
0	0	317

052	IIIIIIII	IIIIIIII	XXXXXX	052
051	IIIIIIII	IIIIIIII	XXXXXX	051
050	IIIIIIII	IIIIIIII	XXXXXX	050
049	IIIIIIII	IIIIIIII	XXXXXX	049
048	IIIIIIII	IIIIIIII	XXXXXX	048
047	IIIIIIII	IIIIIIII	XXXXXX	047
046	IIIIIIII	IIIIIIII	XXXXXX	046
045	IIIIIIII	IIIIIIII	XXXXXX	045
044	IIIIIIII	IIIIIIII	XXXXXX	044
043	IIIIIIII	IIIIIIII	XXXXXX	043
042	IIIIIIII	IIIIIIII	XXXXXX	042
041	IIIIIIII	IIIIIIII	XXXXXX	041
040	IIIIIIII	IIIIIIII	XXXXXX	040
039	IIIIIIII	IIIIIIII	XXXXXX	039
038	IIIIIIII	IIIIIIII	XXXXXX	038
037	IIIIIIII	IIIIIIII	XXXXXX	037
036	IIIIIIII	IIIIIIII	XXXXXX	036
035	IIIIIIII	IIIIIIII	XXXXXX	035
034	IIIIIIII	IIIIIIII	XXXXXX	034
033	IIIIIIII	IIIIIIII	XXXXXX	033
032	IIIIIIII	IIIIIIII	XXXXXX	032
031	IIIIIIII	IIIIIIII	XXXXXX	031
030	IIIIIIII	IIIIIIII	XXXXXX	030
029	IIIIIIII	IIIIIIII	XXXXXX	029
028	IIIIIIII	IIIIIIII	XXXXXX	028
027	IIIIIIII	IIIIIIII	XXXXXX	027
026	IIIIIIII	IIIIIIII	XXXXXX	026
025	IIIIIIII	IIIIIIII	XXXXXX	025
024	IIIIIIII	IIIIIIII	XXXXXX	024
023	IIIIIIII	IIIIIIII	XXXXXX	023
022	IIIIIIII	IIIIIIII	XXXXXX	022
021	IIIIIIII	IIIIIIII	XXXXXX	021
020	IIIIIIII	IIIIIIII	XXXXXX	020
019	IIIIIIII	IIIIIIII	XXXXXX	019
018	IIIIIIII	IIIIIIII	XXXXXX	018
017	IIIIIIII	IIIIIIII	XXXXXX	017
016	IIIIIIII	IIIIIIII	XXXXXX	016
015	IIIIIIII	IIIIIIII	XXXXXX	015
014	IIIIIIII	IIIIIIII	XXXXXX	014
013	IIIIIIII	IIIIIIII	XXXXXX	013
012	IIIIIIII	IIIIIIII	XXXXXX	012
011	IIIIIIII	IIIIIIII	XXXXXX	011
010	IIIIIIII	IIIIIIII	XXXXXX	010
009	IIIIIIII	IIIIIIII	XXXXXX	009
008	IIIIIIII	IIIIIIII	XXXXXX	008
007	IIIIIIII	IIIIIIII	XXXXXX	007
006	IIIIIIII	IIIIIIII	XXXXXX	006
005	IIIIIIII	IIIIIIII	XXXXXX	005
004	IIIIIIII	IIIIIIII	XXXXXX	004
003	IIIIIIII	IIIIIIII	XXXXXX	003
002	IIIIIIII	IIIIIIII	XXXXXX	002
001	IIIIIIII	IIIIIIII	XXXXXX	001

0000000000000000  
0000000001111111  
1234567890123456

MAPA DE LA VARIACION DEL INVENTARIO DE DATOS VISTAS

AREA DE ESTUDIO VALLE DEL PACIFICO

LABORATORIO DE PLANEACION URBANA

0000000000000000  
0000000001111111  
1234567890123456

SUBVARIABLES DATO

- 0 = NO HAY DATOS
- 2 = NO HAY VISTAS IMPORTANTES
- 3 = BUENAS VISTAS AL OCEANO
- 4 = EXCELENTES VISTAS AL OCEANO
- 5 = BUENAS PANORAMICAS DEL LUGAR
- 6 = EXCELENTES PANORAMICAS DEL LUGAR
- 9 = OCEANO

EL TAPADO DE LA CELULA DE LA MALLA P.S. DE 2.5 ACRES

NIVELES

0	1
2	3
4	5
6	9

SIMBOLOS

FRECUENCIA

2	3	4	5
133	87	157	14

6	7	8	9
26	0	0	317

052	0000000000000000	051
051	0000000000000000	050
050	0000000000000000	049
049	0000000000000000	048
048	0000000000000000	047
047	0000000000000000	046
046	0000000000000000	045
045	0000000000000000	044
044	0000000000000000	043
043	0000000000000000	042
042	0000000000000000	041
041	0000000000000000	040
040	0000000000000000	039
039	0000000000000000	038
038	0000000000000000	037
037	0000000000000000	036
036	0000000000000000	035
035	0000000000000000	034
034	0000000000000000	033
033	0000000000000000	032
032	0000000000000000	031
031	0000000000000000	030
030	0000000000000000	029
029	0000000000000000	028
028	0000000000000000	027
027	0000000000000000	026
026	0000000000000000	025
025	0000000000000000	024
024	0000000000000000	023
023	0000000000000000	022
022	0000000000000000	021
021	0000000000000000	020
020	0000000000000000	019
019	0000000000000000	018
018	0000000000000000	017
017	0000000000000000	016
016	0000000000000000	015
015	0000000000000000	014
014	0000000000000000	013
013	0000000000000000	012
012	0000000000000000	011
011	0000000000000000	010
010	0000000000000000	009
009	0000000000000000	008
008	0000000000000000	007
007	0000000000000000	006
006	0000000000000000	005
005	0000000000000000	004
004	0000000000000000	003
003	0000000000000000	002
002	0000000000000000	001

0000000000000000  
0000000001111111  
1234567890123456

### III. MODELOS DE ATRACTIVO

Mediante estos modelos se calcula un índice que expresa el atractivo que tiene cada una de las celdillas del área de estudio para localizar un uso del suelo determinado. Es decir, estos modelos permiten realizar un análisis de la volación del terreno.

Una vez determinados los factores de localización de un uso del suelo específico (variables que intervienen en el modelo), el método consiste primeramente en asignar una ponderación en orden de importancia a cada uno de ellos y en determinar la adecuación o deseabilidad de sus categorías con respecto al uso del suelo considerado.

Al alimentar la computadora con esta información, el programa registra las categorías que toman las variables del modelo en cada celdilla y obtiene un promedio basado en las asignaciones de adecuación o deseabilidad, ponderado por la importancia que tiene cada variable en la localización del uso del suelo en cuestión.

Este índice ó promedio ponderado de atractivo toma valores entre 0 y 9 en tal forma que 0 significa "nada atractivo" y 9 significa "lo más atractivo".

Se debe elaborar un modelo para cada uno de los usos del suelo que se desee localizar en el área de estudio. En la elaboración de un modelo se realizan los siguientes pasos:

#### 1. ENLISTAR LAS VARIABLES DEL MODELO.

Deben identificarse las variables del banco de datos que sean mas representativas de los factores de localización del uso del suelo considerado. Puede incluirse un máximo de diez variables en un modelo. Por ejemplo, en el caso de Valle del Pacífico:

Nombre del modelo: ESTACIONAMIENTOS.

Variables del modelo:

NUMERO	NOMBRE DE LA VARIABLE
2	Porcentaje de pendiente
5	Densidad de arboles
6	Cuencas espaciales
8	Tipo de suelo.

## 2. DIFERENCIAR LA IMPORTANCIA DE LAS VARIABLES

En la mayoría de los casos, no todas las variables tienen la misma importancia en la localización del uso del suelo considerado.

Para dar prioridad a ciertas variables (factores de localización) sobre otras, se deben ponderar de acuerdo a la siguiente escala:

- 1 = Menos importante
- 2 = Importante
- 3 = Muy importante

En el caso de nuestro ejemplo, se puede tener:

NUMERO	NOMBRE DE LA VARIABLE	IMPORTANCIA
2	Porcentaje de pendiente	3
5	Densidad de arboles	2
6	Cuencas espaciales	1
8	Tipo de suelo	1

## 3. ESPECIFICACION DE LA ADECUACION O DESEABILIDAD DE LAS CATEGORIAS

A cada una de las categorías en las que se subdividen las variables se les debe asignar un número entre 0 y 9 para indicar una mayor ó menor adecuación ó deseabilidad de acuerdo a los requerimientos específicos del uso del suelo considerado. Una asignación de 0 significa "nada deseable"; una asignación de 9 significa "lo mas deseable ó adecuado". Por ejemplo:

NUMERO DE LA VARIABLE	CATEGORIAS DE LA VARIABLE									
	0	1	2	3	4	5	6	7	8	9
2	0	0		9		2		0	0	0
5	0		0	7	9	6		5		0
6	0		3	0	0	6		9		0
8	0	9	0	9	0	9	0	9	0	0

Los espacios en blanco son debido a que no existe una categoría con ese número en esa variable.

#### 4. CODIFICACION DEL MODELO.

Para cada modelo de atractivo se deben perforar las siguientes tarjetas:

##### 1a. tarjeta:

Col. 1-2      Número del modelo.

##### 2a. tarjeta:

Col. 1-2      Número de variables que intervienen en el modelo.\*

##### 3a. tarjeta:

Col. 1-2      Número de la primera variable.\*

Col. 10, 12, 14      Preferencias asignadas a las categorías: (Pa-  
16, 18, 20, 22, 24, 26, 28)      Dado en orden ascendente.\*

##### Tarjetas siguientes:

Lo mismo que la 3a tarjeta. Una tarjeta por cada variable incluida en el modelo.

La figura muestra un ejemplo de hoja de codificación para este tipo de modelos.

#### 5. CREACION DE ARCHIVOS DE ATRACTIVO

Para almacenar en disco magnético los índices de atractivo (promedios de las asignaciones de adecuación ponderado por la importancia de las variables), calculados por el modelo para cada una de las celdillas, se debe alimentar el programa ATRACTIVO/IMGRID con la información codificada y perforada en el paso 4 para cada modelo que se desee correr.

En el sistema UNAM/B-6700, esta información debe ir precedida por tres tarjetas de control y por la especificación del contorno del área de estudio (Ver capítulo II, paso 3). Después de las tarjetas del último modelo, debe colocarse una tarjeta de control final para la terminación del trabajo. La forma de preparar el paquete de datos se ilustra en la figura. Al correr el programa se obtiene un listado como el de la figura

\* Los números en estas tarjetas deben ser perforados como enteros justificados a la derecha.

## 6. MAPEAR LOS RESULTADOS.

Mediante el programa MAPAS/IMGRID se debe hacer un mapa que muestre el valor del índice de atractivo en las celdillas del área de estudio para cada uno de los modelos corridos.

La forma de producir los mapas está explicada en el capítulo VI.

Después de revisar y analizar estos mapas, se deben hacer las correcciones y/o modificaciones deseadas.



Programa ATRACTIVO/INGRID

HOJA DE CODIFICACION PARA LOS MODELOS DE ATRACTIVIDAD

Nombre del(os) usuario(s):

Nombre de la actividad (usos del suelo)

para la que se construyó el modelo:

Número del modelo (formato I2): --

Número de variables que intervienen en el modelo (formato I2): --

Columnas 1 y 2

Cuales son las variables relevantes no.

Cual es la preferencia de las subvariables (del 0 al 9)

Cual es la importancia de variable (del 1 al 2)

	Col-10	Col-12	Col-14	Col-16	Col-18	Col-20	Col-22	Col-24	Col-26	Col-28	Col-34

Figura 4.- Hoja de codificación para los MODELOS DE ATRACTIVIDAD.

0000 000

```

09 0 0 2 5 4 2 3 5 2 2 1 (OTRAS VARIABLES)
06 0 0 2 5 4 2 3 5 2 2 1 (OTRAS VARIABLES)
05 0 0 2 5 4 2 3 5 2 2 1 (OTRAS VARIABLES)
02 0 0 2 5 4 2 3 5 2 2 1 (OTRAS VARIABLES)
05 (NUMERO DE VARIABLES QUE INTERVIENEN EN OTRO MODELO)
02 (NUMERO DEL MODELO) (1)
07 0 2 5 9 0 4 2 2 1 0 2 (12-64, 1012-54, 21)
10 3 2 5 4 2 3 5 4 2 2 2 (PRESO DE LA VARIABLE) (1)
09 2 4 6 5 6 7 5 4 2 0 1 (PREFERENCIAS DE SUBVARIABLE) (10) (1)
04 2 5 4 2 5 6 7 5 8 5 0 (NO. DE VARIABLES) (2)
04 (NUMERO DE VARIABLES QUE INTERVIENEN EN EL MODELO) (1)
01 (NUMERO DEL MODELO) (2)
52 20 (TAMANO DE LA MALLA) (2 5)
99999
52 0 0 (CONTORNO IRREGULAR) (2 5)
#DATA
#EXECUTE ATRACTIVO/INGRID
#JOB INGRID ; USER AAB0/AA ; CLASS 3 ; BEGIN ;

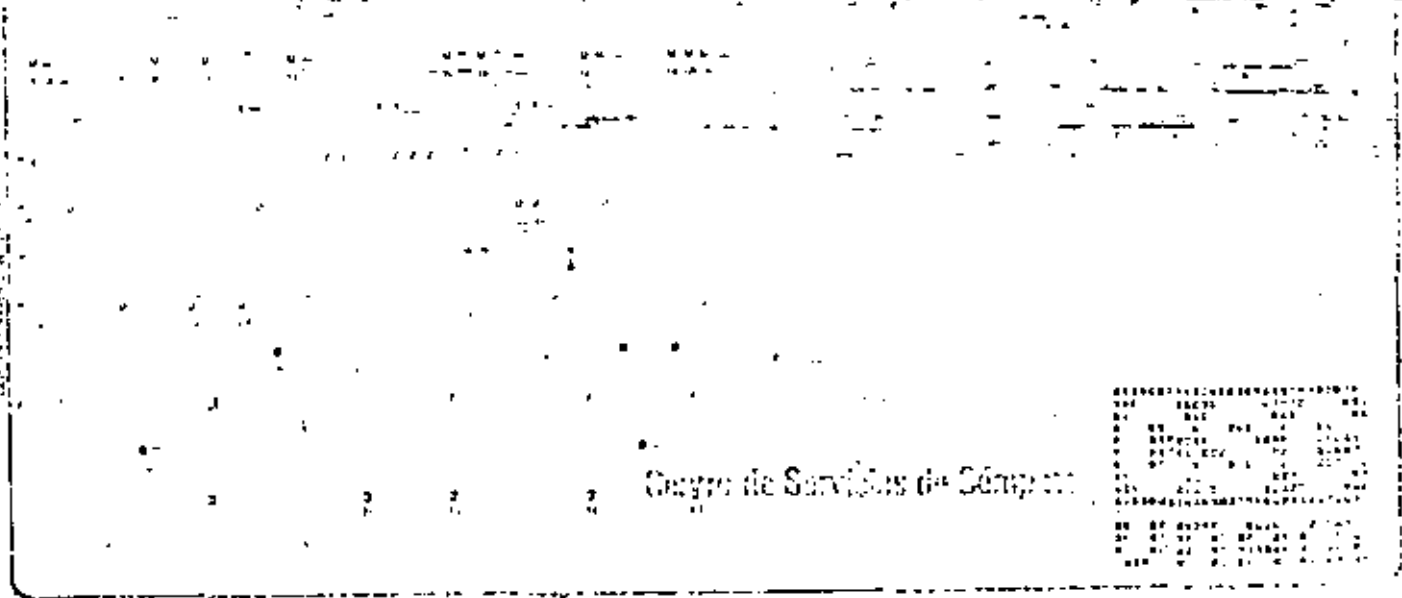
```

Figura . forma de ordenar su paquete de tarjetas para utilizar el programa ATRACTIVO/INGRID en el sistema B-6700 del CSC de la UNAM.

{Tarjetas del programa MAPAS/INGRID ver capítulo VI}

```

MAPAS
00000
52 0 0
IRREGULAR OUTLINE
DATA
FILE FILE40(TITLE=LAEP/ATTR10)
FILE FILE39(TITLE=LAEP/ATTR09)
FILE FILE38(TITLE=LAEP/ATTR08)
FILE FILE37(TITLE=LAEP/ATTR07)
FILE FILE36(TITLE=LAEP/ATTR06)
FILE FILE35(TITLE=LAEP/ATTR05)
FILE FILE34(TITLE=LAEP/ATTR04)
FILE FILE33(TITLE=LAEP/ATTR03)
FILE FILE32(TITLE=LAEP/ATTR02)
FILE FILE31(TITLE=LAEP/ATTR01)
EXECUTE MAPAS/INGRID
JOB=INGRID; USER=AGUOZAR; CLASS=3; BEGIN
  
```



Grupo de Servicios de Computación

--Figura 6.-- Forma de obtener los mapas de los "modelos de actividad" con el programa MAPAS/INGRID. Este ejemplo muestra la manera de obtener 10 mapas {FILE31, ..., 40} a partir de los archivos {LAEP/ATTR01, ..., 10} que fueron generados por el programa ATRACTIVO/INGRID.

\*\*\* MODELO DE ATRACTIVIDAD NUMERICO \*\*\*

LOS INDICES DE ATRACTIVIDAD SON PUESTOS EN EL ARCHIVO EN DISQUETE/ATTR 1  
EL NUMERO DE VARIABLES EN EL MODELO ES 4

ESTAS SON LAS TARJETAS DE LA HOJA DE CODIFICACION USADA EN EL MODELO

10	C 0 0 6 8 7 9 0 0 0	2
6	C 0 0 6 6 9 0 9 0 0	2
9	C 0 7 7 4 6 5 0 0 0	2
2	C 0 0 9 0 8 0 5 0 0	1

TITULO DEL MAPA

MODELO DE ATRACTIVIDAD Y ACTIVIDADES DE ESPACIAMIENTO

AREA DE ESTUDIO: VALLE DE MAGUICO

LABORATORIO DE PLANEACION URBANA

OPCIONES USADAS PARA ESTA MAPA

1 EL TAMAÑO DE LA CELLA ES 32 CARACTERES Y 16 COLUMNAS  
EL TAMAÑO DE LOS CARACTERES ES 12 CARACTERES EN POSICION VERTICAL  
Y 12 CARACTERES EN POSICION HORIZONTAL

7. LOS SIGNOS SON  
SOLIDOS

10 EL TEXTO DEL MAPA SON  
VARIABLES QUE USADAS EN EL MODELO

10 COMPLE

PESI

10 VISTAS

2

6 ZONAS LLANAS

2

9 PROXIMIDAD AL AGUA

2

2 POCIMIENTO DE PENDIENTE

1

13 LA NUMERACION DE LA CELLA COMIENZA EN 1 32

14 SE SUPONE QUE LOS DATOS ESTAN PRE-ESCALADOS



## \*\*\* MODELO DE ATRACTIVIDAD NUMERICO \*\*\*

LUS INDICES DE ATRACTIVIDAD SON MUESTROS EN EL APP-IVO EN DISOLAF:ATTR 2  
 EL NUMERO DE VARIABLES EN EL MODELO ES 4

ESTAS SON LAS TARJETAS DE LA HOJA DE CODIFICACION USADA EN EL MODELO :

2	C C C 9 C 2 C 0 0 0	2
5	C C C 7 9 6 0 5 0 0	1
6	C 0 3 C C 6 0 9 0 0	1
8	C 9 C 0 C 9 0 9 0 0	1

## TITULO DEL MAPA

MODELO DE ATRACTIVIDAD DE ESTACIONARIANTES

AREA DE ESTUDIO VALLE DEL PACIFICO

LABORATORIO DE PLANEACION URBANA

## OPCIONES USADAS PARA ESTIMAR

10 EL TEXTO DEL MAPA ES  
 VARIABLES USADAS EN EL MODELO

AL	NUMERO	DESCRIPCION	RESO
3		PORCIENTO DE PENDIENTE	2
5		PORCIENTO DENSIDAD DE ARBOLES	1
6		ZONAS LLANAS	1
8		RIOS	1

0000000000000000  
 0000000001111111  
 1234567890123456

MODULO DE ACTIVIDADES 2

ESTACIONAMIENTOS

0000																				0000
0001																				0001
0002																				0002
0003																				0003
0004																				0004
0005																				0005
0006																				0006
0007																				0007
0008																				0008
0009																				0009
0010																				0010
0011																				0011
0012																				0012
0013																				0013
0014																				0014
0015																				0015
0016																				0016
0017																				0017
0018																				0018
0019																				0019
0020																				0020
0021																				0021
0022																				0022
0023																				0023
0024																				0024
0025																				0025
0026																				0026
0027																				0027
0028																				0028
0029																				0029
0030																				0030
0031																				0031
0032																				0032
0033																				0033
0034																				0034
0035																				0035
0036																				0036
0037																				0037
0038																				0038
0039																				0039
0040																				0040
0041																				0041
0042																				0042
0043																				0043
0044																				0044
0045																				0045
0046																				0046
0047																				0047
0048																				0048
0049																				0049
0050																				0050

