

INSTITUTO DE CIBERNETICA, MATEMATICA Y FÍSICA

Clasificación no supervisada por métodos lógico-combinatorios en problemas de zonación sísmica.

Dr. Leonardo Alvarez
MC. Ramón Pico Peña
Dr. Mario Cotilla

Diciembre 1995

ICIMAF 95-09
CEMAFIT 95-06

Centro de Matemática y Física Teórica

Calle E # 309 esquina a 15, Vedado, Ciudad Habana, C.P. 10400, telf: 31
Fax: (537) 33 3373, Telex: 51 2230 ICIMAF CU

CLASIFICACION NO SUPERVISADA POR METODOS LOGICO-COMBINATORIOS EN PROBLEMAS DE ZONACION SISMICA

Leonardo Alvarez¹, Ramón Pico², Mario Cotilla³

Resumen

La primera etapa de una zonación sísmica son los mapas con información acerca del comportamiento espacial de indicadores particulares de la peligrosidad (intensidades máximas, intensidades para tiempos de recurrencia fijos o combinaciones de tiempos de espera y probabilidades de no excedencia). Se obtuvo un nuevo tipo de zonación procesando los períodos de recurrencia de cada punto con el sistema lógico-combinatorio de reconocimiento de patrones PROGNOSIS. Dichos períodos de recurrencia para las islas de Jamaica y La Española fueron obtenidos de un trabajo de estimación de la peligrosidad sísmica realizado con anterioridad. Se seleccionó un algoritmo de reconocimiento no supervisado para la clasificación. Se presentan los resultados -grupos de puntos con comportamiento similar de la peligrosidad, caracterizados por uno de ellos llamado "holotipo"- en un mapa de zonación complementado con las curvas de los períodos de recurrencia para los holotipos.

Abstract

First stage on seismic zoning are maps with information about spatial behavior of particular hazard parameters (maximum intensities, intensities for fixed recurrence times or combinations of waiting times and non-excedance probabilities). New kind of zoning was obtained by processing recurrence periods of every point with logical-combinatorial pattern recognition system PROGNOSIS. That recurrence periods for Jamaica and Hispaniola islands were taken from a seismic hazard assessment work previously performed. A non-supervised algorithm was selected for classification. Results -groups of points with similar hazard behavior, characterized by one of them called "holotype"- are presented in a zoning map together with recurrence period curves for holotypes.

Keywords

Seismic hazard, seismic zoning, non-supervised pattern recognition, classification

¹Centro Nacional de Investigaciones Sismológicas

²Instituto de Cibernética, Matemática y Física

³Instituto de Geofísica y Astronomía

Todos del Ministerio de Ciencia, Tecnología y Medio Ambiente

Introducción

El resultado básico de un trabajo de estimación de la peligrosidad sísmica está formado por los períodos de recurrencia (o retorno) de las sacudidas con valores del efecto mayores o iguales que una magnitud dada. Este efecto puede ser intensidad sísmica, aceleración, velocidad o cualquier otro parámetro de interés ingenieril. Cuando se estudian zonas amplias, dichos valores se calculan para todos los puntos de redes regulares, por lo que podemos hablar de resultados en forma de matrices de períodos de recurrencia o retorno $T(\alpha_k)_{ij}$ [T - período de recurrencia, α_i - efecto]. Si se añade un modelo estadístico de ocurrencia de las sacudidas, se pueden hacer estimados de la probabilidad de no excedencia en diferentes intervalos $p(\alpha_k, t)_{ij}$ [p -probabilidad, t -tiempo, llamado de espera]. Sin embargo la forma común de representar gráficamente estos estimados es a través de mapas del tipo $\alpha(T)$, $\alpha(p, t)$, y α_{\max} - tantos como valores de T o de (p, t) se seleccionen -, por lo que se requiere un procesamiento ulterior de los resultados iniciales.

Para la confección de un mapa de zonación sísmica de la forma tradicional se reelabora uno de los mapas obtenidos en la etapa anterior para convertirlo en un mapa de zonas de comportamiento "similar" de las manifestaciones del efecto sísmico. Algunas veces se toman como fronteras de las zonas las isolíneas representadas, en otros casos se delimitan franjas y en numerosas ocasiones las fronteras son "rectificadas" atendiendo a la división político-administrativa del territorio. Sin embargo, cada uno de estos mapas es una realización particular del fenómeno estudiado, lo que hace que el mapa final lleve un carga subjetiva grande, y no se pueda garantizar que dentro de las zonas representadas sean uniformes las manifestaciones del efecto sísmico. Es por ello que en este trabajo se busca realizar la zonación directamente desde las matrices de períodos de recurrencia $T(I)_{ij}$ usando técnicas de clasificación no supervisada mediante el algoritmo lógico-combinatorio denominado HOLOTIPO, incluido en el sistema PROGNOSIS (ICIMAF, 1992) para el caso de las islas de Jamaica y La Española.

Modelo matemático

Los algoritmos lógico-combinatorios para el reconocimiento de patrones fueron introducidos por especialistas de la extinta URSS desde 1965, y fueron desarrollados posteriormente por Zhuravliov en múltiples trabajos, resumidos en (Zhuravliov, 1978). Paralelamente, en la propia URSS se desarrollaron los trabajos de clasificación no supervisada basados en el estudio de la estructura interna de los conjuntos de objetos (Voronin et al., 1968; Sirotinskaya, 1986). Ambas teorías fueron generalizadas para el uso de más tipos de variables y de criterios de comparación lógico-combinatorios más potentes que los de los trabajos iniciales (Ruiz, 1992; Ruiz et al., 1992). La formulación que sigue es tomada de estos últimos trabajos.

Sea U el universo de objetos admisibles, $MI = \{ I(O_1), \dots, I(O_m) \} \subseteq U$ una muestra inicial de objetos descritos en términos de los rasgos $R = \{ x_1, \dots, x_n \}$. A cada x_i se le asocia un conjunto de valores admisibles M_i , $i = 1, \dots, n$ de naturaleza diferente. Estas x_i son llamadas también variables.

Definición 1:

Sea $\Omega = \{x_{S_1}, \dots, x_{S_p}\}$. Por ϖ -parte de una descripción $I(O_i) = (x_1(O_i), \dots, x_n(O_i))$ entenderemos una subdescripción $(x_{S_1}(O_i), \dots, x_{S_p}(O_i))$. A Ω lo llamaremos **conjunto de apoyo**.

Definición 2:

Por **sistema de conjuntos de apoyo** Ω_A entenderemos un conjunto de conjuntos de apoyo.

