

Reconocimiento de elementos de estructuras espaciales

Modelación matemática del pronóstico de magnitudes máximas de los terremotos en la región del Caribe

J. RUIZ SHULCLOPER, R. PICO PERA, L. ALVAREZ
GÓMEZ, Y. CHUY RODRIGUEZ, M. COTILLA RODRIGUEZ

RESUMEN. En el trabajo se expone, siguiendo la metodología expuesta en [1], el proceso de modelación matemática realizado para la solución del pronóstico de magnitudes máximas de terremotos de la región del Caribe. Para su solución fue usado el modelo de algoritmos de votación enriquecido con los conceptos y definiciones que surgieron de dicho proceso de modelación.

1. INTRODUCCIÓN

Uno de los parámetros que caracterizan la ocurrencia de terremotos es su energía. La "magnitud" es la medida logarítmica de la energía que se libera cuando ocurre un terremoto. En lo que pudiéramos denominar "proceso de preparación" de un terremoto se produce una acumulación de energía — cuyas causas no analizaremos aquí — que en cuanto a su medida tiene dos cotas en cada área: una inferior, que en todos los casos está condicionada al tipo de equipamiento que se disponga, a su sensibilidad y una superior, que existe debido a la imposibilidad física de acumular una energía infinita. Es obvio que justamente el conocimiento de la medida de esta cota superior es de sumo interés, lo cual constituye un problema muy difícil que es enfrentado en la actualidad de diferentes formas.

La manera más común para la determinación de la medida de la cota superior de esta energía, o dicho más precisamente, la estimación de la magnitud máxima (M_{max}), es mediante el uso del "criterio sismológico". Este criterio se basa en la hipótesis que los terremotos que ocurrieron en un lugar en cierto momento pueden volver a ocurrir en el futuro. En otras palabras, podemos esperar en un territorio, en una región sísmica activa, que ocurran terremotos al menos tan fuertes como los que habían ocurrido en el pasado. Al usar este criterio podemos esperar que la M_{max} sea igual o que

es un poco mayor que la máxima reportada. Lógicamente este criterio tiene sus limitantes en que se basa en la historia conocida de la sismicidad ocurrida en una región. En algunos lugares, como en la China, existen algunos miles de años de reportes, mientras que en otros, como en nuestro territorio, escasamente contamos con menos de cien años de registros de terremotos en algunas regiones. Esto sin analizar la confiabilidad de cada uno de dichos reportes. Estas razones ponen de manifiesto las grandes limitaciones del método para su aplicación en general.

Otra forma de enfrentar este problema es analizando las condiciones geológicas que son propicias a la generación de terremotos fuertes, basándose en el estudio de dichas condiciones en los lugares donde se han producido los terremotos. De dicho análisis este método concluye que en zonas con características geológicas semejantes a las antes mencionadas, se producirán terremotos de la misma magnitud que las que ocurrieron en las estudiadas. Sin embargo, la práctica ha demostrado que las condiciones geológicas de una región resultan insuficientes, ellas solas, para realizar un pronóstico confiable.

Desde mediados de los años 60, en la Unión Soviética, se desarrolló un grupo de trabajo, bajo la dirección del profesor Ríznichenko cuyo método fue el de estimar las M_{max} de los terremotos a partir de una combinación de la información sismológica, geológica y de tipo tectónica. Esto fue debido a una serie de estudios que comenzaron relacionando la M_{max} con una promediación especial del nivel de ocurrencia de terremotos débiles en una región, la "actividad sísmica". Según avanzaron estos estudios se comprobó que la actividad sísmica por sí sola no era capaz de reflejar exactamente el fenómeno de ocurrencias de M_{max} . Ya eran conocidos algunos trabajos que relacionaban la ocurrencia de terremotos fuertes con ciertas características geológicas, tectónicas y con los campos geofísicos. Por ello se fue a la búsqueda de una función que combinará tanto los parámetros sismológicos (la actividad sísmica) las características geológicas y tectónicas (en este caso cuantificadas de una forma especial) y los campos geofísicos con la M_{max} [2,3].

La utilización de las técnicas de Reconocimiento de Patrón es algo posterior. En la década del 70 en la URSS se desarrollaron fun-

damentalmente dos enfoques diferentes: uno, en el que se utilizaron técnicas del tipo de algoritmos de votación y otro en el que se emplean métodos estadísticos para hallar una función que relacione ciertos parámetros con la M_{max} . En el primer caso, se realizaron una serie de trabajos bajo la dirección del profesor Keilis-Borok en los que el problema inicial se simplifica, no se resuelve totalmente la estimación de la M_{max} , sino que se busca una partición en el espacio de las magnitudes para obtener un mapa de una región dada en el que se exprese qué zonas son capaces de generar terremotos con M_{max} superior o igual a cierto umbral y el resto de la región estaría formada por aquellas zonas que no son capaces de generar esa magnitud. Estos trabajos fueron aplicados por ellos con bastante éxito en prácticamente todas las zonas sismogeneradoras del mundo incluso para la región del Caribe pero, como expresamos anteriormente, no se resuelve totalmente el problema, de hecho sólo nos caracteriza un aspecto del fenómeno "ocurrencias de M_{max} " [4,5].

En el segundo enfoque, un poco más cercano a lo que nosotros esperamos obtener es el debido a trabajos conjuntos realizados entre los Institutos de Problemas de Transmisión de la Información y de Física de la Tierra de la Academia de Ciencias de la URSS y el Ministerio de Geología de la URSS, en los cuales se busca un funcional que relacione a la M_{max} con una serie de parámetros sismológicos, geológicos, geofísicos, geomorfológicos, que son especialmente codificados en el caso que sean cualitativos. En estos trabajos resulta de interés el hecho que el resultado final se da en forma de mapa donde para cada unidad básica de área - la que sea seleccionada - se determina un valor para la M_{max} [6,7]. Los mismos pueden considerarse como un tratamiento matemático más riguroso del problema planteado por la escuela de Ríznichenko.

