ANALISIS Y PREDICCION DE SERIES DE TIEMPO EN GEOLOGIA Y GEOFISICA

J.C.GIANIBELLI (*) y (**)

(*) I.N.I.P.- F.A.M.A. - UNIV. CATOLICA DE LA PLATA
(**) FAC. CS. ASTRONOMICAS Y GEOFISICAS - U.N.L.P.

RESUMEN.- Se realiza un estudio de las componentes de series de tiempo en geología y geofísica, y de los métodos de separación. Se presentan modelos determinísticos para dichas componentes, así como también métodos de estimación de parámetros.

ABSTRACT.- A study on geological aad geophysical time series components and separation methods has been carried out. Also are presented deterministic models for components and parametric estimation methods.

INTRODUCCION.- Dentro de las geociencias encargadas de estudiar el Geosistema y sus relaciones solares y lunares, se efectúan determinaciones del sistema y sus partes, con el objeto de evaluar su comportamiento pasado, presente y futuro. Esto se lleva a cabo a través de modelos que tratan de simular dicho comportamiento, para lo cual es necesario efectuar observaciones de diferentes parámetros del sistema a través del tiempo y del espacio.

Para el caso de observaciones temporales, se contruyen series de tiempo. Como definición diré que: "Una serie de tiempo es un conjunto de determinaciones en algún lenguaje lógico, de una o más características, de un sistema real a través del tiempo". A éste conjunto de series de tiempo que se corresponde a un sistema particular, se lo denomina "ensamble" que puede o no constituirse en series de datos "ergódicos" (ROBINSON, E. A., 1980). La propiedad de ergodicidad se establece mediante la coincidencia de los promedios sobre el "ensamble" (o conjunto de series de datos) respecto del promedio temporal sobre una serie de tiempo perteneciente al "ensamble":

$$P = E^{i}(x_t^{i}) = \lim_{T \to \infty} \frac{1}{2T+1} \sum_{t=-T}^{T} x_t$$

donde E^{i} indica promedio sobre el conjunto de series de datos, T el tiempo y x_t una serie particular sobre la que se efectúa la determinación de su promedio a través del tiempo.

CLASIFICACION DE EVENTOS EN SERIES DE TIEMPO.- La simple observación de registros o gráficas de los fenómenos en el tiempo facilita establecer una clasificación de los eventos en: periódicos, recurrentes y aleatorios, dejando en algunos casos de lado la componente de tendencia o también aperiodicidad. Fenomenológicamente, las componentes que se definen a continuación son parte de una serie de tiempo, de mayor o menor grado, directamente relacionadas con el proceso y estructura del sistema que genera dicho fenómeno.

EVENTOS:

1) Periódicos: corresponden a una fenomenología donde el sistema evidencia una acción del proceso en forma continua.

2) Recurrentes: son fenómenos que aparecen con cierta periodicidad o ciclicidad, pero que están ligados a una ley probabilística determinada en base al registro de la serie de tiempo.

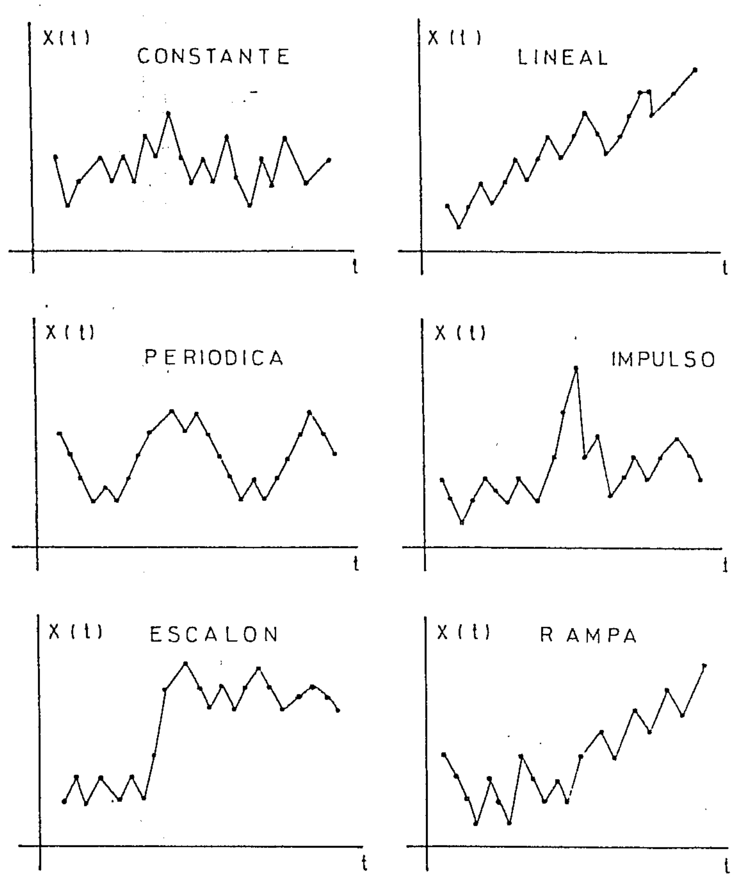
3) Aleatorios: la aleatoriedad corresponde a una respuesta del sistema a procesos que se desarrollan en forma errática e irregular. Corrientemente se la denomina ruido generado por el sistema, pero asimismo está vinculada con leyes probabilísticas o estadísticas del mismo.