Definición 3:

Un **criterio de comparación de valores de un rasgo** x_i es una función $C_t : M_t \times M_t \rightarrow T_t$, $t = 1, \dots, n$ donde M_t es el conjunto de valores admisibles de x_t y $T_t \subseteq [0, 1]$.

Estos criterios, en general, pueden ser de semejanza, de diferencia y en ambos casos booleanos, k-valentes, reales, difusos.

El sistema PROGNOSIS tiene instrumentados diferentes criterios de los tipos mencionados (ICIMAF, 1992), pero en el presente trabajo sólo se usa un criterio de semejanza que es el siguiente:

-Error de semejanza.

$$C_t(x_t(O_i), x_t(O_j)) = \begin{cases} 1 & \text{si } |x_t(O_i) - x_t(O_j)| \leq \varepsilon_t \vee x_t(O_i) = ? \vee x_t(O_j) = ? \\ 0 & \text{en otro caso} \end{cases}$$

donde $\varepsilon_t \geq 0$ es un umbral asociado a cada rasgo x_t . Por supuesto que para poder usar este criterio de comparación es necesario que los valores de los rasgos que estemos considerando admitan las operaciones señaladas, es decir, sean números.

Definición 4:

Sea $\Omega_S = \{x_{S_1}, \dots, x_{S_p}\} \in \Omega_A$. Una **función de semejanza parcial** es un operador $\beta_\varpi : \prod_{t \in \{S_1, \dots, S_p\}} T_t \rightarrow Q_S$ siendo Π el producto cartesiano, donde para cada $t \in \{S_1, \dots, S_p\}$, T_t es la imagen del criterio de comparación C_t , donde Q_S es el conjunto de valores resultante de una función de semejanza parcial para un conjunto de apoyo Ω_S , $S = 1, \dots, k$. Q_S puede ser de dominio booleano o estar en el intervalo $[0, 1]$.

Formalmente β_ϖ también podría definirse sobre $\prod_{t \in \{S_1, \dots, S_p\}} M_t \times \prod_{t \in \{S_1, \dots, S_p\}} M_t$ por lo que usaremos

la notación $\beta_\varpi(I(O_i), I(O_j))$ en lugar de $\beta_\varpi(C_{S_1}(x_{S_1}(O_i), x_{S_1}(O_j)), \dots, C_{S_p}(x_{S_p}(O_i), x_{S_p}(O_j)))$.

Funciones de Semejanza Parcial que se emplean en este trabajo:

Al igual que los criterios de comparación, las funciones de semejanza también pueden ser booleanas o reales. Entre las incluidas en el sistema PROGNOSIS, en el presente trabajo se usan::

1. Función de semejanza con umbral de error admisible.

$$\beta_{\varpi}(I(O_i), I(O_j)) = \begin{cases} 1 & \text{si } \left| \left\{ x_t : C_t(x_t(O_i), x_t(O_j)) = 0 \right\} \right| \leq \varepsilon \\ 0 & \text{en otro caso} \end{cases}$$

donde $|P|$ es el cardinal del conjunto P , siendo C_t un criterio de comparación booleano, de tipo a), y ε un umbral dado por el usuario.

En este caso expresamos que las descripciones de los objetos comparados son o no semejantes.

Dos descripciones de objetos son semejantes si la cantidad de variables cuyos valores en dichos objetos no son semejantes, no excede un número prefijado ε .

2. Suma ponderada del valor de semejanza entre los rasgos.

$$\beta_{\varpi}(I(O_i), I(O_j)) = \frac{1}{n} \sum_{p=1}^n C_p(x_p(O_i), x_p(O_j))$$

siendo n el número de variables que se están considerando y C_p un criterio de comparación cualquiera de los anteriormente definidos.

Definición 5:

Sea $\Omega_A = \{\Omega_1, \dots, \Omega_k\}$. Una **función de semejanza (total)** es un operador $\Gamma : \prod_{s=1}^k Q_s \rightarrow L$, $L \subseteq [0, 1]$.

El PROGNOSIS tiene incluida la siguiente función de semejanza:

$$\Gamma(O_i, O_j) = \frac{\gamma(O_{ij})}{K} \sum_{s=1}^k \rho(\Omega_s) \varpi_s \beta_{ij}$$

siendo $\varpi_s \beta_{ij} = \beta_{\varpi}(I(O_i), I(O_j))$; $\gamma(O_{ij})$ un parámetro de ponderación asociado a la comparación entre los objetos O_i y O_j ; análogamente $\rho(\Omega_s)$ es un parámetro de ponderación asociado al conjunto de apoyo Ω_s y pudiera estar en función del peso informacional de los rasgos que aparecen en dicho conjunto de apoyo, por ejemplo:

$$\rho(\Omega_s) = \frac{\sum_{x \in \Omega_s} \rho(x)}{|\Omega_s|} \text{ donde } \rho(x) \text{ es el peso informacional del rasgo } x.$$

Definición 6:

Dado $O_i, O_j \in MI$, diremos que O_i y O_j son β_0 *semejantes* si y sólo si $\Gamma(O_i, O_j) \geq \beta_0$. Diremos además que $O_i \in MI$ es β_0 *aislado* si y sólo si $\forall O_j \in MI [O_j \neq O_i \rightarrow \Gamma(O_i, O_j) < \beta_0]$.

La magnitud $\beta_0 \in L$, la denominamos *umbral de semejanza* y puede ser definida :

$$a) \beta_0 = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \Gamma(O_i, O_j)$$

$$b) \beta_0 = \frac{1}{n} \sum_{i=1}^n \max_{j=1, \dots, n} \{ \Gamma(O_i, O_j) \}$$

c) valor entre 0 y 1 seleccionado a voluntad

Criterios de clasificación

Criterio β_0 conexo

Definición 7

Sea $\beta_0 \in L$. Un subconjunto $C \neq \emptyset$ de MI es una *componente β_0 conexa* si y sólo si:

- a) $\forall O_i, O_j \in C \exists O_{i_1}, \dots, O_{i_s} \in C [O_i = O_{i_1} \wedge O_j = O_{i_s} \wedge \forall p \in \{1, \dots, s-1\} \rightarrow \Gamma(O_p, O_{p+1}) \geq \beta_0]$.
- b) $\forall O_i \in MI [(O_j \in C \wedge \Gamma(O_i, O_j) \geq \beta_0) \Rightarrow O_i \in C]$.
- c) Todo objeto β_0 aislado es una componente β_0 conexa.