AREA DE SERVICIOS DEL VALLE DEL PACIFICO

~~LABORATORIO DE INVESTIGACIONES~~

AVAILABLE SEPARATE USABASTH TEL - MODELO

CONDICION DE PERDIENTE  
 DIRECTIVO DE SERVICIO DE ARBOLES  
 ZONAS LLANAS  
 SIEMPRE

0000000000000000  
 0000000001111111  
 1234567890123456

DIVISOR	1	2	3	4	5	6	7	8	9
SIEMPRE	X	X	X	X	X	X	X	X	X
FRECUENCIA	60	50	40	30	20	10	00	1	1

## \*\*\*-MODELO DE ATRACTIVIDAD URBANA-3\*\*\*

LOS INDICES DE ATRACTIVIDAD SON PUESTOS EN EL ARCHIVO EN DISCO LAEP/ATIK 3  
 EL NUMERO DE VARIABLES EN EL MODELO ES 4

ESTAS SON LAS TARJETAS DE LA HOJA DE CODIFICACION USADA EN EL MODELO :

2	0 0 0 9 0 7 0 0 0 0	1
5	0 0 0 6 8 9 0 7 0 0	1
6	0 0 0 0 9 0 7 0 0	1
7	0 0 9 0 0 0 0 0 0 0	1

## TITULO DEL MAPA

MODELO DE ATRACTIVIDAD # 3: ESTRUCTURAS

AREA DE ESTUDIO: VALLE DEL PACIFICO

LABORATORIO DE PLANEACION URBANA

## OPCIONES USADAS PARA ESTE MAPA

10. EL TEXTO DEL MAPA ES

VARIABLES DATO USADAS EN EL MODELO

NO	NOMBRE	PESO
2	PORCIENTO DE PENDIENTE	1
5	PORCIENTO DE DENSIDAD DE ARBOLES	1
6	ZONAS PLANAS	1
7	ACCIDENTES GEOLOGICOS	1



MODELO DE ATRACTIVIDAD 3 31

00000000000000000000  
000000000111111111  
1234567890123456

ESTRUCTURAS

052	.....X.....	052
051	.....X.....	051
050	.....D.....	050
049	.....X0.....	049
048	.....000.....	048
047	.....0.....	047
046	.....0.....	046
045	.....X.....	045
044	.....X0X0.....	044
043	.....0X0.....	043
042	.....0000X0X.....	042
041	.....00XX0X.....	041
040	.....00XX0X.....	040
039	.....00XX0X.....	039
038	.....00XX0X.....	038
037	.....00XX0X.....	037
036	.....00XX0X.....	036
035	.....00XX0X.....	035
034	.....00XX0X.....	034
033	.....0000Y.....	033
032	.....0000Y.....	032
031	.....X000000000X.....	031
030	.....0000000000X.....	030
029	.....0000000000X.....	029
028	.....0000000000X.....	028
027	.....X000000000X.....	027
026	.....XX000000000.....	026
025	.....X0000000000.....	025
024	.....0000000000X.....	024
023	.....0000000000X.....	023
022	.....0000000000X.....	022
021	.....0000000000X.....	021
020	.....0000000000X.....	020
019	.....X.....	019
018	.....X.....	018
017	.....X.....	017
016	.....X.....	016
015	.....00X.....	015
014	.....000000X00X.....	014
013	.....0000000000X.....	013
012	.....0000000000X.....	012
011	.....0000000000X.....	011
010	.....00X00000X.....	010
009	.....00X00000X.....	009
008	.....X000000000X.....	008
007	.....X000000000X.....	007
006	.....X000000000X.....	006
005	.....0000000000X.....	005
004	.....0000000000X.....	004
003	.....0000000000X.....	003
002	.....0000000000X.....	002
001	.....0000000000X.....	001

AREA DE ESTUDIO VALLE DEL PACIFICO

LABORATORIO DE PLANEACION URBANA

VARIABLES CANTIDAD USADAS EN EL MODELO

NO	NOMBRE	PESO
2	PORCIENTO DE PENDIENTE	1
3	PORCIENTO DENSIDAD DE ARBOLES	1
6	ZONAS LLANAS	1
7	ACCIDENTES GEOLOGICOS	1

00000000000000000000  
000000000111111111  
1234567890123456

NIVEL	0	1	2	3	4	5	6	7	8	9
SIMBOLOS	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
FRECUENCIA	321	0	159	0	103	84	37	15	0	2

\*\*\* MODELO DE ATRACTIVIDAD NUMERO 4 \*\*\*

LOS INDICES DE ATRACTIVIDAD SON PUESTOS EN EL ARCHIVO EN DISCO LAE/ATTN 4

EL NUMERO DE VARIABLES EN EL MODELO ES 2

ESTAS SON LAS TARJETAS DE LA HOJA DE CODIFICACION USADA EN EL MODELO 1

2	0 0 0 9 0 5 0 0 0 0	1
9	0 0 0 2 5 9 0 0 0 0	1

TITULO DEL MAPA

MODELO DE ATRACTIVIDAD PARA CABALLERIZAS

AREA DE ESTUDIO VALLE DEL PACIFICO

LABORATORIO DE PLANEACION URBANA

OPCIONES USADAS PARA ESTE MAPA

10 EL TEXTO DEL MAPA ES  
VARIABLES DATE USADAS EN EL MODELO

Nº	NOMBRE	PESO
2	PORCIENTO DE PENDIENTE	2
9	PROXIMIDAD AL AGUA	1



\*\*\* MODELO DE ATRACTIVIDAD NUMERO 5 \*\*\*

LOS INDICES DE ATRACTIVIDAD SON PUESTOS EN EL ARCHIVO ESTADISTICO AF/ATR5  
 EL NUMERO DE VARIABLES EN EL MODELO ES 6

ESTAS SON LAS TARJETAS DE LA HOJA DE CODIFICACION USADA EN EL MODELO :

2	0 0 0 9 0 3 0 0 0 0	2
5	0 0 0 6 8 9 0 0 0 0	2
6	0 0 0 0 0 9 0 9 0 0	2
4	0 0 0 0 0 9 0 0 0 0	1
9	0 0 9 9 7 3 0 0 0 0	1
1	0 9 5 0 0 0 5 0 9 0	1

TITULO DEL MAPA

MODELO DE ATRACTIVIDAD Y SU ESTACIONAMIENTO PARA VILLERO  
 AREA DE ESTUDIO VALLE DEL PACIFICO  
 LABORATORIO DE PLANEACION URBANA

OPCIONES DE SELECCION PARA EL MAPA

10 EL TEXTO DEL MAPA ES  
 VARIABLES DATA USADAS EN EL MODELO

NL	NOMBRE	RESU
2	POCIENTO DE PENDIENTE	2
5	POCIENTO DE SENSIDAD DE ARBOLES	2
6	ZONAS LLANAS	2
4	VEGETACION PER SOJA	1
9	PROXIMIDAD AL AGUA	1
1	USOS DEL SUELO ACTUALES	1

00000000000000000000
00000000001111111111
1234567890123456

ESTACIONAMIENTO PARA TRAILERS

Table with 10 columns and multiple rows. Columns are numbered 0-9. Rows contain alphanumeric characters (X, Y, O, B, Z) and symbols (+, -, .) representing data points for the 'ESTACIONAMIENTO PARA TRAILERS' section.

AREA DE ESTUDIO VALLE DEL PACIFICO

LABORATORIO DE FERTILIZANTES

RIABLES PARA USADAS EN EL MODELO

Table with 2 columns: 'NOMBRE' and 'PESO'. It lists variables such as 'PORCENTAJE DE FERTILIZANTE', 'DENSIDAD DE ARBOLES', 'VEGETACION POR ZONA', 'PROXIMIDAD AL AGUA', and 'USOS DEL SUELO ACTUALES'.

00000000000000000000
00000000001111111111
1234567890123456

Table with 10 columns labeled 'NIVELES' (0-9) and 10 rows labeled 'IMPULSOS' (1-10). It displays a grid of alphanumeric characters and symbols (+, -, .) representing data points for different levels and impulses.

\*\*\* MODELO DE ATRACTIVIDAD URBANA \*\*\*

TUS INDICES DE ATRACTIVIDAD SON PUESTOS EN EL ARCHIVO EN DISQUETE/ATP 6  
EL NUMERO DE VARIABLES EN EL MODELO ES 5

ESTAS SON LAS TARJETAS DE LA HOJA DE CODIFICACION USADA EN EL MODELO:

2	C C C 5 C P C A 0 0	2
6	C C C C C 9 0 8 0 0	2
5	C C C 5 4 9 0 - 8 0 0	1
7	C C 9 9 7 0 0 0 0 0	1
4	C C C C 7 5 0 C 0 0	1

TITULO DEL MAPA

MODELO DE ATRACTIVIDAD URBANA - CAMBATA

AREA DE ESTUDIO: VALLE DEL MOTILON

LABORATORIO DE PLANEACION URBANA

OPCIONES PARA ESTE MAPA

10 EL TEXTO DEL MAPA ES  
VARIABLES USADAS EN EL MODELO

NO	NOMBRE	PESE
1	COCIENTE DE PENDIENTE	2
2	ZONAS URBANAS	1
3	PERCENTAJE DENSIDAD DE ARBOLES	1
4	PROXIMIDAD AL AGUA	1
5	VEGETACION POR ZONA	1

CAMINATA

AREA DE ESTUDIO VALLE DEL PACIFICO

LAMORPATRICIO PLANTELAS URBANO

VARIABLES BATA USADAS EN EL MODELO

NO	NOMBRE	PESO
2	PROCESO DE PENDIENTE	2
3	ZONAS URBANAS	2
4	PROB. DE OMBRERA DE ARBOLES	1
5	PROB. DE BATA EL AGUA	1
6	DEFINICION POR ZONA	1

```

050 .....
051 .....X..
052 .....X..
053 .....X..
054 .....X..
055 .....X..
056 .....X..
057 .....X..
058 .....X..
059 .....X..
060 .....X..
061 .....X..
062 .....X..
063 .....X..
064 .....X..
065 .....X..
066 .....X..
067 .....X..
068 .....X..
069 .....X..
070 .....X..
071 .....X..
072 .....X..
073 .....X..
074 .....X..
075 .....X..
076 .....X..
077 .....X..
078 .....X..
079 .....X..
080 .....X..
081 .....X..
082 .....X..
083 .....X..
084 .....X..
085 .....X..
086 .....X..
087 .....X..
088 .....X..
089 .....X..
090 .....X..
091 .....X..
092 .....X..
093 .....X..
094 .....X..
095 .....X..
096 .....X..
097 .....X..
098 .....X..
099 .....X..
100 .....X..

```

NIVEL	1	2	3	4	5	6	7	8	9
SIMPLOS	1	2	3	4	5	6	7	8	9
FRECUENCIAS	507	71	44	70	21	20	19	3	2

#### IV. MODELOS DE VULNERABILIDAD:

Mediante estos modelos es posible analizar los efectos probables que tendrán los usos del suelo propuestos sobre diferentes sistemas biológicos y físicos existentes en el área. Estos efectos se miden de acuerdo al grado de impacto negativo ó daño ambiental, jerarquizándolos en compatible, moderado, severo y terminal. El método consiste principalmente en la asignación y combinación de sensibilidades de las variables que intervienen en el modelo mediante el uso de matrices. Para cada modelo deben realizarse los siguientes pasos:

##### 1. ENLISTAR LAS VARIABLES DEL MODELO.

Por razones prácticas, sólo deben considerarse tres variables para describir un sistema. Estas deben ser enlistadas en orden de importancia decreciente: más importante, importante y menos importante. Por ejemplo, para el caso de Valle del Pacífico:

Nombre del modelo: EROSION

IMPORTANCIA	NUMERO	NOMBRE DE LA VARIABLE
más importante	2	Porcentaje pendiente
importante	4	Tipo de vegetación
menos importante	8	Tipo de suelo

##### 2. DIFERENCIAR LA SENSITIVIDAD DE LAS CATEGORIAS DE CADA VARIABLE.

Las categorías en que se subdivide cada variable, deben agruparse de acuerdo a su sensibilidad al impacto considerado. Los grados de sensibilidad en el modelo se asignan de la siguiente manera:



- 1 = Alta sensibilidad.
- 2 = Media.
- 3 = Baja.

En las hojas para codificar los modelos debe escribirse el número de cada categoría en la casilla correspondiente a su grado de sensibilidad. Por ejemplo:

NOMBRE DE LA VARIABLE	SENSITIVIDAD DE LA CATEGORÍA		
	ALTA	MEDIA	BAJA
Porcentaje de pendiente	8,7	5	0,1,3,9
Tipo de vegetación	2	5	0,4,3,9
Tipo de suelo	4	8,6,2	0,1,3,5,7,9

A continuación, debe anotarse el grado de sensibilidad (1, 2 ó 3) de cada categoría en la columna correspondiente en una tabla como la siguiente:

NUMERO DE LA VARIABLE	NUMERO DE LA CATEGORIA									
	0	1	2	3	4	5	6	7	8	9
2	3	3		3		2		1	1	3
4	3		1	3	3	2				3
8	3	3	2	3	1	3	2	3	2	3

Los espacios dejados en blanco son debidos a que no existe una categoría con ese número en esa variable.

### 3. COMBINAR LAS SENSITIVIDADES DE LAS VARIABLES IMPORTANTE Y MENOS IMPORTANTE.

Se debe asignar un grado de sensibilidad conjunta a cada una de las combinaciones posibles de sensibilidades de las categorías de estas dos variables. La estimación de estos grados de sensibilidad combinada se hace con la jerarquía:

- 1 = Alta sensibilidad

1 = Alta sensibilidad

2 = Media

3 = Baja

Las asignaciones deben ser colocadas en una matriz como la siguiente:

		SENSITIVIDAD DE LA VARIABLE "IMPORTANTE"		
		ALTA	MEDIA	BAJA
Sensitividad de la variable. "menos importante"	ALTA	1	2	3
	MEDIA	1	2	2
	BAJA	1	1	3

#### 4. AGRUPAR LOS USOS DEL SUELO.

Los usos del suelo considerados deben agruparse de acuerdo al grado de impacto potencial ocasionado por su construcción, mantenimiento y actividades de los usuarios. Estos impactos se clasifican como:

1 = Bajo

2 = Medio

3 = Alto

El número del grupo al que pertenece cada uso del suelo considerado en el proyecto, debe ser anotado en una tabla como la siguiente:

		NUMERO DE USO DEL SUELO											
		1	2	3	4	5	6	7	8	9	10	...	20
Grupo al que pertenece	1	2	2	3	3	1	3	2					

## 5. DETERMINAR LA VULNERABILIDAD DEL SISTEMA.

Debe asignarse un índice de vulnerabilidad (grado de impacto negativo: daño ambiental del sistema con respecto al impacto considerado, para todas las combinaciones posibles de los grados de la sensibilidad obtenidos en el paso 3 con los grados de sensibilidad de la variable más importante. - Los índices de vulnerabilidad corresponden a las siguientes categorías:

- 1 = Comparatible
- 2 = Moderado
- 3 = Severo
- 4 = Terminal

Debe hacerse un análisis por separado para cada uno de los grupos de usos del suelo. Las asignaciones deben colocarse en forma matricial como se ilustra a continuación:

## a. Grupo 1 de los usos del suelo.

		SENSITIVIDAD DE LA VARIABLE "MAS IMPORTANTE"		
		ALTA	MEDIA	BAJA
SENSITIVIDAD COMBINADA DE LAS VARIABLES "IMPORTANTE" Y "MENOS IMPORTANTE"	ALTA	3	2	1
	MEDIA	2	1	1
	BAJA	2	1	1

## b. Grupo 2 de usos del suelo.

		SENSITIVIDAD DE LA VARIABLE "MAS IMPORTANTE"		
		ALTA	MEDIA	BAJA
SENSITIVIDAD COMBINADA DE LAS VARIABLES "IMPORTANTE" Y "MENOS IMPORTANTE"	ALTA	4	3	2
	MEDIA	3	2	1
	BAJA	3	2	1

SENSITIVIDAD DE LA VARIABLE  
"MAS IMPORTANTE"

SENSITIVIDAD  
COMBINADA DE  
LAS VARIABLES  
"IMPORTANTE"  
Y "MENOS IM-  
PORTANTE"

	ALTA	MEDIA	BAJA
ALTA	4	3	2
MEDIA	3	3	2
BAJA	3	2	2

## 6. CODIFICACION DEL MODELO.

Para cada modelo se deben perforar las siguientes tarjetas:

1a. tarjeta:

Col. 1-2            Número del modelo. Perforado como entero -  
justificado a la derecha.

2a. tarjeta:

Col. 1-9            LAEP/IMPT para indicar que se crea un archi-  
vo de impactos.  
Col. 10-11          Número del modelo. Perforado como entero --  
justificado a la derecha.  
Col. 12             . Para indicar que termina el nombre del ar-  
chivo.

3a. tarjeta:

Col. 1-3            Número de la variable más importante.  
Col. 4-6            Número de la variable importante.  
Col. 7-9            Número de la variable menos importante.  
Perforados como enteros y justificados a la  
derecha.

4a. tarjeta:

Col. 2, 4, 6        Valores de sensibilidad de las categorías de  
8, 10, 12           la variable más importante, asignados en el-  
14, 16, 18        paso 2.  
20.

5a. tarjeta:

Lo mismo que la 4a. tarjeta para la variable  
importante.

- 6a. tarjeta: Lo mismo que la 4a. tarjeta para la variable menos importante.
- 7a. tarjeta:
- Col. 2, 4, 6 Grados de sensibilidad conjunta de las variables importantes y menos importantes. (matriz 8, 10, 12 del paso 3)\*.
- 14, 16, 18
- 8a. tarjeta:
- Col. 2, 4, 6 Grados de impacto negativo (vulnerabilidad) para el primer grupo de usos del suelo. (1a. matriz del paso 5)\*.
- 8, 10, 12
- 14, 16, 18
- 9a. tarjeta: Lo mismo que la 8a. tarjeta para el segundo grupo de usos del suelo. (2a. matriz del paso 5)\*.
- 10a. tarjeta: Lo mismo que la 8a. tarjeta para el tercer grupo de usos del suelo. (3a. matriz del paso 5)\*.
- 11a. tarjeta:
- Col. 2, 4, 6 Número de grupo de usos del suelo al que pertenece cada uno de los usos considerados en el estudio (paso 4). Se deben utilizar solamente el número de columnas necesario para registrar la clasificación de todos los usos. El programa permite como máximo 10 usos diferentes.
- 8, 10, 12
- ....., 40

Un ejemplo de hoja de codificación se ilustra en la figura

## 7. CREACION DE ARCHIVOS DE VULNERABILIDAD.

Para almacenar un disco magnetico los indices de vulnerabilidad calculados por el modelo para cada una de las celdillas del área de estudio, debe alimentarse el programa IMPACTO/IN GRID con la información codificada en el paso 6 para cada uno de los modelos que se deseen correr.

Los indices de vulnerabilidad calculados por cada modelo para cada celdilla, corresponden a la siguiente jerarquía:

- 0 = Los tres grupos, COMPATIBLE.
- 1 = Grupo 2, MODERADO.
- 2 = Grupo 3, MODERADO; Grupo 2, MODERADO.

\* Las matrices deben ser codificadas empezando por la primera columna de arriba hacia abajo, continuando hasta registrar las nueve casillas.

- 3 = Los tres grupos. MODERADO.  
4 = Grupo 3. SEVERO.  
5 = Grupo 3. SEVERO; Grupo 2. SEVERO.  
6 = Los tres grupos. SEVERO.  
7 = Grupo 3. TERMINAL.  
8 = Grupo 3. TERMINAL; Grupo 7, TERMINAL.  
9 = Los tres grupos, TERMINAL.

En el sistema UNAM/B-6700, esta información debe ir precedida por tres tarjetas de control y la especificación del contorno del área de estudio (ver capítulo II, paso 3). Después de -- las tarjetas del último modelo, debe colocarse una tarjeta de control final para la terminación del trabajo. La forma de -- preparar el paquete de datos se ilustra en la figura

#### 8. MAPEAR LOS RESULTADOS.

Mediante el programa MAPAS/INGRID se debe hacer un mapa que muestre el valor que tiene el índice de vulnerabilidad en las celdillas del área de estudio, para cada uno de los modelos -- corridos.

La forma de producir los mapas está explicada en el capítulo-IV. La opción 10 de texto explicativo en cada mapa debe con- tener la clave de jerarquización de impactos enlistada en el- paso 7.