4) Aperiodicidad: es la parte de la serie de tiempo que corresponde a la tendencia que presenta el fenómeno en general, por medio de las determinaciones, y que no puede evidenciarse como un macroperíodo.

En general, el ruido es generado por el sistema detector y por el método de medida, con sus errores de apreciación que producen una incertidumbre en el dato. Por ésto se mezclan comunmente la aleatoriedad, el ruido y los errores, en un conjunto de elementos que se caracterizan por ser erráticos en su comportamiento, denominándose todo el conjunto "ruido". El ruido compite en magnitud con las partes periódicas, recurrentes y de aperiodicidad (o 'tendencia') lo cual se conoce como 'relación señal/ruido'.

Sistemáticamente se tienen tres elementos en juego (fig.1).

En la fig. 2 se tienen ejemplificadas series de tiempo con características de tendencia (aperiodicidad), presencia de pulsos, fenómenos periódicos, escalones, rampas, y, en todos los casos, superpuesto a éstos eventos, aleatoriedad y ruido.



FORMAS ELEMENTALES DE SERIES DE TIEMPO
Fig. 2

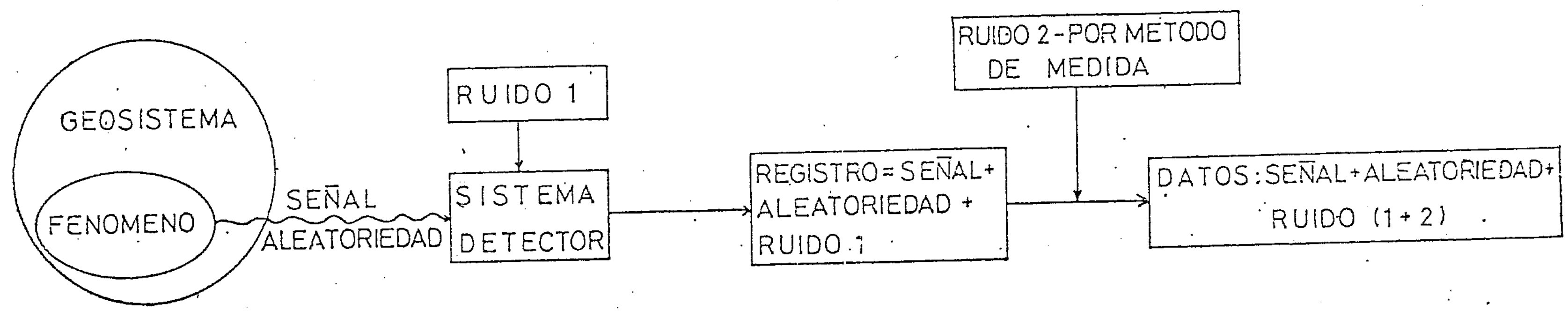

GEOSISTEMA
FENOMENO
SEÑAL
ALEATORIEDAD
RUIDO 1
SISTEMA DETECTOR
REGISTRO = SEÑAL + ALEATORIEDAD + RUIDO 1
RUIDO 2 - POR METODO DE MEDIDA
DATOS: SEÑAL + ALEATORIEDAD + RUIDO (1 + 2)

Fig. 1

METODOS DE ANALISIS.- Los métodos de análisis de series temporalmente variables en geología y geofísica tratan de determinar su estructura periódica, aleatoria y nivel de ruido, entendiéndose por periodicidad la propiedad de repetición en el tiempo de los eventos detectados. Por otra parte, la aleatoriedad es aquella porción de carácter aditiva o multiplicativa, con promedio temporal que puede ser nulo, y que aparece superpuesta a los procesos periódicos; mientras que el ruido tiene autocorrelación nula.