Este criterio es la base del algoritmo HOLOTIPO (Voronin et al., 1968).

Criterio β_0 compacto

Definición 8

Diremos que $C \subseteq MI$, $C \neq \emptyset$ es un *conjunto β_0 -compacto* si y sólo si:

- a) $\forall O_j \in MI [(O_i \in C \wedge \max_{\substack{O_t \in MI \\ O_t \neq O_i}} \{ \Gamma(O_i, O_t) \} = \Gamma(O_i, O_j) \geq \beta_0] \Rightarrow O_j \in C$
- b) $[\max_{\substack{O_i \in MI \\ O_i \neq O_p}} \{ \Gamma(O_p, O_i) \} = \Gamma(O_p, O_t) \geq \beta_0 \wedge O_t \in B] \Rightarrow O_p \in C$
- c) $|C|$ es la mínima
- d) Todo elemento β_0 -aislado constituye un conjunto β_0 -compacto (degenerado).

Este criterio es utilizado en los algoritmos tipo CLASS (Sirovinskaya, 1986).

Con las ideas de componentes β_0 -conexas y conjuntos β_0 -compactos, en general tenemos dos formas de agrupar o clasificar los objetos, susceptibles de diferentes interpretaciones y usos. A partir de estos agrupamientos, se puede continuar con la idea de Karataeva de buscar los holotipos (Voronin et al., 1968).

En ambos criterios se construye una matriz de semejanza (MS) en la que está contenida el resultado de comparar cada objeto con todos los de MI. Obviamente $MS = \|\Gamma(O_i, O_j)\|_{mxm}$ es una matriz simétrica y en la diagonal alcanza el máximo. La determinación del parámetro β_0 para ambos criterios de agrupamiento, no obstante existir algunos métodos de cálculo, resulta un problema engorroso (se calcula exhaustivamente) cuando se busca el adecuado para una clasificación con un número fijo de grupos. Mediante una clasificación jerárquica y su representación gráfica en dendrograma (o árbol de clasificación) es posible, para el caso del criterio β_0 -conexo, determinar el β_0 correspondiente a la partición del conjunto inicial en el número de grupos previamente determinado. Este valor se escoge tomando el nivel de asociación a partir del cual se divide el árbol en el número de ramas que se desee (Pico, 1995).

Holotipos.

Finalmente, dados los agrupamientos obtenidos sobre la base de un criterio de agrupamiento, de la función de semejanza escogida y del valor de umbral de semejanza β_0 , se procede a determinar lo que se denomina el coeficiente de tipicidad de cada objeto, y a partir de éstos, a seleccionar los holotipos de cada grupo.

Definición 9

Por *coeficiente de tipicidad* entendemos el coeficiente:

$$t(O_i) = t_i = \frac{\bar{\Gamma}_i}{\frac{1}{G-1} \sum_{j=1, \dots, G} (\bar{\Gamma}_i - \Gamma_{ij})^2}, \quad \text{donde } \bar{\Gamma}_i = \frac{1}{G-1} \sum_{j \in \{t / t \neq i \wedge O_t \in G(O_i)\}} \Gamma_{ij}$$

$G(O_i)$ es una componente β_0 -conexa o un conjunto β_0 -compacto al que pertenece O_i (tiene como raíz al objeto O_i) y $|G(O_i)| = G > 1$.

Definición 10

Se define como *holotipo* de $G(O_i)$ al objeto O_h tal que $t_h = \max_{O_c \in G(O_i)} \{t_c\}$ es decir, su coeficiente de tipicidad es máximo.

Luego, con la aplicación de uno de estos criterios se obtiene una partición de MI, y además el objeto representativo de cada grupo.

Peligrosidad sísmica de las islas de Jamaica y La Española

Las regiones de La Española y Jamaica han sido objeto de diversos estudios de sismicidad, sobre todo la primera, aunque han sido poco estudiadas desde el punto de vista de la peligrosidad sísmica a que están sometidas, pudiéndose señalar en este sentido como más importantes un trabajo de Reyes (1976) en que se presentan mapas de isoaceleración e isovelocidad con un período de retorno de 50 años para Jamaica, La Española y Puerto Rico, obtenidos a partir de una base de datos muy pobre con un procedimiento estándar de estimación de la peligrosidad sísmica por métodos probabilistas, y otra de Shepherd y Aspinall (1980) donde se presenta un mapa de zonación por intensidades sísmicas en base a datos históricos para Jamaica. La preparación de un catálogo de terremotos desde el siglo XVI hasta 1988 para la región [16°-24° N., 70°-86° W.] (Alvarez et al. inédito), unido a estudios de la sismicidad histórica (Chuy y Alvarez, 1988) y a los datos más recientes de las redes internacionales, permitió pasar a la realización de estudios más profundos, aunque aún con carácter preliminar, de la sismicidad y la peligrosidad sísmica asociada de ambos territorios.

La sismicidad de Jamaica es somera ($h \leq 70$ km.), mientras que La Española se caracteriza por una elevada sismicidad somera ampliamente distribuida pero mayoritaria en su parte norte y por una fuerte presencia de terremotos intermedios ($70 < h \leq 150$ km.) en su mitad centro oriental, que se extiende en profundidad hasta unos 220 km., pero mucho más escasa. Las principales zonas de origen de terremotos (Fig. 1) forman bandas con ancho entre 40 y 60 km; en Jamaica una al norte y otra al sur, en La Española 1 al norte y 2 al sur también, una en la región entre ambas islas (todas ellas de terremotos someros) y una en La Española centro oriental de terremotos intermedios. Las magnitudes máximas fueron estimadas como 0.05 - 0.3 unidades sobre el máximo reportado en cada caso (Cotilla et al., 1991).