##END JOB

010000000000000000000000

(9 I 2)

010000000000000000000000

(9 I 2)

000000000000000000000000

(9 I 2)

de las variables 2 y 3 con la sensibilidad de la variable 1  
(suelo, obtenidos por la combinación de sensibilidades de la matriz  
(Valores de las matrices de impactos, para cada grupo de uso del

comparación de las variables 2 y 3) (formato 9 I 2)

(Valores de la matriz de sensibilidades combinadas, obtenidas de la

0 0 0 1 1 0 0 0 0 (Menos importante) (0 I 1) (variable 1)

0 0 1 0 0 0 0 0 0 (Importante) (10 I 1) (variable 2)

0 0 0 0 0 0 0 1 0 (Muy importante) (1 I 1) (variable 3)

(Valores de sensibilidad de las subvariables)

0 0 0 (Número de variables-estados) (3 I 3)

00000000 (Nombre del archivo donde impacta)

01 (Número del módulo) (1 I 2)

02 10 (Tamaño de la matriz) (2 I 5)

00000

00 (Contorno integral) (3 I 3)

##DATA

000000000000000000000000

000000000000000000000000

000000000000000000000000

000000000000000000000000

000000000000000000000000

000000000000000000000000

000000000000000000000000

000000000000000000000000

000000000000000000000000

000000000000000000000000

000000000000000000000000

000000000000000000000000

000000000000000000000000

000000000000000000000000

000000000000000000000000

000000000000000000000000

000000000000000000000000

000000000000000000000000

000000000000000000000000

000000000000000000000000

000000000000000000000000

000000000000000000000000

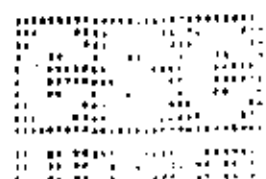
000000000000000000000000

000000000000000000000000

000000000000000000000000

000000000000000000000000

000000000000000000000000



HOJA DE CODIFICACION PARA LOS MODELOS DE IMPACTO

Nombre del modelo: \_\_\_\_\_

Tarjeta 1.- Número del modelo formato (I2): \_\_\_\_\_

Tarjeta 2.- Nombre del modelo: LAEP/IMP01, ..., LAEP/IMP32.

Tarjeta 3.- Número de las variables dato (2I3): \_\_\_\_\_  
cols. 3    6    9

Encontrar los valores de sensibilidad, para cada una de las subvariables de las 3 variables que intervienen en el modelo, de acuerdo a la siguiente escala:

Subvariable más sensitiva = 1 = A

Subvariable moderada = 2 = B

Subvariable menos sensitiva = 3 = C

no. de variable	Variable	Subvariables		
		más sensitiva	moderada	menos sensitiva

más importante  
↓  
menos importante

A                  B                  C

cols. - 2    4    6    8    10    12    14    16    18    20

Tarjeta 4.- \_\_\_\_\_ Variable 1

Tarjeta 5.- \_\_\_\_\_ Variable 2

Tarjeta 6.- \_\_\_\_\_ Variable 3

Reducir las variables 2 y 3 comparando entre sí sus grados de sensibilidad para obtener la matriz (3,3) de sensibilidades combinadas.

		Variable 2		
		A	B	C
Variable 3	A			
	B			
	C			

Valores de:  
 X = Más sensitiva = 1  
 Y = Moderada = 2  
 Z = Menos sensitiva = 3

Tarjeta 7.- Valores de la matriz de sensibilidades combinadas (3 I 2).



Encontrar la matriz de impactos para cada grupo de usos del suelo, estos valores de impactos se obtienen a partir de la combinación de sensibilidades de la matriz de sensibilidades combinadas de las variables 2 y 3 con las sensibilidades de la variable 1.

		Variable 1			Variable 2			Variable 3		
		A	B	C	A	B	C	A	B	C
X	Uso del suelo Grupo I									
Y										
Z										
		Uso del suelo Grupo II			Uso del suelo Grupo III					

Asignación de grupos de usos del suelo

Uso del suelo 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19  
Grupo \_\_\_\_\_

La matriz de impactos tiene tres niveles de impacto por grupo de usos del suelo.

Asignando valores de C- Impacto Compatible

M- Impacto Moderado

S- Impacto Severo

T- Impacto Terminal

Tarjeta 8.- Valores de la matriz de impactos U.S. grupo I (9 I 2)

Tarjeta 9.- Valores de la matriz de impactos U.S. grupo II (9 I 2).

Tarjeta 10.- Valores de la matriz de impactos U.S. grupo III (9 I 2)

Figura 7.- Hojas de codificación para el programa Impacto/INGRID.

TITULO DEL MAPA

MODELO DE IMPACTO N 1:      URBANIZACION  
 AREA DE ESTUDIO      VALLE DEL SACRIFICIO  
 LABORATORIO DE PLANEACION URBANA

OPCIONES USADAS PARA ESTE MAPA

- 1 EL TAMAÑO DE LA MALLA ES 50 DE COLUMNAS Y 10 COLUMNAS  
 EL TAMAÑO DE LA CELDILLA ES 1 CARRILLO EN SENTIDO VERTICAL Y  
 7 LOS SÍMBOLOS SON 1 PLANEACION EN SENTIDO HORIZONTAL  
 \*\* +X00000123456789  
 \* +X00004  
 // \*X  
 +Y
- 10 EL TEXTO DEL MAPA ES -----  
 VARIABLES DATA USADAS EN EL MODELO  
 2 PORCIENTO DE PENDIENTE  
 3 VEGETACION POR ZONA  
 8 SUELOS POR TIPO
- 13 LA NUMERACION DE LA MALLA COMIENZA EN 1 50
- 14 SE SUPONE QUE LOS DATOS ESTAN PRE-ESCALADOS

LEGENDA: EL SÍMBOLO MAS OSCURO EXPRESA EL IMPACTO MAYOR  
 LOS VALORES DEL IMPACTO PARA LOS 3 GRUPOS DEL USO DEL SUELO SON:

U R B A N I Z A C I O N	LOS TRES GRUPOS DE US COMPARTILES
	GRUPO DE US 111 MODERADO
	GRUPO DE US 111 MODERADO, GRUPO 11 MODERADO
	LOS TRES GRUPOS DE US MODERADO
	GRUPO DE US 111 SEVERO
	GRUPO DE US 111 SEVERO, GRUPO 11 SEVERO
	LOS TRES GRUPOS DE US SEVERO
	GRUPO DE US 111 TERMINAL
	GRUPO DE US 111 TERMINAL, GRUPO 11 TERMINAL
LOS TRES GRUPOS DE US TERMINAL	

USO DEL SUELO I

ACTIVIDADES DE ESPACIO  
 CAMINATA

USO DEL SUELO II

CAMINOS  
 ESTACIONAMIENTOS  
 ESTRUCTURAS

USO DEL SUELO III

SERVICIO DE TAXI  
 CABALLOS  
 ESTABLECIMIENTOS

AREA DE ESTUDIO VALLE DEL PACIFICO

LABORATORIO DE PLANEACION URBANA

VARIABLES DATO USADAS EN EL MODELO

- 2 PORCENTAJE DE PENDIENTE
- 3 VEGETACION POR ZONA
- 4 RUILOS POR TIPO

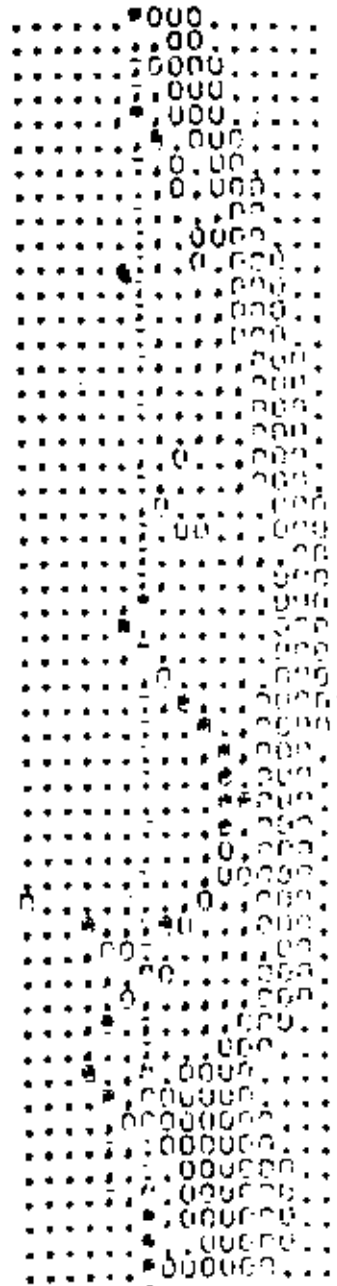
LEGENDA: EL SIMBOLO MAS OSCURO  
CORRESPONDE AL IMPACTO MAYOR

LOS VALORES DEL IMPACTO PARA LOS  
3 GRUPOS DEL USO DEL SUELO SON:

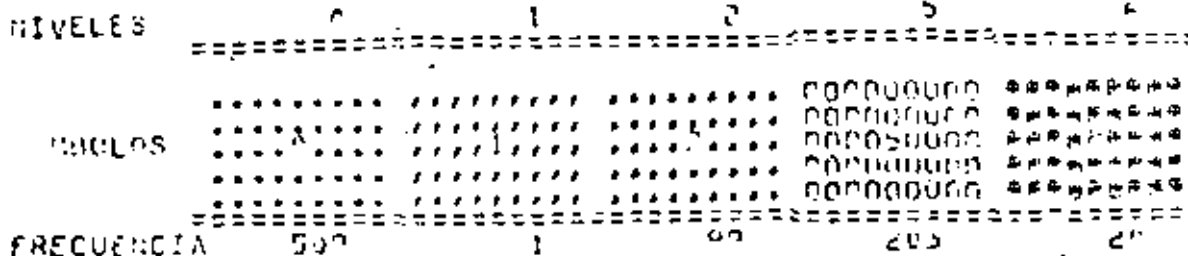
- LOS TRES GRUPOS DE USOS COMERCIALES
- GRUPO DE US 111 COMERCIO
- GRUPO DE US 111 COMERCIO, GRUPO 11 COMERCIO
- LOS TRES GRUPOS DE US INTERMEDIOS
- GRUPO DE US 111 SEVERO
- GRUPO DE US 111 SEVERO, GRUPO 11 SEVERO
- LOS TRES GRUPOS DE US SEVERO
- GRUPO DE US 111 TERMINAL
- GRUPO DE US 111 TERMINAL, GRUPO 11 TERMINAL
- LOS TRES GRUPOS DE US TERMINAL

- USO DEL SUELO 1
- ACTIVIDADES DE ESPACIO COMUNITARIO
- USO DEL SUELO 11
- CANILES
- ESTACIONES DE TRANSITO
- ESTRUCTURAS

USO DEL SUELO 111  
SERVICIO DE ADMISION  
CALLE PEZON  
ESTAC. 7 TRAILERS



0000000000000000  
0000000000000000  
1234567890123456



Vertical text on the right edge of the page, possibly a page number or reference code.

TITULO DEL MAPA

MODELO DE IMPACTO 9 2: CAMBIO EN LA VISUALIDAD  
 AREA DE ESTUDIO VALLE DEL PACIFICO  
 LABORATORIO DE PLANEACION URBANA

OPCIONES USADAS PARA ESTE MAPA

10 EL TEXTO DEL MAPA ES  
 VARIABLES DATO USADAS EN EL MODELO

4 VEGETACION, 50% ZONA  
 5 PORCIENTO DE VISUALIDAD DE ARBOLES  
 2 PORCIENTO DE PASADIZOS

LEGENDA: EL SIGUIENTE TABLA MUESTRA LAS CORRELACIONES DEL IMPACTO PARA  
 LOS VALORES DEL IMPACTO PARA LOS 3 GRUPOS DEL USO DEL SUELO 100:

0	LOS TRES GRUPOS DE US COMPATIBLES
1	GRUPO DE US 111 MODERADO
2	GRUPO DE US 111 MODERADO, GRUPO 11 MODERADO
3	LOS TRES GRUPOS DE US MODERADO
4	GRUPO DE US 111 SEVERO
5	GRUPO DE US 111 SEVERO, GRUPO 11 SEVERO
6	LOS TRES GRUPOS DE US SEVERO
7	GRUPO DE US 111 TERMINAL
8	GRUPO DE US 111 TERMINAL, GRUPO 11 TERMINAL
9	LOS TRES GRUPOS DE US TERMINAL

USO DEL SUELO 1

USO DEL SUELO 11

USO DEL SUELO 111

ACTIVIDADES DE ESPACIO  
 CAMINATA  
 CABALLERIZAS

ESTRUCTURAS  
 VIO

ESTACIONAMIENTOS  
 CENTROS  
 ESTAC. / TRAILERS

### CAMBIO EN LA VISUALIDAD

AREA DE ESTUDIO VALLE DEL PACIFICO

LABORATORIO DE PLANEACION URBANA

VARIABLES DADO USADAS EN EL MODELO

- 4 VEGETACION POR ZONA
- 5 PORCENTAJE DE PENDIENTE DE TERRENO
- 2 PORCENTAJE DE PENDIENTE

LEGENDA: EL SIMBOLO MAS OSCURO CORRESPONDE AL IMPACTO MAYOR

LOS VALORES DEL IMPACTO PARA LOS 3 GRUPOS DEL USO DEL SUELO SON:

LOS TIPOS DE GRUPO DE	GRUPOS DE US COMPATIBLES	
LOS TIPOS DE GRUPO DE	US 111 NOVEDARI	
LOS TIPOS DE GRUPO DE	US 111 NOVEDARI, GRUPO 11 NOVEDARI	
LOS TIPOS DE GRUPO DE	GRUPOS DE US NOVEDARI	
LOS TIPOS DE GRUPO DE	US 111 SEVERO	
LOS TIPOS DE GRUPO DE	US 111 SEVERO, GRUPO 11 SEVERO	
LOS TIPOS DE GRUPO DE	GRUPOS DE US SEVERO	
LOS TIPOS DE GRUPO DE	US 111 TERMINAL	
LOS TIPOS DE GRUPO DE	US 111 TERMINAL, GRUPO 11 TERMINAL	
LOS TIPOS DE GRUPO DE	GRUPOS DE US TERMINAL	

052	0000000000
051	0000000000
050	0000000000
049	0000000000
048	0000000000
047	0000000000
046	0000000000
045	0000000000
044	0000000000
043	0000000000
042	0000000000
041	0000000000
040	0000000000
039	0000000000
038	0000000000
037	0000000000
036	0000000000
035	0000000000
034	0000000000
033	0000000000
032	0000000000
031	0000000000
030	0000000000
029	0000000000
028	0000000000
027	0000000000
026	0000000000
025	0000000000
024	0000000000
023	0000000000
022	0000000000
021	0000000000
020	0000000000
019	0000000000
018	0000000000
017	0000000000
016	0000000000
015	0000000000
014	0000000000
013	0000000000
012	0000000000
011	0000000000
010	0000000000
009	0000000000
008	0000000000
007	0000000000
006	0000000000
005	0000000000
004	0000000000
003	0000000000
002	0000000000
001	0000000000

USO DEL SUELO 1: ACTIVIDADES DE ESPARC. CASINATA CASALLTRIZAS

USO DEL SUELO 11: ESTRUCTURAS VIS

0000000000000000  
0000000011111111  
1234567890123456

USO DEL SUELO 111  
ESTACIONAMIENTOS  
CAMIONES  
ESTAC./ TRAILERS

NIVELES	1	2	3	7
SIMBOLOS	.....	.....	.....	.....
FRECUENCIA	123	11	40	205

TITULO DEL MAPA

MODELO DE IMPACTO N 3: CONTAMINACION DE LA SUPERFICIE DEL AGUA  
 AREA DE ESTUDIO VALLE DEL PACIFICO  
 LABORATORIO DE PLANIFICACION URBANA

OPCIONES USADAS PARA ESTE MAPA

10 EL TEXTO DEL MAPA ES  
 VARIABLES DATO USADAS EN EL MODELO

2 PROXIMIDAD AL AGUA  
 3 SUELOS POR TIPO  
 4 VEGETACION POR ZONA

LEGENDA: EL SIMBOLO MAS GRANDE CORRESPONDE AL IMPACTO MAS ALTO  
 LOS VALORES DEL IMPACTO PARA LOS 3 GRUPOS DEL USO DEL SUELO SON:

USO DEL SUELO	LOS TRES GRUPOS DE USOS COMERCIALES
	GRUPO DE US 111 MODERADO
	GRUPO DE US 111 MODERADO, GRUPO 11 MODERADO
	LOS TRES GRUPOS DE US MODERADO
	GRUPO DE US 111 SEVERO
	GRUPO DE US 111 SEVERO, GRUPO 11 SEVERO
	LOS TRES GRUPOS DE US SEVERO
	GRUPO DE US 111 TERMINAL
	GRUPO DE US 111 TERMINAL, GRUPO 11 TERMINAL
	LOS TRES GRUPOS DE US TERMINAL

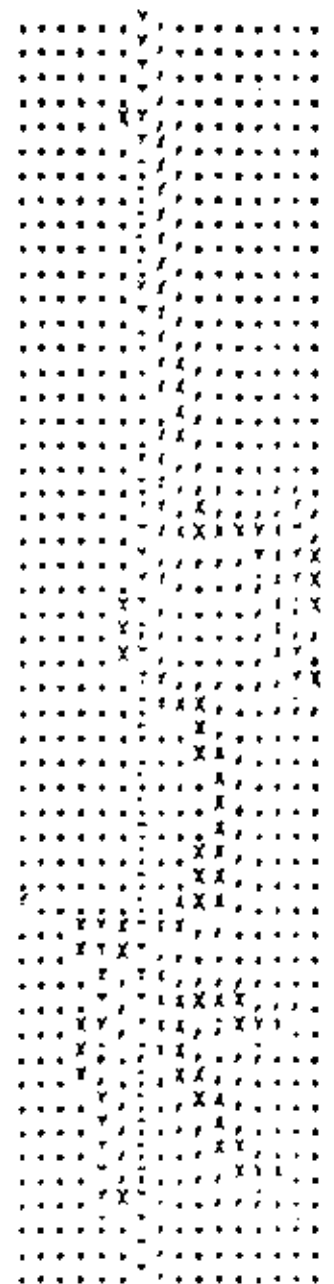
ACTIVIDADES DE ESPACIO ESTACIONAMIENTOS  
 CAMINATA ESTACIONAMIENTOS  
 CAMINOS  
 VIAL

CALLEJONES  
 ESTACIONAMIENTOS

MODELO DE IMPACTO N 3:

CONTAMINACION DE LA SUPERFICIE DEL AGUA

052  
051  
050  
049  
048  
047  
046  
045  
044  
043  
042  
041  
040  
039  
038  
037  
036  
035  
034  
033  
032  
031  
030  
029  
028  
027  
026  
025  
024  
023  
022  
021  
020  
019  
018  
017  
016  
015  
014  
013  
012  
011  
010  
009  
008  
007  
006  
005  
004  
003  
002  
001



AREA DE ESTUDIO VALLE DEL PACIFICO

LABORATORIO DE PLANEACION URBANA

VARIABLES DATO USADAS EN EL MODELO

- 1 PROXIMIDAD AL AGUA
- 2 SUELOS POR TIPO
- 3 VEGETACION POR ZONA

LEGENDA: EL SIMBOLO MAS OSCURO  
CORRESPONDE AL IMPACTO MAYOR

LOS VALORES DEL IMPACTO PARA LOS  
3 GRUPOS DEL USO DEL SUELO SON:

- LOS TRES GRUPOS DE US COMERCIALES
- GRUPO DE US 111 COMERCIO
- GRUPO DE US 111 COMERCIO, GRUPO 11 COMERCIO
- LOS TRES GRUPOS DE US RESIDENCIAL
- GRUPO DE US 111 SEVERO, GRUPO 11 SEVERO
- LOS TRES GRUPOS DE US SEVERO
- GRUPO DE US 111 TERMINAL
- GRUPO DE US 111 TERMINAL, GRUPO 11 TERMINAL
- LOS TRES GRUPOS DE US TERMINAL

ESTACIONAMIENTOS  
OSIPROCTURAS  
CASINOS  
MIS

CARRETERIZAS  
ESTIP. / TRAILERS

ACTIVIDADES DE ESPACIO  
CASINATA

100000 100000 100  
100000 100000 100  
1234567890123456

NIVELES	0	1	2	3
SIMBULOS	=====	=====	=====	=====
	AXYXXYXXYX	AXYXXYXXYX	AXYXXYXXYX	AXYXXYXXYX
	AXYXXYXXYX	AXYXXYXXYX	AXYXXYXXYX	AXYXXYXXYX
	AXYXXYXXYX	AXYXXYXXYX	AXYXXYXXYX	AXYXXYXXYX
	AXYXXYXXYX	AXYXXYXXYX	AXYXXYXXYX	AXYXXYXXYX
FRECUENCIA	581	111	6	131

TITULO DEL MAPA

MODELO DE IMPACTO N. 41: VULNERABILIDAD A RIESGOS  
AREA DE ESTUDIO: VALLE DEL PACIFICO  
LABORATORIO DE PLANEACION URBANA

OPCIONES USADAS PARA ESTE MAPA

- 10 EL TEXTO DEL MAPA ES: -----  
 VARIABLES DATO USADAS EN EL MODELO  
 2 VEGETACION POR ZONA  
 3 ORIENTACION  
 5 PORCENTAJE DE PENDIENTE DE 10 A 20%