FILTRADO (O SUAVIZACION).- las diferentes técnicas en la toma de observaciones, mediciones o lectura de registros, realizan de por sí un filtrado de la información contenida. A pesar de ello, la aleatoriedad y el ruido presente hacen necesario efectuar un proceso de "alisamiento o suavización" que en definitiva conduce a un filtrado de los períodos pequeños y de una forma directa de determinación amortiguada de la intensidad (o amplitud) que presenta el proceso.

Los métodos de la ventana-filtro que se hace correr sobre la serie de datos, son en la mayoría de los casos centrados, de la forma:

$$\hat{F}(t_i) = \bar{K}\left\{ P(0)\,F(t_i) + \sum_{K=1}^{N} P(K)\,\left[F(t_i + K) + F(t_i - K)\right]\right\} \quad (1)$$

donde $F(t_i)$ representa la serie discreta de datos originales, $\hat{F}(t_i)$ representa los datos filtrados, $P(0)$ y $P(K)$ el vector de los pesos del filtro y $\bar{K} = \left[P(0) + \sum_{K=1}^{N} P(K)\right]^{-1}$ la constante de normalización del filtro.

Otra manera de efectuar el filtrado de los datos es produciendo un corte en el campo de las frecuencias presentes en el espectro de Fourier, de tal manera que al reconstituir la serie de datos por medio de la transformada inversa, se obtiene una señal depurada. En otras palabras, se pasa del campo temporal al de las frecuencias; en éste último se aplica un filtro, y se retoma el campo del tiempo.

DETERMINACION DE LAS COMPONENTES NO CICLICAS.- Estas componentes corresponden a la aproximación por medio de una función polinómica de grado n, o a funciones trascendentes no periódicas de mejor ajuste.

El método más apropiado de ajuste de los parámetros es por medio de mínimos cuadrados, tal que las ecuaciones normales para el ajuste polinómico $b_n(t_i)$ de las observaciones y_i en el instante t_i están dadas por:

$$\begin{bmatrix} N & (\Sigma t_i) & \dots & (\Sigma t_i^n) \\ (\Sigma t_i) & (\Sigma t_i^2) & \dots & (\Sigma t_i^{n+1}) \\ \dots & \dots & \dots & \dots \\ (\Sigma t_i^n) & (\Sigma t_i^{n+1}) & \dots & (\Sigma t_i^{2n}) \end{bmatrix} \begin{bmatrix} b_o \\ b_1 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} (\Sigma y_i) \\ (\Sigma y_i t_i) \\ \vdots \\ (\Sigma y_i t_i^n) \end{bmatrix}$$

$$\text{para } y_i = p_n(t_i) = b_o + b_1 t_i + \dots + b_n t_i^n$$

Otros modelos están basados en funciones exponenciales, logarítmicas y potenciales de la forma:

$$\begin{cases} y = a\, e^{bt} \\ y = a + b \ln t \end{cases} \qquad \begin{matrix} y = a\, t^b \\ y = a(t + k)^{-b} \text{ ,con } k{=}cte. \end{matrix} \qquad (3)$$

Es posible reducir estas expresiones a formas lineales mediante una transformación logarítmica y efectuar la determinación de los parámetros por medio de la fórmula (1).

ANALISIS ARMONICO.- Las variaciones periódicas o cíclicas son posibles de modelizar a través de funciones de su misma característica, es decir, combinación entre senos y cosenos de la forma de una serie de Fourier finita:

$$F(t) = \frac{a_o}{2} + \sum_{n=1}^{N/2} \left[a_n \cos\left(\frac{2\pi nt}{N}\right) + b_n \operatorname{sen}\left(\frac{2\pi nt}{N}\right) \right] \quad (4)$$

correspondiendo el corte del número de componentes armónicas a_n y b_n a la mitad de la cantidad de datos digitizados (GECKINLI, N. C. and YAVUZ, D., 1983).

La forma de cálculo de las componentes armónicas es a través de las fórmulas siguientes:

$$\begin{cases} a_n = \dfrac{2}{N} \displaystyle\sum_{j=1}^{N} F(t_j) \cos\left(\dfrac{2\pi n t_j}{N}\right) \\ b_n = \dfrac{2}{N} \displaystyle\sum_{j=1}^{N} F(t_j) \operatorname{sen}\left(\dfrac{2\pi n t_j}{N}\right) \end{cases} \qquad (5)$$

para $n=0, 1, 2, \ldots, N/2$

Para el caso en que los valores $F(t)$ sean datos promediados sobre algún intervalo Δt, es necesario entonces corregir las componentes armónicas para $n \geqslant 1$ por medio de la función:

$$A(n) = \frac{(\pi n/N)}{\operatorname{sen}(\pi n/N)} \qquad (6)$$

denominada 'corrector por achatamiento'. Asimismo la fase de la onda debe ser corregida si se adopta el valor del promedio de los datos a los extremos de este intervalo Δt.