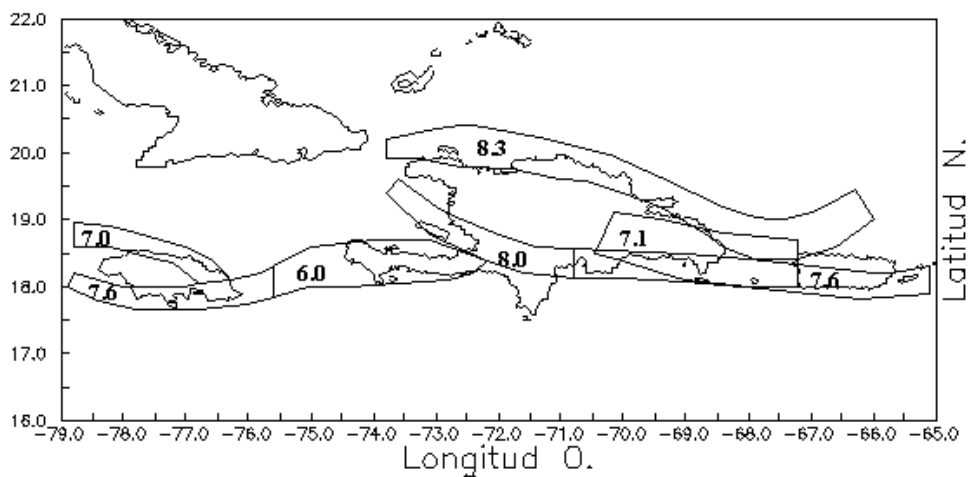


Fig. 1. Esquema de las principales zonas de origen de terremotos para la región de estudio. Sobre las mismas se indican los valores de M_{max} correspondientes.

Alvarez et al. (1994) realizaron estimados de peligrosidad sísmica para ambas islas en términos de intensidad sísmica ($\alpha=I$) usando el programa SACUDIDA (Alvarez y Bune, 1985) que se basa en el cálculo de los períodos de recurrencia de las sacudidas y de las probabilidades de no excedencia asociadas según un modelo poissoniano, a partir de considerar la atenuación de las intensidades en forma de isosistas elípticas y los gráficos magnitud-frecuencia acumulativos como del tipo "distribución exponencial truncada". Como base para los cálculos utilizaron el mapa de la Fig.1 complementado con las zonas para la región oriental de Cuba (Comisión ad hoc, 1991; Alvarez y Bune, 1985), los parámetros del modelo de isosistas para cada zona (Chuy et al., 1992) y los correspondientes al régimen sísmico (Alvarez y Chuy, 1992). Los resultados corresponden a los períodos de recurrencia de las intensidades sísmicas con valores 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10 y 10.5 grados, en redes regulares de puntos separados 0.1° por cada coordenada, a partir de los cuales se confeccionaron diversas variantes de mapas con estimados de la peligrosidad sísmica, uno de los cuales se presenta en la Fig. 2. Ambos territorios presentan un elevado nivel de peligrosidad sísmica asociada. En ellos existen amplias zonas con valor máximo posible $I = X$ grados, mientras que las intensidades de VIII y IX grados tienen un período de recurrencia bajo.

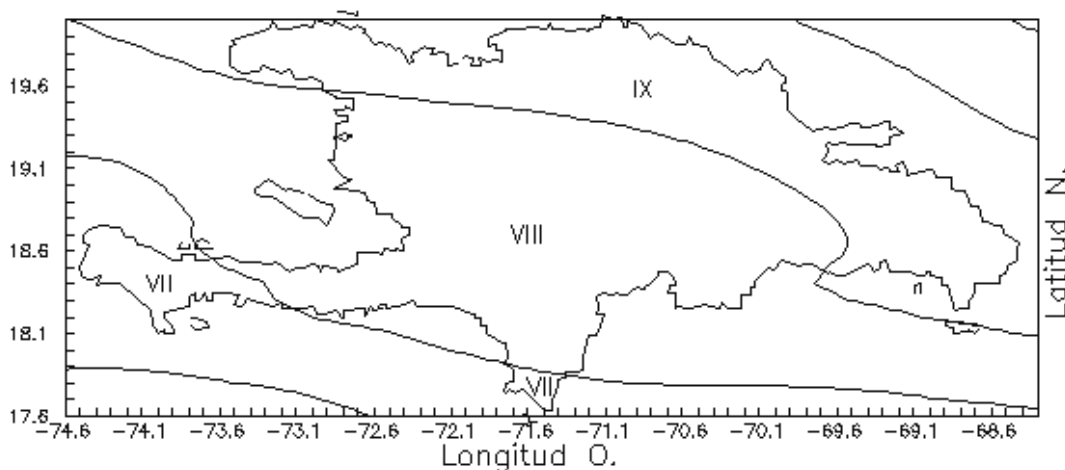


Fig. 2. Mapa de intensidades que no pueden ser igualadas o excedidas en un tiempo de espera de 50 años con una probabilidad de 0.9 para la isla de La Española.

Zonación sísmica de los territorios estudiados

Los períodos de recurrencia normalmente aumentan de forma exponencial en función de la intensidad, por lo que el presente estudio se realizó a partir de sus logaritmos en base 10. Por tanto, para su clasificación cada punto de la red se caracterizó por un conjunto de los 12 valores de los logaritmos de sus períodos de recurrencia:

$$\{\lg T(I_i), I_i = 5, 5.5, \dots, 10, 10.5\}$$

que constituyen las 12 variabes del presente problema de reconocimiento de patrones no supervisado.

Fue necesario realizar una clasificación a priori en función del efecto máximo I_{\max} , pues para valores de $I_i > I_{\max}$ no son posibles las sacudidas en el punto y se tiene $T(I_i) = \infty$; lo que no permite compararlo con otro punto que tenga $T(I_i) < \infty$. Los valores de I_{\max} presentes para ambas zonas eran: 8.5, 9, 9.5, 10, 10.5. Esta primera subdivisión de los puntos a clasificar produjo 5 grupos para La Española y 4 para Jamaica (se obvian los grupos con 1 y 2 elementos). No obstante, en la mayoría de los casos no se realizan los cálculos para valores semienteros de intensidad, y obviando éstos, la subdivisión por I_{\max} sólo conduce a 3 grupos para La Española y 2 para Jamaica.

Para evitar confusiones entre ambos casos, las intensidades del caso de valores enteros solamente se representarán con números romanos y los mapas correspondientes serán llamados de "variante 1". Igualmente, las intensidades del caso de valores enteros y semienteros se representarán con números arábigos y los resultados correspondientes serán llamados de "variante 2". En la tabla 1 se presenta el número de objetos en cada grupo. Aquí se nota una diferencia entre ambos casos. Los grupos correspondientes a La Española tienen por lo general muchos más objetos que los correspondientes a Jamaica, lo que es motivado por la diferencia en extensión territorial entre ambas islas.

El procesamiento de los datos de períodos de recurrencia se llevó a cabo de forma independiente para ambas islas y variantes representadas en la tabla 1. Para su clasificación con el sistema PROGNOSIS se utilizó el algoritmo HOLOTIPO (ICIMAF, 1992) para conjuntos β_0 -conexos, el cual resulta muy cómodo, pues permite representar el agrupamiento mediante árboles de asociación (Pico, 1995), y tomar decisiones sobre el valor de β_0 a seleccionar directamente sobre ellos. La diferencia entre el número de objetos para los grupos de Jamaica y La Española hizo que fueran tratados de forma diferente, tal como se describe a continuación.