2. FORMULACIÓN DEL PROBLEMA

Nosotros nos hemos planteado la solución de un problema semejante. Es nuestro objetivo la determinación de las M_{max} posibles en todos y cada uno de los puntos de la región del Caribe. Para esto nos apoyamos en una serie de presupuestos:

1) Las M_{max} están directamente relacionadas con los valores que tomen algunos parámetros de índole sismológica, geológica, geofísica

es y geomorfológica.

2) El estudio de la región se realizará en base a su descomposición en cuadrículas de dimensiones $0,4^{\circ} \times 0,4^{\circ}$ ($40 \text{ km} \times 40 \text{ km}$ aproximadamente). La región se encuentra comprendida entre los 4° y 24° de latitud Norte y los 57° y 93° de longitud Oeste abarcando la región del Caribe, América Central y una parte importante de América del Sur.

3) Los eventos que estudiaremos y que pretendemos pronosticar son los ocurridos en el intervalo de 0 a 70 km de profundidad.

4) Debemos prever que en las descripciones de las cuadrículas en términos de los parámetros mencionados en 1) no se encuentren todos ellos presentes, es decir, habrán casos en los que no todos esos parámetros sean conocidos en una cuadrícula dada.

5) No compartimos el criterio de que la M_{max} pueda ser obtenida mediante un funcional que la relacione con los parámetros mencionados en el punto 1). Consideramos que la dependencia que exista entre M_{max} y dichos parámetros es mucho más compleja y no puede ser reflejada explícitamente mediante formas analíticas.

6) Los parámetros mencionados en 1) en nuestro caso son: atributos sísmológicos - actividad sísmica, densidad de epicentros, liberación de las deformaciones, distribución de eventos en profundidad; atributos geológicos - presencia de magmatismo cuaternario, número de elementos estructurales, espesor de la corteza consolidada, profundidad de la corteza no consolidada, profundidad de la discontinuidad de Mohorevici (Moho), gradiente de la profundidad de Moho, cercanía a elementos estructurales fundamentales; atributos geofísicos - anomalía de Bouger, gradiente de la anomalía de Bouger, así como una regionalización geomorfológica.

7) Como ya dijimos, la actividad sísmica fue uno de los primeros atributos que se utilizó para el pronóstico de M_{max} , a partir de los trabajos de Rinichenko - creador de esta representación de la sismicidad [3]. Este atributo se representa en mapas de isólinas. En la región de estudio se observa que con isólinas de valores 0; 0,5; 1; 2; 5; 10; 20 y 50 se obtiene una diferenciación adecuada de la misma. Desde el punto de vista sísmológico cada uno de estos valores son diferentes.

8) La densidad de epicentros es otro tipo de promediación de los terremotos en una región y ha sido usada en general para delimitar

a grosso modo las zonas de origen de terremotos. La representación de este atributo es análoga al anterior y en este caso los valores que se tomaron fueron 0; 5; 10; 20; 50; 100; 200 y 500. Cada uno de estos valores son sísmológicamente diferentes.

9) La liberación de deformaciones en un área delimitada se obtiene sumando la raíz cuadrada de la energía liberada por cada terremoto que haya ocurrido en ella. Es una medida de la historia sísmica de esa área. Por su carácter puntual decidimos no trazar mapas de isólinas y utilizamos en su lugar el logaritmo en base 10 de esta magnitud. Por limitaciones en las posibilidades de representación gráfica fueron agrupados los valores en intervalos de 0,5 unidades con aproximación por exceso. Los valores obtenidos fueron 0; 7,5; 8; 8,5; 9; 9,5; 10; 10,5; 11; 11,5; 12; el cero corresponde al caso de ausencia del fenómeno.

10) La distribución de los eventos en profundidad se obtuvo analizando los datos instrumentales de los terremotos reportados hasta 150 km de profundidad y permite describir cualitativamente la ubicación de la capa sismoactiva. Los valores de este atributo para cada cuadrícula se obtienen mediante el siguiente procedimiento: se trazan perfiles con orientación Norte-Sur a un ancho de 5 cuadrículas; se suman todos los eventos que han ocurrido por intervalos de profundidad de 10 km en dichas cuadrículas; se asume para las 5 cuadrículas la suma obtenida y se hace una clasificación cualitativa (en 4 grupos) en función de la distribución de valores siguiendo el criterio: 1 si la cantidad de intervalos de 10 km en los que aparecen al menos un evento en la capa de 0-70 km es igual a la cantidad de intervalos análogos en la capa de 70-150 km; 2 si esa cantidad es mayor en 0-70; 3 si es mayor en 70-150 y 4 si no hay actividad alguna.

11) La presencia de magmatismo cuaternario da una idea de las manifestaciones del vulcanismo de la región al cual se asocia cierto tipo de terremotos.

12) El número de elementos estructurales da una idea de la complejidad tectónica de la corteza de la región.

13) El espesor de la corteza consolidada y la profundidad de la no consolidada dan idea de la complejidad de la corteza en la región. Existen criterios sísmológicos que vinculan la posible aparición de M_{max} con la complejidad de la corteza. Estas se miden en Kilómetros y se obtienen de mapas de isólinas.

14. La profundidad de la discontinuidad de Mohorovičić es el espejo de la corteza y su gradiente da una idea de la variación espacial de este parámetro. También existen criterios sísmológicos que los vinculan a la aparición de terremotos. Este parámetro se obtiene también de mapas de isólinas y en nuestro caso, sus valores cambian con intervalos mínimos de 1 km.