ANALISIS ESPECTRAL.- Cuando los eventos que caracterizan un fenómeno del geosistema se encuentran enmascarados o tienen presunta recurrencia cíclica, se utiliza el análisis espectral con el fin de observar en la función de potencia espectral la presencia y la frecuencia de dicha señal (DAVIS, J. C., 1973; KANASEWICH, E. R., 1981).

Las formas estándares de análisis espectral son esencialmente una generalización del análisis armónico, obteniéndose el espectro de amplitud y fase por medio de:

$$P(f_n) = \int_{-\infty}^{+\infty} F(t)\, e^{-i2\pi f_n t}\, dt = \text{cte. para la forma contínua} \qquad (7)$$

y por $$P(f_n = n f_o) = \frac{1}{T_o} \int_{-To/2}^{+To/2} F(t)\, e^{-i2\pi n f_o t}\, dt \qquad (8)$$

siendo fo= 1/To, para el caso discreto (con $n=0, \pm 1, \pm 2, \ldots$).

Esta última fórmula permite relacionar este espectro discreto con la forma generalizada del análisis armónico.

Sea $\alpha_n = P(f_n = n f_o)$, entonces

$$F(t) = \sum_{n=-\infty}^{+\infty} (\alpha n)\, e^{i2\pi nfot} \qquad (9)$$

con $\alpha n = \frac{1}{2}(n - ib_n)$ y $n=0, \pm 1, \pm 2, \ldots$

De igual manera se vincula el análisis armónico con el espectro de amplitud de Fourier a través de la fórmula (9).

El correspondiente espectro de fase se obtiene mediante la relación:

$$(fn = nfo) = \text{Arc tang}\left(-\frac{bn}{an}\right) \qquad (10)$$

Las fórmulas (9) y (10) se denominan espectros de amplitud y fase discretos respectivamente, y pueden ser calculados para series de hasta 256 datos (N), con un tiempo de procesamiento para computadoras del tipo PC (Texas, IBM) de 25 a 45 minutos. Para microprocesadoras del tipo Sharp, Cassio, HP (con CPU $\leq$ 16 Kb) se utiliza un tiempo de proceso del mismo órden pero tomándose hasta N $\simeq$ 96 datos.

Para series de tiempo cuya digitización brinda más de 256 datos, es posible obtener el espectro de Fourier mediante una forma directa y rápida utilizándose la Transformada Rápida de Fourier, basada en la expresión:

$$P(n) = \frac{2}{N} \sum_{n=0}^{N-1} F(k)\, e^{-i(2\pi nk)/N} \text{, con } n=0,1,2,\ldots,(N-1) \qquad (11)$$

Reemplazando $W = e^{-i2\pi/N}$, la expresión (11) se reduce a:

$$P(n) = \frac{2}{N} \sum_{n=0}^{N-1} F(k)\, W^{nk} \qquad \text{con nk entero} \qquad (12)$$

Para el caso extremo de tener una F(ti) compleja, el número de multiplicaciones es N^2 y el de adiciones es $N(N-1)$, para la determinación del espectro en forma standard.

La aplicación del algoritmo de Cooley y Tukey en la fórmula (12) permite reducir notablemente el número de multiplicaciones y adiciones en el caso de ser N una potencia entera de 2. Un ejemplo de ésto se muestra en la fig. 3.