Tabla 1. Número de objetos en cada uno de los grupos formados para la clasificación

Intensidad Máxima	Jamaica	La Española
8	-	2
8.5	1	27
9	5	118
9.5	33	319
10	105	643
10.5	116	491
VIII	1	29
IX	38	437
X	231	1034

Para el caso de Jamaica, como criterio de comparación entre valores de una variable, se tomó el de que el módulo de su diferencia fuera menor que cierto umbral, el cual, luego de algunas pruebas, fue fijado en un décimo de la diferencia entre sus valores máximo y mínimo dentro del grupo en cuestión. Como sistema de conjuntos de apoyo, en una primera etapa se seleccionó un

solo conjunto, el formado por todas las variables. Igualmente se seleccionó como función de semejanza parcial la que establece la semejanza de todas las variables sin excepción, y como criterio de clasificación empleó el β_0 -conexo, tomando el parámetro β_0 como el promedio de las semejanzas máximas para cada objeto.

Se destacan dos hechos interesantes. El primero es que el número total de holotipos hallado para el caso de intensidades enteras solamente es alrededor del 60 % del hallado para el caso de intensidades enteras y semienteras. El segundo corresponde al número de elementos agrupados alrededor de cada holotipo. En cada grupo coexisten holotipos con un número relativamente grande de elementos, otros con dos o tres elementos y muchos con un solo elemento. Esta característica nos condujo a realizar una nueva subdivisión de los grupos iniciales y que fue la extracción de cada uno de ellos de los elementos que clasificaban en holotipos aislados (de un solo elemento). A ellos nos referiremos como subgrupos "b", mientras que al resto (los holotipos con multiplicidad de elementos) lo haremos como subgrupos "a". El análisis de los datos correspondientes a los holotipos de los subgrupos "b", muestra que con frecuencia ellos sólo se diferencian entre sí significativamente en el valor de la última variable considerada (la correspondiente a la intensidad máxima posible), por lo que se decidió someterlos nuevamente a similar proceso de clasificación con un nuevo conjunto de apoyo, igualmente de un solo elemento, con todas las variables con excepción de esa última y dejando inalterados los demás parámetros usados en la primera clasificación. Como resultado se obtuvo una ligera disminución de la cantidad de holotipos. El número de holotipos hallados para cada grupo en todos los casos se muestra en la tabla 2.

Tabla 2. Resultados de la clasificación para el caso de Jamaica cuando el umbral de semejanza se selecciona como la media de las máximas de las semejanzas entre objetos. El número de holotipos aparece dividido en múltiples (a) y aislados [(b₁) al inicio y (b₂) luego de su reclasificación].

a) Variante 1

Intensidad	# de holotipos			
	Total	(a)	(b ₁)	(b ₂)
8.5	1	-	-	-
9	5	-	-	-
9.5	20	4	16	16
10	21	5	16	13
10.5	20	7	13	13

b) Variante 2

Intensidad	# de holotipos			
	Total	(a)	(b ₁)	(b ₂)
VIII	1	-	-	-
IX	17	4	13	11
X	15	5	10	10

Si se selecciona el β_0 con otra de las opciones implícitas en el PROGNOSIS (la media de todas las semejanzas), al igual que si se usa el algoritmo CLASS para clasificar en grupos β_0 -compactos se obtienen los mismos resultados. Esta forma de enfrentar el problema no produce resultados satisfactorios, ya que para fines prácticos, un mapa de zonación sísmica no puede tener muchas zonas diferentes, pues no es aceptado por aquellos a quienes va dirigido. Por ello se decidió cambiar el método de selección de β_0 , pasando a determinarlo directamente sobre el dendrograma o árbol de clasificación, bajo el criterio de que no se formaran mas de 3 subgrupos

en cada uno de los grupos. En la tabla 3 aparecen los resultados obtenidos en este caso, y en la Fig. 3 se presenta el mapa de zonación sísmica de Jamaica correspondiente.

Tabla 3. Resultados de la clasificación para el caso de Jamaica.

Int. Máxima	Umbral	No. Holotipos	No. objetos/holotipo
8.5	-	1	1
9	0.05	1	5
9.5	0.38	3	31, 1, 1
10	0.55	3	95, 9, 1
10.5	0.38	2	114, 1, 1

Para el caso de La Española inicialmente se realizó el procesamiento con los mismos parámetros que para Jamaica, no obteniéndose resultados satisfactorios para los grupos con mas elementos; incluso para la variante 2 (intensidades enteras solamente) se clasificaban todos los elementos en un solo grupo. Por ello se realizaron pruebas con diferentes criterios de semejanza entre variables y sistemas de conjuntos de apoyo. En la tabla 4 se muestran los resultados de las pruebas realizadas sobre el grupo de $I_{\max}=10.5$.