15) La litosfera terrestre se encuentra dividida en placas tectónicas que se desplazan relativamente entre sí. Estas placas, por lo general de grandes dimensiones, se encuentran en muchos casos subdivididas en grandes bloques. Las estructuras fundamentales corresponden a las fronteras entre placas o entre bloques y a algunos fenómenos asociados a la dinámica del movimiento de las placas, por ejemplo el vulcanismo. La distancia geográfica (en cuadrículas) a uno de los elementos puede darnos una medida de la capacidad de generar terremotos fuertes en una región. Este atributo lo consideraremos cualitativamente, siguiendo el criterio: cuadrícula coincidente con la estructura, si la distancia es menor que una cuadrícula; muy cercana cuando la distancia es de una cuadrícula; cercana cuando es de 2 y lejana cuando es 3 ó más. Nosotros consideraremos la distancia al más cercano de los elementos estructurales.

16) Han sido establecidos hechos que permiten afirmar que las anomalías gravimétricas tienen una relación fuerte con la ocurrencia de terremotos y sus M_{max} . En nuestro caso disponimos de un mapa de anomalías gravimétricas en la corrección llamada de Bouguer, en isólinas de valores múltiples de 10, que varían entre -50 y 300 miligales.

17) El gradiente de la anomalía de Bouguer nos da una idea de la variación espacial del parámetro antes descrito. Este parámetro, así como el gradiente de la profundidad de Moho, se obtuvo según el siguiente método:

$$\nabla f|_0 = \sqrt{\frac{(f_1 - f_3)^2}{\Delta y^2} + \frac{(f_2 - f_4)^2}{\Delta x^2}}$$

	f_1	
f_3	f_0	f_4
	f_2	

Esta magnitud se multiplica posteriormente por un factor de escala conveniente. Obviamente las cuadrículas de la frontera de la zona de estudio tendrán un valor para este parámetro menor que las restantes.

18) Es común para grandes regiones preparar mapas donde se reflejan

las principales características del relieve, que están relacionadas con la estructura geológica. Así se representan cuencas, crestas, escudos, plataformas, etc. Estos factores a su vez se relacionan en una u otra medida con la ocurrencia de terremotos. A partir de un mapa de este tipo se agruparon las diferentes características geomorfológicas en 10 grupos principales.

19) El presupuesto 5 nos lleva a suponer la existencia de intervalos de las M_{max} cada uno de los cuales pueda ser considerado como un criterio para agrupar las cuadrículas de la región de estudio. En nuestro caso consideraremos 3 intervalos de las M_{max} que desde el punto de vista sísmológico tienen significación:

Intervalo

1. $7 \leq M_{max}$
2. $6 \leq M_{max} < 7$
3. $5 \leq M_{max} < 6$
4. $4 \leq M_{max} < 5$
5. $M_{max} < 4$

Este último intervalo tiene incluido el caso de magnitud 0, es decir, no hay terremotos reportados en la cuadrícula. Se considera que una cuadrícula puede relacionarse a lo sumo con uno de dichos intervalos.

20) Desde el punto de vista sísmológico los errores de clasificación no tienen el mismo valor. Para establecer un orden de "prioridad" entre los mismos utilizamos los siguientes criterios:

Sea α el intervalo en el que se encuentra una cuadrícula y β el intervalo con la que es asociada la cuadrícula por algún procedimiento. Diremos que dicho procedimiento puede abstenerse de clasificar y eso lo denotaremos por $\beta = *$.

1. Por "pérdida de objetivo" indicaremos la situación en la que $\alpha < \beta$ y por "falsa alarma" $\beta < \alpha$. La pérdida de objetivo es un error más costoso que la falsa alarma en cualquier caso.

2. Se consideran los efectos que se producen en un sismo de magnitud del intervalo α implicado. Para $\alpha = 1$ es la peor situación, es menor cuanto es el $\alpha = 2$ y así sucesivamente.

3. La abstención de clasificación es preferible a la clasificación

errónea.

En virtud de los criterios enunciados podemos obtener la siguiente sucesión decreciente respecto a la magnitud del error, entendiendo que el par de números d_i^j denota que una cuadrícula que sabemos se encuentra asociada al intervalo α_i ha sido ubicada por algún procedimiento en el intervalo β_j .

15	
14	51
13	25 24
41	12 1+ 35
52	31 23 42 21
2+	33 34 32 3+ 45
43	4+ 54 5+

Como se ve tienen un peso mucho mayor los errores de pérdida de objetivo, lo que es debido a la naturaleza del problema. No pronosticar un terremoto fuerte acarrea daños irreparables, mientras que pronosticar uno superior a lo sumo ocasiona gastos.

21. Las fuentes informativas utilizadas fueron:

- para los mapas de anomalías de Bouguer, espesor de corteza consolidada, profundidad de corteza no consolidada y de discontinuidad de Mohorovicic el informe [8];

- para la información que se usó en la preparación de los mapas de presencia de magnetismo cuaternario, número de elementos estructurales, distancia a elementos estructurales fundamentales y regionalización geomorfológica el mapa [9];

- para la preparación de los mapas de actividad sísmica, densidad de epicentros, liberación de deformaciones así como de distribución de eventos en profundidad, el fichero de terremotos [10];

- para la preparación de las muestras de aprendizaje y control se utilizó el fichero mencionado, así como un catálogo preparado por los autores [11].