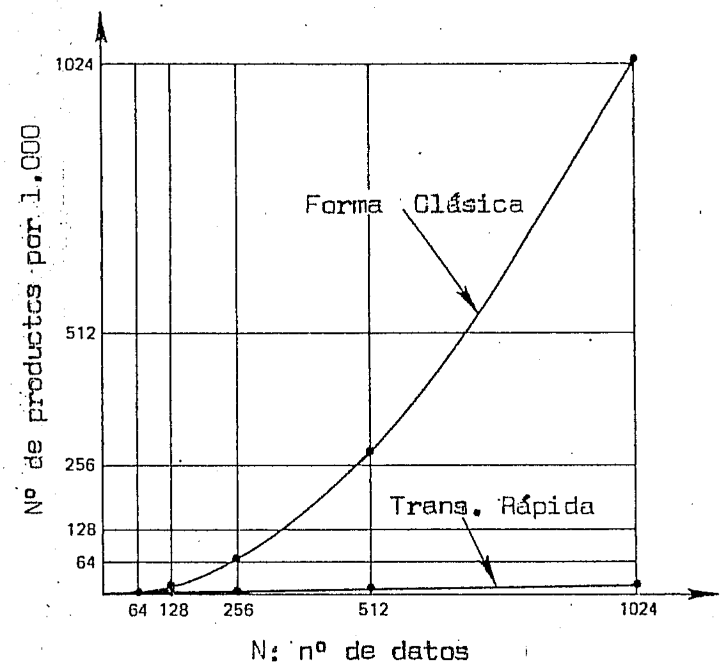


Fig. 3

La forma de P(n) es compleja, de tal manera que la relación con la forma generalizada del análisis armónico (9) está dada por:

$a_n = \mathrm{Re}(P(n))$ Re es parte real

$b_n = \mathrm{Im}(P(n))$ Im es parte imaginaria

Se hace notar que el espectro de potencia es $P(n)=\frac{2}{N}p(n)$, donde p(n) es la transformada rápida de Fourier de la función del tiempo F(k), es decir que $p(n)=\sum_{k=0}^{N-1} F(k)\, W^{nk}$.

Con éste método es posible determinar la transformada rápida de Fourier de dos señales digitizadas de N valores, simultáneamente, utilizándose la posibilidad de que F(k) sea compleja.

Sean entonces d(k) y t(k) (k=0,1,2,...,N-1), los datos de las funciones. Se forma luego la función $F(k) = b_k + it(k)$. Se efectúa después la transformada $X(n) = \sum_{k=0}^{N-1} F(k)\, e^{-i2\pi nk/N} = A(n) + iB(n)$ (n=0,1,2,...,N-1) donde A(n)=Re(X(n)) y B(n)=Im(X(n)).

La transformada rápida de t_k y d_k está dada por:

$$D(n) = \frac{1}{2}\left\{\left[A(n)+A(N-n)\right] + i\left[B(n)-B(N-n)\right]\right\}$$
$$T(n) = \frac{1}{2}\left\{\left[B(n)+B(N-n)\right] - i\left[A(n)-A(N-n)\right]\right\}$$

donde $D(n)$ y $T(n)$ son las transformadas discretas de $d(k)$ y $t(k)$, (BRIGHAM, E. O., 1974).

También se puede, para el caso de series de tiempo extremadamente extensas $(N > 2^{12}=4096)$, utilizar éste método cortando la serie en la mitad y rearmando una función $F(k)$ con la mitad en su parte real y la otra en su parte imaginaria, según el siguiente esquema:

Sea $F(k)$ una función real con $k=0,1,2,\ldots,2N-1$ datos; se efectúa una división de los datos en dos, formándose así dos funciones:

$d(k)=F(2k)$ y $t(k)=F(2k+1)$ $(k=0,1,2,\ldots,N-1)$

Se genera entonces una nueva función compleja

$f(k)= d(k) + it(k)$

Se aplica la transformada rápida de Fourier

$$X(n)= \sum_{n=0}^{N-1} f(k)\, e^{-i2\pi nk/N} = XR(n) + iXI(n) \qquad n=0,1,2,\ldots,N-1$$

donde $XR(n)=Re(X(n))$ y $XI(n)=Im(X(n))$.

Efectuando el cálculo siguiente se tiene la parte real e imaginaria de la transformada de $F(k)$:

$$TF(n)= TFr(n) + iTFi(n)$$

$$TFr(n) = \left[\frac{XR(n)+XR(N-n)}{2}\right] + \cos\left(\frac{n\pi}{N}\right)\left[\frac{XI(n)+XI(N-n)}{2}\right] -$$
$$- \operatorname{sen}\left(\frac{n\pi}{2}\right)\left[\frac{XR(n)-XR(N-n)}{2}\right]$$

$$TFi(n) = \left[\frac{XI(n)-XI(N-n)}{2}\right] - \operatorname{sen}\left(\frac{n\pi}{N}\right)\left[\frac{XI(n)+XI(N-n)}{2}\right] -$$
$$- \cos\left(\frac{n\pi}{N}\right)\left[\frac{XR(n)-XR(N-N)}{2}\right] \qquad n=0,1,2,\ldots,N-1$$

TEORIAS Y MODELOS: CONCLUSIONES.- Toda serie de tiempo está constituida por componentes de la forma siguiente:

TENDENCIA + COMPONENTE PERIODICA + COMPONENTE RECURRENTE +
+ COMPONENTE ALEATORIA + RUIDO (13)

a dichas compontes es posible aislarlas en mayor o menor grado por los métodos descriptos. Para éste caso la serie de tiempo presenta una forma aditiva de las componentes, pero éstas pueden aparecer relacionadas multiplicativamente entre sí, de manera parcial o total.