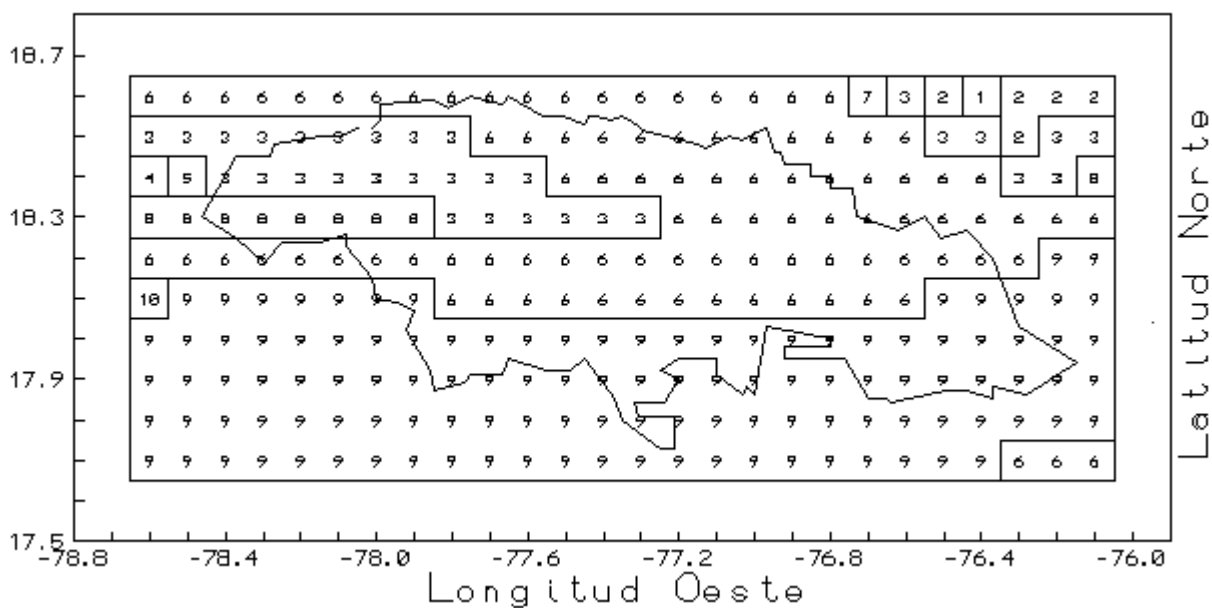


Fig. 3. Resultados de la clasificación para la isla de Jamaica. En cada celda se indica el número de orden del holotipo que la caracteriza.

Tabla 4. Pruebas de clasificación realizadas sobre el grupo de $I_{\max}=10.5$ para La Española.

CC	SCA	FS	β_0 (No. de holotipos)
1/20	todas	suma ponderada	0.75(14), 0.675(11), 0.66(3), 0.5(1)
1/20	CF3	suma ponderada	0.004(84), 0.0035(35), 0.0032(14), 0.003(3)
1/20	CF4	suma ponderada	0.0015(14), 0.0012(3), 0.001(1)
1/20	todas	de 1 umbral (2 var.)	35 holotipos para cualquier valor de β_0
1/20	todas	de 1 umbral (1 var.)	84 holotipos para cualquier valor de β_0
1/30	todas	suma ponderada	0.7(82), 0.6(34), 0.55(4), 0.5(1)

CC-criterio de comparación entre los valores de una variable (el mismo para todas, representado en fracción de su recorrido total), **SCA**-sistema de conjuntos de apoyo, **CFn**-cardinal fijo de n variables, **FS**-función de semejanza parcial.

Primeramente resalta el hecho de que para un mismo criterio de comparación (CC), cualesquiera sean los sistemas de conjuntos de apoyo y la función de semejanza se mantiene el mismo número de holotipos a diferentes niveles del árbol de clasificación. Más aún, aunque no aparece en la tabla, cada holotipo mantiene el mismo número de elementos en cada caso, siendo estos elementos los mismos también. Además, al cambiar el criterio de comparación, varía muy levemente la composición de los holotipos; por ejemplo, la cifra de 3 holotipos que se repite para CC=1/20, en comparación con el caso de 4 holotipos para CC=1/30 solo se diferencia en que un elemento del grupo mayor pasó a formar un holotipo independiente. Se tomó la decisión de realizar la clasificación con el sistema de conjuntos de apoyo formado por todas las variables, para un criterio de comparación de los valores de una variable de que el módulo de su diferencia sea menor que 1/20 del recorrido de la variable en el grupo, y como función de semejanza parcial la suma ponderada. Igualmente el valor de β_0 se seleccionó sobre el dendrograma bajo el criterio de que no se formaran más de 3 subgrupos en cada grupo. Los resultados se presentan en la tabla 5, y el mapa correspondiente de zonación sísmica de La Española en la Fig. 4. Una vez concluido el proceso de clasificación fueron confeccionadas las curvas I vs. T(I) de todos los holotipos determinados. En las Fig. 5 y 6 se presenta una selección de éstas.

Tabla 5. Resultados de la clasificación para el caso de La Española.

Int. Máxima	Umbral	No. Holotipos	No. objetos/holotipo
8.5	0.20	2	14, 13
9	0.40	2	114, 4
9.5	0.70	2	318, 1
10	0.60	2	627, 16
10.5	0.65	3	486, 4, 1