LEGENDA: EL SIMBOLO MAS OSCURO CORRESPONDE AL IMPACTO MAS ALTO  
 LOS VALORES DEL IMPACTO PARA LOS 3 GRUPOS DEL USO DEL SUELO SON:

USO DEL SUELO

- LOS TRES GRUPOS DE USOS CONVENCIONALES
- GRUPO DE USOS III MODERADO
- GRUPO DE USOS III MODERADO, GRUPO II TRAZADO
- LOS TRES GRUPOS DE USOS CONVENCIONALES
- GRUPO DE USOS III SEVERO
- GRUPO DE USOS III SEVERO, GRUPO II SEVERO
- LOS TRES GRUPOS DE USOS CONVENCIONALES
- GRUPO DE USOS III TRAZADO
- GRUPO DE USOS III TRAZADO, GRUPO II TRAZADO
- LOS TRES GRUPOS DE USOS TRAZADO

USO DEL SUELO I  
ESTACIONAMIENTOS  
CASILLERIZAS  
CAMPOS

USO DEL SUELO II  
ACTIVIDADES DE ESPAZO  
ESTRUCTURAS  
VIA

USO DEL SUELO III  
ESPACIO / PARQUEO  
CAMINATA



AREA DE ESTUDIO VALLE DEL PACIFICO

MODELO DE IMPACTO N 4: VULNERABILIDAD A FUEGOS 0000000000000000  
 0000000001111111  
 1234567890123456

VARIABLES DATA USADAS EN EL MODELO

- 4 VEGETACION POR ZONA
- 3 ORIENTACION
- 5 PORCIENTO DENSIDAD DE ARBULES

LEGENDA: EL SIMBULO MAS OSCURO

CORRESPONDE AL IMPACTO MAYOR

LOS VALORES DEL IMPACTO LARGO LOS

3 GRUPOS DEL USO DEL SUELO SON:

- 0 LOS TRES GRUPOS DE US DUMMABLES
- 1 GRUPO DE US 111 MODERADO
- 2 GRUPO DE US 111 MODERADO, GRUPO 11 MODERADO
- 3 LOS TRES GRUPOS DE US MODERADO
- 4 GRUPO DE US 111 SEVERO
- 5 GRUPO DE US 111 SEVERO, GRUPO 11 SEVERO
- 6 LOS TRES GRUPOS DE US SEVERO
- 7 GRUPO DE US 111 TERMINAL
- 8 GRUPO DE US 111 TERMINAL, GRUPO 11 TERMINAL
- 9 LOS TRES GRUPOS DE US TERMINAL

USO DEL SUELO I

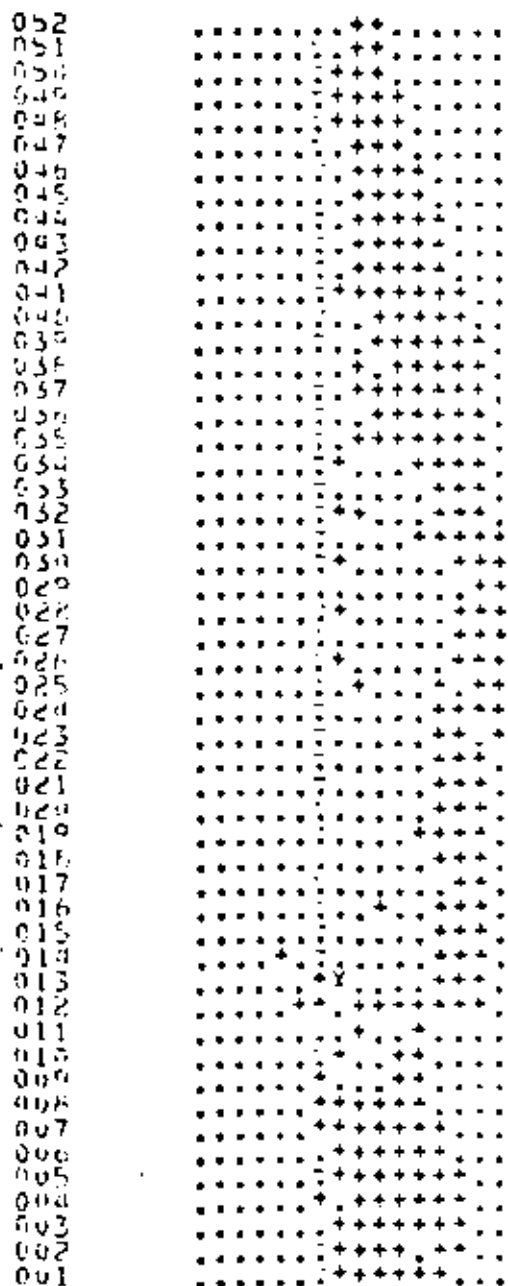
ESTACIONAMIENTOS  
 CANALIZACIONES  
 CAMINOS

USO DEL SUELO II

ACTIVIDADES DE ESPACIO  
 ESTRUCTURAS  
 PIS

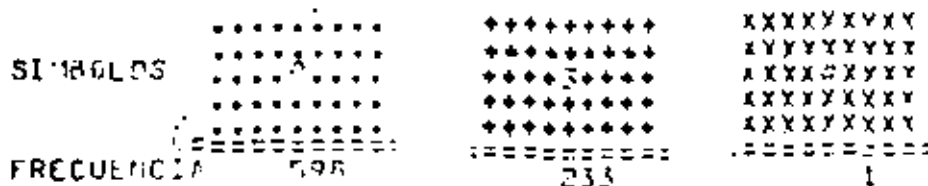
USO DEL SUELO III

ESTAC./ TRAILERS  
 CANALIZ



0000000000000000  
 0000000001111111  
 1234567890123456

NIVELES 0 3 4  
 =====



TITULO DEL MAPA

MODELO DE IMPACTO N 5: DESLIZAMIENTO DE TIERRA,  
 AREA DE ESTUDIO: VALLE DEL PACIFICO  
 LABORATORIO DE PLANEACION URBANA

OPCIONES USADAS PARA ESTE MAPA

10 EL TEXTO DEL MAPA ES  
 VARIABLES DATO USADAS EN EL MODELO,  
 7 ACCIDENTES GEOLOGICOS  
 2 POZOS DE AGUA  
 4 VEGETACION POR ZONA

LEGENDA: EL SIMBLO MAS URGENTE CORRESPONDE AL IMPACTO MAYOR  
 LOS VALORES DEL IMPACTO PARA LOS 3 GRUPOS DEL USO DEL SUELO SON:

0	LOS TRES GRUPOS DE US COMPATIBLES
1	GRUPO DE US 111 MODERADO
2	GRUPO DE US 111 MODERADO, GRUPO 11 MODERADO
3	LOS TRES GRUPOS DE US MODERADO
4	GRUPO DE US 111 SEVERO
5	GRUPO DE US 111 SEVERO, GRUPO 11 SEVERO
6	LOS TRES GRUPOS DE US SEVERO
7	GRUPO DE US 111 TERMINAL
8	GRUPO DE US 111 TERMINAL, GRUPO 11 TERMINAL
9	LOS TRES GRUPOS DE US TERMINAL

USO DEL SUELO :	USO DEL SUELO 11	USO DEL SUELO 111
ACTIVIDADES DE ESPARO, CAMINATA	CARALLERIZAS	ESTACIONAMIENTOS ESTRUCTURAS VIS CAMINOS ESTACOS/ TRAILERS

DESPLAZAMIENTOS DE TIERRA

AREA DE ESTUDIO ... VALLE DEL PACIFICO

LABORATORIO DE PLANEACION URBANA

VARIABLES DATO USADAS EN EL MODELO

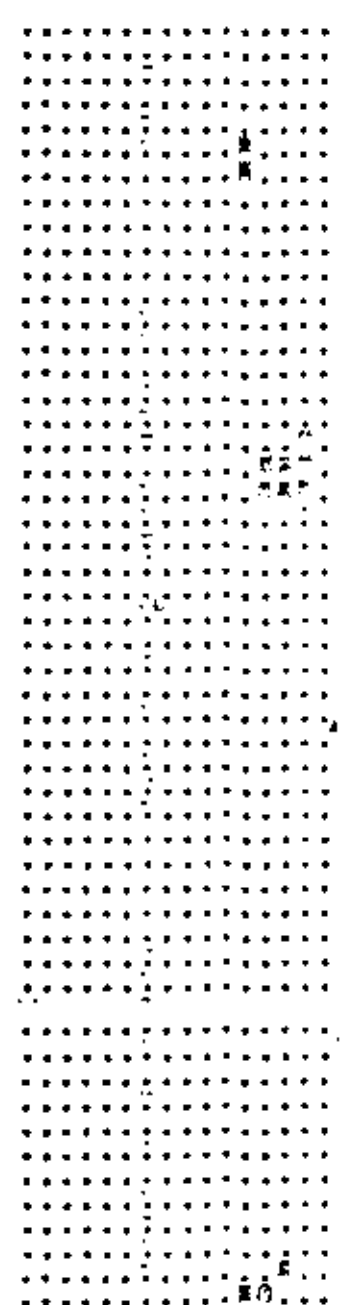
- 7 ACCIDENTES GEOLOGICOS
- 2 PORCIENTO DE PENDIENTE
- 4 VEGETACION POR ZONA

LEENDAS: EL SIMBOLO MAS OSCURO  
CORRESPONDE AL IMPACTO MAYOR

LOS VALORES DEL IMPACTO PARA LOS  
3 GRUPOS DEL USO DEL SUELO SON:

- LOS TRES GRUPOS DE US COMPATIVOS
- GRUPO DE US 111 MODERADO
- GRUPO DE US 111 MODERADO, GRUPO 11 MODERADO
- LOS TRES GRUPOS DE US MODERADO
- GRUPO DE US 111 SEVERO
- GRUPO DE US 111 SEVERO, GRUPO 11 SEVERO
- LOS TRES GRUPOS DE US SEVERO
- GRUPO DE US 111 TERMINAL
- GRUPO DE US 111 TERMINAL, GRUPO 11 TERMINAL
- LOS TRES GRUPOS DE US TERMINAL

052  
051  
050  
049  
048  
047  
046  
045  
044  
043  
042  
041  
040  
039  
038  
037  
036  
035  
034  
033  
032  
031  
030  
029  
028  
027  
026  
025  
024  
023  
022  
021  
020  
019  
018  
017  
016  
015  
014  
013  
012  
011  
010  
009  
008  
007  
006  
005  
004  
003  
002  
001



052  
051  
050  
049  
048  
047  
046  
045  
044  
043  
042  
041  
040  
039  
038  
037  
036  
035  
034  
033  
032  
031  
030  
029  
028  
027  
026  
025  
024  
023  
022  
021  
020  
019  
018  
017  
016  
015  
014  
013  
012  
011  
010  
009  
008  
007  
006  
005  
004  
003  
002  
001

USO DEL SUELO 1  
ACTIVIDADES DE ESPACIO  
CAMINATA

USO DEL SUELO 11  
CABALLETRIZAS

0000000000000000  
0000000011111111  
1234567890123456

USO DEL SUELO 111  
ESTACIONAMIENTOS  
ESTRUCTURAS  
VIA  
CANTONS  
ESTAC./ TRAILERS

NIVELES

0

7

9

SIMBOLOS

FRECUENCIA



TITULO DEL MAPA

MODELO DE IMPACTO N. 6: BASURA EN EL MAR  
 AREA DE ESTUDIO VALLE DEL PACIFICO  
 LABORATORIO DE PLANEACION URBANA

OPCIONES USADAS PARA ESTE MAPA

10 EL TEXTO DEL MAPA ES -----  
 VARIABLES DATO USADAS EN EL MODELO

4 VEGETACION POR ZONA  
 6 ZONAS LLANAS  
 8 PROXIMIDAD AL AGUA

LEGENDA: EL SIMBOLO MAS OSCURO CORRESPONDE AL IMPACTO MAYOR  
 LOS VALORES DEL IMPACTO PARA LOS 3 GRUPOS DEL USO DEL SUELO SON:

0	LOS TRES GRUPOS DE US COMPATIBLES
1	GRUPO DE US 111 MODERADO
2	GRUPO DE US 111 MODERADO, GRUPO 11 MODERADO
3	LOS TRES GRUPOS DE US MODERADO
4	GRUPO DE US 111 SEVERO
5	GRUPO DE US 111 SEVERO, GRUPO 11 SEVERO
6	LOS TRES GRUPOS DE US SEVERO
7	GRUPO DE US 111 TERMINAL
8	GRUPO DE US 111 TERMINAL, GRUPO 11 TERMINAL
9	LOS TRES GRUPOS DE US TERMINAL

USO DEL SUELO I

USO DEL SUELO II

USO DEL SUELO III

ACTIVIDADES DE ESPARC. ESTRUCTURAS  
 CANTINAS  
 CABALLERIZAS

ESTACIONAMIENTOS  
 VIS  
 CANTINAS  
 ESTIC./ TRAILERS

AREA DE ESTUDIO VALLE DEL PACIFICO

LABORATORIO DE PLANEACION URBANA

VARIABLES DATO USADAS EN EL MODELO

- 4 VEGETACION POR ZONA
- 4 ZONAS ALTIAS
- 9 PROXIMIDAD AL AGUA

~~EL USO DEL SUELO EN LAS ZONAS DE~~  
~~COMERCIO EN EL CENTRO URBANO~~

LOS VALORES DEL IMPACTO PARA LOS  
3 GRUPOS DEL USO DEL SUELO SON:

- 1 LOS TRES GRUPOS DE US COMPATIBLES
- 2 GRUPO DE US 111 RECREACION
- 3 LOS TRES GRUPOS DE US MODERADOS
- 4 GRUPO DE US 111 SEVERO
- 5 GRUPO DE US 111 SEVERO GRUPO 11 SEVERO
- 6 LOS TRES GRUPOS DE US SEVERO
- 7 GRUPO DE US 111 TERMINAL MODELO 11 TERMINAL
- 8 LOS TRES GRUPOS DE US TERMINAL

USO DEL SUELO 1

USO DEL SUELO 11

ACTIVIDADES DE ESPORO  
CAMINATA  
CAFETERIA

ESTRUCTURAS

USO DEL SUELO 111

ESTACIONAMIENTOS  
VIA  
ESTACION TRAILERS

NIVELES

SIGUIENTES

FRECUENCIA

052	.....	052
051	.....	051
050	.....	050
049	.....	049
048	.....	048
047	.....	047
046	.....	046
045	.....	045
044	.....	044
043	.....	043
042	.....	042
041	.....	041
040	.....	040
039	.....	039
038	.....	038
037	.....	037
036	.....	036
035	.....	035
034	.....	034
033	.....	033
032	.....	032
031	.....	031
030	.....	030
029	.....	029
028	.....	028
027	.....	027
026	.....	026
025	.....	025
024	.....	024
023	.....	023
022	.....	022
021	.....	021
020	.....	020
019	.....	019
018	.....	018
017	.....	017
016	.....	016
015	.....	015
014	.....	014
013	.....	013
012	.....	012
011	.....	011
010	.....	010
009	.....	009
008	.....	008
007	.....	007
006	.....	006
005	.....	005
004	.....	004
003	.....	003
002	.....	002
001	.....	001

10000000000000000000  
10000000000000000000

.....

.....

.....

## V. EVALUACION DE PLANES DE USO DEL SUELO

El objetivo de los planes es localizar los usos del suelo propuestos en los sitios de mayor atractivo y al mismo tiempo minimizar el impacto ambiental negativo en el área.

El resultado de la evaluación de cada plan se expresa en dos tablas sumario. Una con el atractivo total del plan y la otra con los impactos producidos en los sistemas considerados. Además, se obtienen un mapa de atractivo del plan y uno para cada impacto.

La evaluación de planes de uso de suelo se lleva a cabo mediante los siguientes pasos:

### 1. LOCALIZAR LOS USOS DEL SUELO

Cada uno de los usos del suelo considerados en el proyecto debe ser localizado en el área de estudio. Esta localización debe seguir los lineamientos de diferentes conceptos de desarrollo espacial desarrollados por los planificadores.

Cada uso del suelo debe tener un número para su identificación, este deberá ser el mismo que el utilizado en los modelos de atractivo y vulnerabilidad.

El número del uso del suelo localizado debe ser anotado en la celdilla correspondiente. Se permite solamente un tipo de uso del suelo en cada celdilla.

Deben ocuparse tantas celdillas como sea necesario para alojar toda el área requerida en el proyecto destinada a un uso del suelo específico.

La figura muestra un ejemplo de localización en un plan.

### 2. CODIFICAR LOS PLANES

El procedimiento es bastante similar al utilizado en la preparación del banco de datos. Se debe comenzar por la orilla superior izquierda del área de estudio y recorrer horizontalmente, celdilla por celdilla e hilera por hilera. El número del uso del suelo debe perforarse en la columna correspondiente, dejando en blanco aquellas en donde no se localice ningún uso. Las hileras de la retícula se registran mediante una o más tarjetas, dejando 2 columnas por cada celdilla ya que algunos

números de usos del suelo necesitan dos cifras.

Para cada uno de los planes que se deseen evaluar debe prepararse un paquete como el que se ilustra en la figura

### 3. CREAR ARCHIVOS DE PLANES

Cda uno de los planes es almacenado en un archivo en disco -- magnético mediante el programa PLANES/INGRID. Estos archivos serán utilizados posteriormente en la evaluación de los planes.

En el sistema UNAM/B-6700, la información codificada y perforada en el paso 2, debe ir precedida por tres tarjetas de control, el contorno del área de estudio (ver capítulo II, paso 3), y el número de hileras y columnas de la retícula. Después de las tarjetas del último plan se debe colocar una tarjeta de control final para la terminación del trabajo. La forma de preparar el paquete de datos esta indicada en la figura

### 4. EVALUACION DE LOS PLANES EN LA COMPUTADORA

Durante la fase de evaluación, la computadora suma todos los atractivos que corresponden a las celdillas designadas para localizar un uso del suelo particular. La suma total se divide entre el número total de celdillas distribuidas para ese uso del suelo. El resultado es un indicador de atractivo promedio. Este es multiplicado por 10 para ponerlo en una escala de 1 a 100.

Un marcador final alto indica que el uso del suelo fue puesto en áreas que cumplen muchos de los criterios descados para ese uso del suelo. Un marcados bajo indica una pobre selección en la localización de los usos del suelo con respecto al atractivo de las celcillas.

Todos los usos del suelo que se van a incluir en el área de estudio son evaluados en forma similar y reciben un marcador final de atractivo promedio.

La segunda parte en la evaluación de un plan consiste en dar una estimación del grado de impacto negativo creado por el plan de uso del suelo en su totalidad, sobre cada uno de los sistemas previamente definidos por los modelos de vulnerabilidad.

Cuando la computadora encuentra una celdilla a la que se ha asignado un uso del suelo, determina el grado de impacto que

tendrás ese uso del suelo sobre el sistema que está siendo analizado. El grado de impacto dependerá de la combinación de categorías encontrada en la celdilla y del grupo de usos del suelo al que pertenezca.

El marcador final de cada modelo indica el grado de impacto - promedio creado por la totalidad del plan sobre el sistema -- considerado en toda el área de estudio. La media numérica es una media ponderada. El número de celdillas que reciben un -- indicador de impacto 1, se multiplica por 1. El número de -- celdillas que reciben un indicador de impacto 2 (impacto moderado) se multiplica por 2, etc. El total final de los cuatro impactos se divide entre el número total de celdillas donde -- se localizan los nuevos usos del suelo. La media resultante -- estará entre un máximo de 4 y un mínimo de 1. Un marcador -- cercado a 4 indican que los usos del suelo han sido puestos -- donde crean una gran cantidad de impacto negativo. Un marcador cercado a 1 indica poco impacto negativo en el sistema.

Para la evaluación de los planes deben perforarse las siguientes tarjetas para alimentar el programa EVALUA/IMGRID.

1a. tarjeta:

Col. 1-5      Número de hileras de la retícula.\*  
Col. 6-10     Número de columnas de la retícula.\*

2a. tarjeta:

Col. 4-5      Número de modelos de atractividad realizados en el estudio.\*  
Col. 9-10     Número de modelos de impacto realizados en el -- estudio.\*  
Col. 14-15    Número del plan que se desea evaluar.\*

3a. tarjeta:

Col. 2, 4, 6 Agrupación de los usos del suelo en el primer --  
8, 10, modelo de impacto (paso 1). Debe anotarse el --  
12, 14, número de grupo al que pertenece cada uno de --  
16 los usos considerados en el estudio dados en --  
orden creciente.\*

Tarjetas siguientes:

Lo mismo que la 3a. tarjeta.  
Una para cada modelo de vulnerabilidad realizado en el estudio.  
Deben darse en el mismo orden en que se corrieron los modelos.

En el sistema UNAM/B-6700, esta información debe ir precedida por tres tarjetas de control y por el contorno del área --

\*Deben perforarse como números enteros justificados a la derecha.



de estudio (Ver capítulo II, paso 3).