22. El presupuesto 5 también nos lleva a la conclusión que la base sobre la cual cualquier procedimiento relacione una cuadrícula con α lo sumo uno de los intervalos mencionados es la analogía con la experiencia acumulada. Esta analogía por supuesto se debe realizar en términos de los atributos mencionados en el presupuesto 6. Para ello contamos con la descripción de un número de cuadrículas en

las que sabemos se han reportado M_{max} correspondiente al intervalo en el que cada una es ubicada. Obviamente este presupuesto está sujeto al error que un caso dado esa M_{max} fuese sobrepasada en el futuro, es decir, trabajemos en base a las magnitudes reportadas ya que no es posible determinar la M_{max} en otros términos.

23) Como información inicial sobre el fenómeno objeto de estudio se tienen las magnitudes de los terremotos desde inicios de siglo, que corresponden a determinaciones instrumentales. A pesar de que existen determinaciones de magnitud por datos macrosísmicos para terremotos ocurridos desde el siglo XVI, no fueron usadas por considerar que están sometidas a un rango de incertidumbre mayor que las instrumentales.

3. FORMALIZACIÓN DEL PROBLEMA

Resulta claro que el camino a seguir en el proceso de formalización del problema formulado en el epígrafe anterior debe iniciarse en virtud del presupuesto 1 con la suposición que las cuadrículas sean descritas en un espacio "14-dimensional" en el que cada coordenada denota el valor del correspondiente atributo en la respectiva cuadrícula.

24) En virtud del presupuesto 4 deberos admitir en la descripción de las cuadrículas la ausencia de información de algunos valores de sus atributos, lo que denotaremos por "x". Obviamente esto condiciona la selección de los procedimientos de clasificación a utilizar.

25) En virtud del presupuesto 7, el conjunto de valores admisibles para el atributo "actividad sísmica" lo conforman los valores de las escaleras mencionadas, es decir, 0; 0.5; 1; 2; 5; 10; 20; 50. En este caso y todos los demás, el conjunto de valores admisibles de las variables definiremos el símbolo α . Todos esos valores son diferentes.

26) Análogamente, la "densidad de epicentros" tendrá como conjunto de valores admisibles 0; 5; 10; 20; 50; 100; 200; 300; α . Todos esos valores son diferentes.

27) La "distribución de los eventos en profundidad" es una variable cualitativa positivamente con valores 1, 2, 3, 4, α en virtud del presupuesto 10.

Este presupuesto nos lleve a restringir el conjunto de procedimientos a utilizar en la solución de nuestro problema a aquel en el que aparecen los que admiten ausencia de información y el uso de variables cualitativas y cuantitativas simultáneamente y además, en virtud del presupuesto 22, el análisis de toda la descripción debe ser cualitativo (en términos de la analogía).

28) La "presencia de magnetismo cuaternario" es una variable booleana.

29) El "número de elementos estructurales" es una variable natural que en nuestro caso toma valores inferiores a 50.

30) El "espesor de la corteza consolidada" se mide en números naturales y en nuestra región aparecen sólo valores inferiores a 30.

31) Análogamente la "profundidad de la corteza no consolidada" se mide en números naturales inferiores a 20.

32) Como quiera que la "profundidad de Moho" se obtiene a partir de un mapa de isofleas en el que existen zonas con valores que crecen en una unidad a partir de la isoflea de valor 10, el valor mínimo en la región, consideraremos esta variable como un número natural que toma los valores entre 10 y 40. Atendiendo a las características de esta variable se tomó el criterio de dividir su dominio de existencia en 2 intervalos. En el primero, que incluye los valores entre 10 y 20 Km, dos valores se consideran iguales si el módulo de su diferencia es ≤ 2 . En el segundo, que incluye los valores mayores de 20 Km, dos de estos se consideran iguales si dicho módulo es ≤ 4 .

33) En virtud del presupuesto 18, la "certanía a elementos estructurales fundamentales" se da como una variable cualitativa polivalente con valores 0, 1, 2 y 3.

34) La "anomalía de Bouguer", por el presupuesto 16, puede considerarse una variable entera que toma valores entre -60 y 120 y dado el paso de crecimiento de las isofleas, dos valores se consideran iguales si el módulo de su diferencia es menor o igual a 10.

35) Para el "gradiente de la anomalía de Bouguer", así como para el "gradiente de la profundidad de Moho" que son obtenidos según (1), se usa un factor de escala que los convierte en números naturales. Para la región de nuestro estudio el primero de ellos toma valores entre 0 y 350 y el segundo entre 0 y 300. En ambos casos se vio que desde el punto de vista sismológico cualesquiera dos va-

lores cuya diferencia no excediera de 25 podrían ser considerados como valores análogos.

36) De acuerdo al presupuesto 18, la regionalización geomorfológica presenta 10 valores (1 al 10) todos diferentes entre sí.

37) Dados los criterios expuestos en el presupuesto 19, al evaluar la calidad de cualquier procedimiento de clasificación debemos definir un funcional de error generalizado 12 para clases disjuntas, cuyos parámetros se obtengan de la siguiente tabla:

Tabla 1

$\alpha \backslash \beta$	1	2	3	4	5	*
1	-	3.75	5.0	6.0	7.0	3.9
2	2.2	-	2.6	4.33	4.67	2.0
3	2.8	1.49	-	1.66	3.25	1.32
4	4.0	2.4	1.0	-	1.15	0.75
5	5.5	3.0	1.83	0.5	-	0.25

Estos valores se obtuvieron a partir de la sucesión de importancia del error expuesta anteriormente en forma de "escalera" (presupuesto 20) a partir de los criterios siguientes:

- de un piso a otro de la escalera el error se incrementa en una unidad
- dentro de un mismo piso el incremento de 1 se divide en partes iguales comenzando de derecha a izquierda.