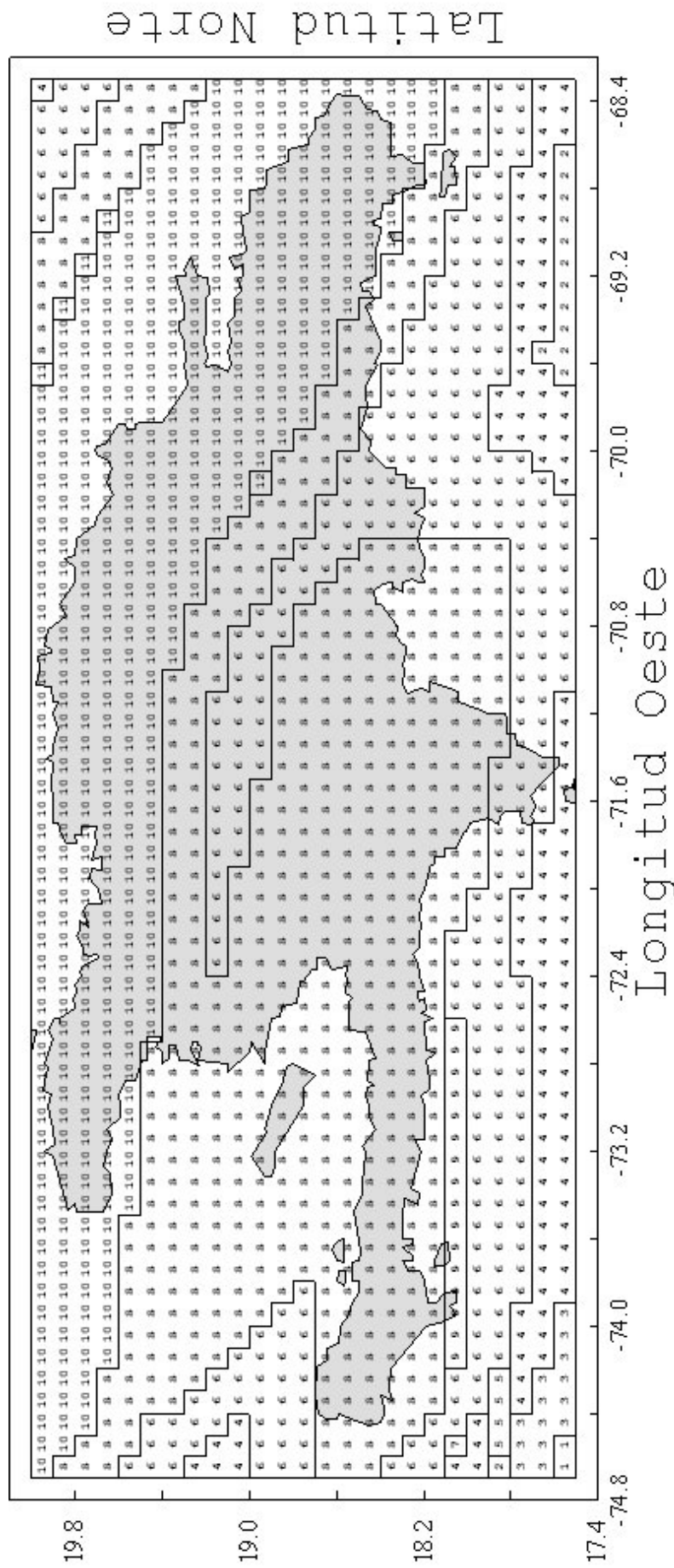


Fig. 4. Resultados de la clasificación para el caso de La Española. En cada celda se indica el número de orden del holotipo que la caracteriza.

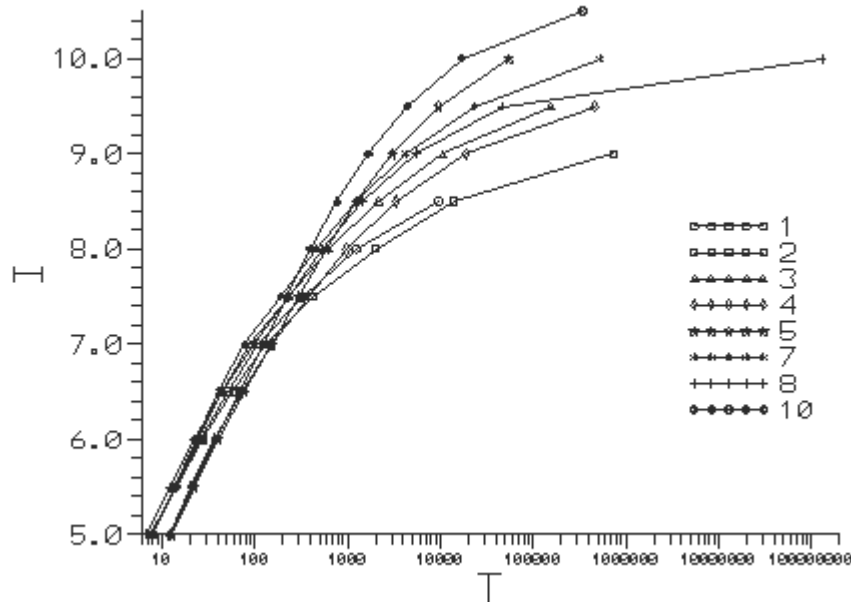
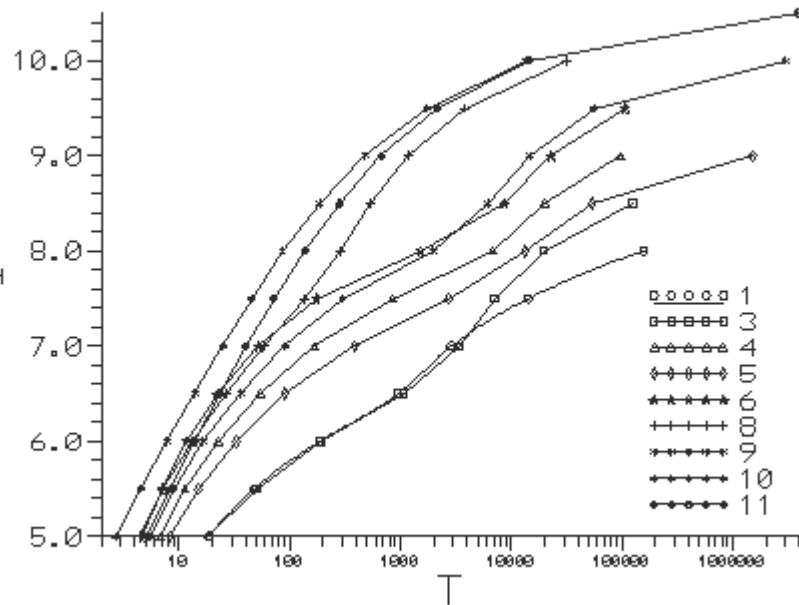


Fig. 5. Gráficos de intensidad vs. períodos de recurrencia para algunos de los holotipos determinados para la isla de Jamaica. Los números en la leyenda corresponden a los que aparecen en el mapa de la Fig. 3.

Fig 6 Gráficos de intensidad vs. períodos de recurrencia para algunos de los holotipos determinados para la isla de La Española. los números en la leyenda corresponden a los que aparecen en el mapa de la Fig. 4.



Discusión

Obviando los "grupos" con $I_{\max} = 8.5$ y VIII, que solo tienen muy pocos puntos, vemos que la subdivisión de la muestra en intensidades máximas enteras y semienteras redunda, cuando el valor de β_0 se determina por una de las opciones implícitas en el sistema, en un incremento sustancial del número de holotipos. Esto puede ser explicado por el hecho de que al no exigírsele

a los elementos de un grupo ser todos semejantes entre sí, sino al menos a uno de los restantes, al subdividir la muestra se rompen muchos enlaces, y los elementos se agrupan alrededor de nuevos holotipos formando más grupos, menos numerosos que los anteriores.

Por otra parte existe una relación inversa entre el volumen de la muestra y el número de holotipos para un mismo valor de β_0 , y las muestras mayores tienen la tendencia a agruparse alrededor de un solo holotipo, dejando sólo algunos elementos independientes. Igualmente, tal relación existe entre el número de holotipos y el criterio de comparación para valores de una variable, por lo que resulta evidente que para lograr una mayor diversidad en la clasificación es imprescindible disminuir los valores de β_0 en los criterios de comparación según se incremente el volumen de la muestra.

Con relación a los mapas, ellos muestran la diversidad obtenida en la clasificación; lamentablemente, la misma fortuitamente está ubicada con preferencia en las zonas marítimas, en las cuales carece de sentido hablar de zonación sísmica, por lo que sólo presentan interés metodológico. Para fines prácticos, en ambas islas sólo hay 4 zonas diferentes. Los gráficos representados muestran en primer lugar una diferencia acusada entre los grupos en que inicialmente se dividió la muestra, lo que tiene la explicación física de que la tendencia general de comportamiento de los estimados de peligrosidad sísmica es que según sea mayor la intensidad máxima posible, son menores los períodos de recurrencia de las sacudidas, por lo que las curvas en un gráfico I vs. $\lg(T)$ se desplazan hacia arriba. En segundo lugar, se nota también a simple vista la diferencia entre las curvas pertenecientes a diferentes holotipos del mismo grupo. Si bien es cierto que no se representan toda las curvas (para no perder claridad), las restantes se diferencian también, pero se entremezclan unas con otras, cruzándose, lo que dificulta su representación conjunta.