La expresión (13) constituye un modelo de comportamiento de series de tiempo y a partir de él es posible establecer las formas de funciones explícitas para constituir un modelo determinístico (problema directo). En geología y geofísica la mayoría de los fenómenos se presentan en forme de modelos aditivos, tales como las series de datos meteorológicos, hidrogeológicos, hidrológicos, geomagnéticos,etc.

La fig. 4 muestra un esquema de las relaciones existentes entre las funciones, el modelo, su síntesis, el pronóstico, las observaciones y los parámetros del modelo sustentado por relaciones funcionales.

Todo modelo determinístico pretende pronosticar en un número finito de períodos, valores de la serie de tiempo que podría generar el fenómeno del geosistema.

La comparación en tiempo real de los datos pronosticados por el modelo y los observados, facilita la adaptación dinámica del modelo mediante los cambios en sus parámetros característicos.

La fig. 5 muestra en bloques el comportamiento del sistema de pronóstico para series de tiempo, que coincide con el método de Teoría Inversa por análisis de datos brindados por las Geociencias, indicándose que el bloque "Modificación de los parámetros del modelo" significa los cambios y la estimación de los parámetros que intervienen directa e indirectamente en el modelo (análisis factorial).

De la metodología directa del análisis de los datos geofísicos y geológicos (MENKE, W., 1984), se tiene en base a las fórmulas:

PARAMETROS DEL MODELO $\longrightarrow$ MODELO $\longrightarrow$ PRONOSTICO DE LOS DATOS

y, de la metodología inversa:

DATOS $\longrightarrow$ MODELIZACION $\longrightarrow$ ESTIMACION DE LOS PARAMETROS

Es posible cambiar los dos métodos en uno adaptativo que permita obtener un sistema de tratamiento de los datos a través de modelos, síntesis y simulación en forma directa e inversa de manera organizada (fig.4). Este sistema se denominará entonces "Modelo Directo - Inverso de Análisis de Datos Geomómicos", basado en la combinación de las teorías Directas e Inversas del tratamiento de dichos datos.