Después de las tarjetas para evaluar el último plan, debe colocarse una tarjeta de control final para la terminación del trabajo.

La forma de preparar el paquete de datos se ilustra en la figura

## 5. MAPEO DE LOS RESULTADOS DE LA EVALUACION

Se debe producir un mapa del atractivo del plan y uno de cada impacto considerado.

El mapa de atractivo indica el grado de atraktividad logrado - al haber localizado los usos del suelo en los sitios que se pensaron convenientes.

Se obtendrá un mapa impreso para cada sistema descrito por un modelo de vulnerabilidad. Cada mapa representa un sistema diferente e indica cuantas celdillas del total distribuidas tienen impacto compatible, moderado, severo ó terminal.

La impresión del mapa estará basada en un rango numérico de 1 a 4. El número 1 indica que el uso del suelo es compatible; un número 4 indica que el uso del suelo tiene un impacto terminal sobre el área.

La forma de especificar los mapas con el programa MAPAS/INGRID se explica en el capítulo VI.

Para producir los mapas de evaluación debe prepararse un paquete como el de la figura.

## 6. ANALISIS DE RESULTADOS Y GENERACION DE NUEVOS PLANES

Con esta información a mano y las áreas problema localizadas, se puede preparar un segundo plan.

La preparación de este segundo plan implica la relocalización de los usos del suelo que crean impactos mayores ó que reciben bajos indicadores de atractivo.

El objetivo ginal es elevar la atractivo y disminuir los impactos. Se debe lograr un equilibrio en el que la atraktividad total no puede ser elevada sin elevar también los impactos negativos.

Deben identificarse las áreas problema cuyos indicadores de --

atractividad son mas bajos que los esperados ó sus indicadores de impacto son mayores que los deseables.



TÍTULO DEL MAPA

UBICACIÓN DEL IMPACTO DEL PLAN 3 EN LA VULNERABILIDAD A FUEGOS

DE ESTUDIO VALLE DEL PACÍFICO

ÁREA DE INFORMACIÓN

OTROS USOS PARA ESTE MAPA

EL TÍTULO DEL MAPA ES

LOS USOS DEL SUELO SE AGRUPAN EN LOS GRUPOS:

ESTACIONAMIENTOS  
CAMPOS

ACTIVIDADES DE ESPARO  
VIA

ESTACIONAMIENTOS

TÍTULO DEL MAPA

UBICACIÓN DEL IMPACTO DEL PLAN 3 EN LA DESLIZAMIENTOS DE TIERRA

DE ESTUDIO VALLE DEL PACÍFICO

ÁREA DE INFORMACIÓN

OTROS USOS PARA ESTE MAPA

EL TÍTULO DEL MAPA ES

LOS USOS DEL SUELO SE AGRUPAN EN LOS GRUPOS:

ACTIVIDADES DE ESPARO  
CAMALLERIZAS

ESTACIONAMIENTOS

TÍTULO DEL MAPA

UBICACIÓN DEL IMPACTO DEL PLAN 3 EN LA BASURA EN EL MAR

DE ESTUDIO VALLE DEL PACÍFICO

ÁREA DE INFORMACIÓN

OTROS USOS PARA ESTE MAPA

EL TÍTULO DEL MAPA ES

LOS USOS DEL SUELO SE AGRUPAN EN LOS GRUPOS:

ACTIVIDADES DE ESPARO  
CAMALLERIZAS

ESTACIONAMIENTOS  
CAMPOS  
ESTACIONAMIENTOS

4

COMPUTACION

APLICADA

A LA

PLANEACION

URBANA

DECFI, AGOSTO 10-14, 1981.

- 1.- Descripción del formato MPS
- 2.- Ejemplos de :
  - 2.1.- Programación lineal . . . . .
  - 2.2.- Programación entera . . . . .
  - 2.3.- Programación mixta . . . . .
- 3.- Descripción del formato SPSS
- 4.- Ejemplo de uso del paquete SPSS  
a través de problema de regresión .

PADILLA MIRANDA JUAN M.

S E C C I O N      I I

En esta sección se especifica el formato utilizado - por el paquete de programación matemática TEMPO/MPS/ALL , el - cual para desarrollar la solución de un problema , utiliza dos archivos de datos necesariamente .

1/ CCARD .- En este archivo están contenidas las instrucciones sobre el problema .

2/ CARDIN .- En este archivo están contenidos los datos del problema , los cuales se dividen en 5 secciones ROWS , COLUMNS , RHS , RANGES , BOUNDS .

Cada tarjeta de datos consiste de 80 columnas y esta dividida en 6 campos , un \* en la columna 1 indica se trata de una tarjeta de comentarios .

FORMATO DE LOS DATOS DE ENTRADA

(Formato MPS)

CAMPO		1		2		3		4		5		6	
1	2	3		5 - 12		15 - 22		25 - 36		40 - 47		50 - 61	
N A M E						DATOS							
*				Tarjeta de comentarios									
R O W S													
N E G L				NOM. RES.									
C O L U M N S				NOM. COL.		NOM. RES. 1		COEF. 1		NOM. RES. 2		COEF. 2	
R H S				NOM. RHS		NOM. RES. 1		COEF. 1		NOM. RES. 2		COEF. 2	
R A N G E S				NOM. RANGO		NOM. RES. 1		COEF. 1		NOM. RES. 2		COEF. 2	
B O U N D S				NOM. R. COT.		NOM. COL.		COEF.		NO UTILIZADOS			
E N D A T A													

Organización de los datos de entrada del problema de programación matemática .

1/ Una tarjeta de NAME sera la primer tarjeta en los datos y una de ENDATA al final , ya que con esta se indica fin de -- archivo de datos , la tarjeta de NAME también tiene en el campo numero 3 un nombre especificado por el usuario . Las secciones de ROWS , COLUMNS , y RHS son necesarias , y las secciones -- RANGES , BOUNDS serán opcionales y no es necesario declararlas o especificarlas .

2/ La sección de ROWS es en la que se define el tipo de restricción de cada renglón , y se utilizan 4 indicadores para especificar el tipo de restricción en el modelo , seguido por algun nombre especificado por el usuario en el campo dos,

- a/ N Renglón de la función objetivo .
- b/ E Igualdad en la restricción .
- c/ G Mayor o igual en la restricción .
- d/ L Menor o igual en la restricción .

3/ La sección de COLUMNS , es en la que se especifican los coeficientes de las restricciones y los nombres de cada variable , el nombre de la columna es dado en el campo dos , y los nombres de los renglones (restricciones) estarán especificados en los campos 3 y 5 , en los campos 4 y 6 se declaran los --- coeficientes correspondientes , [El sistema trata como cero a los coeficientes omitidos en el modelo , es decir solo es --- necesario declarar los coeficientes que sean diferentes de -- cero.

4/ La sección de RHS , es donde se especifica el vector de -- recursos ( o vectores ) empezando en el campo dos , el formato es el mismo conque fue definido columns , aunque un nombre para el vector de recursos debiera ser declarado y esto para cada -- vector de recursos . con que se desea resolver el modelo .

5/ La seccion de RANGES , es utilizada para condensar los datos de entrada .

Esto se hace cuando una restricción es mayor o igual y menor o igual , El renglón original necesita ser especificado en la sección de rows y columns solamente con uno de sus límites superior o inferior  $b(i)$  especificado en la sección de rhs.

EL rango es utilizado cómo se especifica en la siguiente tabla , donde  $r(i)$  es el rango sobre el renglón  $i$  dado en la sección de rangos , si  $r(i)$  es negativo para un renglón  $G$  o un renglón  $L$  , se utiliza el valor absoluto.

Tipo de renglón	Signo de $b(i)$	Límite superior que resulta en la rest.	Límite inferior que resulta en la rest.
G	+	$b(i) + r(i)$	$b(i)$
L	+	$b(i)$	$b(i) - r(i)$
E	+	$b(i) + r(i)$	$b(i)$
E	-	$b(i)$	$b(i) - r(i)$

6/ La sección de BOUNDS , es en donde se especifica que variables estarán acotadas , si esta sección no se especifica se considera que todas las variables son mayor o igual que cero , se utilizan 6 indicadores para especificar el tipo de cota que será impuesto sobre la actividad .

- a/ LO Cota inferior.
- b/ UP Cota superior.
- c/ FX Valor fijo.
- d/ MI Cota inferior es - infinito.
- e/ PL Cota superior es + infinito.
- f/ FR Variable libre ( - infinito a + infinito )

En los campos 3 y 4 se especifica el nombre de la variable que será acotada y su cota finita asociada-respectivamente , dejar el campo 4 en blanco si la cota es infinito. El campo 2 identifica un nombre asociado con un conjunto de cotas esto es , puede haber más de un conjunto de cotas declaradas en la sección de bounds , lo cual es similar a la opción de definir varios vectores de recursos en la sección rhs .



Dentro de las secciones de ROWS y COLUMNS las opciones 'MARKER' son provistas con el siguiente formato.

- i/ El campo 1 permanece en blanco excepto cuando es utilizado para datos GUB .
- ii/ El campo 2 identifica el nombre de el marker , el cual - debe ser distinto de los nombres de rows y columns .
- iii/ El campo 3 contiene la palabra 'MARKER' (incluyendo los apostrofos ) .
- iv/ El campo 4 permanece en blanco .
- v/ El campo 5 contiene las palabras claves :
  - a/ 'INTORG' indica que las variables que están en las tarjetas que siguen , son enteras hasta donde se -- encuentre 'INTEND' .
  - b/ 'BIVORG' indica que las variables que están en las tarjetas que siguen , son binarias (solo toman el - valor cero o uno ) hasta donde se encuentre 'BIVEND'.
  - c/ 'SEPORG' y 'SEPEND' son reservadas para programación separable.
  - d/ 'GUBORG' y 'GUBEND' se reservan para definir renglon es tipo GUB.

Las palabras incluyen los apostrofes , los nombres asignados a marker , renglones , variables , terminos del vector de recursos , cotas , y rangos son asignados por el usuario y son a lo más de ocho caracteres .

- vi/ El campo 6 permanece en blanco .

La forma de ejecutar el programa por tarjetas es :

```

?JOB LINEAL;USER=SFB2/HP;CLASS=5;BEGIN
?RUN*SERVICIO/MP5/ALL;DATA CARD
      .
      Instrucciones      } Tarjetas de
      .                  } procedimiento
      .
?DATA CARDIN
      .
      Tarjetas de datos
      .
?END JOB
  
```

Explicación de las tarjetas de procedimiento (Archivo CARD) :

ZDATA="DATOS"

Esta instrucción asigna al conjunto de datos el nombre "DATOS" , el cual debe ser a lo más de 8 caracteres .ZDATA debe aparecer con el mismo nombre con el que se identifica el conjunto de datos en el archivo CARDIN (Tarjeta de NAME) .

ZNAME="LINEAL"

Esta instrucción asigna el nombre del problema a -- ZNAME .

ZOBJ="FD"

Esta instrucción asigna el nombre de la función --- objetivo .

ZRHS="Ii"

Esta instrucción asigna el nombre del vector de --- recursos del lado derecho .

ZRNGST="RANGO"

Esta instrucción asigna el nombre para el conjunto de restricciones con rangos .

ZBNDST="COTA"

Esta instrucción asigna el nombre para el conjunto de variables acotadas , es opcional y si no hay variables acotadas no es necesario declararla lo mismo que ZRNGST en caso de que - no haya restricciones con rangos , los nombres asignados a las instrucciones que empiezan con Z , que aparecen entre comillas deberán ser a lo más de 8 caracteres .

INPUT(CARD,SUMMARY)

Este procedimiento es utilizado para leer los datos del problema y los convierte a un formato binario en el archivo ZPROF , SUMMARY ocasiona que las estadísticas de los reng - lones y columnas se imprimán .

BCDDUT

Imprime el problema en el formato de los datos de -- entrada .

SETUP(MAX)

Genera la matriz de trabajo para el problema identificado por ZNAME, MAX indica se trata de optimizar maximizando en caso de querer minimizar reemplazar MAX por MIN .

PRIMAL

Este procedimiento obtiene la solución óptima si -- existe .

PICTURE

Este procedimiento imprime una grafica de la matriz original en la que los coeficientes que son cero permanecen en blanco y en el lugar en que hay un coeficiente distinto de cero aparece un caracter alfabetico .

RANGE

Este procedimiento produce un analisis de sensibilidad de la solución óptima .

OUTPUT

Este procedimiento imprime la solución incluyendo - niveles de actividad ; costos reducidos , coeficientes de los costos originales , elementos del vector de recursos , y ----- actividades duales .

La impresión es listada cómo sigue :

- 1/ Nombre del problema .
- 2/ El nombre del conjunto de rangos , si hay alguno.
- 3/ El nombre del conjunto de cotas , si lo hay .
- 4/ La función .
- 5/ La restricción .
- 6/ El estado de la solución :
  - a/ Infactible .
  - b/ No-óptimo
  - c/ Óptimo .
- 7/ Número de iteración .
- 8/ Valor de la función .

Las secciones de ROWS y COLUMNS son similares en -- estructura y contenido .

Una línea de salida es impresa para cada variable - con la siguiente información impresa .

1/ NUMBER .- El número entero asignado a la variable . A la - primer columna es asignado el número  $(n+1)$  , donde  $n$  es igual al número del último renglón , a la segunda columna es asignado el número  $(n+2)$  , y así sucesivamente , este número puede ser utilizado para identificar la variable en la iteración corrien te .

2/ NAME.- El nombre de entrada del renglón o la columna .

3/ STATUS.- Dos caracteres indicarán el estado de solución en el renglón o columna :

a/ BS En la base y factible .

b/ \*\* En la base y infactible .

c/ FR No-básica , libre .

d/ EQ No-básica , artificial o fija .

e/ UL No-básica , actividad en límite superior .

f/ LL No-básica , actividad en el límite inferior .

g/ IV No-básica , variable entera .

4/ ACTIVITY.- El valor de actividad del renglón o columna en la solución .

5/ SLACK ACTIVITY/INPUT COST.- Es el costo de entrada .

6/ LOWER LIMIT.- El más pequeño valor factible que la actividad puede tomar .

7/ UPPER LIMIT.- El mayor valor factible que la actividad puede tomar .

8/ DUAL ACTIVITY/REDUCED COST.- En la sección ROWS representa la actividad dual , esto es también el costo reducido de la -- variable lógica asociada . En la sección de COLUMNS es el costo reducido de la columna , los costos reducidos son conocidos e-- cómo los  $d_j$ 's . Las actividades duales son conocidas cómo los valores  $p_i$  o multiplicadores simplex . El costo reducido de una variable es la tasa de incremento en la función objetivo por - unidad de incremento en la actividad de la variable .

La presencia de una variable no-básica con costos reducidos de cero en la solución óptima indica que la solución alternativa óptima existe. Tales variables son identificadas por una A sobre el lado izquierdo de la tabulación.

Para algun renglón compuesto o columna compuesta, - los valores impresos para el renglón base o la columna base son los valores compuestos. Si la función objetivo es compuesta - las actividades duales, costos de entrada y costos reducidos son para la función objetivo compuesta. Si el lado derecho es compuesto, los límites inferior y superior en la sección de - ROWS son para el lado derecho compuesto.

Cero o cantidades muy cercanas a cero serán impresas como un punto decimal.

#### TRANCOL

Este procedimiento imprime el última tabla por columnas.

#### TRANROW

Este procedimiento imprime la tabla actualizada por renglones.

#### TRANCOL (INVERSE)

Este procedimiento imprime la inversa.

#### TRANCOL (MATRIX)

La matriz original es impresa.

## S E C C I O N      I I . 1

a/ Considere el siguiente problema , el cual es un problema de programación lineal .

$$\text{Min } Z = -20x_1 - 10x_2 - x_3$$

Sujeto a.

$$3x_1 + 2x_2 + 10x_3 \leq 10$$

$$2x_1 + 4x_2 + 20x_3 \leq 15$$

$$x_1 \geq 0; \quad x_2 \geq 0; \quad x_3 \geq 0$$

Se resolvera este problema usando el paquete TEMPO/MPS/ALL via tarjetas (Batch) . En la hoja de codificación anexa se muestra :

i/ Tarjetas de control del sistema.

ii/ Tarjetas de instrucciones sobre el problema .

iii/ Tarjetas de datos (Formato MPS) . .

Esto es en el formato especificado en la sección anterior , a continuación en las hojas siguientes se da la --- secuencia utilizada para resolver el problema por terminal .

Zprinter - .true. es utilizado para que los resultados aparezcan tanto en la terminal como en la impresora que se localiza en el C.S.C. en caso de omitirla se obtendrán los --- resultados en la terminal solamente .

En la hoja se muestra la secuencia seguida con los -- comandos del paquete tempo y posteriormente la forma en que se interactua con la computadora a través de la terminal .

La ventaja de utilizar una terminal es que la información se transmite en formato libre y no en el formato ---- convencional mencionado en la sección anterior .



SFB2/MP	(R)
RUN*SERVICIO/MPS/ALL	(R)
ZDATA = "DATOS"	(R)
ZNAME = "LINEAL"	(R)
ZRHS = "T1"	(R)
ZOBJ = "OF"	(R)
INPUT(REMOTE, SUMMARY)	(R)
ROWS	(R)
N OF	(R)
L R1	(R)
L R2	(R)
COLUMNS	(R)
X1 OF -20	(R)
* R1 3	(R)
* R2 2	(R)
X2 OF -10	(R)
* R1 2	(R)
* R2 4	(R)
X3 OF -1	(R)
* R1 10	(R)
* R2 20	(R)
ENDATA	(R)
RHS	(R)
T1 R1 10	(R)
* R2 15	(R)
ENDATA	(R)
SETUP(MIN)	(R)
BCDOUT	(R)
PRIMAL	(R)
OUTPUT	(R)
EXIT	(R)

Nota.- (R) indica teclear return .



#B0700126 CANDE 30.1407 YOU ARE SCHEDPCC1(99)  
#DEFAULT PRINT DESTINATION=SITE  
#SESSION 3028 17:32:20 01/14/81  
#UNSERVICED/MPS/ALL  
#PLANNING 3029

#?  
#7700/#0700 TEMP0(28.600,000 = PARC1 71) 01/14/81 17:33:26  
READY

#PRINTER = .TRUE.

READY  
#DATA = "LATS"

READY  
#NAME = "FCLIND"

READY  
#MPS = "TEMP1"

READY  
#OBJ = "LPT"

READY  
INFLY(REFLECT,SUMMARY)

--- INFLY ---

#  
#MPS  
#  
# CF

L R1  
A  
L R2  
A  
COLLAPS  
A  
A1 CF 20

A R1 3  
A  
A R2 2  
A

A2 CF 10  
A  
A R1 2  
A  
A R2 4  
A

A3 CF 1  
A  
A R1 10  
A  
A R2 20  
A

MMS  
A  
IERMI R1 10  
A  
A R2 15  
A

ENDATA

OLD ENTI CLINE DELETED ON ZPROF (OR ZSELF)

NEW ENTRY MLLINE ENTERED ON ZPROF (OR ZSELF)

NUMBER OF ELEMENTS BY COLUMN ORDER

49 X1 .....3 X2 .....3 X3 .....3

NUMBER OF ELEMENTS BY ROW ORDER, EXCLUDING RHS, INCLUDING SLACK ELEMENT

1 A CF .....4 L R1 .....4 L R2 .....4

READY

SETUP(MIN)

--- SETUP ---

READY

PRIMAL

--- PRIMAL ---

--- INVERT ---

FACELEN FEASIBLE.

ITER NO	OBJ. VALLE	NEG DV	LLT	IN
1	-06.66667	3	2	49

EXIT CONDITION: OPTIMAL SOLUTION.

FUNCTION VALUE \* 65.06607

READY  
LUTFLT

\*\*\* OUTPUT \*\*\*

SECTION STATUS \* OPTIMAL  
ROWS SECTION

NAME	STAT	ACTIVITY	LOWER LIM	UPPER LIM	DUAL ACTIVITY
R1	BS	50.000	NONE	NONE	1.000
R2	UL	10.000	NONE	10.000	0.000
R3	BS	0.000	NONE	15.000	.

COLUMNS SECTION

NAME	STAT	ACTIVITY	LOWER LIM	UPPER LIM	REDUCED COST
X1	BS	3.333	.	NONE	.
X2	LL	.	.	NONE	3.333
X3	LL	.	.	NONE	65.000

READY

EXIT

07700/06700 TEMPC TERMINATED.

\*ET=4324.4 FT=7.4 LC=5.0

Para el problema considerado se obtuvieron los siguientes resultados :

a/ Solución óptima factible , la cual se obtuvo en la primera iteración y resulto igual a  $-66.66667$  , esto es que

$$\text{Min } Z = -20(3.33) - 10(0) - 1(0) = -66.66$$

b/ Se obtuvo como resultado  $x_1 = 3.333$  , con  $x_1$  variable básica factible y  $x_2 = x_3 = 0$  y son no-básicas con actividad en el límite inferior .