Si denotamos por R un algoritmo de clasificación, su funcional de calidad lo daremos en términos del funcional de error generalizado para clases disjuntas

$$E(R) = \sum_{i=1}^2 \sum_{j=1}^2 E_{ij} X_{ij} + \sum_{i=1}^2 \omega_{i*} Y_{i*} \quad (2)$$

siendo X_{ij} la cantidad de cuadrículas del intervalo i ubicadas en el intervalo j , E_{ij} el costo de dicho error, Y_{i*} la cantidad de cuadrículas del intervalo i sobre las cuales R se abstuvo y ω_{i*} el costo de ese error. E_{ij} y ω_{i*} se obtienen de la Tabla 1.

4. SELECCIÓN DEL MODO DE SOLUCIÓN DEL PROBLEMA A

Es necesario hacer un énfasis de los vectores que conforman la matriz inicial de datos con vistas a determinar el grado de informatividad de cada uno de los atributos considerados, si existe o no dependencia de algún tipo entre los mismos, si los dominios y criterios de comparación de valores de cada atributo no conllevan contradicciones con las expectativas de solución al problema.

Sabemos que nuestra matriz de datos está conformada por variables cuantitativas y cualitativas. Además, desde los presupuestos 6 y 19, quedó claro que estamos ante un problema de clasificación en 5 clases disjuntas, que se suscita la necesidad de clasificación y que disponemos de aprendizaje para nuestros algoritmos. Ante es el problema A' que debemos enfrentar [1]. En base a todos estos presupuestos y consideraciones nosotros concluimos que eran idóneos para la solución del problema A', la familia de algoritmos de reconocimiento basados en los modelos lógico-combinatorios [12]. Para proceder, como dijimos al inicio de este epígrafe, nuestra matriz inicial, utilizamos el sistema SELOCTER [13] y el programa PRONDS [14]. Con estos se determinó la frecuencia informacional de cada uno de los atributos y se conformaron las matrices de aprendizaje y de control.

Para la determinación del peso informacional de las cuadrículas se utilizó la fórmula:

$$P_v(o_j) = \frac{1}{\sum_{i=1}^n P(x_i)} \sum_{i \in \{i | x_i(o_j) \neq * \}} a_i^v(o_j) P(x_i) \quad (3)$$

donde $P(x_i)$ es el peso informacional de la variable X_i y $a_i^v(o_j)$ se calcula como en [13].

Las matrices de aprendizaje y de control quedaron conformadas como se indica en la Tabla 2.

Tabla 2

Intervalo	Matriz inicial	Matriz de aprendizaje	Matriz de control
1	40	20	20
2	56	29	27
3	12	17	13
4	50	33	27
5	50	25	25

Para los sistólogos el concepto de tester típico [12] resultó muy claro y adecuado para describir la importancia informacional de los rasgos y de los objetos y también para hacer el análisis de las analogías (parciales) que se presentan entre las descripciones de dos cuadrículas dadas. Por tal razón el sistema de conjuntos de apoyo que utilizamos fue el formado por todos los testers típicos de la matriz de aprendizaje. En cuanto al segundo parámetro de los algoritmos de votación [12], la función de semejanza, nos dará la analogía entre subdescripciones irreducibles de las cuadrículas y esta analogía se obtiene sobre la base de un análisis similar de los atributos que conforman dicha subdescripción. En nuestro caso no fue posible determinar por razones sistológicas, cuántos de estos atributos podían no ser análogos sin perderse la analogía entre las subdescripciones, por lo que decidimos proceder heurísticamente. Una vez calculado el valor de la función de semejanza entre una subdescripción irreducible de una cuadrícula de la matriz de aprendizaje y la correspondiente subdescripción de la cuadrícula sujeta a estudio, el mismo se pondera con un parámetro que denota la importancia informacional de la cuadrícula de la matriz de aprendizaje; posteriormente se promedia todos los valores que se obtienen de manera análoga a la antes descrita respecto a todas las cuadrículas de cada clase y finalmente se promedian valores que se obtienen al repetir todo el conjunto de cálculos anteriores para cada uno de los conjuntos de apoyo, es decir, en nuestro caso, para cada tester típico de la matriz de aprendizaje. Así se obtiene una magnitud que denota la votación que recibe la cuadrícula sujeta a estudio para pertenecer a cada una de las clases. En nuestro caso esta votación se puede representar mediante:

$$\Gamma_j(o) = \frac{1}{|\mathcal{R}_A|} \sum_{\omega \in \mathcal{R}_A} \frac{1}{|K_j|} \sum_{o_i \in K_j} \delta(o_i) \beta^E(\omega o, \omega o_i) \quad (4)$$

donde $\mathcal{R}_A$ denota al conjunto de los conjuntos de apoyos; $\omega \in \{x_1, \dots, x_n\}$; $\delta(o_i)$ el peso informacional de o_i ; β^E la función de semejanza que depende del parámetro E (la cantidad de atributos $X_i \in \{x_1, \dots, x_n\}$ que pueden no ser análogos y mantenerse la analogía entre las subdescripciones).

El sexto parámetro, en nuestro caso, fue la regla de solución "mayoría simple".

Para una matriz inicial de 233 objetos se realizó el análisis comparativo entre objetos para la búsqueda de filas iguales en clases diferentes. Como resultado del mismo fueron eliminados 9 objetos. En la matriz resultante no existen filas iguales en clases diferentes. Dicha matriz fue dividida en las matrices de aprendizaje y control con 114 objetos cada una. El criterio usado fue, tratando de mantener la representación geográfica (distribución de puntos por toda la región de estudio), el seleccionar para la matriz de aprendizaje aquellos objetos para los cuales existía mayor seguridad de pertenencia a la clase correspondiente. La determinación de la pertenencia de un objeto a una clase específica tiene implícita una carga de subjetividad. Como se usan las magnitudes máximas reportadas, estamos seguros que no puede pertenecer a una clase superior ($M_{\max}$ menor), pero nunca estamos completamente seguros que no pueda pertenecer a una clase inferior ($M_{\max}$ mayor); y en ese caso se toma la decisión analizando la ubicación geográfica del objeto en relación con los objetos de clases inferiores.