La zonación sísmica está constituida por los mapas y gráficos en conjunto. Los primeros indican la distribución espacial del tipo de comportamiento del efecto sísmico, y los segundos particularizan el comportamiento temporal y por intensidad de dicho efecto. Una zonación realizada de tal forma es cualitativamente superior a las que habitualmente se emplean en los códigos de construcción en zonas sísmicas.

Conclusiones

1) Se realizó la zonación sísmica de las islas de Jamaica y La Española mediante el procesamiento de los períodos de recurrencia de las sacudidas de diferente intensidad (tomados de un estudio de estimación de la peligrosidad sísmica) con algoritmos lógico-combinatorios de reconocimiento de patrones no supervisado, incluidos en el sistema PROGNOSIS.

2) La zonación consiste en un mapa con indicación del tipo de comportamiento del efecto sísmico (los holotipos de la clasificación) en celdas de $0.1^\circ \times 0.1^\circ$, y en un conjunto de curvas (una para cada holotipo) que indican la frecuencia de ocurrencia de los efectos de diferente intensidad. Dicha zonación es cualitativamente superior a las se realizan por los métodos tradicionales.

3) Se comprobó la utilidad del método propuesto por Pico (1995) de selección del valor de β_0 par el agrupamiento o clasificación en algoritmos β_0 -conexos directamente sobre un árbol de clasificación.

4) Se estableció la existencia de dependencias inversas entre el número de holotipos y (a) el volumen de la muestra a clasificar, (b) el umbral en los criterios de comparación para valores de una variable.

Bibliografía

ALVAREZ, L. (1995): SACUDIDA, versión 2.0. Un programa para la estimación de la peligrosidad sísmica. La Habana, Editorial Academia, 64 p.

ALVAREZ, L.; BUNE, V.I. (1985): A computer program for seismic hazard estimation. Proceedings of the 3rd. International Symposium on the Analysis of Seismicity and on Seismic Risk, Liblice Castle, Czechoslovakia, June 17-22, pp. 432-439.

ALVAREZ, L.; CHUY, T. (1992): Régimen sísmico. En: Informe de las investigaciones sismológicas complejas para el Complejo Hidroenergético Toa-Duaba. Capítulo 3. Informe Científico- Técnico, Fondos del CENAI, CITMA.

ALVAREZ, L.; COTILLA, M.; CHUY, T. (1994): Sismicidad y peligrosidad sísmica de La Española y Jamaica. Resúmenes, III Encuentro Nacional de Sismología, Santiago de Cuba.

ALVAREZ, L.; MIJAILOVA, R.S.; CHUY, T. (inédito): Catálogo de terremotos fuertes de la región [16°-24° N., 70°-86° W.] desde el siglo XVI hasta 1988. Fondos del CENAI, CITMA.

CHUY, T., ALVAREZ, L. (1988): Sismicidad histórica de La Española. Comunicaciones Científicas sobre Geofísica y Astronomía, No. 16, 14 p.

CHUY, T.; ALVAREZ, L.; GUASCH, F.; MOJENA, R.; SIERRA, L. (1992): Características de la sismicidad de la región. En: Informe de las investigaciones sismológicas complejas para el Complejo Hidroenergético Toa-Duaba. Capítulo 2. Informe Científico- Técnico, Fondos del CENAI, CITMA.

COMISION AD HOC (1991): Dictamen de la Comisión Ad-hoc para la determinación de las zonas sismogeneradoras de la región oriental de Cuba y zonas aledañas. La Habana, p., 2 fig. Fondos del CENAI, CITMA.

COTILLA M., ALVAREZ, L., CHUY, T. (1991): Determinación de las zonas de origen de terremotos de La Española y Jamaica. Informe, 13p., 12 fig. Fondos del CENAI, MCTMA.

ICIMAF (1992): PROGNOSIS, versión 2.1. Software para las geociencias. Manual de usuario, 154 p.

- PICO, R. (1995): Determinación del parámetro β_0 para los algoritmos de agrupamiento lógico-combinatorios. Conferencia Internacional "Ciencia y tecnología para el desarrollo", CIMAFA'95. I Taller Iberoamericano de reconocimiento de patrones. Resúmenes, pp. 99-100.
- REYES, L. (1977): Sismicidad y riesgo sísmico de La Hispaniola, Jamaica y Puerto Rico. Boletín IMME, Año XV, No 59-60, 1977.
- RUIZ, J. (1992): Modelos de algoritmos de reconocimiento con aprendizaje parcial. En: 3er Congreso Iberoamericano de Inteligencia Artificial, IBERAMIA-92, La Habana, 17 al 22 de enero de 1992, Memorias, pp. 541-559. México, LIMUSA, 1992.
- RUIZ, J.; PICO, R.; LÓPEZ, R.; ALAMINOS, C.; LAZO, M.; BOGGIANO, M.; BARRETO, E.; SANTANA, A.; ALVAREZ, L.; CHUY, T. (1992): PROGNOSIS y sus aplicaciones a las geociencias. En: 3er Congreso Iberoamericano de Inteligencia Artificial, IBERAMIA-92, La Habana, 17 al 22 de enero de 1992, Memorias, pp. 561-586. México, LIMUSA, 1992.
- SHEPHERD, J.W.; ASPINALL, W.P. (1980): Seismicity and seismic intensities in Jamaica, West Indies; a problem in risk assessment. Earthq. Eng. Struct. Dyn., vol 8, No.4, p.315-335.
- SIROTINSKAIA, S.V. (1986): Métodos lógicos de análisis de información geológica (en ruso). Nedra. 1986.
- VORONIN, YU. A.; KARATAEVA, G. N. ; EPSHTEIN, E.N. (1968): Programas HOLOTIPO para la solución de problemas de reconocimiento de patrones (en ruso). Alma Atá.
- ZHURAVLIOV, Yu. I. (1978): Acerca del enfoque algebraico para la solución de problemas de reconocimiento y clasificación (en ruso). Problemy Kibernetiki, No. 33, pp. 5-68.