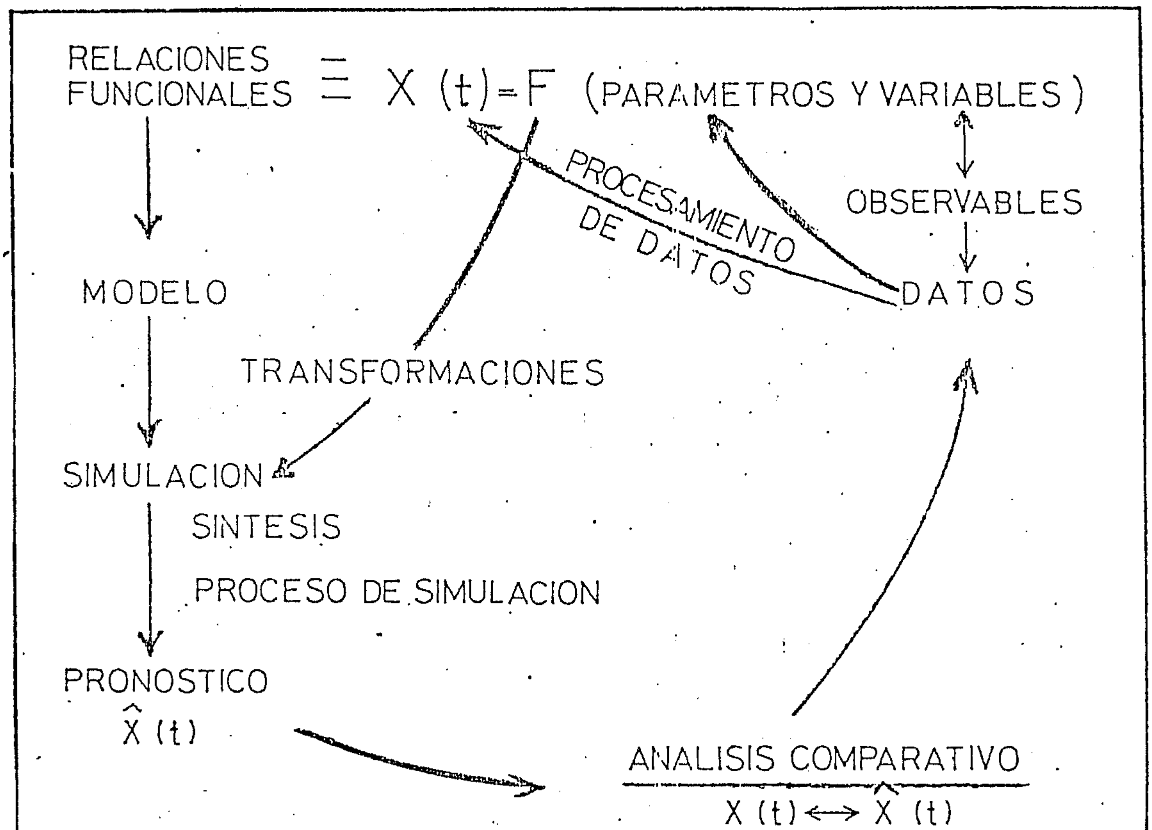


Fig. 4

La forma de aplicación estaría basada en el diagrama de la fig. 6.

Se podría modificar el modelo ya sea por sus parámetros o por sus relaciones funcionales en la Teoría Inversa. Mientras que en la Teoría Directa el proceso de síntesis permite evaluar datos predictivos, en la Teoría Inversa se busca adaptar el modelo a los datos existentes y estimar parámetros cuyo valor es desconocido pero que se encuentran en las relaciones funcionales del modelo.

El sistema propuesto vincula las dos metodologías a través de un procedimiento adaptativo sobre el modelo y de un proceso de síntesis, en contínuo intercambio comparativo con el sistema real existente.

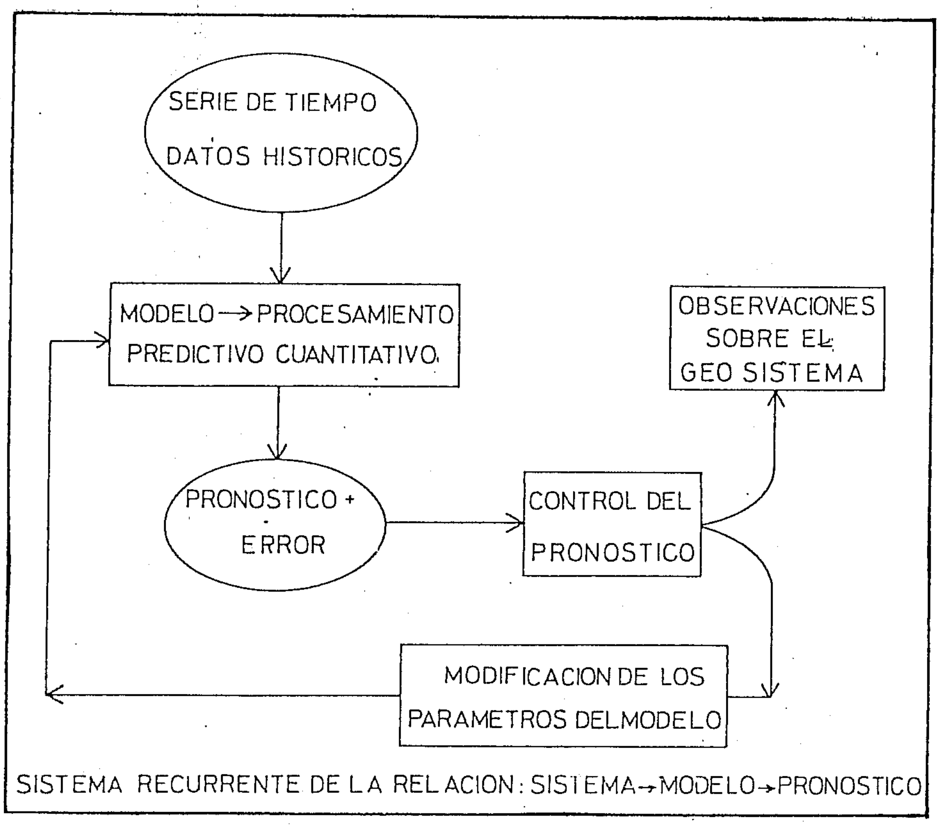

SERIE DE TIEMPO
DATOS HISTORICOS
MODELO → PROCESAMIENTO
PREDICTIVO CUANTITATIVO
OBSERVACIONES
SOBRE EL
GEO SISTEMA
PRONOSTICO +
ERROR
CONTROL DEL
PRONOSTICO
MODIFICACION DE LOS
PARAMETROS DEL MODELO
SISTEMA RECURRENTE DE LA RELACION: SISTEMA → MODELO → PRONOSTICO