En la restricción R1 se obtuvo

$$3(3.333) + 2(0) + 10(0) = 10 \leq 10$$

y en la restricción R2 obtuvo

$$2(3.333) + 4(0) + 20(0) = 6.666 \leq 15 \quad (\&)$$

c/ Termina la ejecución .

S E C C I O N      11.1

b/ Considere el siguiente problema , el cual es un problema de programación entera .

$$\text{Max } Z = 10x_1 + x_2 + 12x_4 + 7x_5 + 8x_6 + 6x_7 + 3x_8$$

Sujeto a.

$$6x_1 + .8x_2 + x_3 + 10x_4 + 5x_5 + 6x_6 + 5x_7 + 3x_8 \leq 25 \quad . \quad 30$$

$$x_1, x_3, x_4, x_5, x_6, x_7, x_8 = 0,1$$

$$x_2 \in \mathbb{Z}^+$$

Se resolvera este problema a través de una terminal - utilizando el paquete TEMPO/MP5/ALL , note que en el problema las variables son binarias a excepción de  $x_2$  que es entera , - este problema es conocido como el problema de la mochila , en la restricción se considera primero acotada superiormente por 25 y se resuelve el problema , posteriormente se resuelve el - problema con cota superior en la restricción de 30 , las ---- instrucciones TEMPO utilizadas y su secuencia se muestra en la siguiente hoja .

RUN*SERVICIO/MPS/ALL	(R)
ZPRINTER=.TRUE.	(R)
ZNAME="ENTERA"	(R)
ZDATA="DATOS"	(R)
ZOBJ="OF"	(R)
ZRHS="TI"	(R)
INPUT(REMOTE, SUMMARY)	(R)
ROWS	(R)
N OF	(R)
L R	(R)
COLUMNS	(R)
EMPICZA 'MARKER' 'BIVORG'	(R)
X1 OF 101	(R)
X1 R 6	(R)
X3 R 1	(R)
X4 OF 12	(R)
X4 R 10	(R)
X5 OF 7	(R)
X5 R 5	(R)
X6 OF 8	(R)
X6 R 6	(R)
X7 OF 6	(R)
X7 R 5	(R)
X8 OF 3	(R)
X8 R 3	(R)
TERMINA 'MARKER' 'BIVEND'	(R)
EMP 'MARKER' 'INTORG'	(R)
X2 OF 1	(R)
X2 R .8	(R)
TERM 'MARKER' 'INTEND'	(R)
RHS	(R)
TI R 25	(R)
TI2 R 30	(R)
ENDATA	(R)
SETUP(MAX)	(R)

PRIMAL	(R)
MXINT	(R)
OUTPUT	(R)
ZRH5="T12"	(R)
MXINT	(R)
RETURN	(R)
OUTPUT	(R)
EXIT	(R)



RB6700=126 CANDE 30.14) YOU ARE SCHEDULED(102)  
R DEFAULT PRINT DESTINATION=SITE  
SEMINARIUS PARA MAYO/JUNIO TACLEE NEWS  
SESSION 3343 15:13:14 05/06/81  
RUN\*SERVICIO/MPS/ALL  
NRUNING 3344  
R?  
87700/86700 TEMPO(23.6)C(00) - MARCH 77) 05/06/81 15:13:19  
READY  
ZPRINTER=.TRUE.  
READY  
ZNAME="ENTERA"  
READY  
ZDATA="DATOS"  
READY  
ZOBJ="OF"  
READY  
ZRHS="TI"  
READY  
INPUT(REMOTE,SUMMARY)

--- INPUT ---

Z  
ROWS  
X  
N OF

L R

Z

COLUMNS

Z

EMPIEZA 'MARKER' 'BIVOPG'

Z

X1 OF 101

Z

X1 R 6

Z

X3 R 1

Z

X4 OF 12

Z

X4 R 10

Z

X5 OF 7

Z

X5 R 5

Z

X6 OF 8

Z

X6 R 6

Z

X7 OF 6

Z

X7 R 5

Z

X8 OF 3

Z

X8 R 3

Z

EMPIEZA 'MARKER' 'BIVOPG'

MP 'MARKER' 'INTORG'

2 OF 1

12 R 20

ERM 'MARKER' 'INTEND'

RHS

1 R 25

12 R 30

DATA

NEW ENTRY ENTERED ENTERED ON ZPROF (OR ZSOLF)

NUMBER OF ELEMENTS BY COLUMN ORDER

49	X1	.....2	X3	.....1	X4	.....2
52	X5	.....2	X6	.....2	X7	.....2
55	X8	.....2	X2	.....2		

NUMBER OF ELEMENTS BY ROW ORDER, EXCLUDING RHS, INCLUDING SLACK ELEMENT

1 N OF .....8 L R .....9

READY

SETUP(MAX)

--- SETUP ---

ERPUP \* INTEGER X2 BOUNDS RESET TO: UPPER= 1022 LOWER= 0

\*\*\* CONTINUE DEMING SEE \*\*\*

READY  
PRIMAL

--- PRIMAL ---

--- INVERT ---

PROBLEM FEASIBLE.

ITER NO	OBJ VALUE	NEG DJ	OUT	IN
1	101.00000	7	490	490
2	113.00000	7	510	510
3	121.00000	7	530	530
4	125.20000	7	2	52
5	125.60000	1	520	510
6	126.00000	1	51	56

EXIT CONDITION: OPTIMAL SOLUTION.

FUNCTION VALUE = 126.00000

READY  
MXINT

--- MXINT ---

\* X I N T B R A N C H & B O U N D  
L P S O L U T I O N I S I N T E G R A L

READY  
OUTPUT

SOLUTION 5 DS = OPTIMAL  
ROWS SECTION

NAME	STAT	ACTIVITY	LOWER LIM	UPPER LIM	DUAL ACTIVITY
OF	BS	126.000	NONE	NONE	1.000
R	UL	25.000	NONE	25.000	-1.250

COLUMNS SECTION

NAME	STAT	ACTIVITY	LOWER LIM	UPPER LIM	REDUCED COST
X1	UL	1.000	.	1.000	93.500
X3	LL	.	.	1.000	-1.250
X4	LL	.	.	1.000	-0.500
X5	UL	1.000	.	1.000	0.750
X6	UL	1.000	.	1.000	0.500
X7	LL	.	.	1.000	-0.250
X8	LL	.	.	1.000	-0.750
X2	BS	10.000	.	1022.750	.

READY

ZRHS="TJZ"

READY

MXINT

--- MXINT ---

MXINT BRANCH & BOUND

INTEGER SOLUTION

NODE \*\* 12 \*\* OPTIMIZED INTEGER NODE  
FIRST INTEGER SOLUTION REJECTIVE = 129.0000  
CUTOFF NOW AT 129.000 POSTPONE AT 129.000

ACTIVE INTEGER VARIABLES

Para el problema considerado en esta sección se obtuvieron los siguientes resultados .

a/ Los resultados obtenidos cuando la restricción está acotada superiormente por 25 son los siguientes :

$x_2 = 10$  , y se trate de una variable básica factible ;

$x_1 = x_5 = x_6 = 1$  son variables no-básicas con actividad en el límite superior ;

$x_3 = x_4 = x_7 = x_8 = 0$  son variables no-básicas con actividad en el límite inferior .

La solución es óptima y el valor de la función objetivo es de 126.00 ya que

$$\text{Max } Z = 10(1) + 1(10) + 7(1) + 8(1) = 126$$

y en la restricción

$$6(1) + .8(10) + 5(1) + 6(1) = 25 \leq 25$$

b/ Los resultados obtenidos cuando la restricción está acotada superiormente por 30 son los siguientes :

$x_1 = x_2 = x_4 = x_5 = x_7 = x_8 = 1$  son variables enteras ;

$x_3 = x_6 = 0$  son variables enteras también ;

El problema tiene solución óptima y es cuando la función objetivo es igual a 130

$$\text{Max } Z = 10(1) + 1(1) + 12(1) + 7(1) + 6(1) + 3(1) = 130$$

y en la restricción se obtiene a través del paquete

$$6(1) + .8(1) + 10(1) + 5(1) + 5(1) + 3(1) = 29.8 \leq 30 ; (\&)$$

(\&) ver páginas 60,61 .

PROCEDURE.

READY  
OUTPUT

--- OUTPUT ---

SOLUTION STATUS = OPTIMAL  
ROWS SECTION

NAME	STAT	ACTIVITY	LOWER LIM	UPPER LIM	DUAL ACTIVITY
DF	DS	150.000	NONE	NONE	1.000
R	BS	29.900	NONE	30.000	.

COLUMNS SECTION

NAME	STAT	ACTIVITY	LOWER LIM	UPPER LIM	REDUCED COST
X1	IV	1.000	.	1.000	101.000
A X3	IV	.	.	1.000	.
X4	IV	1.000	.	1.000	12.000
X5	IV	1.000	.	1.000	7.000
X6	IV	.	.	1.000	8.000
X7	IV	1.000	.	1.000	6.000
X8	IV	1.000	.	1.000	3.000
X2	IV	1.000	.	1022.000	1.000

READY  
EXIT

B7700/B6700 TEMPO TERMINATED.  
RET=40.7 PT=10.6 IQ=5.3

## SECCION 11.1.

2/ Considere el siguiente problema , el cual es un problema de programación mixta .

$$\text{Min } Z = 3x_1 + 2x_2 - 6x_3$$

Sujeto a.

$$9 \leq 3x_1 + 6x_2 \leq 20$$

$$4x_1 + 2x_2 - 5x_3 = -15$$

$$3 \leq x_2 \leq 5 , \quad x_2 \text{ variable entera}$$

$$x_3 = 0 , 1$$

$$x_1 \geq 0 , \quad x_1 \text{ variable real}$$

Se resolvera este problema usando el paquete TEMPO/MPS/ALL a través de una terminal remota .

Note que en la primer restricción se considera para este problema cómo una restricción con limite inferior 9 y rango 11 . la variable  $x_1$  pertenece a los reales y es no-negativa , la variable  $x_2$  es entera y esta acotada con limite inferior 3 y limite superior 5 , y la variable  $x_3$  es binaria , es decir solo puede tomar el valor cero o uno .

En las hojas a continuación se muestra la forma de resolver este problema a través de la terminal , primero se dan los comandos utilizados en la secuencia seguida y posteriormente la forma en que se interactua con la computadora a través de la terminal .



RUN*SERVICIO/MPS/ALL	(R)
ZPRINTER = .TRUE.	(R)
\$FILE ZPROF = PROBLEMA	(R)
\$FILE ZSOLF = SOLUCION	(R)
ZNAME = "N"	(R)
ZDATA = "N"	(R)
INPUT(REMOTE)	(R)
ROWS	(R)
N Z	(R)
G R1	(R)
E R2	(R)
COLUMNS	(R)
X1 Z 3	(R)
X1 R1 3	(R)
X1 R2 4	(R)
EMP1 'MARKER' 'INTORG'	(R)
X2 Z 2	(R)
X2 R1 6	(R)
X2 R2 2	(R)
TERM1 'MARKER' 'INTEND'	(R)
EMP2 'MARKER' 'BIVORG'	(R)
X3 Z -6	(R)
X3 R2 -5	(R)
TERM2 'MARKER' 'BIVEND'	(R)
RHS	(R)
B R1 9	(R)
B R2 15	(R)
RANGES	(R)
R R1 11	(R)
BOUNDS	(R)
UP C X2 5	(R)
LO C X2 3	(R)
ENDATA	(R)
BCDOUT	(R)
ZOBJ = "Z"	(R)

ZSOLNM = "SOLUCIO"	(R)
ZRHS = "B"	(R)
ZRNGST = "R"	(R)
ZBNDST = "C"	(R)
SETUP(MIN)	(R)
PRIMAL	(R)
MXINT	(R)
OUTPUT(FILE)	(R)
EXIT	(R)

#06(001126 CANDE 30+140) YOU ARE SCHED#001(102)  
# DEFAULT PRINT DESTINATION=SITE  
SEMINARIOS PARA MAYU/JUNIO ECLEA NEWS  
#SESSION 3582 15134105 05/06/81  
RUN#SERVICIO/MYS/ALL  
#RUNNING 3583  
#?  
#7700/06700 TEMPC(20+000+000 - MARCH 77) 05/06/81 15134110  
READY  
ZPRINTER=.TRUE:  
READY  
ZFILE ZPROF= PROBLEMA  
READY  
ZFILE ZSOLF= SOLUCION  
READY  
ZNAME = "N"  
READY .  
ZDATA = "N"  
READY  
INPUT(REMOTE)  
--- INPUT ---  
X  
RONS  
Z  
N Z

X

G R1

X

E R2

X

COLUMNS

X

X1 4 3

X

X1 01 3

X

X1 02 4

X

EMP1 2MARKER2 4INTORG2

X

X2 4 2

X

X2 01 6

X

X2 02 2

X

TERM1 2MARKER2 4INTEND2

X

EMP2 2MARKER2 4DIVORG2

X

X3 4 0

X

X3 02 05

X

TERM2 2MARKER2 4DIVEND2

X

RHS

X

B R1 9

X

B R2 15

X

RANGES

X

R R1 11

X

BOUNDS

X

UP 5 X2 5

X

LO 4 X2 3

X

ENDATA

OLD ENTRY N

DELETED ON ZPROF (OR ZSOLF)

NEW ENTRY N

ENTERED ON ZPROF (OR ZSOLF)

READY

BCDOUT

--- BCDOUT ---

NAME N

RHS

N 4

G R1

E R2

COLUMNS

X1	4	3.00000	R1	0.00000
X1	R2	4.00000		
EMP1	ZMARKRZ		ZTATDFGZ	
X2	Z	2.00000	R1	6.00000
X2	R2	2.00000		
TERM1	ZMARKRZ		ZTATENRZ	
EMP2	ZMARKRZ		ZBIVCFGZ	
X3	4	-6.00000	R2	-5.00000
TERM2	ZMARKRZ		ZBIVENDZ	

RHS

b	R1	9.00000	R2	15.00000
---	----	---------	----	----------

RANGES

R	R1	11.00000		
---	----	----------	--	--

BOUNDS

UP C	X2	5.00000		
------	----	---------	--	--

LO C	X2	3.00000		
------	----	---------	--	--

ENDATA

READY

ZOBJ = "Z"

READY

ZRHS = "b"

READY

ZRNGST = "R"

READY

ZBNVST = "C"

READY

SETUP(MIN)

\*\*\* SETUP \*\*\*

READY

PRIMAL

\*\*\* PRIMAL \*\*\*

\*\*\* INVERT \*\*\*

PROBLEM INFEASIBLE  
 NUMBER OF INFAS \* 1  
 SUM OF INFAS \* 9.00000

ITER NO	SUM OF INF	NO INF	CUT	IN
1	6.33333	1	2U	49

EXIT CONDITION: INFEASIBLE SOLUTION

NUMBER OF INF \* 1  
 SUM OF INF \* 6.33333

\*\*\* ZDUMPS LEHAND SET \*\*\*

TCL COMMANDS MAY NOW BE ENTERED - ENTER ZRETURN TO RESUME AN ITERATIVE  
 PROCEDURE

READY  
 MXINT

\*\*\* MXINT \*\*\*

MXINT BRANCH & HOLD

CONTINUOUS PROBLEM INFEASIBLE -- PRIMAL CALLED

\*\*\* PRIMAL \*\*\*

NUMBER OF INF = 1  
 SUM OF INF = \*6.33333

\*\*\* ZDQNF5 DEMAND SET \*\*\*

TCL COMMANDS MAY NOW BE ENTERED - ENTER ZR F1 UR A2 TO RESUME AN ITERATIVE PROCEDURE

READY  
 ZSDIHM = "SOLUCIO"

READY  
 OUTPUT(FILE)

--- OUTPUT ---

SOLUTION STATUS = INFEASIBLE  
 ROWS SECTION

NAME	STAT	ACTIVITY	LOWER LIM	UPPER LIM	DUAL ACTIVITY
R1	BS	0.000	NONE	NONE	1.000
R2	UL	20.000	9.000	20.000	-1.000
R3	**	0.667	10.000	15.000	.

COLUMNS SECTION

NAME	STAT	ACTIVITY	LOWER LIM	UPPER LIM	REDUCED COST
X1	BS	0.667	.	NONE	.
X2	LL	3.000	3.000	5.000	-4.000
X3	LL	.	.	1.000	-6.000

NEW ENTRY SOLUCIO ENTERED ON ZPROF (OR ZSOLF)

READY

EXIT  
 07700/66700 TEMPC TERMINATED



## SECCION II.2

Los archivos más utilizados por el paquete TEMPO/MPS  
 /ALL son :

- 1/ ZPROF .- Este archivo almacena los problemas generados por INPUT , REVISE , SAVE en un formato binario empa-  
gado , por lo que BCDOUT es utilizado para obtener  
la información contenida en el archivo en imagen  
de tarjetas .
- 2/ ZSOLF .- Este archivo almacena la solución en un formato  
empacado , el cual en otra corrida puede ser --  
impreso al utilizar SOLOUT .
- 3/ DISKIN .- Este archivo es utilizado para leer de disco los  
datos de un problema , grabados en el formato --  
especificado en la sección II , primero se le --  
asigna el nombre de ZRHS , ZOBJ , ZDATA , ZNAME  
y en caso de que se utilice rangos y cotas . -  
ZRNGST , ZBNDST , que deben coincidir con los --  
nombres utilizados en el archivo , el cual se --  
lee con INPUT(DISK) , despues se utilizón las -  
instrucciones de procedimiento.

Cómo ejemplo considerece el problema de programación  
 mixta de la sección anterior , el cual sera modificado usando  
 el procedimiento REVISE , para resolver el problema siguiente:

$$\text{Min } Z = 2x_2 - 6x_3 - x_4$$

Sujeto a.

$$7 \leq 6x_2 - x_4 \leq 18$$

$$-x_2 + 3x_4 + 4x_3 \leq 15$$

$$x_2 \leq 666; 2 \leq x_4 \leq 8 : x_2 , x_4 \in Z$$

$$x_3 = 0 , 1$$

El procedimiento seguido fue el siguiente :

1/ Para que el problema original se grabara con su solución se declaró el archivo ZPROF y ZSOLF de la forma siguiente :

i/\$FILE ZPROF = PROBLEMA , con lo que los datos se ---  
grabarán en un archivo llamado problema .

ii/\$FILE ZSOLF = SOLUCION , con lo que la solución del  
problema quedó grabada en un archivo llamado solución  
y el nombre asignado a la solución dentro del archivo  
fue solucio , el que fue asignado a través de ZSOLNM.

2/ Para modificar el problema de la sección IIIc , se volvió a correr el paquete declarando ZPROF y ZSOLF con los mismos--- nombres en que se declaró la versión anterior y que resultó ser no factible , y lo que hizo fue lo siguiente :

a/ Listar los datos del problema original (el de programación mixta ) con el procedimiento ECDOUO .

b/ Asignar el nombre de la solución requerida que en este caso es SOLUCIO asignado en ZSOLNM y se lista lo que se tenía como solución con SOLUO .

c/ Se modifica el problema con el procedimiento REVISE de la siguiente forma :

i/ Se modifica la segunda restricción de igual a menor o igual .

ii/ Se borra la columna  $x_1$  .

iii/ Se inserta el vector  $x_4$  después del vector  $x_2$  para que  $x_4$  sea una variable entera .

iv/ Se modifican los coeficientes de  $x_2$  y  $x_3$  en la - segunda restricción .

v/ Se modifica la restricción 1 .

vi/ Se modifican las variables acotadas .

vii/ Se lista lo hecho para verificar los cambios , y posteriormente se lista todo el problema .

d/ Se dan las instrucciones de procedimiento ya conocidas en las secciones anteriores de este capítulo .

3/ A continuación se da la secuencia seguida en las instrucciones , y posteriormente el listado con la ejecución en la terminal.

RUN*SERVICIO/MP5/ALL	(R)
ZPRINTER=.TRUE.	(R)
\$FILE ZPROF = PROBLEMA	(R)
\$FILE ZSOLF = SOLUCION	(R)
ZNAME = "N"	(R)
ZDATA = "N"	(R)
BCDOUT	(R)
ZSOLNM = "SOLUCIO"	(R)
SOLOUT	(R)
ZONAME = "N"	(R)
ZNAME = "N1"	(R)
REVISE(RENDTE)	(R)
ROWS	(R)
MODIFY	(R)
L R2	(R)
COLUMNS	(R)
DELETE	(R)
X1	(R)
AFTER X2	(R)
X4 Z -1	(R)
X4 R1 -1	(R)
X4 R2 3	(R)
MODIFY	(R)
X2 R2 -1	(R)
X3 R2 4	(R)
RHS	(R)
MODIFY	(R)
B R1 7	(R)
BOUNDS	(R)
MODIFY	(R)
UP C X4 8	(R)
LD C X4 2	(R)
UP C X2 666	(R)
LIST	(R)
ENDATA	(R)

BCDDUT	(R)
ZBNDST = "C"	(R)
ZRUGST = "R"	(R)
ZOBJ = "Z"	(R)
ZRHS = "B"	(R)
SETUP(MIN)	(R)
PICTURE	(R)
TRANCOL	(R)
PRIMAL	(R)
MXINT	(R)
RANGE	(R)
OUTPUT(BASIS)	(R)
OUTPUT(INFEAS)	(R)
TRANCOL	(R)
TRANCOL(MATRIX)	(R)
TRANCOL(INVERSE)	(R)
OUTPUT	(R)
EXIT	(R)

#B6700#126 CANDE 30#140# YOU ARE SCHED#001(102)

# DEFAULT PRINT DESTINATION#SITE

SEMINARIUS PAR#MATH/JUNIO TECLEE NEWS

#SESSION 3747 15121'57 05/06/81

#RUN#SERVICIO/MPS/ALL

#RUNNING 3749

??