La subfamilia de algoritmos de votación antes descrita fue la utilizada en este trabajo. Con esta subfamilia, las matrices de aprendizaje y de control y el funcional de error expresado en (2) se procedió a encontrar el algoritmo de mayor eficiencia respecto a esos parámetros.

Se tomó la decisión de emplear solamente los testores típicos, y se probaron diferentes funciones de cercanía. Finalmente se seleccionó el caso en el que el 80% de los elementos del testor típico fueron semejantes.

Los resultados fueron los siguientes:

Cantidad de testores típicos de la matriz de aprendizaje: 8

Valor mínimo del funcional de error: 111.23

Comportamiento de la clasificación por intervalos:

Intervalo	Bien clasificados	Mal clasificados	Abstenciones
1	9	10	1
2	18	8	1
3	5	10	-
4	21	6	-
5	25	-	-

Debe señalarse que del análisis de los testores típicos resulta que las variables, por orden de importancia son, en primer lugar (aparecen en todos los testores típicos), la liberación de deformaciones, la distancia a las estructuras fundamentales y la regionalización geomorfológica, en segundo lugar, la actividad sísmica y el número de elementos estructurales (aparecen en 7 y 6 testores típicos respectivamente). Las anomalías de Bouguer no entraron a formar parte de ningún testor típico, y las demás variables sólo de uno, y en el caso de la densidad de epicentros de dos.

5. SOLUCIÓN DEL PROBLEMA A'

El problema matemático a resolver, una vez hallado el mejor algoritmo de reconocimiento respecto a las matrices de aprendizaje; de control y el funcional de calidad (2), es en nuestro caso, el de determinar la pertenencia a uno de los cinco intervalos definidos de las 5022 cuadrículas que cubren toda la zona de estudio, antes especificada. Los resultados alcanzados con el algoritmo seleccionado fueron:

Intervalo	Cantidad de Cuadrículas
1	136
2	258
3	240
4	412
5	3858
no clasificados	117

El resultado de esta clasificación, en forma de mapa se presenta en la Figura 1. En el mismo cada cuadrícula está sombreada de acuerdo a la clase a que pertenece, quedando en blanco los casos de no clasificación. Para facilitar la interpretación se incluyen 2 detalles ampliados correspondientes a las regiones Cuba-Jamaica y La Española-Puerto Rico (Figuras 2 y 3), de las cuales se posee un mayor volumen de información sobre terremotos fuertes ocurridos desde el siglo XVI hasta la fecha. Todos estos mapas corresponden a la impresión desde pantalla de microcomputadora de las representaciones gráficas que se obtienen usando el programa MAPAS [15].

a. distribución

En primer lugar podemos asegurar que en términos generales los resultados obtenidos tienen una buena correspondencia con el cuadro de ocurrencia de terremotos fuertes en la región de estudio. No obstante hay algunos aspectos que merecen ser comentados.

Los objetos no clasificados (2,1% del total) se distribuyen fundamentalmente a lo largo de América Central, en América del Sur y el Norte de las Antillas Mayores. Las dos primeras corresponden a zonas donde numerosas variables presentan ausencia de información, lo cual puede conducir a la no clasificación. Sin embargo, para la zona al norte de las Antillas Mayores existe prácticamente una cobertura total de datos para todas las variables, por lo que no resulta fácil determinar la causa de esa concentración de objetos no clasificados.

Viendo el detalle de la Figura 2 (Cuba-Jamaica, encontramos zonas al norte de Cuba centro oriental con M_{max} en el rango [5,6] que salvando las imprecisiones de la escala utilizada, se corresponden con lo que esperábamos. Sin embargo, zonas similares deberían aparecer en el centro de la provincia de Pinar del Río, y no aparecen. Igualmente en la región sur oriental la magnitud máxima que aparece es en el intervalo [6,7], y en específico, al sur de Santiago de Cuba, la M_{max} resulta en el intervalo [5,6].

Por otra parte, en el detalle de la Figura 3 (La Española-Puerto Rico) observamos fenómenos similares. El extremo este de La Española (Puerto Rico), está bastante bien caracterizado, no ocurriendo lo mismo en el resto de La Española donde aparecen zonas con muy baja magnitud en lugares donde históricamente han ocurrido terremotos fuertes.

Las causas de estas incongruencias están localizadas en dos factores: imprecisiones en las variables utilizadas e insuficiencia de la matriz de aprendizaje para describir completamente el fenómeno de ocurrencia de magnitudes máximas en la región.

Con respecto a las variables, las geológico-geotécnicas y geofísicas son tomadas de materiales regionales con un alto grado de generalización, y las sismológicas procesando un catálogo de terremotos que no ha sido objeto de un trabajo especial de estudio de la predicción de las ocurrencias. Como resultado tenemos variables sujetas

a un margen de indeterminación difícil de determinar.