Fig. 5

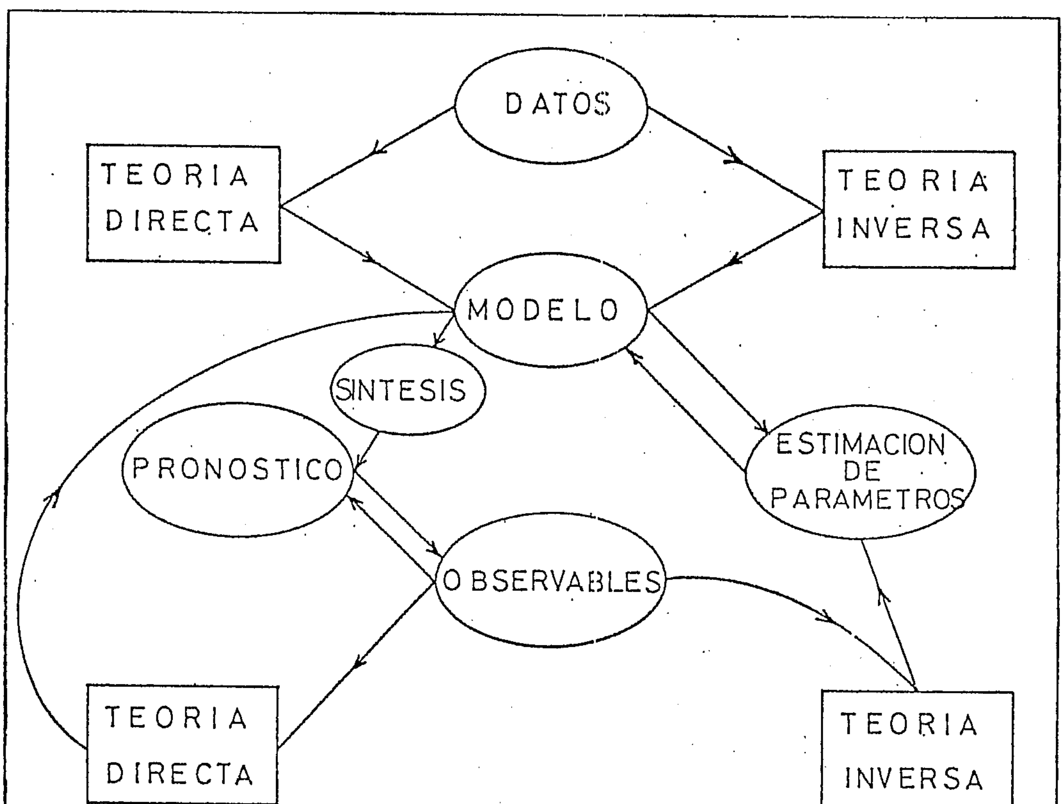

DATOS
TEORIA
DIRECTA
TEORIA
INVERSA
MODELO
SINTESIS
PRONOSTICO
ESTIMACION
DE
PARAMETROS
OBSERVABLES
TEORIA
DIRECTA
TEORIA
INVERSA

Fig. 6

BIBLIOGRAFIA.-

- BRIGHAM, E. O., 1974.- The Fast Fourier Transform.- Ed. Prentice Hall.
- COOLEY, J. W. and TUKEY, J. W., 1965.- Math. Comp., vol.19, p.297-301.
- DAVIS, J. C.,1973.- Statics and Data Analysis in Geology.- Ed. J. Wiley, N. Y.
- GECKINLI, N. C. and YAVUZ, D., 1983.- Discrete Fourier Transformation and its applications to Power Spectra Estimation.-Elsevier Pub. Co.
- KANASEWICH, E. R.,1981.- Time Sequences Analysis in Geophysics.- The University of Alberta.
- MENKE, W.,1984.- Geophysical Data Analysis Discrete Inverse Theory.- Academic Press.
- ROBINSON, E. A.,1980.- Physical Applications of Stationary Time-Series.- Charles Griffin Co.,London.