37700/B6700 TEMP(20#600#000 - MARCH 77) 05/06/81 151521 2

READY

#PRINTER#TRUE#

READY

#FILE #PROF# PROBLEMA

READY

#FILE #SOLF# SCLUTION

READY

#NAME = "N"

READY

#DATA = "N"

READY

COUOUT

-- BCDOUT -->

AME

A

ONS

N 2

G R1

E R2  
COLUMNS

X1	Z	3.00000	R1	3.00000
X1	R2	4.00000		
EMP1	ZMARNRZ		ZINTORGZ	
X2	Z	2.00000	R1	6.00000
X2	R2	2.00000		
TERM1	ZMARNRZ		ZINTEACZ	
EMP2	ZMARNRZ		ZBIVORGZ	
X3	Z	-6.00000	R2	-5.00000
TERM2	ZMARNRZ		ZBIVENDZ	

RHS

B	R1	9.00000	R2	15.00000
---	----	---------	----	----------

RANGES

R	R1	11.00000		
---	----	----------	--	--

BOUNDS

UP C	X2	5.00000		
LO C	X2	3.00000		

ENDATA

READY

ZSOINM = "SOLUTION"

READY

SOLVED

--- SOLUOT ---

S O L U O T

PROBLEM IDENTIFICATION

PROBLEM NAME = N

RANGE SET NAME = M  
 BOUND SET NAME = C  
 FUNCTIONAL NAME = Z  
 RESTRAINT NAME = B

SOLUTION STATUS = INFEASIBLE  
 ITERATION NUMBER = 1  
 FUNCTIONAL VALUE = 0.00000

ROWS SECTION

NAME	STAT	ACTIVITY	LOWER LIM	UPPER LIM	DUAL ACTIVITY
Z	BS	0.000	NONE	NONE	1.000
R1	LL	20.000	9.000	20.000	+1.000
R2	**	8.667	10.000	15.000	.

COLUMNS SECTION

NAME	STAT	ACTIVITY	LOWER LIM	UPPER LIM	REDUCED COST
X1	BS	0.667	.	NONE	.
X2	LL	3.000	3.000	5.000	-4.000
X3	LL	.	.	1.000	-6.000

READY  
 ZONAME = "N"  
 READY  
 ZNAME = "N1"  
 READY  
 REVISE(REMOTE)

\*\*\* REVISE \*\*\*

OWD

ODIFY

R\*

OLYKNS

ELETE

1

FTER X2

4 4 =1

4 N1 =1

4 N2 3

ODIFY

2 N2 =1

3 N2 4

HS

ODIFY



```

B R1 7
*
BOUNDS
*
MODIFY
*
UP 4 X4 6
*
LO 4 X4 2
*
UP 4 X2 666
*
LIST
NAME      A
RHS
  MODIFY
  L R2
COLUMNS
  DELETE
  X1
  AFTER    X2
  X4      2      -1
  X4      R1      -1
  X4      R2      3
  MODIFY
  X2      R2      -1
  X3      R2      4
RHS
  MODIFY
  B      R1      7
BOUNDS

```

MODIFY

UP C	X4	8
LD C	X4	2
UP C	X2	666

ENDATA

\*

ENDATA

REVISING N            IN N1            ACCORDING TO N  
 NEW ENTRY N1            ENTERED ON ZPROF (OR ZSOLF)  
 READY  
 BCDOUT

--- BCDOUT ---

NAME            A

ROWS

N 2

G R1

L R2

COLUMNS

EMP1	ZHARER2		ZINTOR02	
X2	Z	2.00000	R1	6.00000
X2	R2	-1.00000		
X4	Z	-1.00000	R1	-1.00000
X4	R2	3.00000		
TEHM1	ZHARER2		ZTATEND2	
EMP2	ZHARER2		ZRIVOR02	
X3	Z	-6.00000	R2	4.00000
TEHM2	ZHARER2		ZRIVEND2	
RHS				
B	R1	7.00000	R2	15.00000

RDS 11.0000  
 C 666.0000  
 C X4 8.0000  
 C X4 2.0000

AT A  
 DY  
 DST = "C"  
 DY  
 GST = "F"  
 DY  
 J = "Z"  
 DY  
 IS = "B"  
 DY  
 UP(MIN)

\* SETUP ---

PROBLEM STATISTICS

	NUMBER	FREE	FIXED	BOUNDED	NORMAL
ROWS :	3	1	0	1	1
COLUMNS:	3	0	0	3	0

TRIX IN CORE : MEMORY ALLOCATION = 75 WORDS.  
 VERSE: MEMORY ALLOCATION = 450 WORDS. RECORD LENGTH = 225 WORDS.

NUMBER OF INTEGER VARIABLES = 3

ADY  
CTUPE

- PICTUR ---

OUTPUT FOR PICTURE WILL BE PRODUCED ON SYSTEM PRINTER ONLY.

700/7700 TEMPO  
 VERSION: 28.000-000

N1

PAGE 7

PART 1 OF 1  
 PAGE 1

X	X	X	B	R	S
2	4	1		A	A

UPPER	LOWER	SCALE	BOUND	BOUND	TYPE	C	A	A	A	U	R1	R2
7		1	A	-1	A							
R1		5	A	-1	A							
R2		L	-1	A	A	B						

66

SUMMARY OF MATRIX RANGE

SYMBOL	RANGE	COUNT	INCL	RHS
Z	LESS THAN	.000001		0
Y	.000001 THRU	.000009		0
X	.000011 THRU	.000099		0
W	.000101 THRU	.000999		0
V	.001001 THRU	.009999		0
U	.010001 THRU	.999999		0
T	.100001 THRU	.999999		0
I	1.000001 THRU	1.000000		3
A	1.000001 THRU	15.000000		6
U	10.000001 THRU	15.000000		1
C	100.000001 THRU	1.00000000		0
D	1.000000001 THRU	15.00000000		0
E	10.00000001 THRU	15.00000000		0
F	100.00000001 THRU	1.0000000000		0
G	GREATER THAN	1.0000000000		0

MINIMUM = -6.000000 MAXIMUM = 15.000000

READY  
 PRIMAL

--- TRANCO ---

OUTPUT FOR TRANCO WILL BE PRODUCED ON SYSTEM PRINTER ONLY.

--- INVERT ---

CURRENT INVERSE =  
 CURRENT BASIS = EQUALITY = 0. ETA RECORDS = 1. ETA VECTORS = 0. ELEMENTS = 0.  
 FUNCTION VALUE = -2.00000 SLACKS = 2. STRUCTURALS = 0. ELEMENTS = 2.

TRANCO L

TABLEAU

BOUND	X2 LOWER	X4 LOWER	X3 LOWER	B LOWER
2	2.00000	-1.00000	-6.00000	
X1	-6.00000	1.00000		-7.00000
X2	-1.00000	1.00000	4.00000	15.00000

READY  
 PRIMAL

--- PRIMAL ---

ITERATION NUMBER = 7  
 TYPE NUMBER INFEAS = 1  
 SUM OF INFEAS = 0  
 ZRHS = 0  
 NUMBER INFEAS = 1  
 REDUCED COST = -6.00000  
 PIVOT INDEX = 2  
 VECTOR OUT = 2  
 VECTOR IN = 49  
 FUNCTION VALUE = 1.00000

SOLUTION FEASIBLE

100

CURRENT IN : EQUALITY = 0. ETA RECORDS = 1. A VECTORS = 2. ELEMENTS = 4.  
 CURRENT EAS : FORWARD TRIANGULAR VECTORS = 1. STRUCTURALS = 1. ELEMENTS = 4.  
 NEW INVERSE : FUNCTION VALUE = 1.00000  
 700J = 7 ZPHS = 0  
 ITERATION NUMBER SUM OF NUMBER REDUCED PIVOT VECTOR VECTOR FUNCTION  
 TYPE NUMBER INFEAS INFEAS NEG DJ COST INDEX OUT IN VALU  
 PB2 1 2 0 . 2 -6.00000 3 51U 51J -6.00000  
 1 3 0 . 2 -7.65667 3 5U 5B -6.52941

EXIT CONDITION: OPTIMAL SOLUTION.  
 FUNCTION VALUE = -6.52941

READY  
 MXINT

--- MXINT ---

M X I N T    B R A N C H   &   B O U N D

I N T E G E R   S O L U T I O N

NOTE \*\* A \*\* OPTIMIZED INTEGER NOTE  
 FIRST INTEGER SOLUTION OBJECTIVE = -6.0000  
 CUGHT NOW AT -0.0000 POSTPONE IT -6.0000

A C T I V E   I N T E G E R   V A R I A B L E S

NAME	NUMBER	ACTIVITY
1. X2	40	2.00
2. X4	50	4.00
3. X3	51	1.00

EXIT PRINT ON INTEGER SOLUTION DEMAND

\*\*\* ZINTSOL DEMAND SET \*\*\*

TCL COMMANDS MAY NOW BE ENTERED - ENTER ZRETURN TO RESUME AN ITERATIVE  
 PROCEDURE.

READY  
RANGE

\*\*\* RANGE \*\*\*

COLUMNS AT LIMIT LEVEL

COLUMN ST	ACTIVITY INPUT COST	LO LIMIT UP LIMIT	LO ACTIV UP ACTIV	UNIT COST	UP COST LO COST	LIMITING PROCESS	ST
X2	.	.	-0.167	-2.000	INFINITY	R1	LL
LL	2.000	664.000	1.667	2.000	.	R1	UL
X4	4.000	2.000	-6.000	1.000	.	R1	LL

UL	*1.000	4.000	4.333	*1.000	INFINITY R2	.LL
X3	1.000	.	*INFINITY	6.000	.	NONE
UL	*6.000	1.000	1.250	*6.000	INFINITY R2	UL

ROWS AT INTERMEDIATE LEVEL

ROW	ACTIVITY	LO LIMIT	LO ACTIV	UNIT	LIMITING
ST	SLK ACTIV	UP LIMIT	UP ACTIV	COST	PROCESS ST
R1	6.000	7.000	8.000	INFINITY	NONE
BS	*1.000	10.000	INFINITY	0.333 X2	.LL
R2	14.000	NONE	*10.000	0.333 X4	UL
BS	1.000	15.000	14.000	INFINITY	NONE

READY  
OUTPUT(BASIS)

--- OUTPUT ---

SOLUTION STATUS : OPTIMAL  
ROWS SECTION

NAME	STAT	ACTIVITY	LOWER LIM	UPPER LIM	DUAL	ACTIVITY
Z	BS	*6.000	NONE	NONE		1.000
R1	BS	6.000	7.000	10.000		.
R2	BS	14.000	NONE	15.000		.

COLUMNS SECTION

NAME STAT ACTIVITY LOWER LIM UPPER LIM REDUCED COST  
 READY  
 OUTPUT(INFEAS)

--- OUTPUT ---

SOLUTION STATUS : OPTIMAL  
 ROWS SECTION

NAME	STAT	ACTIVITY	LOWER LIM	UPPER LIM	DUAL ACTIVITY
Z	BS	-6.000	NONE	NONE	1.000
R1	BS	8.000	7.000	18.000	.
R2	BS	14.000	NONE	15.000	.

COLUMNS SECTION

NAME	STAT	ACTIVITY	LOWER LIM	UPPER LIM	REDUCED COST
X2	IV	2.000	.	666.000	-2.000
X4	IV	4.000	2.000	8.000	-1.000
X3	IV	1.000	.	1.000	-6.000

READY  
 TRANCO

--- TRANCO ---

OUTPUT FOR TRANCO WILL BE PRODUCED ON SYSTEM PRINTER ONLY.  
 READY  
 TRANCO(MATRIX)

--- TRANCO ---

OUTPUT FOR TRANCO WILL BE PRODUCED ON SYSTEM PRINTER ONLY.



READY  
TRANCL<INVERSE>

--- TRANCO ---

OUTPUT FOR TRANCL WILL BE PRODUCED ON SYSTEM PRINTER ONLY.

READY  
OUTPUT

--- OUTPUT ---

SOLUTION STATUS = OPTIMAL  
ROWS SECTION

NAME	STAT	ACTIVITY	LOWER LIM	UPPER LIM	DUAL ACTIVITY
Z	BS	-6.000	NONE	NONE	1.000
R1	BS	7.000	7.000	18.000	.
R2	BS	14.000	NONE	15.000	.

COLUMNS SECTION

NAME	STAT	ACTIVITY	LOWER LIM	UPPER LIM	REDUCED COST
X2	IV	2.000	.	666.000	2.000
X4	IV	4.000	2.000	8.000	-1.000
X3	IV	1.000	.	1.000	-6.000

READY

EXIT

87700/06700 TEMPC TERMINATED.

#ET=1109.2 PT=15.0 IO=0.5

#END SESSION 2747 LT=1112.7 PT=15.0 IO=0.5

#USER = SF82 15153110 05/06/81

## CAPITULO V

SPSS es un conjunto de algoritmos de análisis estadístico, diseñado para el análisis de datos de ciencias sociales (Statistical Package for the Social Sciences) el cual está constituido por los siguientes 18 análisis.

AGGREGATE	.- Agrega subpoblaciones estadísticas y las imprime en un archivo en disco .
ANOVA	.- Análisis de varianza .
BREAKDOWN	.- Descripción de subpoblaciones y prueba de linealidad .
CANCORR	.- Análisis de correlación canónica .
CONDESCRIPTIVE	.- Estadísticas básicas .
CROSSTABS	.- Tablas cruzadas .
DISCRIMINANT	.- Análisis discriminante .
FACTOR	.- Análisis factorial .
FREQUENCIES	.- Tablas de frecuencias .
GUTTMAN SCALE	.- Análisis de escalogramas .
NONPAR CORR	.- Correlación spearman o kendall .
ONEWAY	.- Análisis de varianza one-way .
PARTIAL CORR	.- Cálculo de coeficientes de correlación parcial .
PEARSON CORR	.- Coeficiente de correlación de pearson .
REGRESSION	.- Regresión lineal múltiple .
SCATTERGRAM	.- Gráficas de parejas de datos .
T-TEST	.- Prueba t-student .
WRITE CASES	.- Para escribir algunas o todas las variables en un archivo en disco o cinta en formato bcd (formato de fortrán) .

## S E C C I O N V

En esta sección se indicara el procedimiento para el acceso a los análisis del paquete SPSS . A continuación se --muestrán unos de las instrucciones utilizadas por el paquete --de las cuales se considerara que las que se encuentran subrayadas serán necesarias y las demás opcionales , y su formato es:

1	16
RUN NAME	Ejemplo del uso de SPSS Nombre asignado a la corrida .
<u>FILE NAME</u>	LECTURA , archivo con la información Nombre de archivo y su etiqueta .
<u>VARIABLE LIST</u>	Lista de variables (nombre de las variables a lo más de ocho caracteres) <u>i/</u> Los nombres de las variables senarados por comas , por ejemplo VAR1,VAR2,VAR3,VAR4,VAR5 y se considerán 5 variables . <u>ii/</u> VAR1 TO VAR10 indica que se considera un conjunto de variables en donde la primer --variable se llama VAR1 y la última VAR10 .
<u>INPUT MEDIUM</u>	CARD si el archivo de datos está en tarjetas. DISK si está en disco . TAPE si se encuentra en cinta .
SUBFILE LIST	Lista de subarchivos (a lo más 100) SUBF1(19),SUBF2(24),SUBF3(37) Se considera que el archivo esté constituido por tres subarchivos y el primer subarchivo consta de 19 casos , el segundo de 24 y el --tercero de 37 .
<u>INPUT FORMAT</u>	Formato en que se encontrarán los datos . <u>i/</u> FREEFIELD indica que estén en formato libre es decir separados por comas , <u>ii/</u> FIXED( formato de los datos BCD ) , note que se indica el formato utilizado en fortrán.
<u>N OF CASES</u>	Número de casos <u>i/</u> n entero indicando el número de casos . <u>ii/</u> UNKNOW cuando se desconoce el número de casos.

1 16

MISSING VALUES VAR1 TO VAR10 (3), se considera en el conjunto de variables que cuando tomén el valor de 3 - se tratán cómo valores perdidos .

VAR LABELS Lista de variables con sus etiquetas . .  
cómo por ejemplo las siguientes :  
VAR1 NIVEL ECONOMICO/VAR2 INDICADGR/  
VAR10 ETIQUETADO/

VALUE LABELS Lista de variables con sus valores etiquetados.  
cómo por ejemplo :  
VAR1 (5) ocasional (6) otro/  
VAR3 (2) nada /

PRINT FORMATS Lista de variables .  
VAR1(4) / VAR4(A) , en VAR1 se indica se usén en la impresión 4 dígitos decimales , y VAR4 se imprima cómo variable alfanúmerica , en el caso de impresión de dígitos decimales se usén a lo más cinco .

RECODE  
SELECT IF  
IF  
COMPUTE Són utilizados para modificar variables , calcular nuevas variables , seleccionar variables , las que quedán en el SAVE/FILE .

\*RECODE  
\*SELECT IF  
\*IF  
\*COMPUTE Similar a los anteriores pero no quedán en el SAVE/FILE .

DO REPEAT .  
:  
END REPEAT Repite transformaciones de datos sobre un conjunto de variables , utilizando las instrucciones anteriores .

RUN SUBFILES ALL , todos los subarchivos són tratados cómo un solo archivo .  
EACH , cada subarchivo es tratado independientemente de los demás .

1 16  
**TASK NAME** TITULO  
 este título aparecera al principio de cada corrida hasta encontrar otro task name .  
**ASSIGN MISSING** Asignación de valores faltantes a nuevas variables .  
**COMMENT** Para comentar el programa spss no queda en el save-file .  
**DOCUMENT** Similar al anterior y si queda en el save-file .

Nombre de algún procedimiento (ver pags. 555-561 del manual)

ANOVA , FACTOR , etc .

**OPTIONS** Números de las opciones requeridas .

**STATISTICS** ALL y se dan todas las estadísticas o bien los números de las estadísticas requeridas .

**READ INPUT DATA**

⋮ Datos en el caso de tarjetas

Nombre de algún procedimiento (puedén seguir varios procedimientos o bien ninguno)

**SAVE FILE** Para terminar la corrida SPSS , y que quede

**FINISH** en el save-file todo el archivo .

Otro conjunto de instrucciones utilizadas frecuentemente por SPSS són las siguientes :

**EDIT** Se utiliza para verificar si las instrucciones están correctas en su sintaxis (es la primer tarjeta si se utiliza )

**PAGESIZE** n número entero indicando el número de líneas que se imprimirán por hoja o NOEJECT para -- impresión continue .

**PRINT BACK** NO y suprime la impresión de las tarjetas de control , también es utilizada para el control de formato .

<u>L</u>	16
LIST FILEINFO	COMPLETE , lista información de la definición del archivo .
NUMBERED	YES y loé solo hasta la columna 72 , de la 73-80 se puede usar para numerar las tarjetas.
LIST CASES	CASES - n/VARIABLES - lista de variables o ALL / y se listán los primeros n casos de - cada variable .
WRITE CASES *	Genera un deck de tarjetas .
WRITE FILEINFO	Genera un bloque de tarjetas con la definición del archivo .
ADD VARIABLES	Lista de variables que se van a añadir en un archivo .
ADD DATA LIST	Formato de las variables que se van a añadir.
ADD CASES	Número de casos que se añaden al archivo del sistema .
ADD SUBFILES	Añade nuevos subarchivos al sistema .
DELETE SUBFILES	Para borrar subarchivos .
DELETE VARS	Lista de variables que no quedarán en el save-file .
KEEP VARS	Lista de variables que se conservarán en el archivo del sistema .
MERGE FILES	Para intercalar archivos .
SORT FILES	Para ordenar casos .
RAW OUTPUT UNIT	n número entero indicando la unidad asignada .

## S E C C I O N V.1

## b/ Análisis de regresión .

El procedimiento regression es utilizado para ajustar modelos de regresión lineal simple y múltiple , consta de 15 opciones y 7 estadísticas .