El caso de la matriz de aprendizaje es diferente. En ella deliberadamente no se incluyeron magnitudes determinadas por datos macrofísicos, lo cual nos privó de contar con numerosos elementos de la clase-1 ($M_{max} \geq 7$), incluso en la zona al sur de Santiago de Cuba que mencionamos como mal clasificada. Sin embargo, la necesidad de dividir en matriz de aprendizaje y de control nuestra información inicial es responsable de algunos errores, ejemplo de ello es una cuadrícula al suroeste de Jamaica con M_{max} en el intervalo [6,7] para la cual existen datos de una $M=7.1$ reportado en [94], solo que el objeto correspondiente se incluyó en la matriz de control. Una posible solución a este problema sería usar la división en MA y MC sobre la base de la función de cercanía que utilizamos en el algoritmo de votación, dejando en MA a aquellas cuadrículas que menos se parecen entre sí, con los que obtendríamos el máximo de diversas representaciones de objetos de una misma clase.

7. CONCLUSIONES Y RECOMENDACIONES

De los resultados discutidos anteriormente se pueden extraer las siguientes conclusiones:

1. Se obtuvo un mapa de M_{max} de los terremotos para la región del Caribe que en términos generales corresponde al cuadro de ocurrencia de terremotos fuertes en la región.
2. El modelo de los algoritmos de votación es aplicable a la solución del problema de determinación de las magnitudes máximas de los terremotos, pero deben tomarse en consideración algunos factores:
 - Es imprescindible que los parámetros que se utilizan como variables para el trabajo de reconocimiento de patrones sean depurados al máximo de las fuentes de incertidumbre;
 - Debe tomarse un cuidado especial al preparar la matriz de aprendizaje, ya que en ella deben cumplirse diferentes requerimientos como son: representación geográfica máxima, confiabilidad elevada del criterio de pertenencia a una clase, etc.
3. El trabajo en esta dirección debe continuarse, buscando mejorar los resultados obtenidos hasta el presente, así como realizar estudios a escala más detallada en regiones más pequeñas.

- [1] Cherenasina, Eugenia N.; Ruiz Shulcloper, José (En prensa). "Cuestiones metodológicas de la aplicación de modelos matemáticos de reconocimiento de patrones en zonas del conocimiento poco formalizado". Revista Ciencias Matemáticas.
- [2] Ryznichenko, Yu.V. (1964). "Sobre la relación entre la energía de los terremotos máximos con la actividad sísmica" (en ruso). Doklady Akademii Nauk. SSSR, Vol. 157, No.6, pp. 1352-1354.
- [3] _____; Dzhibladze, E.A. (1974). "Determinación de los terremotos máximos posibles por datos complejos para el Cáucaso" (en ruso). Fizika Zemli, No.5, pp. 64-65.
- [4] Gelfand, I.M.; Guberman, Sh.A.; Izvekova, M.L.; Kailis-Borok, V.I.; Mantman, E.Ya. (1973). "Diagnóstico del lugar de posible ocurrencia de terremotos fuertes I. Pamir y Tianshan" (en ruso). Vichislitel'naya Seismologiya, No.6, pp. 107-113.
- [5] Orishiani, A.D.; Zelyvinski, A.V.; Kailis-Borok, V.I.; Kosobokov, V.G. (1960). "Diagnóstico de los lugares de surgimiento de los terremotos más fuertes del cinturón del Pacífico (m 8.2)" (en ruso). Vichislitel'naya Seismologiya, No.13, pp. 30-43.
- [6] Burd, V.I.; Furbovich, I.T.; Borisov, B.A.; Gitis, V.G.; Reizner, G.I.; Yurkov, E.P. (1975). "Método de pronóstico de la magnitud máxima de los terremotos" (en ruso) Fizika Zemli, No.10, pp. 31-43.
- [7] Gitis, V.G.; Mironov, M.A.; Burd, V.I.; Kalerix, V.N.; Schulkin, Yu.V. (1981). "Max earthquake prediction map development: methods, software, results". En: Proceedings of the 2nd International Symposium on the Analysis of Seismicity and on Seismic Hazard. Praha, Academy of Sciences, pp. 437-466.
- [8] Levchenko, V.A.; Gurevich, G.S.; Zhabreliva, V.V.; Kitis, N.A.; Parfionova, S.I. y otros (1976). "Características comparativas de la construcción geológica de Cuba y otros países de la cuenca del Mar Caribe y del Golfo de México con el objetivo de evaluar perspectivas de posibilidades petrolíferas y caudíferas de Cuba y las Américas adyacentes". Informe, Fondo Geológico Nacional.
- [9] Case, J.E.; Holcombe, T.H. (1980). "Geologic-geographic map of the Caribbean region". Dept. of Interior, U.S. Geological Survey.
- [10] Catálogo de terremotos fuertes hasta Abril de 1983 preparado por la NOAA de los EEUU. Selección para la región del Caribe obtenida por Cortesía del Centro Mundial de Datos B-2 de Moscú.
- [11] Guay Rodríguez, Tomás; Alvarez Gómez, Leonardo (Inédito). "Catálogo unificado macro sísmico de las Antillas Mayores".
- [12] Ruiz Shulcloper, José (En prensa). "Modelos matemáticos para el reconocimiento de patrones". México.
- [13] Pico Peña, Ramón; Ruiz Shulcloper, José (En prensa). "MAPAS: TCR: Sistema herramienta de selección de variables y objetos". Proceedings de Informática'90. Febrero de 1990. Habana, Cuba.
- [14] PROMOS es un módulo del sistema PROGNOSIS por APAFECAR.
- [15] Pico Peña, Ramón (1990). "MAPAS": Manual de Usuario.