## OPCIONES :

- 1.- Incluir valores faltantes en el cálculo de coeficientes de correlación .
- 2.- Eliminación de datos faltantes en forma de parejas.
- 3.- Suprimir la impresión de las etiquetas de las variables .
- 4.- La matriz de correlación será dada por el usuario.
- 5.- Medias y desviaciones estándar serán dadas por el usuario , precediendo a la matriz de correlación , (se usa cuando se use la opción 4)
- 6.- Suprimir la impresión step-by-step y solo el resumen de la tabla de regresión se imprime .
- 7.- Suprimir el resumen y solo el step-by-step se imprime.
- 8.- La matriz de correlación o matrices utilizadas en los cálculos serán impresas en una unidad , la que es especificada en RAW OUTPUT UNIT por el usuario.
- 9.- La matriz de correlación es indexada por la lista de variables (no se puede usar sin la opción 4) .
- 10.- Ocasiona se secuencia la información de la columna 1 a la 20 de cada registro sobre el raw-output-data file .
- 11.- y 12.- Los residuales estandarizados se grabarán en una unidad especificada por el usuario sobre el raw-output-data file .
- 13.- Predicciones estandarizadas , las cuales son un producto pesado de los datos existentes .
- 14.- Suprimir la impresión de ejes sobre las gráficas de predicciones estandarizadas contra residuales estandarizados .
- 15.- Imprimir medias y desviaciones estándar en el -- raw-output-data file especificado por el usuario.

1	16
REGRESSION	VARIABLES = V1,V2/ REGRESSION = V1 WITH V2(1)/
STATISTICS	ALL
READ INPUT DATA	
FINISH	

La forma de ejecutar este problema es la siguiente :

RUN*SERVICIO/SPSS6;FILE FILE6(REMOTE),%	(R)
FILES(KIND=DISK,TITLE=REGRESION),%	(R)
FILE8(KIND=DISK,TITLE=DATOSR)	(R)

Por lo que en las paginas siguientes se muestran los resultados de la ejecución .



STRIETED FOR THE BUREAUXS 86700 BY THE  
 CIAL SCIENCE DATA SERVICE  
 UNIVERSITY OF CALIFORNIA, DAVIS  
 DEFAULT WORKSPACE FOR THIS RUN = 20000 WORDS  
 REGRESSION LINEAL MULTIPLE  
 INFORMACION A AJUSTAR EN MODELO LINEAL  
 A CONTINUACION LA LISTA DE VARIABLES  
 VI V2  
 SE ACCESA POR TARJETAS ES CARU  
 EN DISCO DISK \* EN CINTA TAPE  
 DISK  
 LOS DATOS ESTARAN EN FORMATO LIBRE  
 FREEFIELD  
 EL NUMERO DE CASOS ES 11  
 11  
 ETIQUETADOS LAS VARIABLES V1 COMO Y \* V2 COMO X  
 V1 Y/V2 X/  
 UTILIZAMOS EL PROCEDIMIENTO DE REGRESION  
 VARIABLES = V1, V2/  
 REGRESSION = V1 WITH V2(1)/  
 ALL

\*\*\*\*\* REGRESSION PROBLEM REQUIRES 34 WORDS OF WORKSPACE, NOT INCLUDING RESIDUALS \*\*\*\*\*

READ INPLI DATA  
READ OF DATA INPLI READ COUNT = 11 DATA FROM COUNT = 0.

081

LE INFCMPAC (CREATION DATE = 06/02/81) A AJUSTARLE UN MODELU LINEAL

VARIABLE	MEAN	STANDARD DEV	CASES
	17.8102	5.8621	11
	19.2897	5.9991	11

TOT

REGRESSION LINEAL MULTIPLE

06/02/81

PAGE 3

LE IMPACTO (CREACION DATE = 06/02/81) A AJUSTARLE UN MODELO LINEAL

RELACION COEFICIENTES

VALUE OF 99.00000 IS PRINTED  
A COEFFICIENT CANNOT BE COMPUTED.

V1	V2
1.00000	0.99840
0.99840	1.00000

REGRESSION LINEAL MULTIPLE

LE INTERFAC (CREATION DATE = 02/02/81) A AJUSTABLE LN MODELL LIN  
\*\*\*\*\* MULTIPLE REG  
DEPENDENT VARIABLE: V1 Y

\*\*\*\*\*  
VARIABLE(S) ENTERED ON STEP NUMBER 1. V2 X

MULTIPLE R 0.99840 ANALYSIS OF VARIANCE  
R SQUARE 0.99679 REGRESSION  
ADJUSTED R SQUARE 0.99444 RESIDUAL  
STANDARD ERROR 0.34549

----- VARIABLES IN THE EQUATION -----

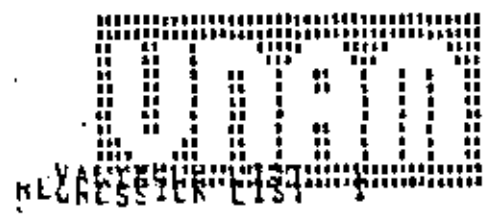
VARIABLE	B	BETA	STD ERROR B	F
V2	0.97555	0.99840	0.01845	2797.148
(CONSTANT)	-1.00070			

MAXIMUM STEP REACHED

02/02/81 PAGE 4

EAL

REGRESSION \*\*\*\*\*



\*\*\*\*\*

DF	SUM OF SQUARES	MEAN SQUARE	F
1.	342.53424	342.53424	2797.14779
9.	1.10213	0.12246	

----- VARIABLES NOT IN THE EQUATION -----

VARIABLE	BETA IN	PARTIAL TOLERANCE	F
----------	---------	-------------------	---

FILE INTERFAC (CREATION DATE = 06/02/81) A AJUSTABLE LN MODELO LINEAL

\*\*\*\*\* MULTIPLE REGRESSION \*\*\*\*\*

DEPENDENT VARIABLE: V1 Y

SUMMARY TABLE

VARIABLE	MULTIPLE R	R SQUARE	RSG CHANGE	SIMPLE R
2 CONSTANT)	0.99840	0.99679	0.99679	0.99840

DATA TRANSFORMATION DONE UP TO THIS POINT..

NO OF TRANSFORMATIONS	0
NO OF RECORD VALLES	0
NO OF ARITHM. OR LOG. OPERATIONS	0

FINISH

## SECCION IV.1

a/ Análisis De Regresión .

Conceptos del marco teorico utilizados .

$$1/ \text{ Promedio muestral} = \bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

$$2/ \text{ Desviación standard} = S = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{Y})^2}{(n-1)}}$$

$$3/ \text{ Varianza} = \hat{V}^2 = S^2$$

$$4/ \text{ Coeficiente de correlación muestral} = r_{x,y} =$$

$$= \left[ \frac{\left[ \frac{\sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})}{n} \right]}{\sqrt{\left[ \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n} \right] \left[ \frac{\sum_{i=1}^n (y_i - \bar{Y})^2}{n} \right]}} \right]$$

$$i/ r_{x,y} \rightarrow 1 \text{ entonces } y = a + bx ; b > 0$$

$$ii/ r_{x,y} \rightarrow -1 \text{ entonces } y = a + bx ; b < 0$$

$$iii/ r_{x,y} \rightarrow 0 \text{ no indica que no están asociadas las variables, quiere decir que no están asociadas linealmente .}$$

$$5/ \text{ Coeficiente de determinación} = r_{x,y}^2$$

$$6/ \text{ Covarianza} = \text{Cov}(x,y) = \frac{\sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})}{n}$$

$$i/ \text{ Cov}(x,y) > 0 \text{ entonces si } x \text{ aumenta } y \text{ aumenta}$$

$$\text{si } x \text{ disminuye y disminuye .}$$

$$ii/ \text{ Cov}(x,y) < 0 \text{ entonces si } x \text{ aumenta y disminuye}$$

$$\text{si } x \text{ disminuye y aumenta .}$$

$$iii/ \text{ Cov}(x,y) = 0 \text{ inconcluyente .}$$

En estadística uno de los modelos lineales más utilizados es ; El Modelo De Regresión lineal multiple ,el cual es de la forma siguiente :

$$y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \dots + b_r x_{ri} + e_i$$

Y que en forma matricial se representa por :

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1r} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \dots & x_{nr} \end{bmatrix} \begin{bmatrix} b_0 \\ \vdots \\ b_r \end{bmatrix} + \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix} ; e_i \sim N(0, \sigma^2)$$

o también  $Y = XB + E ; E \sim N(\underline{0}, \sigma^2 I)$

En donde  $x_{ij}$  representa el  $j$ -ésimo valor observado de las  $r$  variables, con el  $i$ -ésimo valor de  $Y$ , y  $b_j$  es el parámetro que especifica cómo  $Y$  está relacionado con los  $x_j$ . La variable  $Y$  es conocida como la variable dependiente y las variables  $x$  son conocidas como variables independientes,  $b_j$  es la derivada parcial  $\partial Y / \partial x_j$  por lo que los  $b_j$  son conocidos como coeficientes de regresión parcial y indican el cambio en la variable dependiente  $Y$  asociado con una unidad de cambio en la correspondiente variable independiente ( $x_j$ ), mientras el resto de variables independientes permanecen constantes,  $b_0$  es el valor de intersección con el eje  $Y$ , cuando todos los  $x_j$  son cero.

Un caso particular es el modelo de regresión lineal simple que es de la forma:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix} ; e_i \sim N(0, \sigma^2)$$

Considere el siguiente modelo el cual se resolverá utilizando el paquete :

$$\begin{bmatrix} 10 \\ 12 \\ 13 \\ \vdots \\ \vdots \\ 18 \\ 21 \end{bmatrix} = \begin{bmatrix} 1 & 11.6 \\ 1 & 13.87 \\ 1 & 13.972 \\ \vdots & \vdots \\ \vdots & \vdots \\ 1 & 18.985 \\ 1 & 22.675 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ \vdots \\ e_{10} \\ e_{11} \end{bmatrix} ; e_i \sim N(0, \sigma^2)$$

El modelo en forma matricial es de la forma siguiente:

$Y = XB + E$  ;  $E \sim N(0, \sigma^2 I)$  . El modelo de regresión lineal estimado es  $\hat{Y} = X\hat{B}$  , donde  $Y - \hat{Y}$  es el vector de residuales , y la suma de cuadrados de los residuales es  $\sum_{i=1}^n E_i^2 = (Y - \hat{Y})'(Y - \hat{Y})$  , la cual cuando se minimiza con respecto a B produce el sistema de ecuaciones  $(X'X)B = X'Y$  , lo que es la forma equivalente de las ecuaciones normales en forma matricial .  $X'X$  es simétrica y consiste del producto cruz de las variables x ,  $X'Y$  es el vector de los productos cruz xy . Al resolver el sistema se obtiene el estimador  $(\hat{B})$  .

$$\hat{B} = (X'X)^{-1}X'Y = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

Ahora bién la información de que nos provee el paquete en la rutina de regresión lineal múltiple (MULTR) es :

$\bar{X} = 19.28$	Promedio de las $x_i$
$S_x = 5.999$	Desviación standard de las $x_i$
$\bar{Y} = 17.81$	Promedio de las $y_i$
$S_y = 5.862$	Desviación standard de las $y_i$

Matriz normal  $X'X = \begin{bmatrix} 11 & 212.13 \\ 212.13 & 4450.3 \end{bmatrix} \approx \begin{bmatrix} 1 & 19.28 \\ 0 & 359.8891 \end{bmatrix}$  .

Matriz de correlación  $\begin{bmatrix} r_{xx} & r_{xy} \\ r_{yx} & r_{yy} \end{bmatrix} = \begin{bmatrix} 1 & .9984 \\ .9984 & 1 \end{bmatrix}$

Opcionalmente se dan las inversas de la matriz normal y de la de correlación con OPTION INVERSE :

$$\hat{B} = (X'X)^{-1}X'Y = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} -1.0007 \\ 0.9755 \end{bmatrix}$$

Error standard estimado  $= \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{(n-2)}} = .34994$



El modelo obtenido es:  $y_i = -1.0007 + .9755x_i + e_i$

$$\text{Coeficiente de determinación} = \frac{\sum_{i=1}^n (y_i - \bar{Y})^2}{\sum_{i=1}^n (y_i - \bar{Y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

$$= .9967$$

$$\text{Coeficiente de correlación múltiple} = R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{Y})^2}$$

$$= .9984$$

La tabla de análisis de varianza es utilizada para la prueba de hipótesis,  $H_0 : B = 0$  contra  $H_a : B \neq 0$

TABLA DE ANALISIS DE VARIANZA

fuentes de variación	grados de libertad	suma de cuadrados	cuadrado medio	F
Regresión	1	$SSR = \frac{\sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})^2}{\sum_{i=1}^n (x_i - \bar{X})^2}$ $= 342.5342$	$MSR = \frac{SSR}{1}$ $= 342.53$	$\frac{MSR}{MSE}$ $= 2797.14$
Residual	$n-2$ $= 9$	$SSE = SST - SSR$ $= 1.1021$	$MSE = \frac{SSE}{n-2}$ $= .1225$	
Total	$n-1$ $= 10$	$SST = \sum_{i=1}^n (y_i - \bar{Y})^2$ $= 343.6364$	— — —	

Regla.- Si  $F > F_{1-\alpha, 1, n-2}$  (de tablas), rechazar  $H_0$ , es decir si F es muy grande rechazamos la hipótesis  $H_0$ , esto quiere decir que la explicación al modelo de nuestras variables es muy buena.

$$F_{0.01, 1, 9} = 5.12 < F \quad \text{significativa}$$

$$F_{0.05, 1, 9} = 10.6 < F \quad \text{muy buena}$$

Lo que quiere decir es que la prueba es altamente significativa.

En la tabla de residuales se tiene, observación  $i$ ,  $y_i$  observación real,  $\hat{y}_i = b_0 + b_1 x_i$  observación estimada,  $y_i - \hat{y}_i$  residuales.

Por ultimo tenemos la estadística d Durbin-Watson que es la base para probar autocorrelación en análisis de regresión. La prueba se basa sobre la acepción de que los errores constituyen una serie recursiva de primer orden.

$$u_i = \rho u_{i-1} + e_i \quad \dots (a) \quad |\rho| < 1$$

$$e_i \sim N(0, \sigma^2) \quad ; \quad E \sim N(0, \sigma^2 I)$$

En otras palabras el residual para el periodo  $t$  es una función del residual del periodo previo  $(t-1)$  más un error aleatorio, la existencia de un modelo recursivo se puede detectar por un análisis de los residuales del modelo, al utilizar la estadística d Durbin Watson.

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} = 1.80735$$

Para probar la hipótesis nula  $H_0 : \rho = 0$  contra la alternativa  $H_a : \rho > 0$ ; note que cuando  $\rho = 0$  en (a) los  $u$ 's son no correlacionados, el parametro se puede estimar por  $r$ .

$$\rho = r = \frac{\sum_{i=2}^n e_i e_{i-1}}{\sum_{i=1}^n e_i^2}$$

Donde existe una relación de aproximación entre  $d$  y  $r$ ,  $d \approx 2(1-r)$   $d$  toma valores entre cero y cuatro, cuando  $d \rightarrow 2$  es cuando  $\rho = 0$  y  $d \rightarrow 1$  cuando  $\rho = 1$ , al hacer tender el valor muestral  $d$  a 2 se afirma la evidencia de que no se tiene autocorrelación en el error, una evidencia de autocorrelación es indicada por la desviación de  $d$  de 2, la prueba formal para autocorrelación opera de la siguiente forma:

- i/ Si  $d < d_L$  ; rechazar  $H_0 : \rho = 0$
- ii/ Si  $d > d_U$  ; no rechazar  $H_0 : \rho = 0$
- iii/ Si  $d_L < d < d_U$  ; no decidir

## S E C C I O N V.1

## b/ Estadísticas básicas .

El procedimiento condensativo es utilizado para - obtener estadísticas básicas , consta de 4 opciones y 9 estadísticas .

## OPCIONES :

- 1.- Incluir todos los casos a excepción de los declarados faltantes .
- 2.- Suprimir impresión de las etiquetas de las variables.
- 3.- Imprimir valores estandarizados (debe indicarse la unidad en RAW OUTPUT UNIT , que debe ser de 15 a - 20 , y se asigna título , maxrecsize , kind , etc. por omisión es cinta,tape y maxrecsize 22 ) .
- 4.- Imprime un diccionario en forma alfabética de las variables utilizadas .

## ESTADÍSTICAS :

- |                          |                     |
|--------------------------|---------------------|
| 1.- Media .              | 8.- Simetría .      |
| 2.- Error estándar .     | 9.- Rango .         |
| 5.- Desviación estándar. | 10.- Valor mínimo . |
| 6.- Varianza .           | 11.- Valor máximo . |
| 7.- Curtosis .           |                     |

En este caso el problema se ejecutara de la siguiente forma :

```

RUN*SERVICIO/SPSS6;FILE FILE6(REMOTE);%      (R)
FILE FILE5(DISK,TITLE=PROGSPSS);%           (R)
FILE FILE8(DISK,TITLE=DATOSSPSS);%         (R)
FILE FILE4(DISK,TITLE=ESTADISTICA)         (R)

```

Se considero que el programa SPSS se encontraba en - disco con el nombre PROGSPSS , y los datos también se tenían en disco con el nombre DATOSSPSS , y se utilizo la instrucción - SAVE FILE para almacenar el programa spss y los datos en disco con el nombre ESTADISTICA para posteriormente accederlo sin -- tener que generarlo de nuevo .

1	16
PRINT FORMATS	V8,V9(4)
RUN SUBFILES	ALL
TASK NAME	ESTADISTICAS BASICAS
COMMENT	
COMMENT	SE UTILIZARA EL PROCEDIMIENTO CONDESCRIPTIVE
COMMENT	
CONDESCRIPTIVE	V2 TO V8
OPTIONS	1,4
STATISTICS	ALL
READ INPUT DATA	
SAVE FILE	
FINISH	

En las paginas siguientes se muestra la ejecución .

TRIBUTED FOR THE BUREAUX B6700 BY THE  
IAL SCIENCE DATA SERVICE  
UNIVERSITY OF CALIFORNIA, DAVIS

DEFALL WORDSPACE FOR THIS RUN = 20000 WORDS  
EJEMPLO DE ESTADISTICAS BASICAS  
ESTADISTICA ARCHIVO CCF LA INFORMACION

ESTE ARCHIVO SERA ALMACENADO EN UN ARCHIVO

QUE SE LLAMA ESTADISTICA PARA DESPUES ACCESARLO

CON OTRO PROCEDIMIENTO DE S F S S

VI TO V4, V8, V9  
DISK

SE UTILIZARAN TRES SUBARCHIVOS EN ESTA CORRIDA

SUB(15), SUB(15), SUB(15)  
EL PRIMER SUBARCHIVO CONSTA DE 15 CASOS

EL SEGUNDO DE 15 CASOS Y EL TERCERO DE 13 CASOS

FREEFIELD

VI ZONA ASIGNADA/ V2 RIVEL/ V3 PESO ASIGNADO/  
V4 CANTIDAD DE COMPRA/ V8 VOLUMEN DE PRODUCCION/  
V9 CREDITO/

SE ETIQUETAN VALORES EN LAS VARIABLES

V1 (0) NO TUVO (1)PENDIENTE/  
V2 (2) FUMADOR (3) EN EL CESO/  
V3 (0)

LAS VARIABLES V8 Y V9 SE IMPRIMIRAN CON 4 DECIMALES

V8, V9(4)  
ALL

ESTADISTICAS BASICAS

SE UTILIZARA EL PROCEDIMIENTO CONDESCRIPTIVE

V2 TO V8

188

EJEMPLO DE ESTADISTICAS BASICAS  
ESTADISTICAS BASICAS  
STATISTICS 1,4  
STATISTICS ALL  
LAD INPUT DATA  
NO OF DATA INPUT HEAD COUNT = 13 DATA ERROR COUNT = 0.

VARIABLE V2 NIVEL

MEAN	7.349	STD ERROR	1.088	STD DEV	7.134
VARIANCE	50.869	KURTOSIS	10.202	SKEWNESS	3.230
RANGE	34.000	MINIMUM	2.000	MAXIMUM	36.000

VALID OBSERVATIONS = 43 MISSING OBSERVATIONS = 0

VARIABLE V3 PESO ASIGNADO

MEAN	-3.465	STD ERROR	4.749	STD DEV	11.140
VARIANCE	969.083	KURTOSIS	5.864	SKEWNESS	12.800
RANGE	118.000	MINIMUM	-95.000	MAXIMUM	15.000

VALID OBSERVATIONS = 43 MISSING OBSERVATIONS = 0

VARIABLE V4 CANTIDAD DE COMPRA

MEAN	7.140	STD ERROR	0.470	STD DEV	3.083
VARIANCE	9.504	KURTOSIS	3.243	SKEWNESS	1.819
RANGE	16.000	MINIMUM	2.000	MAXIMUM	18.000

VALID OBSERVATIONS = 43 MISSING OBSERVATIONS = 0

VARIABLE V6 VOLUMEN DE PRODUCCION

MEAN	7.372	STD ERROR	0.464	STD DEV	3.237
VARIANCE	10.477	KURTOSIS	6.569	SKEWNESS	2.035
RANGE	18.000	MINIMUM	3.000	MAXIMUM	21.000

VALID OBSERVATIONS = 43 MISSING OBSERVATIONS = 0