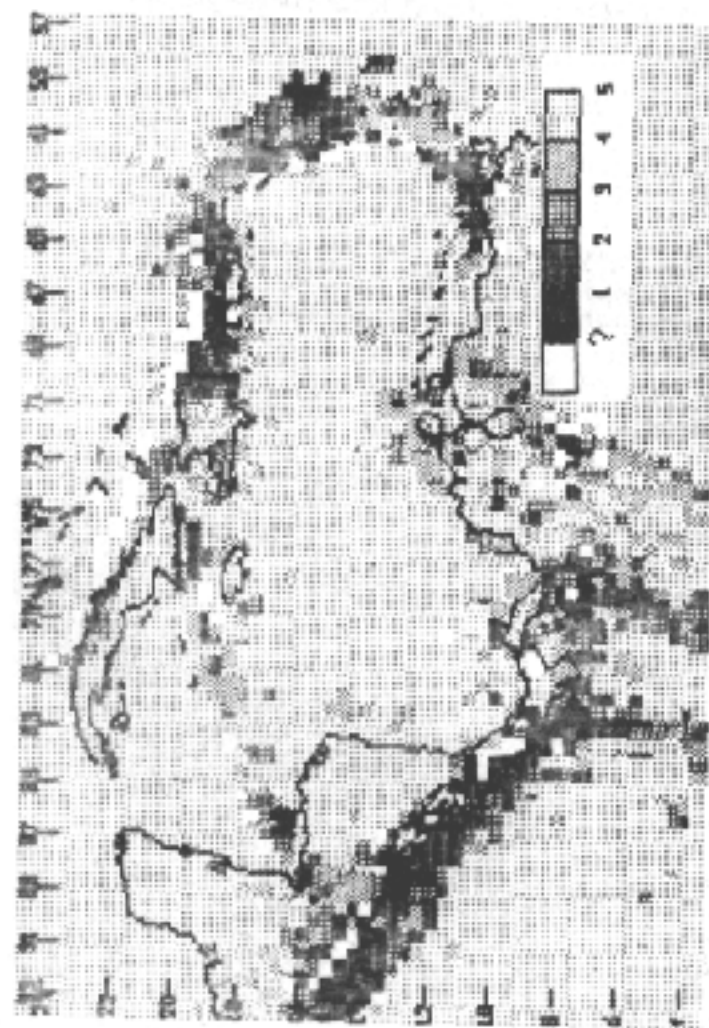


Fig. 1. Mapa de Maniz para la región estudiada. 1 - sin edificación, 2 - Maniz 7, 3 - 6 Maniz 7, 4 - 4 Maniz 5, 5 - Maniz 4.

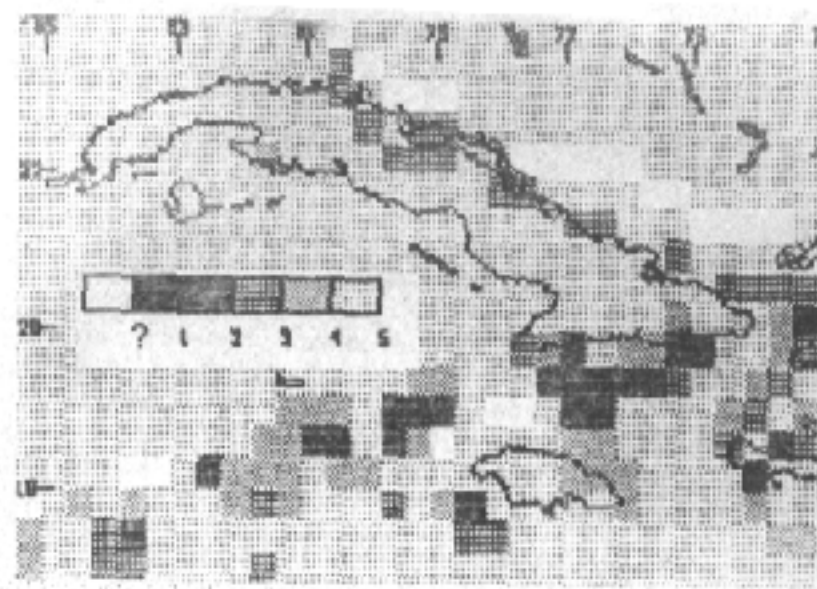


Fig. 2. Detalle ampliado para el sector Cuba-Jamaica.

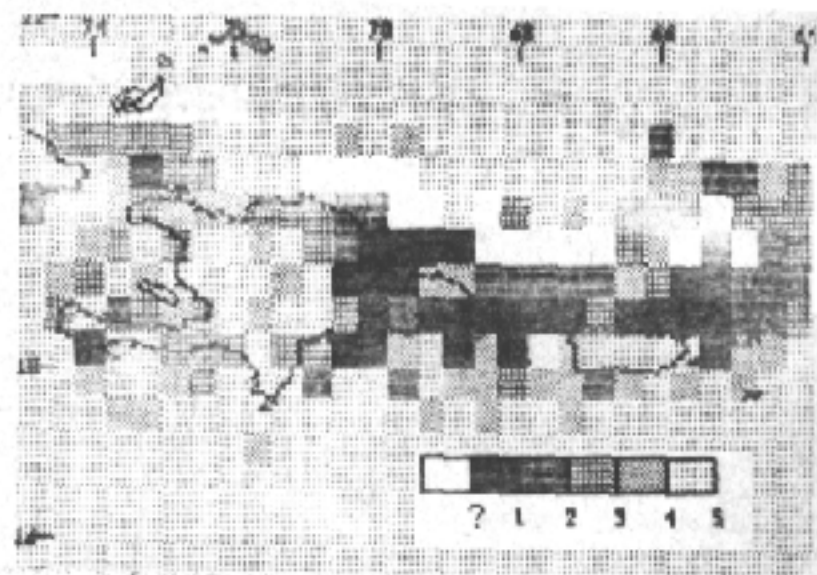


Fig. 3. Detalle ampliado para el sector Española-Puerto Rico.